mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Epidemiologia e Bioestatística

Introdução

Existem inúmeros livros e de epidemiologia e de bioestatística de excelente qualidade e completos em detalhes. O objetivo desta apostila é combinar os conhecimentos de fundamentos de epidemiologia (métodos) e de bioestatística mostrando a razão de existirem e ressaltando a necessidade que temos desses conhecimentos cominados na prática de uma profissão da área de saúde.

A base inicial desta apostila que foi escrita em 2003 para servir de material didático dos curso de Epidemiologia e Bioestatística criado especificamente para a graduação e pós-graduação em Odontologia os quais seguiam uma sequência diferente dos livros existentes tando de Epidemiologia como de Estatística Básica. Portanto, esta apostila tem uma apresentação bem diferente de livros tradicionais de epidemiologia e bioestatística por combinar as duas disciplinas, e por ter uma sequência de epidemiologia diferente, pois considero que as sequências padrão em livros tradicionais não favorece o aprendizado de iniciantes. Apenas em 2014, dois autores, Keyes e Galeo, questionaram as sequências de livros de Epidemiologia como um obstaculo no aprendizado de epidemiologia num artigo publicado na revista American Journal of Epidemiology. No mesmo ano, estes autores publicaram um livro interessante chamado Epidemiology Matters com uma abordagem e sequência um pouco mais coerentes e com algumas familiaridades com esta apostila que está escrita desde 2003. No entanto, considero que a sequência desta apostila seja mais lógica do que o livro de Keyes e Galeo (2014), pois começa com o objetivo geral da ciência que é de desvendar processos causais mesclando de forma intuitiva princípios de estatística. É claro que o objetivo da apostila tem um propósito claro de ensino para alunos de graduação em área que não é a saúde pública ou epidemiologia pura, ela foi construída com o propósito de auxiliar o ensino de ciência e princípios de estatística por meio da epidemiologia, além de subsidiar conhecimentos necessários para a prática de odontologia/medicina/enfermagem baseada em evidências além de leitura crítica de artigos científicos.

A seguir descrevo alguns diferenciais mais importantes da apostila em relação aos livros tradicionais de epidemiologia. Em geral os livros de epidemiologia têm como primeiro capítulo a história da epidemiologia, seguido de medidas de frequência e associação que são utilizados em estudos epidemiológicos, que são abordados em capítulos posteriores junto com viéses. No entanto, as medidas de frequência dependem da forma como os estudos são montados, isto é, como as pessoas são selecionadas para entrarem num estudo e como elas são observadas, se pontualmente ou ao longo do tempo. Desta forma, não faz sentido abordar medidas de frequência de doenças antes do conhecimento dos tipos de estudos. Por outro lado, não faz sentido apresentar tipos diferentes de estudos sem fundamentar sua origem e necessidade de

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

seu desenvolvimento. Os desenhos de estudos e metodologias aplicadas evoluiram aos poucos apartir da necessidade de desvendar causa de doenças ou eventos por meio de observação e não de experimentos. Apartir de hipóteses ou relatos de casos que davam indícios de possíveis associações causais houve necessidade de se buscar evidências que fundamentassem tais associações e deste desafio os tipos de estudos foram desenvolvidos e aprimorados. Desta forma, causalidade é o foco inicial da pesquisa cientifica na área da saúde, assim como em qualquer área. A física se preocupa com a causa de fenomenos na natureza e a origem do universo (que não deixa de ser a causa da existência do universo), a sociologia com a causa de fenomenos sociais, a ciência politica com as causas de eventos politicos, a econômia com a causa de fenomenos econômicos e assim por diante.

Portanto, causalidade é o primeiro tópico a ser abordado apartir do qual os desenhos de estudos (experimentais e observacionais) serão intuitivamente construídos, explorando suas limitações e desafios para serem realizados. No entanto, o foco não é apresentar discussão formal e teórica sobre causalidade em profundidade, o que é abordado em diversos excelentes livros incluindo modelos teóricos como desfechos em potencial (pontential outcomes) e modelos causais estruturados incluindo gráficos de diagramas acíclicos (DAG-diagram acyclical graph). Nossa abordagem de causalidade será bem elementar, com a finalidade de estimular a curiosidade para que o aluno procure outros fontes de aprendizado posteriormente. A minha experiência ao ensinar profissionais da área da saúde é a falta completa de conhecimento to tema causalidade ou mesmo da necessidade de abordagem do tema.

Causalidade é introduzida por meio da apresentação de um relato de casos do Dr. Gregg, médico que primeiro identificou a possível associação entre rubéola durante a gravidez e catarata congênita relatada em um artigo de 1944. Este exemplo será o principal a partir do qual serão discutidos os problemas que podemos enfrentar ao tentar concluir causalidade com base em simples observações e também e quais as alternativas que poderiam ser utilizadas para verificar a veracidade da associação proposta. Essa discussão nos levará de forma lógica ao desenvolvimento de desenhos de estudos alternativos e mais confiáveis para se elucidar a associação entre rubéola e catarata congênita. Em seguida para cada alternativa de desenho de estudo veremos que medidas de frequência de eventos e medidas de associação específicas com interpretações especificas foram desenvolvidas.

Ao abordar os estudos transversais o processo de amostragem (não probabílistica e probabilística incluindo amostras simples e complexas) é apresentado com devenvolvimento dos conceito de erro amostral e Intervalo de Confiança. Esses conceitos são apresentados de forma bem intuitiva por meio de simulação (bootstrapping), portanto não necessitando de fórmulas para serem compreendidos. Apartir do conceito de intervalo de confiança, introduzimos comparação de dois ou mais estimadores (médias, proporções, razões) iniciando assim princípios de Estatística básica, sem a necessidade de utilização de fórmulas. Em sequência, são apresentados princípios de análise de regressão linear e análise de variância entre outros tópicos. Após os conceitos, as fórmulas são apresentadas de forma intuitiva.

Um diferencial desta apostila é simplificar denominações atribuídas e incorporadas aos nomes dos estudos por diversos autores. Alguns termos como, por exemplo, retrospectivo, concorrente e não concorrente são utilizados com diversos

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

significados em diferentes livros o que dificulta o aprendizado dos iniciantes na área. Por exemplo, é muito comum a confusão entre uma coorte retrospectiva e caso controle. Vandenbroucke em 1991 (BMJ 1991; 302:249-50) publicou um artigo excelente sobre o assunto em que sugere abandonar determinadas terminologias como retrospectivo e prospectivo porque em geral causam confusão. Não sei se concordo completamente com o abandono, mas acho que seria necessário unificar a interpretação dos termos. Existe um dicionário de Epidemiologia, porém o mesmo não serve para estabelecer normas, pois apenas apresenta diversas formas como os termos são utilizados por diversos autores. Infelizmente alguns profissionais ligados a Epidemiologia, passaram a considerar tal dicionário como se fosse uma fonte de definições exatas de termos da área.

A falta de concordância de significados entre os livros leva à dificuldades como por exemplo a atribuição do nome “caso-controle” á qualquer estudo que tenha um controle. Por exemplo, encontramos alguns artigos que consideram de caso-controle apartir de dados de estudo transversal populacional, pelo simples fato dos autores considerarem que doentes e não doentes são grupos comparados na análise estatística. No entanto, os estudos recebem suas denominações não baseados em na análise estatística e sim da forma de seleção da amostra. Outra confusão comum é atribuir o nome de “estudo de coorte” a todo e qualquer estudo que tenha acompanhamento, mesmo na presença de uma intervenção o que caracterizaria um quasi-experimento. O termo quasi-experimento é outro dilema nas últimas decadas, pois foi abandonado num dos livros principais da Epidemiologia (Modern Epidemiology), no corpo da apostila irei apresentar melhor meus argumentos em favor do uso de quasi-experimentos. Em defesa do termo, eu o considero extremamente útil no ensino, além de ser um termo utilizado amplamente em outras áreas como Sociologia, Ciência Política e Econômia. Numa momento em que anti-disciplinaridade se torna um novo movimento de educação, seria interessante que diversas áreas utilizassem terminologias comuns.

Nesta apostila ainda temos o diferencial de introdução de conceitos de estatística logo que o estudo transversal ou de prevalência é apresentado, pois o mesmo depende de amostragem. A amostragem serve de ligação entre os tipos de estudo (epidemiologia) e a estatística inferencial, a qual é abordada antes mesmo do que a estatística descritiva. Parece loucura, mas sempre achei que esta é a melhor forma de se iniciar o ensino de estatística e assim ensino desde 2003. Embora eu tenha começado a ensinar estatística desta forma invertida em 2003 contra todos os livros existentes na época, em 2005 Joan Garfield publicou um livro com sugestões semelhantes (Innovations in Teaching Statistics) e posteriomente em 2013 foi publicado o livro Unlocking the Power of the Data dos autores Lock et al que se inicia pela inferência estatística com ajuda de bootstrapping e applicativos disponíveis no site “statkey”.

Embora fundamentos de Epidemiologia e Estatística sejam matérias densas espero que a apresentação destas disciplinas de forma despretensiosa mostre ao aluno a importância e a interligação das duas disciplinas na prática clínica e para a pós-graduação.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

1. Causalidade e Tipos de estudos

A definição mais usual de Epidemiologia é que é a ciência que estuda a as causas e a distribuição de estados de saúde e doença nas populações. O seu grande objetivo final é poder por meio das informações obtidas em estudos promover intervenções com a finalidade de previnir doenças (vacinas, medicamentos de controle, e outras medidas preventivas como fluoretação de águas) e promover saúde (nutrição, meio ambiente). A epidemiologia tem origem no estabelecimento de métodos científico para o estudo de causas de doenças. O evento considerado marco da Epidemiologia Moderna foi a descoberta da água contaminada causando a transmissão da cólera em Londres, Inglaterra, por John Snow. No entanto, a simples indagação sobre causas de doenças por Hipócrates faz do mesmo o pai da Epidemiologia, inclusive com as atribuições das definições dos termos epidemia (aumento de uma doença recem introduzida ou não numa população) e endemia (presença constante de uma doença numa população). A Epidemiologia, portanto, nada mais é do a ciência que se debruça sobre os problemas metodológicos na descoberta de causas de doenças sejam elas crônicas ou infecciosas, e desta forma, é essencial para toda e qualquer área da saúde. Embora em geral Epidemiologia seja abordada em áreas de saúde humana, ela também é aplicada a animais e portanto veterinários estudam métodos epidemiológicos para estudar causas de doenças em animais.

Pelo exposto, é evidente que estudar causas das doenças é objetivo principal da Epidemiologia, e o estudo da distribuição e monitoramento de doenças na população também fazem parte deste processo de conhecimento das causas, sejam elas diretas ou indiretas. As vezes para o clínico é difícil compreender que estudando um indivíduo isoladamente é quase impossível ter evidências científicas sobre as causas de uma doença. Espero que ao final da apostila isso fique mais claro para aqueles que assim

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

pensam. É difícil para muitos também compreender que causas de doenças têm componentes ambientais e especialmente sociais além agentes biológicos. Essa dificuldade é particularmente decorrente da maneira tradicional de ensino na área da saúde focando em tratamentos de indivíduos isolados e fora de seu contexto social. Por ter que abordar componentes sociais e ambientais, a Epidemiologia engloba conhecimentos de sociologia, econômia, meio ambiente e ciências exatas como estatística, matemática, física, fora é claro áreas relacionadas a saúde, química, bioquímica, anatomia e patologia. Portanto, pode-se dizer que é Epidemiologia é multidisciplinar (envolve várias disciplinas no estudo de um estado de saúde) e transdiciplinar pois o objetivo de entender uma doença permeia várias disciplinas que na verdade tem o método científico em comum. Embora eu defenda o ensino da disciplina de Epidemiologia, ela nada mais é do que uma disciplina que se preocupa com o método científico dentro da área de saúde e portanto deve estar presente em todas as disciplinas. Um outro aspecto importante para quem está começando na epidemiologia é compreender sua suposta fragmentação em várias Epidemiologias como Epidemiologia Molecular, Epidemiologia do Meio Ambiente, Epidemiologia Social, Farmaco Epidemiologia entre outras. Na verdade não são disciplinas diferentes, todas utilizam métodos desenvolvidos ou aplicados em Epidemiologia em situações específicas e por vezes com particularidades. Por exemplo, medir exposição a poluição tem seus desafios metodológicos em relação a medir exposição a medicamentos ingeridos por uma pessoa.

Essa suposta fragmentação se deu com a evolução dos estudos populacionais que no passado em sua maioria incluiam especialmente levantamentos de condições de saúde por meio de questionários e alguns exames físicos. Conforme os aspectos sociais se mostravam importantes também para determinação de doenças estabeleceu-se uma disciplina destinada a estes aspectos criando-se a Epidemiologia Social, que se preocupa não somente em saber a educação do indivíduo como do bairro em que vive entre outras condições como violência, urbanização, presença de áreas de lazer, mobilidade etc. De maneira semelhante foram criadas a Epidemiologia Molecular, Epidemiologia Genética, Farmaco Epidemiologia, Epidemiologia do MeioAmbiente entre outras. Na verdade são todas Epidemiologia com aspectos diferentes de coletas de dados,

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

processamento de amostras, e que precisam por vezes de conhecimentos específicos de genética, de técnicas de laboratório, de farmacos e de aspectos sociais.

É importante também ressaltar que não apenas causas de doenças fazem parte da Epidemiologia, mas também métodos de prevenção de doenças, vacinas e tratamentos de doenças. Embora essas atividades por vezes sejam realizadas em diversas disciplinas clínicas, os métodos utilizados para tanto foram desenvolvidos e/ou aprimorados pela Epidemiologia.

Nesta primeira parte da apostila, vamos começar com a discussão de causalidade e quais as possibilidades de montagem dos principais estudos (chamado tipos de estudos) para desvendar um processo causal. Apartir de um exemplo real, vamos apresentar as limitações de evidências que o ser humano em geral tende a acreditar, mas que pode levar a conclusões erradas. Discutindo a limitação das evidências iniciais propostas por um pesquisador em 1941, vamos propor alternativas a seu estudo discutindo suas limitações. A finalidade deste texto é que o aluno compreenda intuitivamente a importância do estudo de causalidade, e perceba que os tipos de estudos foram elaborados com o tempo, para suprir a necessidade de evidências mais fortes de processos causais. Como os tipos de estudos são apresentados de acordo com a necessidade de informações mais específicas, ao final você terá uma visão da qualidade e aplicabilidade da cada estudo. Ainda, ao apresentar os estudos falaremos qualidade de coleta de informações e como selecionar indivíduos para os estudos (amostragem).

O exemplo sobre a associação entre rubéola e catarata congênita será apenas utilizado como motivação para as discussões. O objetivo não é contar a história da associação, mas utilizar a mesma para discutir causalidade. Desta forma, a todo o momento no decorrer do texto deste capítulo, voltaremos a situação do relato de Greeg sobre a associação entre Rubéola e catarata congênita. O texto é um vai e vai no exemplo do Greeg.

1.1 O relato de casos de Normal M. Gregg : catarata congênita e rubeola

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Começando nosso exemplo: em 1941, Gregg, um médico oftalmologista, observou maior ocorrência de catarata em sua clínica na Austrália. Curioso com o aumento repentino, Gregg contactou seus amigos oftalmologistas que confirmaram o aumento de ocorrência. Um dia Gregg ouviu a conversa de duas clientes que tiveram filhos com catarata dizendo que elas tiveram rubéola durante a gravidez. Rubéola, na época, não era muito comum, porém o retorno dos pracinhas infectados durante a segunda guerra mundial provocou uma série de epidemias da doença.

Greeg, então, partiu para sua pesquisa tentando desvendar o mistério. Primeiro contactou vários colegas de profissão que entrevistaram todos os casos de catarata congênita ocorridos na época em seus consultórios. Dos 78 casos levantados, Greeg observou a seguinte ocorrência.

O levantamento de Greeg encontra-se na tabela 1.

Tabela 1. Casos de Catarata relatados por Gregg , 1941

Condições Apresetadas	**Número de casos**
Rubéola durante a gravidez	68
Doença Renal	1
Sem história de rubéola	9
Total	78

Com os resultados da tabela acima Gregg concluiu que a causa do aumento de crianças com catarata era a rubéola. O que você acha desta conclusão de Gregg? Antes de prosseguir com a leitura, tente refletir sobre o que Gregg concluiu. Faça anotações sobre suas reflexoes, sejam elas quais forem (concorda, não concorda, não concorda por quê?). Anote o que vir a cabeça sobre a conclusao do Gregg.

Desvendar um fator causal e compreendê-lo em geral não é nada fácil, se fosse, nós já saberíamos a causa de todas as doenças e eventos no mundo. O processo causal é discutido em várias áreas especialmente a filosofia e a física. A literatura em causalidade é vasta e alguma informação pode ser obtida na internet como, por exemplo, no site do

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Quando observam estes números pela primeira vez, em geral, as pessoas acreditam na associação, sem nenhum quesitonamento, uma vez que a maioria das crianças com catarata congênita tiveram mães com rubéola. Uma parcela menor de pessoas irá dizer que não, com o argumento de que 10 casos não foram associados a rubéola, ou ainda com o argumento de que o número de casos foi pequeno (n= 78), ou que o estudo é restrito a uma cidade e que deveria incluior o país inteiro.

Vamos analisar algumas destas argumentações.

"Existirem casos de catarata cogênita sem que a mãe tenha relatado doença".

Essa condição pode ser explicada de várias maneiras como, por exemplo, é possível que todas as mães tenham tido rubéola, mas não se lembrem do fato, porque os sistomas eram bem leves ou inexistentes (1). Devemos lembrar que na época não havia testes laboratoriais para se determinar/confirmar se a mãe tinha tido ou não rubéola. Algumas doenças podem provocar sintomas leves e não serem percebidos. Portanto, poderia ser que todas as mães de fato apresentassem rubéola. Outra explicação que também não anularia a possibilidade de associação é que existem outras causas de catarata congênita (fatores genéticos e outros) e, portanto conseguiriamos explica os 10 casos em que a mãe não havia tido rubéola.

Uma das grandes dificuldades para se desvendar um processo causal é a expectativa de se encontrar um fator único que leva a um **desfecho** (doença, ou situação). Há alguns anos, presenciei um aluno perguntando a um professor de psicologia, se o estresse era causa da depressão. O professor mais do que de pressa respodeu que não, pois a causa verdadeira da depressão seria algo de dentro do

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

indivíduo, possívelmente uma susceptibilidade genética. Será? Na verdade nem sempre sabemos o que algumas pessoas chamam de "susceptibilidade genética". Seria muito bom que pudessemos explicar todas as doenças por suceptibilidade genética, mas não sabemos se é assim que funciona, e provávelmente não é.

A ideia de causa única é fascinante e confortável, mas sabemos que, por exemplo, o câncer de pulmão pode ter como causa principal o fumo, mas também pode ser decorrente de exposição ao asbesto (amianto) e ou ao benzeno entre outras causas. Esta idéia de causa única suficiente tem origem, provávelmente, nos estudos de doenças infecciosas que dominaram os séculos passados culminando nos Postulados de Koch. Deixo aqui a tarefa para o aluno procurar os Postulados de Koch na internet. Sem dúvida uma gripe somente é possível com a presença do vírus específico, porém, a simples presença do vírus não desencadeará a doença. Além do vírus, há a necessidade de o indivíduo ter alguma fragilidade do sistema imunológico no momento propiciando que o vírus se desenvolva, caso contrário, o vírus deve ser eliminado e não vai provocar a gripe. Portanto, nessa situação embora o vírus seja essencial para causar a gripe, a susceptibilidade do hospedeiro é essencial para o desenvolvimento da doença e, portanto, faz parte destre processo causal. Assim, causa em geral é um conjunto de fatores que resultam num **desfecho** (algum resultado como doença, acidente, ou qualquer evento final). Podemos dizer então que numa gripe o vírus é a parte da causa que é **necessária** (essencial), mas que não é **suficiente** para desencadear a doença. Assim, o vírus é a parte necessária, mas não suficiente, e a susceptibilidade do hospedeiro não é necessário, mas é parte integrante da causa. Portanto, esta ideia de causa única e suficiente nem sempre existe nem mesmo na AIDS. O vírus HIV é extremamente importante, mas ele sozinho sem a "susceptibilidade" do hospedeiro não dará origem a doença AIDS.

Vejamos outro exemplo: uma determinada força é aplicada a uma caneta e a caneta se quebra. A princípio, pode parecer que a força é o fator causal único neste processo, porém, não é bem assim. Uma mesma força de igual magnitude pode ser aplicada sobre a caneta em outro dia sem que ela se quebre. Por quê? Talvez, a temperatura possa estar maior no outro dia e a caneta entorta, mas não quebra. Também, se o ponto exato em que se colocou a força não for o mesmo pode não ser

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

suficiente para quebrar a caneta. Assim, compõem a causa da quebra da caneta a força aplicada, mais a temperatura do ambiente, umidade, local de aplicação da força, idade do plástico etc. Desta forma, compreendemos que é possível que a simples presença da rubéola pode não desencadear catarata congênita em todas as crianças. Além da Rubéola, uma série de outros fatores têm que acontecer para que a catarata ocorra.

Esta noção de causa suficiente e não suficiente é muito discutida em filosofia. Assim uma condição necessária para se ter um evento nem sempre é suficiente, assim condição suficiente pode ser um conjunto de condições que precisam estar presentes para que um evento aconteça. Na Epidemiologia, o autor Kenneth Rothman apresentou esta discussão de causas suficientes como modelo causal de componente-suficiente (sufficient-component cause model). Rothman se fundamental no trabalho de John Mackie denominado de condições INUS (INUS conditions) isto é insuficiente, mas não redundante parte de uma condição que é por si só não é necessária, mas suficiente para o resultado. Ou ainda um evento A é a causa de um evento B, se A for não redundante parte de uma condição complexa C, que embora suficiente não é necessária para causar B.

"Amostra pequena para se concluir que a rubéola causaria catarata congênita"

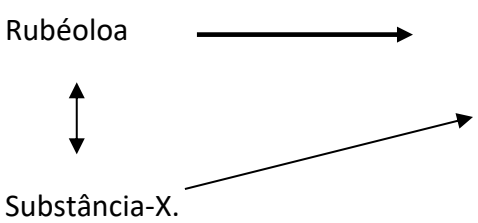
Outra opnião frequente sobre os dados apresentados pelo Gregg é que a amostra seria pequena, pois são apenas 78 casos de catarata congênita. Para alguns, a percepção é que é mais do que suficiente, mas para outros o número é pequeno. Vamos supor que na Austrália inteira tivessemos 7800 (sete mil e oitocentos casos) sendo 6800 deles relacionados às mães que tiveram rubéola durante a gravidez. Agora seria convincente? Muitos dizem que agora sim, aceitam que a rubéola é a causa da catarata, enquanto outros ainda achariam que o número deveria ser maior e abranger não somente a Austrália, mas os casos do mundo inteiro. Realmente nossa percepção nos engana e nos faz acreditar que se tivessemos os casos do mundo inteiro poderiamos acreditar. Note que mesmo que tenhamos os casos do mundo inteiro, a proporção de exposição, isto é proporção de mães que tiveram rubéola para os casos de catarata, será a mesma. A proporção de casos com rubéola impressiona (87,2%), porém deveria impressionar se e somente se essa porcentagem de mães com rubéola for maior do que

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

aquela encontrada em mães de crianças **sem catarata congênita**. Porque isso? Pense numa situação de epidemia com todas as mães pegando rubéola, logo a porcentagem de rubéola entre as mães de crianças sem catarata congenita seria a mesma. Assim, preciso saber como é a ocorrência de rubéola entre as mães de crianças sem catarata congênita.

Independente da porcentagem de mães afetadas por rubéola no estudo do Greeg ou do número de casos relatados por ele, poderiamos elaborar várias outras hipóteses para explicar o achado de Greeg. Uma possibilibidade seria outra causa para a catarata congênita que não fosse a rubéola. Vamos supor que estas mães tivessem em comum o uso de um medicamento que foi tomado durante a Rubéola, e este sim seria a causa da catarata. Ainda, podemos imaginar que as mães poderiam ter em comum certo tipo de emprego, como em alguma fábrica ou várias fábricas que as expusessem a alguma substância ou radiação. Trabalhando em serviços semelhantes o contágio por rubéola seria apenas uma coincidência.

A seguir temos uma figura com conexão entre rubéola e catarata congênita que vamos chamar de diagrama causal simplificado. Falaremos de diagrama causal de forma bem simplificada. Diagramas causais de verdade podem ser bem complexos e sofisticados como o atual DAG que tem aparecido em muitos artigos de diversas áreas. DAG (direct acyclic graph) é um tipo de diagrama causal chamado de diagrama acíclico direto, nele se podem representar fatores diretos e indiretos na cadeia causal seguindo alguns princípios. Não é nossa intensão abordar o DAG, nossa finalidade é de apenas introduzir alguns conceitos e para tanto nossa forma simplificada ajudará.



mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Neste desenho que chamamos de diagrama causal simplificado, Rubéola está ligado com uma seta única em direção à catarata congênita, porque a hipótese é que a Rubéola causaria a catarata congênita. A seta é única porque não estamos admitindo que a catarata congênita venha a provocar ou facilitar a ocorrência da rubéola na mãe. No entanto, nós levantamos a suposição de que outra substância estaria ao mesmo tempo associada à rubéola e que esta seria a verdadeira causa da catarata. Assim, a substância X também está ligada a Catarata congênita por uma seta de única direção. A rubéola e a susbtância X, estão unidas por uma seta dupla, apenas representando que existe associação entre elas, sem nenhuma causar a outra, a seta dupla apenas indica que as duas apenas se associam e podendo ou não ser causal. Note que a hipótese principal a ser testada é que a rubéola é a causa, então a linha que liga rubéola a catarata congênita é mais grossa.

Note que a associação que existe entre ser contaminado por uma substância tóxica e ter rubéola (doença contagiosa) poderiam ser apenas circunstâncial, porque as pessoas trabalhariam próximas. A nossa mente às vezes nos leva a certos racíocinios equivocados. Não pense que é apenas seu raciocício, talvez 90% das pessoas tenham concordado com Gregg sem argumentar sobre outras possíveis explicações. Não é difícil imaginar que mesmo pesquisadores acabem concordando com Gregg no primeiro momento, especialmente aqueles não familiarizados com estudos em humanos e que realizam estudos de laboratório em que vários fatores externos podem ser controlados.

Outra explicação que poderia ser dada para os números encontrados por Greeg, é que as mães poderiam pertencer a uma mesma grande família susceptivel a ter catarata congênita e que por contato tiveram rubéola durante a gravidez. Veja que qualquer fator associado ao mesmo tempo á rubéola e à catarata pode ser considerado o verdadeiro fator causal da associação. Portanto, Greeg jamais poderia concluir com certeza apartir de seus dados que a Rubéola era a causa da catarata congênita, no máximo poderia levantar a hipótese de que talvez a rubéola fosse o resposável e que novos estudos seriam necessários.

Existem diversos exemplos na literatura de associações equivocadas. Se você pesquisar na internet (hilarious spurious correlations ou spurious associations) irá encontrar vários exemplos engraçados como correlação entre número de nascimentos

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

e retorno de cegonhas na Inglaterra no pós-guerra, diminuição de consumo de margarina e de divórcios no Maine nos EUA entre varios outros. São correlações, mas claramente não são causais.

Até agora levantamos problemas e hipóteses que poderiam explicar os dados observados por Gregg. Esta parte de fazer critica aos outros é a mais facil de todas, porém o que Gregg deveria ter feito para que aceitassem melhor sua hipótese causal?

Quando fazemos esta pergunta em sala de aula são várias as respostas. Uma das respostas mais frequentes é que Gregg deveria fazer um estudo em animais. Para a maioria das pessoas um **estudo experimental** em animais resolveria tudo e esclareceria de vez a questão. Mas vamos supor que façamos tal estudo e os ratos não desenvolvam catarata. Então descartaríamos o achado do Gregg? E se achassemos associação, isso confirmaria os achados de Gregg? Bem, temos que encarar estes resultados com cautela. Se não encontrarmos associação entre rubéola e catarata congênita nos ratinhos, poderiamos explicar esta falta de associação pelas diferenças entre a fisiologia dos ratos e humanos. É possível que doenças que afetam os humanos se manifestem de forma de diferente em ratos e vice-versa. Se por outro lado a associação for observada em animais, será mais um indício de que talvez a hipótese de Gregg seja verdadeira. O estudo de ratinhos poderia também nos ajudar a entender como o virus da Rubéola causaria a catarata, pelo simples fato de que é mais facil eticamente matar o ratinho, e fazer uma necropsia.

Quer dizer então que experimentos em animais não são conclusivos, mas podem nos ajudar, porém temos que ter evidencia em humanos e, portanto, precisamos procurar maneiras de colocar à prova a afirmação de Gregg.

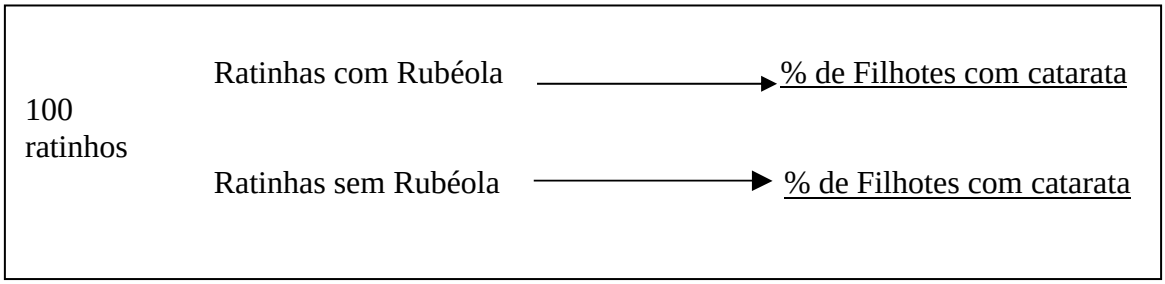


Fig.1

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

O desejo de se fazer um estudo em animais, é uma herança que temos dos estudos experimentais de laboratório a que estamos acostumados. Especialmente o estudo em animais é desejo da maioria dos alunos e professores de áreas básicas, porque é o tipo de estudo que em geral fazem em seus laboratórios. Ainda o experimento em animais tem o atrativo de supostamente ser mais fácil e prático de ser realizado porque podemos controlar várias características dos animais como idade das mães ratininhas, peso, genética, comida, temperatura, atividade física etc. Esta facilidade se contrapoe aos estudos em humanos nos quais inúmeros fatores não podem ser controlados. Estudando mulheres grávidas contaminadas com rubéola não podemos controlar sua alimentação, atividade fisica, exposição a outros fatores e tudo que já aconteceu no passado e mesmo durante a gravidez. Com certeza o estudo em animais é mais controlado, mas o que acontece na realidade em humanos é diferente e isso pode influenciar os efeitos da Rubéola. Esta "falta de controle" nos estudos em humanos leva as pessoas a acreditarem que os estudos em humanos não são bons, e apenas os estudos em animais são confiáveis. Desta forma, é comun que pesquisadores de laboratório acostumados a ter controle "absoluto" de seus experimentos, tenham aversão a estudos em humanos. No entanto, em determinados momentos precisamos estudar as doenças em humanos, mas por vezes não da maneira como poderiamos fazer se fossem com animais. Não podemos criar humanos controladamente para usar em experimentos, pois seria antiético. Também, talvez não devessemos criar animais para isso segundo protetores de animais. Para quem é apaixonado por animais não achamos correto usá-los em experimentos, porém considera-se que por vezes seja necessário. Existem pesquisadores que sofrem ao matar animais e preferem se dedicar a estudos que não envolvam animais. A questão ética de uso tanto de animais e humanos levou a criação de comitês de ética tanto para discutir o assunto como para supervisionar estudos realizados em humanos como também em animais. No Brasil o Comitê de Ética em Pesquisa (CONEP) lida com estas questões e leis na área. Para maiores informações acesse o site: (http://conselho.saúde.gov.br/web_comissoes/conep/index.html).

Embora o ideal fosse realizar experimentos em humanos, como por exemplo, inocular o virus da rubéola em mulheres grávidas para se observar as consequências, um estudo deste tipo não seria nada ético. Não sendo possível realizar experimentos em

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

humanos nos resta apenas observá-los! Somente **podemos** testar em humanos, em medidas preventivas ou em tratamentos para uma determinada doença, jamais podemos causar qualquer mal a humanos. Esta noção de não poder causar qualquer mal a humanos em estudos científicos, no entanto, é relativamente novo. Lamentavelmente, encontramos, no passado, histórias constrangedoras de estudos antieticos como o estudo de Tuskegee (http://www.cdc.gov/tuskegee/timeline.htm) nos Estados Unidos. Este estudo tinha o objetivo de seguir o desenvolvimento da sífilis em humanos quando o tratamento para a doença ainda não existia. No entanto, enquanto o estudo ainda estava sendo realizado descobriu-se o tratamento e este foi oferecido apenas aos indivíduos brancos, deixando os negros sem tratamento para que pudessem observar a evolução da doença. Uma evidência extrema de racismo e despreso pelas minorias!

Na odontologia também temos exemplos de estudos antiéticos que hoje em dia jamais seriam realizados como o estudo chamado História Natural das Doenças Periodontais. Neste estudo pesquisadores examinaram plantadores de chá do Sri Lanka e acompanharam a evolução da doença por mais de 40 anos sem oferecer tratamento a nenhum dos participantes. Hoje em dia toda doença diagnosticada num estudo deve ser tratada ou pelo menos o indivíduo deve ser encaminhado para o tratamento. Outro estudo importante na Odontologia é o Estudo de Vipholm em que se testou se a frequência de ingestão de açúcar era mais importante do que a quantidade ingerida para origem de cárie dentária. O estudo foi realizado em indivíduos com deficiência mental.

Assim, a utilização de indivíduos vulneráveis, como pobres e deficientes mentais era comun no passado, mas inadimissível nos dias de hoje. Estas atrocidades levaram a criação de comites de ética em todas as instituições de pesquisa para avaliar o potencial de lesão aos indivíduos participantes de pesquisas.

Voltando às alternativas para o estudo do Gregg, como explicado anteriormente, os estudos em ratinhos poderiam ajudar, mas mesmo assim, estudos em humanos deveriam ser realizados. No entanto, se não podemos realizar experimento no qual inoculamos a rubéola, para ter certeza de que a rubéola não está associada a outro evento, como radiação, o que podemos fazer? Bom, em humanos o estudo de fatores que podem causar doenças ou maleficios somente pode ser realizado observando quem

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

já tem o fator, no caso, a rubéola. Estes estudos em que não inoculamos o fator causal a ser estudado são denominados de **estudos observacionais**. Em contraposição os estudos em que vamos inocular o fator causal são chamados de **experimentos** (veremos posteriormente a definição precisa de experimento).

Os estudos em animais, quando puderem ser realizados, são interessantes principalmente para nos ajudar a entender o mecanismo de causa de um determinado fator. Quando nos deparamos com algo novo, como a proposta do Gregg de que a rubéola era a causa da catarata congênita, a primeira pergunta ser feita deveria ser, mas será que tem sentido, isto é, é plausível que a rubéola afete a formação dos olhos levando a catarata? Esta primeira evidência de **plausibilidade biológica** é muito importante para nos convencer da associação entre rubéola e catarata congênita. O que não é plausível é difícil de ser aceito. Por exemplo, posso fazer um estudo numa sala de aula para ver o que pode estar associado a asma, e a única informação comum a todos os asmaticos é o fato de todos eles estarem utilizando meia de cor amarela. Assim, eu concluo que meia de cor amarela está provocando asma. Ou seriam os asmáticos que tem preferência pelo amarelo? Parece loucura esta associação e, portanto não vamos dar muito valor para ela e esquecer o assunto. No entanto, pode ser que ainda não entendemos o porquê que a meia de cor amarela pode afetar as pessoas e provocar a ama. Plausibilidade é importante, mas podemos encontrar associações que no momento não conseguimos entender o mecanismo de ligação entre causa e efeito.

Um exemplo classico, é a associação entre estresse e doenças em geral que por muito tempo não foi aceita por não se conseguir explicar o mecanismo de ação. Hoje em dia sabe-se bem que estresse altera o sistema imunológico. Essa evidência de alterar o sistema imunológico apenas veio com a evolução de técnicas e aparelhos. Isso nos leva a refletir que a ciência que fazemos está em constante evolução, e temos num determinado momentos sofremos com as limitações de observação (aparelhos e tecnicas) que temos. Ainda podemos concluir que o que parece implausível num momento pode se tornar plausível em outro e vice-versa. Bom, isso é algo inerente a ciência que nós, dentistas, médicos, físicos, engenheiros e químicos fazemos. Nossa ciência é chamada CIÊNCIA EMPÍRICA que significa “ciência baseada em observação”.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Apenas temos ciência não empírica na matemática e na lógica que utilizam da dedução, dedução de fórmulas etc.

Agora você deve estar lembrando-se de ouvir pessoas, até mesmo na universidade, dizendo que uma associação é empírica se referindo a algo que não seria comprovado cientificamente. Por outro lado, está lendo aqui que a ciência que fazemos é empírica. Pois bem, na Universidade de maneira formal apartir de agora o correto é que a ciência que fazemos é chamada ciência empírica, e quando dizemos “nao existe evidência empírica que esta associação seja verdadeira” singifica dizer que “nao existe evidência científica”. Vamos deixar a palavra empírico como sendo “pelo achismo” de alguém para a televisão e jornais. È importante entendermos que a ciência que fazemos é sim empírica (baseada em observações) sejam elas em experimentos (onde temos controle de tudo) ou em estudos observacionais. Mas mesmo a medida de um aparelho que conta leucócitos é algo empírico? Sim, é empírico porque o aparelho foi desenvolvido por humanos por meio do conhecimento gerado por observação. O fato de a nossa ciência ser empírica é muito importante para deixar nossa mente aberta ao fato de que o que observamos hoje pode não ser verdade amanha.

Outro exemplo recente de conflito entre plausibilidade biológica e o fato de nossa ciência ser empírica é a história do descobrimento do H. Pylori (bactéria associada à gastrite e úlcera de estômago). Até algumas décadas atrás se acreditava que nenhuma bactéria conseguia se instalar e colonizar o estômago porque este era muito ácido. Todas as tentativas de se cultivar alguma bactéria de material coletado do estômago sempre dava negativo. Isso acontecia por causa das técnicas precárias. Quando um cientista, Barry Marshall, começou a levantar evidências de que o H. Pylori poderia estar associado à úlcera ninguém conseguiu acreditar. Além do mais o cientísta fez como Gregg, ele isolou a bactéria de 80% de seus pacientes com gastrite/úlcera e concluiu que a associação era causal. Da mesma forma que aconteceu com Gregg o cientista, que por coincidência também era Australiano, estava certo, mas muito tempo se passou até que as pessoas acreditassem o que dizia. Primeiro, porque o estudo que fizera era apenas um relato de casos, e outra que ninguém acreditava que seria possível uma bactéria colonizar o estômago. Barry Marshall tentou infectar ratinhos, mas estes não desenvolveram úlcera. Desta forma, para mostrar que estava correto, ele fez um

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

experimento com ele mesmo, bebendo uma emulsão de *H. Pylori*. Assim que ele desenvolve gastrite , tomou atibióticos, eliminou o *H. Pylori* curando sua gastrite.

O que podemos concluir com estes exemplos é que plausibilidade biológica é algo importante, mas às vezes não a conhecemos porque não temos meios de verificá-la. Claro que precisavamos de ter evidência empírica de que o H. Pylori provocava gastrite para prescrever antibióticos para seu tratamento na prática clínica, mas talvez os cientistas pudessem ter sido mais pacientes com Barry Marshall. No entanto, o Marshall poderia ter feito um estudo melhor para convencer seus colegas!

Voltando ao nosso assunto de rubéola e catarata congênita, outra sugestão comum entre os alunos é que deveríamos seguir mães com rubéola até o nascimento da criança para ver se realmente rubéola levaria a catarata congênita. Vamos tentar entender o que este seguimento nos daria de evidência científica (ou evidência empírica)

Mães com Rubéola ⟶ % com Catarata Congênita

Suponhamos que temos 100 mães com Rubéola, e ao final, 80 crianças nasçam com catarata congênita. O que podemos concluir sobre isso? Muitos acharão que este resultado demonstraria definitivamente que a Catarata seria decorrência da Rubéola ao passo que outros acharão que não, pois nem todas as crianças nasceram com catarata. Novamente, estamos numa situação muito semelhante às dúvidas quanto à afirmação de Gregg. Lembre-se de que não necessáriamente todos têm que desenvolver uma dõença. E ainda, o que falta novamente é um grupo controle para servir de comparação. Se observarmos apenas as mães com rubéola, não saberemos se a proporção de casos de catarata entre as mães que não tiveram rubéola seria o mesmo.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Mães com Rubéola ⟶ % com Catarata Congênita

Mães sem Rubéola ⟶ % com Catarata Congênita

Fig.2

Assim, precisamos de um grupo controle de mães sem rubéola para comparar com as mães com rubéola. Outra observação que pode ser levantada é sobre a comparabilidade deste grupo controle de mães sem rubéola. Vamos supor que coletamos dados de mães com rubéola de uma cidade e mães sem rubéola de outra cidade. Pode até ser que as mães sejam comparáveis, mas pode ser também que não. Vamos supor que na cidade das mães sem rubéola estas também são expostas a radiação constante que vaza de uma usina atômica e por isso elas também tem elevada ocorrência de catarata congênita. Se compararmos os dois gupos propostos chegaremos a conclusão de que a rubéola não é responsável pela catarata congênita. Esta conclusão seria incorreta, porque o que temos de errado, na verdade é, a falta de comparabilidade do grupo controle. Portanto, **a comparabilidade do grupo controle é fundamental na montagem do estudo.**

Até este ponto você deve ter aprendido que (1) a ciência pode ser empírica e não empírica, a não empírica inclui somente a lógica e matemática, o restante, portanto é empírica; (2) Quando se diz que algum fator leva a alguma doença ou evento, antes de acreditar e utilizar a informação, nós devemos verificar se existe alguma evidência de plausibilidade biológica (além de outros critérios de causalidade que veremos posteriormente); (3) Existem duas maneiras gerais de se realizar estudos por meio de experimentos (vamos ver definição precisa depois) onde o pesquisador inocula o fator causal a ser estudado ou se não for possível inocular o fator este deve ser observado na população. Assim, temos de forma geral estudos chamados experimentos e estudos observacionais.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

De forma intuitiva concluímos que o estudo de Gregg possuia um grande problema que era a falta de um grupo de crianças sem catarata (grupo controle). Na discussão foram propostos algums tipos de estudos. Num deles se testaria inoculando rubéola em grávidas (que seria antiético) e comparando-as a um grupo de grávidas em que não inoculariamos rubéola, este seria chamado experimento. O experimento, no entanto não seria possível em humanos apenas em animais por questões éticas. Uma vez que o experimento seria antiético, nos sobraria apenas observar (estudos observacionais) a população, e esta observação, para ser eficiente, deverá ser de forma sistemática (com critérios). Complementando o estudo do Gregg, para se tornar sistemático, poderiamos adicionar o grupo controle de crianças sem catarata. Outra forma de orbservar sistematicamente seria organizar um grupo de grávidas que se contaminaram com rubéola e outro sem rubéola e acompanha-las durante a gravidez até o nascimento das crianças.

Atribuindo nomes apropriados ao que acabamos de discutir

Boa parte do que fazemos na escola, na faculdade e também na vida é aprender a dar nomes às coisas. Fazemos isso para podermos nos comúnicar com mais eficiência.

O que estudamos em ciência basicamente é o fenomeno de causalidade, isto é sempre estamos perguntando se um evento leva a outro evento. Chamamos isso de estudo de causalidade, ou causação, ou ainda estudo de **causa e efeito**. Na epidemiologia adotamos a expressão **exposição** ao invés de causa e desfecho (outcome em inglês) ao invés de efeito. O nome exposição pode parecer estranho a princípio, pois pensamos em exposição ao sol, ou algo que esteja um pouco longe, mas aqui usamos este termo para expressar toda causa estudada. Dizemos que as mães que efetivamente tiveram rubéola são expostas a rubéola e não aquelas que tiveram contato com uma pessoa com rubéola. Exposto aqui não é que teve contato com alguém que teve rubéola, mas é efetivamente ser exposta aos efeitos da doença. Se estudarmos a saúde de um fumante, então o fumo será tambem chamado de exposição. Num estudo sobre um gene e uma doença, o gene é também a exposição. Da mesma forma, se estivermos

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

estudando o efeito da ingestão de refrigerantes na obesidade, o consumo de refrigerantes será chamado de exposição ao refrigerante.

O resultado da exposição chama-se **desfecho**. Logo, catarata congênita seria o desfecho da rubéola durante a gravidez. Considerando um estudo que investiga se o consumo de açúcar levaria a cárie dentária, a exposição seria o açúcar e o desfecho a cárie dentária. Num estudo sobre diabetes como possível causador de problemas renais, a exposição seria a diabetes e o problema renal o desfecho. Nunca decore que exposição é **causa** e desfecho **é doença**, pois você pode-se confundir, porque às vezes uma doença é estudada causando outra doença. Por exemplo, Síndrome de Down pode ser exposição para maior ocorrência de doença periodontal e doença periodontal pode aparecer num estudo como exposição para doença cardiovascular. E ainda uma doença pode levar a um comportamento a ser estudado, e assim a exposição seria uma doença e o comportamento a exposição. Pense sempre em causa e efeito.

Causa ⟶ Efeito

=

Exposição ⟶ Desfecho.

As expressões exposição e desfecho por vezes podem receber o nome de ***variável independente*** e ***variável dependente*** respectivamente. Esta terminologia é mais utilizada na estatística, mas é bom que já seja introduzida aqui para que fique familiarizado. O desfecho é chamado de **variável dependente** porque ele depende de algo no caso a exposição. A exposição (causa) então é chamada de **variável independente** porque ela existe independentemente do desfecho que irá acontecer. O termo variável é dado uma unidade que é observada e que pode assumir mais de um valor (ou condição), caso contrário seria uma constante. No nosso exemplo, a exposição rubéola na verdade se refere a variável que pode ser do tipo sim ou não (ter ou não ter a rubéola). Quanto a variável catarata seria o mesmo (ter ou nao ter catarata).

Observando a forma como o estudo apresentado na fig. 2 foi montado, podemos identificar muitas semelhanças com o estudo dos ratinhos (fig. 1.). A diferença, é que com os ratinhos o pesquisador começa com um grupo de ratinhas no início da gravidez

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

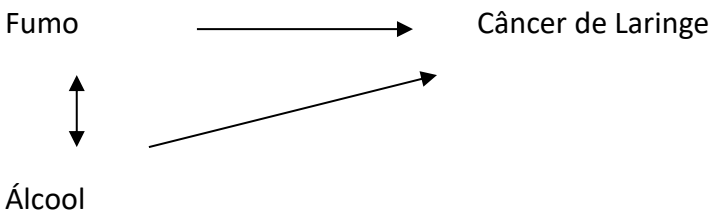
e assim será o responsável por expor às ratinhas a rubéola. Chamamos isso **de intervenção**, isto é o pesquisador intervem expondo (no caso contaminando) parte das ratinhas a rubéola. Assim, o pesquisador terá certeza de quando e como as ratinhas foram infectadas com rubéola, e ainda terá a possibilidade de seguir as ratinhas até que os bebês nasçam. Na figura 2 temos o seguimento de mães que já possuem rubéola na população, portanto o pesquisador não foi responsável pela intervenção. Note que uma diferença marcante é que na figura 1, o pesquisador primeiro reúne um grupo de ratinhas grávidas e ele intervem inuculando rubéola num grupo e deixando o outro livre. Já no caso do seguimento das mães com rubéola, temos que identificar estas mães com e sem rubéola na população. Portanto, neste caso não houve **intervenção,** o pesquisador apenas **observa e** separa as grávidas com e sem rubéola em grupos diferentes e por isso chamamos este tipo de estudo de **estudo observacional**. Especificamente chamamos este tipo de estudo observacional em que temos dois grupos de expostos e não expostos que serão acompanhados até o desenvolvimento do desfecho (que no caso é catarata congênita) de **estudo de coorte**. O nome pode parecer estranho, mas é coorte (com dois oo) mesmo. Note que no estudo observacional o pesquisador não faz nenhuma intervenção, isto é ele não aplica a exposição. Podemos então dizer que a exposição aconteceu de forma natural na população. Por ser a exposição de forma natural ela pode estar associados à varios fatores que podem também estar associados ao desfecho.

Vamos imaginar um estudo de coorte para estudar o efeito do fumo na ocorrência de câncer de laringe. Começamos separando um grupo de mil homens fumantes aos 35 anos de idade e comparamos com outro grupo de não fumantes de 35 anos, e os seguimos até os 70 anos. No entanto, sabemos que beber álcool também está associado ao aumetno de câncer de laringe. E ainda sabemos que quem fuma tende a beber mais do que quem não fuma. Logo, como vou saber se o câncer desenvolvido foi devido ao fumo ou álcool? Numericamente vamos verificar que entre os fumantes existem proporcionalmente muito mais indivíduos que ingerem álcool do que entre os não-fumantes. Se o número fosse igual, isso também signficaria que quem fuma não é mais provável de beber. Se pudessemos fazer um experimento sorteando pessoas para serem expostas ao fumo e outro para não serem expostas, poderiamos controlar a

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

quantidade de pessoas que bebem álcool em cada grupo, mas se escolhemos estas pessoas da população em geral isso não é possível.

Este fator externo à associação principal de estudo (fumo e cancer de laringe) é chamado de **fator de confusão**. Este fator de confusão é assim chamado, pois pode confundir (atrapalhar) a visualização correta da associação. Vamos supor que ao final de 30 anos de acompanhamento dos mil homens expostos ao fumo, 150 deles tenham desenvolvido câncer de laringe, e entre os não-fumantes apenas 50. Logo temos que os fumantes desenvolveram 3 vezes mais câncer do que os não fumantes. No entanto, uma vez que grande parte dos fumantes também bebe, a associação correta entre fumantes e não fumantes deve ser menor do que 3 vezes. Portanto, não é verdadeiro que "3" seja o valor correto. Se tivessmos num experimento em que houve alocação das pessoas para se tornarem fumantes e não fumantes, ai sim, poderiamos ter controlado a quantidade de pessoas que bebem álcool em cada grupo (imaginando que humanos iriam obedecer corretamente às ordens de fumar e de beber ou não álcool ao longo de 30 anos). Óbvio que eticamente este experimento em humanos não poderia ser realizado, mas se possível diriamos que equilibrando o número de consumidores de álcool nós estariamos controlando os **fatores de confusão**. Assim, no experimento, é possível controlar os fatores de confusão ao se distribuir igualmente os indivíduos que bebem nos grupos que sofreram intervenção, no entanto, no estudo observacional isso não é possível neste momento do estudo. Os fatore de confusão devem ser controlados nos estudos observacionais posteriomente durante a análise estatística.



Uma forma prática de equalizar os fatores de confusão no experimento é atribuir a exposição de forma aleatória. Aleatório significa sorteado ao acaso, o que em geral é feito por meio de sorteio usando um programa de computador, ou mesmo usando bolinhas de bingo/loteria. Há necessidade de ser um processo aleatório (sorteio mesmo)

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

correto. É preciso ficar claro o que **é alocação aleatória**. Vamos supor que queremos testar duas pastas de dentes diferentes (A e B), e vamos realizar o experimento numa escola. Chegando a escola reunimos todos os alunos no pátio e pedimos para que formem uma fila única. Ai, vamos distribuindo uma pasta A para um aluno e uma pasta B para o seguinte na fila. Para muitos pode parecer um processo aleatório, mas não é, porque as crianças podem ter tendências diferentes ao formar a fila. Pode ser que crianças formem a fila única obedecendo a algum critério como amizade, crianças mais pobres e ricas, e no final não teremos dois grupos homogeneos de crianças. Para ser aleatório, é preciso dar um número a cada criança e, ai sim, sortear metade dos indivíduos para o grupo da pasta A e o restante para a pasta b. Este sorteio deve usar tabela de números aleatórios ou um computador. Usar papelzinho para sorteio, não da certo, porque os papeizinhos podem ser dobrados de formas diferentes, e terem pesos diferentes de quantidade de tinta. Fechar o olho e ir colocando uma criança para um lado e outra para outro também não torna o processo aleatório, para ser aleatório (ou seja, ao acaso) tem que ser com sorteio.

Chamamos **de experimento somente** os estudos que tem **intervenção** e que esta intervenção seja **alocada aleatóriamente** entre os indivíduos. **Se não houver alocação aleatória da intervenção não é um experimento!** Com a alocação aleatória, nos podemos durante o desenho do estudo (montagem do estudo) controlar os fatores de confusão. Este controle não é possível para os estudos observacionais nos quais as pessoas não são alocadas a comportamentos, mas estes já existem e, portanto, podem estar atrelados a outros comportamentos e características que podem estar ligadas diretamente ao desfecho estudado. No entanto, os efeitos dos fatores de confusão podem ser controlados posteriormente na análise estatística que veremos posteriormente. No entanto, é obvio que o controle feito na aleatorização de um experimento é muito mais eficaz.

Você deve estar confuso porque as pessoas de laboratório e em geral chamam tudo de experimento, no entanto, aqui vamos chamar de experimento sempre que tivermos alocação aleatória da exposição. Na epidemiologia dos ultimos anos este temos experimento também passou a não ser utilizado, especialmente depois que Rothman e colaboradores no livro Modern Epidemiology, se posiciona contra o termo

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

experimento utilizado como estudo que contem intervenção alocada aleatoriamente. O argumento de Rothman é que a palavra Experimento sempre existiu e a aleatorização foi introduzida por Fisher em 1925. Meu argumento no entanto para utilizar o termo experimento é que antes de Fisher propor a aleatorização, os pesquisadores achavam que fatores de confusão não existiam e interpretavam os resultados dos estudos com intervenção como se os fatores de confusão não existissem. Sendo contra o termo experimento, Rothman atribui outros termos diferentes para cada detalhe de um experimento como vamos descrever agora superficialmente. Para Rothman o experimento com alocação aleatória em humanos pacientes (com alguma doença) deve ser chamado de "***Ensaio Clinico Aleatório (ou Randomizado) Controlado"*** portanto, este tipo de estudo deve se restringir a tratamentos de uma doença ou recuperação de sequelas. Quando o investigador quiser testar uma medida preventiva em pacientes sadios deve chamar o estudo de **Ensaio de Campo** (Field Trial) mesmo que este seja realizado com indivíduos atendidos numa clínica ou escola. Se o estudo de uma intervenção for realizado em um grupo de indivíduos pertencentes a comunidades ou cidades que serão alocadas a intervenção então o estudo é chamado de **Ensaio de Intervenção Comunitária Aleatória** (ex: cidades com e sem fluoretação de água). Lembrando que apenas podemos pensar nessa situação em exposição que seja medida preventiva. Se o ensaio for realizado com grupos, mas não toda a cidade então é chamada **de Ensaio de Conglomerado Aleatório**. Uso aqui o termo aleatório ao invés do popular "randomizado" porque temos a palavra em português e portanto não temos motivo para usar randomizado, apesar de ser encontrado nos dicionários da lingua portuguesa.

Bom, no fundo, a base de todos esses ensaios é o experimento que deve ser definido pela alocação aleatória da intervenção, independente se em humanos, corpos de prova, ratos, em comunidades, em humanos doentes ou saudáveis. Os princípios de um experimento são os mesmos (intervenção e alocação aleatória da intervenção), tendo é claro, particularidades se você faz uma intervenção em comunidades, cidades, ou famílias. Imagine agora se alguem trabalha com animais, não pode falar de ensaio clínico, e vai chamar o experimento em animais de que? Bom, eu acho muito confusa a nomeclatura que Rothman tem utilizado, e na minha opnião os alunos passam a

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

confundir tudo. Porém, lembre-se que os livros especialmente os mais recentes aderiram a essa nomeclatura de Rothman et al.

Para simplificar aqui vamos estabelecer que experimento é apenas o estudo que tem intervenção alocada aleatóriamente, e se a interevenção não for aleatória ou se a intervenção estudada não foi promovida pelos pesquisdores (implantação de leis em determinados estados ou cidades) então chamaremos de **quasi-experimentos**. Este termo quasi-experimento foi cunhado dois pesquisadores muito importantes Julian C. Stanley e o famoso Donalt T. Campbell e falaremos sobre quasi-experimentos mais adiante.

Vamos a mais um exemplo para entender bem o significado da alocação aleatória. Ao fazer um estudo de ratinhos, em geral o pesquisador pede ao biotério que mandem ratinhos semelhantes e com mesmo peso e idade. Muitos pesquisadores somente pegam estes ratinhos e separam para os grupos que sofreram intervenção e controle sem nenhum critério. Acontece que embora pareçam iguais os ratinhos não são todos identicos, e nem vieram de uma mesma mãe. È importante, portanto, dar um número aos ratinhos e sorteá-los. Vamos supor que cheguem 30 ratinhos no laboratório, sendo que a cada 10 vindo de uma mãe ratinha diferente. Uma das mães era um pouco diferente com sistema imunológico mais fraco. Ao separar em dois grupos para o experimento, vamos supor que 10 destes ratinhos da mãe com sistema imunológico deficiente caiam num mesmo grupo. Assim, um grupo tem 10 ratinhos imunologicamente deficiente e 5 normais, enquanto o outro grupo vai ter 15 ratinhos imunologicamente bem. Se o grupo com mais ratinhos imunologicamente comprometidos recebem uma nova droga para cicatrização que estaria sendo comparada com uma droga antiga já conhecida, mesmo sendo a nova droga melhor ela ou vai se apresentar pior ou será igual à antiga. Este é um exemplo do possível efeito de um fator de confusão quando este não é controlado. Por isso, é primordial, essencial a alocação aleatória.

Porém, você pode perguntar se sempre que fizer a alocação aleatória resolveremos o problema de fatores de confusão? A resposta é que não é 100% garantido, porque durante o processo aleatório, ainda assim, os grupos podem ficar desbalanceados. Desta forma, se eu soubesse a origem das ratinhas seria bom checar

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

depois da aleatorização se os números ficaram balanceados. Outra opção seria que sabendo da importância da origem dos ratinhos poderiamos fazer o sorteio estratificado. Isto é pegaria 10 ratinhos da primeira mãe e sortearia 5 para cada grupo, em seguida os outros 10 ratinhos da segunda mãe e da terceira mãe.

Outro fator que afeta o resultado da alocação aleatória é o número de indivíduos que estão sendo sorteados. Por exemplo, se temos 10 indivíduos sendo 2 homens e 8 mulheres e sortearmos dois grupos de 5 indivíduos, a probabilidade dos 2 homens cairem num grupo será grande. Se, no entanto o sorteio for realizado com a mesma proporção de homens (20%), porém em 100 indivíduos, a probabilidade de que os 20 homens serão sorteados para o mesmo grupo será bem pequena. Assim, quanto maior o número de participantes num sorteio melhor deve ser a distribuição dos fatores de confusão.

Além da alocação aleatória os fatores de confusão podem ser controlados limitando a participação de um grupo no estudo. Isso pode ser realizado tanto no experimento como nos estudos observacionais. Por exemplo, se sexo pode ser um fator de confusão, posso restringir o estudo a apenas homens ou apenas mulheres. Se idade também é um fator importante, posso restringir o estudo a apenas um grupo de idade. Note, porém que se eu excluo um dos sexos, eu não posso estrapolar meus resultados para todos os sexos. Portanto, limitando uma caracteristica do meu estudo eu restrinjo o estudo àqueles indivíduos que participaram do mesmo. Outro detalhe importante é que posso pensar então que num estudo posso restringir para todos os fatores de confusão, mas ai será difícil encontrar os participantes que queremos.

Desde o último quadro até o parágrafo acima adicionamos alguns termos e conhecimentos: (1) causa-efeito; (2) exposição-desfecho; (3) variável dependente e independente; (4) experimento (intervenção e alocação aleatória); (5) importância da alocação aleatória na diluição de fatores de confusão; (6) importância do número de indivíduos na diluição dos fatores de confusão; (7) estudos observacionais.

Agora você deve entender o que é o fator de confusão, o que é um experimento e uma coorte, e o que significa e como deve ser a alocação aleatória. Deve entender bem a diferença entre estudos observacionais e experimentos, sendo que até o momento falamos apenas do estudo de coorte.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Demos nome até agora apenas ao estudo observacional do tipo Coorte. Havíamos comentado que o estudo do Gregg seria melhorado se tivesse um grupo controle de crianças sem catarata congênita. Desta forma, se tivesse um grupo controle, o estudo do Gregg que é considerado apenas um **relato de casos,** seria chamado de **estudo de caso-controle.** O que caracteriza o estudo de caso-controle e o diferencia do estudo de coorte, é que o estudo começa pela seleção de pessoas que tem o desfecho a ser estudado. Na coorte o estudo começa pela seleção de expostos e não expostos, isto é o possível fator causal a ser estudado. Já no caso-controle, o estudo começa pela seleção de pessoas com o desfecho **(casos**) que devem ser comparados com um grupo chamado **controle** formado por pessoas que não tem o desfecho. Na coorte acompanhamos os expostos e não expostos até o desenvolvimento do desfecho, enquanto no caso-controle comparamos casos e controles e não os acompanhamos porque o desfecho já aconteceu, e, portanto, apenas podemos coletar informações sobre a exposição que nos interessa. Fundamental é que o estudo de coorte tem característica **prospectiva** e o de caso-controle **restrospectiva**. Prospectivo significa que se acompanha (o pesquisador acompanha) o desenvolvimento da causa até o desfecho, e restrospectivo signfica que tanto a exposição como o desfecho aconteceu no passado e o pesquisador não foi capaz de acompanhar esta transição de exposto sem desfecho até o desfecho. Estas são as caracterésticas e diferenças básicas entre coorte e caso-controle. Existem muitos detalhes sobre estes dois tipos de estudos que comentaremos posteriomente, também existem modificações destes tipos de estudos básicos, mas por enquanto o importante é diferenciá-los. Apenas mais um comentário importante, no estudo de coorte, ao montar os grupos de expostos e não expostos é importante ter a certeza de que não existem pessoas com o desfecho em nenhum dos grupos. Por exemplo, se vamos realizar um estudo de coorte em adultos de 30-35 anos de idade expostos e não expostos ao fumo para verificar a incidência de câncer de pulmão, temos que ter certeza de que nenhum deles já tem cancer de pulmão, devemos então realizar algum exame em todos os participantes no inicio, para certificar que nenhum tem diagnóstico de câncer. Óbvio que sempre teremos a possibilidade de ter feito o exame e o câncer estar em estágios tao iniciais que nenhum exame conseguiu detectar. Porém tecnicamente ninguém pode ter o desfecho ao se iniciar um estudo de coorte.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Chegou a hora de voltarmos a discussão dos quasi-experimentos. Anteriormente dissemos que se num estudo tiver exposição, mas esta não for aleatória não termos um experimento verdadeiro mas um **quasi-experimento**. Ressaltamos também que Rothman et al não concordam com o termo experimento e consequentemente com quasi-experimentos. No passado epidemiologistas utilizavam as terminologias de experimento e quasi-experimento, mas o modismo de Rothman dominou a área concordando com os termos estudos clinicos randomizados que os clínicos vinham utilizando. A terminologia de quasi-experimentos foi elaborada por Campbell & Stantley (Experimental and Quasi-experimental designs for Reseach) em 1963. Na justificativa, de Campbell & Stanley ressaltam que a definição de experimento por Fisher 1925 (intervenção e aleatorização) não se aplicava a várias situações quando é impossível realizar aleatorização. Apesar da falta de aleatorização em algumas circunstâncias existe a possibilidade de se analisar estes estudos com cautela e apartir dos mesmos tirar conclusões úteis. Na área de saúde muito frequentemente nos deparamos com tais situações. Por exemplo, num estudo de tratamento de mordida cruzada em que o objetivo é verificar se a ultilização de aparelho para descruzar a mordida afeta o crescimento fascial, e aleatorização do tratamento não pode ser feita. Se a criança tem mordida cruzada ela não pode servir de controle sem nenhum tratamento. Esta pergunta não pode ser respondida a não ser por um quasi-experimento, isto é um grupo terá intervenção, sendo observado antes e depois do tratamento e durante todo o crescimento fascial. O grupo controle deverá ser um grupo de crianças sem mordida cruzada em que se espera um crescimento fascial normal. Este quasi-experimento tem o nome de **quasi-experimento do tipo antes e depois com controle externo.** Note que uma vez que não houve alocação aleatória, eu não sei como os fatores de confusão estão agindo. Eu como pesquisadora posso limitar os pacientes do grupo experimental quanto a algumas características como sexo, idade, etc que poderiam ser possíveis fatores de confusão, no entanto, muitos estarão presentes e, portanto, tenho **que levá-los em consideração na interpretação dos resultados**. Quando for possível em alguns tipos de quasi-experimentos posso tentar ajustar os fatores de confusão na analise estatística, mas é bem complicado.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Outro exemplo de quase experimento clássico é o primeiro estudo para avaliação dos efeitos da fluoretação de águas em Grand Rapids e Musquegon. A realização deste primeiro estudo foi uma oportunidade, porque nenhuma cidade queria fazer parte do estudo. Assim, alguma conveniência politica existiu que estas duas cidades aceitaram participar do mesmo. Embora fossem apenas duas cidades os pesquisadores levaram o estudo à frente. Uma das cidades ficou sem agua fluoretada (Musgkegon) e a outra recebeu a mesma (Grand Rapids). As duas cidades tinham originalmente número de crianças, níveis de cárie dentária e características socioeconomicas semelhantes. No entanto, como eram apenas duas cidades, o processo de alocação aleatória não funcionou, mesmo que se jogue uma moeda para decidir quem receberia a fluoretação. Assim, como houve intervenção, mas não houve alocação aleatória, este estudo é na verdade um **quasi-experimento com comparação antes e depois**. Em praticamente todos os livros, este estudo é chamado de experimento de campo, no entanto, não é um experimento de campo é um **quasi-experimento com comparação antes e depois.** Existem vários tipos de quasi-experimentos, sempre que se tiver dúvida do nome do quasi-experimento existe um livro chamdo Quasi-Experimental Design que é livro de referência.

Se você estiver refletindo sobre os estudos até agora mencionados deve estar meio confuso sobre a diferença entre o quasi-experimento e coorte. Por exemplo, se um pesquisador segue indivíduos que foram submetidos por ele ao tratamento para descruzar mordida e os comparam indivíduos que não foram submetidos ao tratamento, porque não chamá-lo de coorte de expostos e não expostos ao tratamento? A razão de considerar este estudo um quasi-experimento está no fato do pesquisador ter feito a intervenção (tratamento para descruzar mordidas). Se o pesquisador selecionasse crianças que já estivessem com o aparelho numa cidade e comparasse o grupo com outras crianças sem aparelho ai eu poderia chamar de coorte de expostos ao aparelho e não expostos.

Existem autores que ignoram o termo quasi-experimento e consideram como coorte em que houve intervenção. Eu prefiro manter o termo quasi-experimento que caracteriza melhor as limitações do estudo e ainda conseguimos dialogar com outras

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

áreas em que a terminologia é sempre empregada como economia e ciência politica entre outras.

Nestes últimos parágrafos acrescentamos e definição de coorte de caso-controle, e aprendemos a diferenciá-los. Ainda, vimos o que vem a ser quasi-experimento. Até agora foram apresentados o experimento, e os estudos observacionais que incluem relatos de caso, coorte, caso-controle. Além disso, o quasi experimento.
Resumo dos estudos (até o momento) Experimentos Quasi-experimentos Observacionais - relatos de caso - coorte - caso-controle

Cuidado! Apesar das definições dos tipos de estudos parecerem simples, a confusão é grande principalmente quando se trata do estudo de caso-controle! Os estudos têm dois momentos. O primeiro momento os participantes são selecionados e é neste momento que os nomes dos mesmos são definidos. O segundo momento é quando informações são coletadas. Por exemplo, para o caso controle a seleção deve ser de casos e controles, o segundo momento é de coletar informações sobre exposição. Começa-se o estudo sem saber se os indivíduos são expostos ou não, não interessa para a seleção de casos e controles se estes são ou não expostos. A coletad e informação faz parte da segunda fase.

Da mesma forma, para a coorte o primeiro momento é a seleção de indivíduos expostos e outro grupo de não expostos. Nenhum indivíduo pode ter o desfecho que se quer estudar. Num segundo momento, acompanham-se os grupos para coletar informações sobre incidência de desfecho.

No experimento, a seleção inclui um grupo homogênio num primeiro momento que num segundo momento vai ser alocado de forma aleatória a grupos experimentais e controles. O que acontece neste segundo momento é a intervenção de forma aleatória.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

A palavra **controle** dos estudos de caso-controle e a utilzação da mesma palavra nos experimentos (grupo controle) faz com que as pessoas confundam o que é caso-controle. Muitos acham que um experimento pode ser chamado de caso-controle, porque o experimento tem um **controle,** mas devemos lembrar que caso-controle é observacional e não tem intervenção. Além disso, o que define os estudos é a seleção dos participantes. O caso-controle é um estudo observacional, enquando o experimento não o é (os indivíduos sao alacocados para a exposição). No caso-controle comparam-se pessoas que tiveram o desfecho com pessoas que não tiveram o desfecho. Você poderia até substituir o nomed e caso-controle para "desfecho-não desfecho". Se tiver com dúvida se o estudo pode ser chamado de caso-controle veja se poderia ser chamado de desfecho-controle.

Mais um exemplo para ficar mais claro. Num protocolo de estudo um pesquisador com a intensão de verificar se o padrão de citocinas associado à doença periodontal crônica seria influenciado pela AIDS definiu seu estudo como sendo um caso-controle. Vamos analisar se ele denominou adequadamente seu estudo. Para tanto, o pesquisador relata que um grupo de pacientes com AIDS e doença periodontal crônica seria comparado a um grupo controle de pacientes com doença periodontal crônica, mas sem HIV ou AIDS. Por ter um grupo que ele chamou de controle, logo ele concluiu ser um caso-controle. No entanto, o caso-controle deve ser montado com desfecho-controle. Neste estudo a doença periodontal é comum a todos, o HIV é a exposição, o desfecho é representado pelos níveis de citocina. Para ser um caso controle, pacientes com altos níveis de citocina (casos) deveriam ser comparados com pacientes com baixos níveis de citocinas e ai sim verificar se teriam AIDS ou não (exposição). Logo este estudo proposto não era um caso-controle e sim um estudo transversal comparando expostos e não expostos a AIDS.

Mais sobre fator de confusão e fator modificador (ou de interação) – as terceiras variáveis.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Voltando ao fator de confusão, ele é sempre um problema que devemos refletir, analisar, estudar e interpretar sua ação em qualquer relação de causa e efeito ou seja exposição e desfecho. Devemos ser claros que sempre que testamos uma associação entre exposição e desfecho, nós queremos saber se eles estão associados *independente da interferência de qualquer fator de confusão.* Por exemplo, se temos a hipótese de que contaminação por chumbo está associada à cárie dentária, queremos na verdade testar se "chumbo está associado à cárie dentária independente de qualquer fator de confusão". Para estudar esta associação precisamos ir numa população de expostos ao chumbo e compara-los com não expostos, mas isso não é suficiente, é necessário coletar informações sobre os fatores de confusão. Sabemos que crianças contaminadas com chumbo em geral são mais pobres (pois se contaminam em locais onde o chumbo foi despejado em terrenos ou perto de oficinas de bactérias de carro) e também sabemos que crianças mais pobres em geral tem mais cárie dentária. Se eu fizer apenas um estudo comparando crianças expostas ao chumbo e não expostas, eu posso encontrar associação quando não levar em consideração se ela tem o mesmo nível socioeconomico, mesma alimentação e mesmo atendimento odontológico. Todas estras outras variáveis consideradas de confusão são diferentes das duas primeiras variáveis mais importantes chumbo e cárie dentária (causa-efeito), desta forma constuma-se chamar estas outras variáveis de **terceira variável,** na verdade podemos ter "varias terceiras variaveis". As terceiras variáveis que podem interfeir na associacao causa-efeito podem ser fatores de confusão e também **fatores modificadores** (ou também chamados de fatores de interação).

Agora introduzimos mais um termo, fator modificador ou de interação. Vamos supor que álcool seja um possível fator de confusão para associação entre fumo e câncer de laringe (CL). Se ele for apenas um fator de confusão o que vai acontecer é que se medissemos a associação entre fumo e CL e vissemos que as pessoas que fumam são 7 vezes mais prováveis de terem CL, sem levar em consideração o álcool (isto é tanto faz se as pessoas bebem álcool ou não). Para entender como o álcool pode afetar esta associação, o que fazemos é remover todo mundo que bebe tanto dos expostos como não expostos ao fumo e calcular a associação novamente. Depois deixamos apenas os que bebem tanto entre expostos como não expostos e ai calculamos de novo. Se entre

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

as pessoas que bebem, a associação entre fumo e CL for 7, e entre os que não bebem também for de 7, isso quer dizer que beber não afetou a probabilidade do indivíduo ter CL. Em outras palavras, podemos dizer que a associação entre fumo e CL foi da mesma magnitude tanto para quem bebe como para quem não bebe. Porém, se o valor de 7 cair ou ficar diferente como por exemplo o valor 8,2 para quem bebe e 5,4 quem não bebe, concluimos que o álcool foi um fator de confusão e que a verdadeira associação entre fumo e CL é algo menor em torno de 3 e 4 e não o inicial 7 que tinha imbutido o efeito do álcool sobre CL. Os valores que aqui aprensentei foram todos inventados, apenas para formar o exemplo.

Outra situação seria se os valores resultantes forem muito diferentes como, por exemplo, associação de valor 8 para quem bebe e de valor 2 para quem não bebe, isso significa que quando não se bebe a associação entre fumo e CL é relativamente pequena e se o indivíduo beber esse valor vai para 7, quer dizer aumenta muitíssimo! Dizemos que o álcool é um fator modificador da associação entre fumo e CL, isto é na presença do álcool o fumo causa muito, muito mais CL.

Tabelas mostrando como se faz a estratificação para estudar se existe fator de confusão ou modificador.

	Câncer	Não Câncer			Razão
Fumante	700	300	1000	70%	7
Não fumantes	100	900	1000	10%	

Entre os que bebem álcool

	Câncer	Não Câncer			Razão
Fumante	300	100	400	75%	8,2
Não fumantes	50	500	550	9,1%	

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Entre os que não bebem álcool

	Câncer	Não Câncer			Razão
Fumante	350	200	600	58,3%	6,4
Não fumantes	50	500	550	9,1%	

Portanto a terceira variável pode ser tanto um fator de confusão como um fator modificador, é possível também que um mesmo fator seja considerado de confusão e modificar ao mesmo tempo. No entanto, sempre que tiver o papel de modificador este será o preponderante. A discussão de fatores de confusão e modificadores é um pouco mais complexa do que estamos aprensentando, mas no geral o fator de confusão é aquele associado ao mesmo tempo com o desfecho e a variável de exposição estudada, e ele modifica o valor da associação bruta (isto é antes de levar em consideração o fator de confusão) em pelo menos 10% do valor original. Se ao estratificar verificarmos que os valores em cada estrato forem muito diferentes então consideramos fator modificador. Este diferente tem que ser tanto em número como estatisticamente (como veremos quando estudarmos estatística). Outra possibilidade é que ao estratificar não se observe associação entre a exposição e desfecho que se está estudando. Por exemplo, a associação bruta do entre fumo e cancer foi de 7, mas se ao estratificar verificassemos que a associação entre fumo e câncer fosse igual a 1 para os que bebem álcool e também para os que não bebem álcool, isso significaria que a associação entre fumo e câncer na verdade estava sendo totalmente explicada pelo álcool. Nesta condição fumo não estaria em nada associado ao câncer, porque uma vez levado em consideração o álcool o efeito do fumo desapareceu.

Até o momento talvez você esteja se perguntando o que fazer se tiver um fator de confusão para se chegar ao real valor da associação. No exemplo da associação entre fumo e CL, o valor bruto era de 7, já entre os que bebem foi de 8,2 e entre os não bebem de 6,4. O valor depois de ajustado será diferente de 7 e próximo de 6,4 que é o valor para as pessoas que não fumam. Posteriormente, faremos este cálculo correto, que leva em consideração a distribuição ponderada do número de pessoas que bebem álcool.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Este procedimento de cálculo ponderado pelo fator de confusão é o que chamamos de **valor ajustado** pelo fator de confusão, que significa valor indepentende do efeito do fator de confusão. Ao ponderar, e levar em consideração o fator de confusão teoricamente eliminamos o efeito do fator de confusão por isso é chamado de independente.

Nesta última parte acrescentamos mais informações sobre fatores de confusão e como reconhecê-los estratificando o estudo. Ainda, adicionamos a noção de fator modificador e o significado da terceira variável.

Mais um tipo de Estudo: o Transversal (Estudo de Prevalência)

Até o momento abordamos três tipos de estudos observacionais sendo eles os estudos de casos, a coorte e o caso controle, e existe ainda outro estudo observacional chamado de estudo de **corte transversal** (agora **não é coorte**, é corte - com um "o" apenas-, signficando secção). Na coorte o estudo começa formando especificamente um grupo de expostos e um grupo de não expostos, no estudo de caso-controle formam-se dois grupos um de casos e outro de controles. No estudo de corte transversal (iremos chamar de transversal apenas) não começa bem escolhendo expostos/expostos e nem casos/controles, a escolha é de uma **população definida** de uma área geográfica, e depois de escolher esta população é que se começa o estudo verificando se os indivíduos são expostos ou não e doentes ou não. Note que sempre os estudos têm duas etapas, sendo a primeira a seleção das pessoas, e depois a coleta de informações.

	Seleção/Recrutamento	**Intervenção**	**Coleta de dados Durante o Estudo**	**Medida de Frequência**	**Medida de Associação**
Experimento	Seleção de grupo para a participação: exemplo com uma doença para ser tratado.	Alocar a exposição	Desfecho	Risco Taxa	Risco Relativo Razão de Taxa
Coorte	Expostos Não Expostos	Não existe	Desfecho	Risco Taxa	Risco Relativo Razão de Taxa
Caso-Controle	Casos Controles	Não existe	Exposição	Não existe	Odds Ratio
Transversal	População geográfica definida toda ou de forma aleatória	Não existe	Exposição/Desfecho	Prevalência	Razão de Prevalência

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

No estudo transversal **sempre temos** que definir muito bem a população geograficamente e os participantes **não são** seguidos ao longo do tempo; o participante fornece informações tanto sobre a existência do defecho como da exposição. Este "oferece informação" não necessáriamente é verbal por questionários e pode também ser feito por meio de exames para detectar a presença do desfecho. Vamos pensar num estudo para se avaliar a associação entre consumo de açúcar e cárie dentária em crianças. Poderíamos pensar num estudo experimental, mas seria antiético. Considerando a "qualidade dos estudos", poderíamos pensar numa coorte, mas esta levará muito tempo. Mesmo levando tempo se justifica desde que já se tenha evidências suficientes da associação, faltando evidências de temporalidade e talvez alguns aspectos mão conhecidos sobre a doença que somente seriam revelados num estudo de coorte. Como a cárie é uma doença relativamente comum (embora em alguns lugares já seja considerada rara), um caso-controle talvez seja difícil de ser realizado, especialmente porque teremos que examinar muitas crianças para triar aquelas com cárie. Assim, a alternativa de selecionar uma população definida seja o mais "facil" e correto para o estágio de conhecimento para uma determinada doença. Na verdade como já sabemos bastante sobre a associação de açúcar e cárie o estudo de coorte seria o melhor. Mas vamos supor que ainda não tivessemos muito conhecimento, então o estudo transversal seria o mais indicado. Como tanto consumo de açúcar como a ocorrencia de cárie dentária são comuns o estudo transversal é recomendado. Para tanto, o primeiro passo é determinar os indivíduos que vão participar do estudo que neste caso determinanos que seriam as crianças de 7 a 12 anos da cidade de Ribeirão Preto. Para participar do estudo então não precisa de nenhum conhecimento previo sobre o indivíduo, apenas que ele pertence às crianças qde 7 a 12 anos de idade que vivem em Ribeirão Preto.

Lembre-se que no estudo de coorte o motivo de entrar para o estudo é ser exposto ou não a alguma coisa (porém nenhum deve ter o desfecho estudado), e no estudo de caso-controle o motivo para entrar para o estudo é ter ou não o desfecho e assim no estudo transversal é apenas participar do grupo geograficamente definido independente de ser ou não exposto ou com desfecho.

Assim como em qualquer outro estudo observacional, os fatores de confusão não são controlados no desenho do estudo trasnversal e, portanto, é necessário coletar

informações sobre os mesmos que devem ser levados em consideração na análise estatística. Lembre-se que apenas no experimento os fatores de confusão são controlados no desenho, isso é ao se alocar aleatóriamente a exposição.

Um erro comum é a classificação de estudos como sendo caso-controle quando na verdade são estudos transversais populacionais. Se a amostra realizada é um transversal não podemos chamá-lo de caso-controle. Um exemplo que me deparei foi um estudo para estudar associação entre o desempenho escolar e cárie dentária numa determinada cidade. O autor havia realizado um estudo transversal representativo da população. Ao descrever o estudo ele menciona que fez um estudo caso-controle em que os casos seriam as crianças com desempenho escolar baixo e o controle as com bom desempenho escolar. Para ser um caso-controle ele deveria separar os casos e sortear controles para estes casos. No entanto, o pesquisador tem uma amostra representativa da população o que é muito melhor do que montar um caso-controle. Chamar o estudo de caso-controle embora pareça mais bonito, está na verdade menosprezando o estudo transversal que é muito melhor. Ainda, por ter considerado o estudo como caso-controle o autor calculou Odds Ratio e não razão de Prevalência que seria muito melhor. Vamos entender melhor o que é Odds Ratio e Razão de Prevalência mais adiante.

Até o momento foram definidos os desenhos dos principais estudos básicos. Existem mais tipos de estudos combinando os estudos básicos. Veremos a seguir mais informações específicas sobre cada tipo de estudo.

Detalhes sobre seleção de indivíduos para os estudos – <u>viés de seleção</u>

Até agora já descrevemos as características básicas dos estudos observacionais e experimentos, mas faltam muitos detalhes sobre a seleção em cada um deles. Existem muitas particularidades (problemas e desafios) na seleção e que às vezes se modificam dependendo do estudo, da doença e da exposição que se está estudando. Apartir de agora vamos comentar sobre problemas de seleção para cada um dos estudos começando com a coorte.

Já descrevemos bem a característica básica dos estudos de **coorte** que sempre começa selecionando grupos de expostos e não expostos. Um grande desafio do estudo de coorte é selecionar estes grupos, como chegar a estes indivíduos que são expostos e

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

não expostos. Num estudo de fumantes e não fumantes como devemos recrutar esses indivíduos? Ao formar grupos de expostos e não expostos, estes grupos não podem ser completamente diferentes em tudo além da exposição porque se não seberiamos se algum outro fator ou a exposição seriam responsáveis pelas diferenças na ocorrência do desfecho estudado. Vamos supor que sabendo que um grupo de indígenas não fuma, decidimos comparar estes indígenas com trabalhadores rurais que fuman na região. Sim temos expostos e não expostos, mas os grupos são muito diferentes em relação ao que se alimentaram a vida toda, genetica, estilo de vida etc, então talvez não possamos atribuir ao fumo a associação encontrada. Assim, quando montamos um estudo de coorte os grupos têm que ser de certa forma homogênios e comparáveis. Os estudos de coorte para desvendar efeito de exposições ocupacionais são bons exemplos de dificuldade de se montar um estudo de coorte. Imaginem um estudo sobre os efeitos de exposição ao benzeno numa fábica tendo como expostos os empregados braçais da fabrica, qual seria o melhor grupo de não expostos? Em geral as pessoas pensam em trabalhadores da mesma fábrica que ficam nos escritórios. Bom, estes dois grupos são extremamente diferentes, os trabalhadores braçais são bem mais pobres, tem estilos de vida completamente diferentes dos trabalhadores do escritório. Assim, o melhor grupo de não expostos seria composto de trabalhadores de nível socioeconomico e hábitos de semelhantes, mas de outra fábrica onde os trabalhadores não são expostos ao benzeno (e mesmo nenhum outro fator que leve ao mesmo desfecho que está sendo estudado). Por exemplo, se estamos estudando câncer de pulmão não adianta comparar a incidência de câncer de expostos ao benzeno de uma fábrica com expostos ao asbesto de outra fábrica porque as duas exposições levam ao câncer de pulmão. Assim, ao se montar um estudo de coorte há necessidade de muito planejamento para a seleção dos melhores grupos de expostos e não expostos. Se o grupo de não expostos não for comparável à conclusão será falha e dizemos que o resultado é enviesado, devido ao **viés de seleção**. Portanto, devemos fazer de tudo para que não exista viés de seleção num estudo.

Outra forma de se montar um estudo de coorte é elaborando um estudo de coorte populacional. Numa coorte populacional seleciona-se um grupo de indivíduos baseado numa localização geografica onde todos são incluídos e daí separa-se expostos

e não expostos. Um exemplo é a coorte de Framingham nos EUA que incluiu todos os indivíduos da cidade com mais de 35 anos de idade e começou em 1948. O objetivo do estudo era estudar o efeito do colesterol na incidência de doenças cardiovasculares. Mas este tipo de estudo não é eficiente se o estudo se restringe a um fator ocupacional como benzeno, pois os únicos expostos são os trabalhadores de uma ou algumas fábricas, assim não tem o porquê de amostrar uma cidade inteira. Outro exemplo é a coorte populacional de nascidos vivos de Ribeirão Preto e de São Luiz, Maranhão, cujo objetivo é estudar a saúde de prematuros comparados aos não prematuros. Nesta coorte de nascidos vivos, todos os nascidos vivos de Ribeirão Preto em 2010 (de 1º de janeiro de 2010 a 31 de dezembro do mesmo ano), cerca de 7 mil, foram incluídos no estudo. Equipes de pesquisadores ficaram de prontidão em cada hospital da cidade, e assim que uma gestante era admitida, a equipe abordava os responsáveis, e coletavam informações sobre as gestantes por meio de entrevistas e examinavam as crianças (peso, comprimento, idade gestatcional, teste APGAR, circunferência de cabeça). Em São Luis onde o número de nascimentos era muito maior, uma amostra com 1/7 dos nascimentos naquele ano foram recrutados. A amostra foi sistemática, isto é, a cada três nascimetno em ordem um era recrutado para o estudo.

As coortes populacionais são consideradas as melhores evitando-se que tenha viés de seleção. Porém, podem ser mais caras e trabalhosas. Além do processo de seleção, a coorte tem outro desafio que é o seguimento dos indivíduos (*follow-up)*. Durante um estudo longo é difícil manter todos os participantes e se estes vão desistindo do estudo isso pode afetar a validade do estudo introduzindo viés. Por exemplo, se o desfecho da coorte de prematuros de Ribeirão Preto é cárie dentária, e se mais não prematuros especialmente os ricos (que tem menos cárie em geral) desistem do estudo, talvez não consigamos encontrar associação entre prematuridade e aumento de cárie dentária, pois os não-prematuros que permaneceram no estudo são os que tinham mais cárie. O estudo então será enviesado pela perda seletiva de determinadas pessoas. Se a perda for aleatória entre expostos e não expostos nada provávelmente vai acontecer exceto diminuição de participantes (que acarreta problemas de poder do estudo que veremos posteriormente). Mas se a perda for maior em um dos grupos de expostos e principalmente associada ao desfecho o viés irá

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

acontecer. Este tipo de viés por perda de seguimento não deixa de ser um tipo de viés de seleção, mas chamamos especificamente de viés de aderência (ou seguimento). Portanto, manter a participação no estudo de coorte é muito importante e requer muito esforço, tempo, e dinheiro para manter a coorte motivada em participar do acompanhamento. A permanência de indivíduos num estudo é camado de **aderência** a um estudo. Qualquer estudo que precise que as pessoas sejam seguidas é sujeito a viéses por perda de aderência, por exemplo, no experimento. Em inglês o nome dado à aderência é *compliance.*

Outro detalhe de seleção dos estudos de coorte é que o desfecho precisa ser muito bem especificado. Estamos falando de viés de seleção e o desfecho vai ser medido depois da seleção, no entanto, ao montar o estudo com expostos e não expostos nenhum dos indivíduos que entra para o estudo pode ter o desfecho no inicio. Assim, se o estudo de coorte tem como objetivo o estudo da associação entre fumo e câncer de pulmão, nenhum indivíduo pode estar com diagnóstico de câncer de pulmão. Há, portanto, a necessidade de verificar a saúde de todos os participantes e aqueles que têm indícios de já estarem com câncer de pulmão devem ser excluídos. O princípio básico num estudo de coorte é que expostos e não expostos tem supostamente a mesma chance (teórica) de ter a doença no futuro exceto pelos fatores de exposição estudados. Por exemplo, se já sabemos que quem tem um polimorfismo que aumenta a probabilidade de se desenvolver câncer em muito (muito mesmo) como o BRC1A associado a câncer de mama e ovário, devemos excluir previamente estes indivíduos do estudo.

Com o mesmo princípio se algum fator faz com que o indivíduo não tenha como desenvolver a doença este deve ser eliminado do estudo. Por exemplo, mulheres com histerectomia não podem participar de um estudo de câncer de útero. Se esses detalhes não forem observados vamos ter ao final um estudo enviesado.

Outro problema de seguimento em estudos de coorte é a mudança de exposição ao longo do estudo. Por exemplo, fumantes podem diminuir a quantidade de cigarros consumidos ou ainda podem parar de fumar, e não fumantes podem se tornar fumantes. Assim, embora os indivíduos sejam classificados como expostos e não expostos no início do estudo de coorte, dependendo da exposição deve-se manter o

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

monitormento da mesma durante todo o estudo, assim como de fatores de confusão que podem se modificar ao longo do estudo.

Ainda em relação às coortes realizadas em locais de trabalho deve-se observar a possível existência do viés de seleção chamado **viés do trabalhador sadio.** Este viés pode acontecer porque em geral os trabalhadores que estão na ativa são mais sadios do que aqueles que não estão trabalhando. Imagine uma coorte que será estabelecida numa fábrica com exposição ao asbesto durante 15 anos. Os indivíduos que estao trabalhando no início do estudo e também aqueles que permaneceram no trabalho serão aqueles com saúde melhor. Se por um acaso alguém muito susceptivel ao asbesto começou a trabalhar e passou mal logo no inicio, provávelmente ele irá pedir demissão e trocar de emprego.

Evitar o **viés de seleção** é particularmente complicado nos **estudos de caso-controle**. Selecionar casos de certa forma é relativamente fácil. Em geral os estudos de caso-controle são realizados para estudar doenças raras, e em geral que as tem procura um hospital em algum momento. Assim, os casos em geral são obtidos em hospitais ou clínicas ou centros de referência. No entanto, os controles são difíceis de serem encontrados. Os primeiros estudos de caso-controle como os de fumo-câncer de pulmão foram realizados com controles do próprio hospital que não possuiam câncer de pulmão. Posteriomente descobriu-se que outras pessoas internadas em um hospital eram mais prováveis de serem fumantes do que outras pessoas na população em geral. Isso acontece porque o fumo não somente leva ao câncer de pulmão, mas a problemas cardiovasculares e outros que propiciam os indivíduos a serem internados ou permanerecem mais tempo internados devido a outras causas. Assim, inúmeros estudos foram realizados para se determinar como melhor escolher um controle. A melhor e menos enviesada maneira de se escolher um controle é escolhê-lo da própria população de onde saiu o caso. Por exemplo, se estamos estudando câncer de pulmão no Hospital das Clínicas, ao invés de escolher o controle do hospital, devemos estabelecer onde mora o paciente e escolher o controle entre seus vizinhos. Em geral demarcam-se várias quadras ao redor da casa do caso, e identifica-se todos os invíduos de mesma idade e sexo, e sortea-se 1 controle. É muito trabalhoso para se montar um caso controle, mas é a forma de se evitar viés de seleção e invalidar o estudo.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Um exemplo real é um caso-controle realizado em um hospital de referência de São Paulo para se testar se fatores nutricionais estavam associados a câncer de cabeça e pescoço. Os casos eram pacientes com câncer em tratamento vindo de várias regiões do país como Bahia, Minas Gerais que não encontraram tratamento adequado em seus estados. Os controles, no entanto eram pacientes do ambulatorio do hospital, em geral paulistanos que consultavam por motivos variados. Ao final do estudo concluiu-se que comer muita carne de porco estava associado ao câncer de boca. O problema é que grande parte dos casos vinham de áreas rurais da Bahia e Minas Gerais e a alimentação destes indivíduos foi comparada com paulistanos. Sabemos que o consumo de carne de porco e linguiças é muito maior em regioes rurais. Desta forma, não podemos acreditar nos resultados do estudo. O estudo de fato foi negado publicação numa revista de epidemiologia por causa do viés de seleção, mas foi aceito numa revista de oncologia especifica da odontologia. Provávelmente o trabalho foi aceito porque a maioria dos pesquisadores de odontologia não conhece este problema metodológico importante e muitos não creditam. Bom acreditar não é questão de ciência é questão religiosa e epidemiologia e ciência não são religiões. Com certeza realizar um caso-controle bem feito da muito trabalho, mas se não obedecermos as regras vamos terminar com um estudo enviesado, com problema metodologico tão grande que não podemos acreditar no estudo. Por exemplo, este estudo sobre carne de porco e câncer de cabeça e pescoço nunca poderá ser utilizado para contribuir com avaliação de causalidade entre carne de porco e câncer de cabeça e pescoço, porque é enviesado e, portanto não tem validade.

Infelizmente, é muito comum que dentistas e médicos achem difícil fazer um caso-controle de maneira adequada e terminam ultizando controles de hospitais e clínicas. Em geral eles alegam que não acreditam ser problema, sem nunca terem estudado toda a literatura extensa que existe demonstrando que os resultados não terão validade. O último pesquisador que confrontei com o assunto me disse “não acredito que tenha problema algum de usar o controle da clínica, e já publiquei trabalho assim”. Não consegui fazer com que a pessoa lesse as centenas de trabalhos que abordam o viés de seleção de controles em estudos de caso-controle. Este viés especificamente é chamado de **viés de Berkson**. Existem mais detalhes em relação ao

viés de Berkson, e outros nomes dados a viés de seleção, mas por enquanto isso é o suficiente.

Ainda sobre casos-controles, uma dúvida comum é a quantidade de casos para controles. Em geral temos um controle para cada caso, mas por vezes encontramos estudos com mais de um controle para cada caso. O maior número de controles em geral é ultilizado quando o número de casos não é muito grande e se agrega um maior número de controles para simplesmente aumentar o número de indivíduos total. Posteriormente vamos dizer que aumenta o poder de teste do estudo, mas por enquanto, ficamos apenas com esta explicação superficial. No entanto, não compensa ter mais do que 4 controles para cada caso. Algo que tem que ser observado é que se deve evitar ter menos casos do que controles. O motivo para tanto é que estamos comparando casos investigados e precisamos ter uma boa base do que acontece entre controles. O primeiro estudo de Offenbacher (J Periodontol 1996:) testando a hipotese de associação entre periodontite em grávidas e prematuridade e baixo peso foi um caso-controle, em que haviam 91 casos e apenas 31 controles. Uma das grandes críticas que eu faço do estudo é esta instabilidade criada por ter tão poucos controles. È interessante que no abstract os autores apenas citam que foram observados 124 nascimentos e não especifica quantos casos ou controles. Interessante é que é muito mais difícil achar os prematuros que acontecem em cerca de 10-12% dos nascimentos do que bebês normais.

Embora seja simples entender o que se considera um caso, quando se prepara um estudo começam os problemas práticos de como conseguir estes casos. Se vamos estudar câncer de pulmão quais casos posso incluir no meu estudo? Posso ir a um hospital verificar todos os casos diagnosticados nos últimos dois anos e incluí-lo. Neste caso, teremos casos diagnosticados há dois anos que já estão em tratamento, casos que tiveram sucesso até então e casos que talvez não estejam reagindo bem. Ainda neste período de dois anos podemos ter casos que já falesceram. Então qual a forma ideal de se selecionar os casos? Podemos incluir estes casos todos que aconteceram nos últimos anos (deve ser determinado de acordo com o tipo de doença ou desfecho), e desta forma chamamos de casos prevalentes (que já existem). De forma mais correta, mas que pode levar mais tempo para a realização do estudo, podemos selecionar apenas os casos

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

incidentes (casos novos) que forem ocorrendo. Assim, para cada novo caso, devemos na época, ir atrás de um controle oriundo da mesma população. Incluindo casos incidentes evitamos perder os casos que já falesceram e casos que não estão reagindo a tratamentos. A perda destes casos mais graves ou que morrem mais rápido pode levar a viés na estimativa de OR, portanto, enviesando os resultados. Alguns detalhes podem ser pensados como incluir apenas os casos incidentes que foram diagnosticados em estágios iniciais da doença diminuindo viés, mas os casos que já estiverem mortos pode-se pensar se vale a pena entrar para o estudo entrevistando um parente próximo. O viés de se incluir casos prevalentes é chamado de **viés de incidência-prevalência** ou **viés de Neyman**.

O viés de seleção nos estudos transversais vai ser abordado quando começarmos o assunto sobre como fazer uma amostra (amostragem).

Quanto aos experimentos, em geral são considerados quase imunes ao viés de seleção, porque a aleatorização deveria distribuir os fatores e confusão uniformemente entre expostos e não expostos. No entanto, temos que nos certificar que os fatores foram distribuidos verificando estes fatores após a aleatorização. Ainda, não podemos esquecer que quanto menor o número de participante maior a chance da aleatorização não funcionar. Outro fator importante também é que ao seguir os indivíduos num experimento, podemos ter perda de seguimento e desbalancear a distribuição de fatores de confusão. Ainda, podemos perder seguimento associado a presença de doença ou efeitos colaterais de um medicamento ou tratamento que esteja sendo testado. Por exemplo, se um medicamento começa a dar nausea ou desenvolver o desfecho mais rápido, pode ser o que participante desista sem avisar ao pesquisador. A perda de indivíduos

Um outro tipo de viés que pode ser decorrente da seleção ou ainda surgir devido análise estatística de ajuste não apropriado é chamado de viés de colisão (colider bias). De forma geral, este viés surge quando a seleção de indivíduos está atrelado a exposição. O exemplo clássico é o equivoco relatado por Sacket que estudando 257 indivíduos internados em um hospital concluiu que pacientes de doenças locomotoras teriam maior ocorrência de doenças respiratórias. No entanto, como os indivíduos eram pacientes hospitalizados e sabendo-se que pacientes com doenças locomotoras e

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

respiratórias são mais prováveis de serem hospitalizados a associação surgiu pela intermediação da hospitalização. A princípio este foi denominado d**e viés de admissão**. Por isso que devemos ser cuidadosos ao estudar pacientes hospitalizados ou que procuram especificamente um serviço médico. Quando Sacket estudou a associação numa amostra populacional tal associação não se sustentou. Dois outros exemplos clássicos são denominados de paradoxos da obesidade e paradoxo do baixo peso com mortalidade e também o da obesidade e mortalidade e mais recente temos o exemplo de fumantes e covid-19.

Enquanto um fator de confusão causa tanto a exposição como o desfecho, a variável colisora é **causada** tanto pela exposição como pelo desfecho e se estiver presente a análise entre exposição e desfecho não deve ser adjustado pela variável colisora, porque ela levará ao viés. O viés de colisão em geral é criado durante a analise ajustada pelo colisor, ao passo que nao deveria ser ajustado. Ou se ele está influenciando na seleção ele resultará em associacao espúria.

Quanto a hipótese de que fumar poderia proteger de Covid-19 se deve ao fato de que fumantes tendem a ter mais tosse, e covid-19 leva a tosse também, e ainda recomenda-se que quem está com algum sintoma de tosse e coriza deva fazer o teste de covid. Logo, podemos ter maior detecção de covid entre fumantes e ao mesmo tempo muitos dos testes negativos serão de fumantes e pode-se concluir erradamente que fumar seria protetor da covid. Mais uma vez, olhando-se apenas para quem procura pelo teste podemos chegar a uma conclusão errada. O correto para se testar esta hipótese seria uma amostra populacional aleatória na qual o teste seria realizado.

Um recente estudo sobre colider bias de Griffith et al (2020), entitulado “Collider bias undermines our understanding of COVID-19 disease risk and severity” ressalta o risco de se realizar estudos com amostras não representativas.

Viés de informação

Viés de informação também é chamado de viés de observação é o nome que se dá por falha na coleta de informações num estudo, pode ser tanto informação verbal ou por meio de exames. A informação falha pode acontecer tanto em relação a exposição como em relação ao desfecho.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

O viés de informação mais clássico é o associado ao estudo de caso-controle. Uma vez que o caso-controle começa pelo desfecho, a informação sobre exposição em geral é informada pelo participante que tem que se lembrar de quando foi exposto, como e em que quantidade. O esquecimento e o relato impreciso são comums tanto aos casos como controles, mas o pior viés é quando casos se lembram de forma diferente dos controles. Se as lembranças são melhores entre os casos do que entre os controles, o efeito da exposição parecerá maior do que na verdade deveria ser, isto é, inviesado. Portanto, ao se preparar um caso-controle tem que se tentar resgatar da melhor maneira as informações tanto dos casos como dos controles. Como em geral os casos, por terem uma doença, se lembram de coisas anteriores melhor do que o controle então tem que estimular os controles a se lembrarem de detalhes da exposição. Enquanto fazer uma pergunta de forma simples sobre algo que aconteceu de estranho para uma mãe que teve um filho com uma doença congênita seja suficiente para recuperar uma informação, talvez não seja suficiente para a mãe de uma criança saudável. Embora a forma de se perguntar deva ser a mesma para casos e controles, devem-se preferir maneiras mais detalhadas que estimulem os controles a se lembrarem. Este viés especificamente é chamado de viés de memória (recall bias) que é um viés de informação.

Ainda como viés de informação, pode acontecer que o erro seja tão grande que pessoas possam ser classificadas de forma errada como expostas enquanto são não expostas ou erradamente diagnosticadas como desfecho enquanto não tem desfecho. Por exemplo, se estamos investigando o efeito do consumo de carne no desenvolvimento de câncer de intestino podemos obter informação tão errada de alguns pacientes que ao invés de classificá-lo como grande consumidor de carne, nós classificamos como não consumidor. Neste caso temos um viés que é de informação, mas especificamente é chamado de **viés de erro de classificação** (*misclassification bias*). Qual o efeito disso? Pode ser desastroso como os demais. Por exemplo, se temos 1000 consumidores de carne e 1000 não consumidores, e sabemos que consumir carne leva ao câncer de intestino. Vamos imaginar que aumenta o risco de cancer de intestino em 10 vezes. Se diagnosticarmos 200 consumidores de carne como não consumidores e vice versa o que acontecerá? Bom, os classificados para o estudo como consumidores vão

exibir menos câncer, e os não consumidores vão ter mais cancer do que o esperado, logo a relação deve ser bem menor do que 10 ou mesmo pode se invalidar levando os dois grupos a terem a mesma quantidade de cancer. Se este erro de classificação estiver apenas atrelado à exposição, no máximo que vai acontecer é não encontrar associação alguma. A nao ser claro que todos os expostos sejam classificados como não expostos e ai o estudo vai mostrar que quem não come carne (mas que na verdade todos comem) teria menos cancer de intestino do que os que comem. Assim, seria erro de mais.

Se por um acaso, as pessoas que comem muita carne e tem parentes com cancer na familia (logo já são mais propensos a terem cancer) mentirem sobre o consumo de carne por receio que sejam questionados, qual seria o resultado? Aumento de cancer no grupo classificado como não consumidor, e diminuição entre os que foram classificados como consumidores. Logo diminui o efeito e vamos concluir erradamente que o risco de desenvolver câncer entre consumidores de carne é menor do que seria, ou mesmo inexistente. Mas se as pessoas que já tem familiares com câncer de intestino são tão estressadas e paranoicas que ao relatar o consumo exageram e são classificadas como consumidores de carne quando não deviam ser? Neste caso, vamos ter um adicional de gente muito susceptivel ao cancer de intestino entre os que foram classificados como consumidores de carne, e assim, vamos exagerar o risco de ter cancer entre os consumidores. Quando o erro de classificação atinge os dois grupos (expostos e não expostos) sem estar atrelado ao desfecho, chamamos de erro não diferencial, e em geral o que acontece é diminuir o efeito da exposição. Quando o erro (viés) de classificação estiver associado ao desfecho chamamos de erro (viés) diferencial e ai se isso acontecer pode tanto aumentar como até diminuir o efeito da exposição.

Comentamos sobre viés de classificação da exposição, mas o mesmo pode acontecer em relação ao desfecho.

Um exemplo de viés de informação referente ao desfecho, famoso na odontologia, é o erro cometido no diagnóstico de doença periodontal durante o estudo NHANES III (National Health and Nutrition Examination Survey III). Este é um estudo populacional transversal realizado nos Estados Unidos nos anos de 1988 a 1992. Para este estudo 8 dentistas fizeram os exames de perda de inserção periodontal. Quando se realiza um estudo com vários examinadores é importante que sejam treinados e que

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

avaliem de forma equivalente a doença. Posteriormente abordaremos este treinamento (chamado de calibração) e como fazer sua avaliação por meio de teste de repetibilidade. Por enquanto, o importante é entender o que aconteceu no NHANES III. Para se poder conferir se os dentistas fizeram de forma adequada os exames, os pacientes são aleatoriamente designados para cada dentista. Se os pacientes foram atribuídos aleatórimanete significa que cada um deles deve ter examinado quantidades semelhandes de pessoas de todas as idades, sexo, cor, e nível de doença. Se por um acaso um dentista acabou diagnosticando muito mais doença do que o outro, deve signficar que ele fez algo errado atribuindo mais doença do que se devia. Ao terminar o estudo, notou-se que dois dentistas havia atribuido mais doença do que os demais, porém os grupos eram equivalentes em relação a várias características. No entanto, observou-se que ao examinar Afro-americanos os dois dentistas atribuiram mais doença aos mesmos. Provávelmente sabendo que os Afro-americanos tem mais doença, ao medir a perda de inserção é possível que quando em duvida, por exemplo, entre 3 ou 4 milimetros (a sonda é medida em milimetros) estes dois dentistas devem ter atribuido o valor maior. Uma vez sendo detectado o problema, embora não se pudesse corrigir, pois o exame estava realizado, pode-se levar em consideração na análise estatística o erro. Este é um exemplo de viés de observação atribuindo mais doença a Afro-americanos. Se o objetivo de uma analise fosse verificar a diferença de perda de inserção entre Afro-americanos e Brancos o resultado seria enviesado, neste caso maior do que deveria ser.

Viéses e Terceiras Variáveis (fator de confusão e fator modificador)

A dúvida sempre surge sobre a diferença entre viés e terceiras variáveis ou seja fatores de confusão e fatores modificadores. Reservamos o termo viés para expressar erros de coleta de dados seja seleção ou informação. Se acontecer um erro (viés de seleção ou informação), nós não temos como concertá-los. Já os fatores de confusão e modificadores, desde que tenhamos coletado informações sobre eles, nós podemos levá-los em consideração na análise estatística. Tanto viéses como os fatores de confusão ou modificadores (estes últimos quando não levados em consideração na

análise estatística) levam a resultados errados sobre associação entre exposição e desfecho. Dizemos que **viéses e terceiras variáveis ameaçam a validade interna** de um estudo e podem acabar com ela. Portanto devemos evitar os viéses e devemos levar em consideração nas análises estatisticas a presença de fatores de confusão ou modificadores.

Para evitar os viéses temos que planejar muito bem os estudos, começando por elaborar adequadamente a pergunta que se quer responder, ou melhor, a hipótese que se quer testar. Uma vez elaborada e hipótse deve-se construir o diagrama causal considerando todos os fatores de confusão e fatores modificadores que se tem conhecimento. Uma vez feito isso, deve-se estabelecar qual o mellhor tipo de estudo para testar a hipótese dado o conhecimento atual sobre o assunto. Uma vez determinado o tipo de estudo que se vai realizar, deve-se planejar exatamente quem serão os grupos, e como evitar viés de seleção. Ainda têm que se planejar todos os instrumentos de coleta de dados e diagnósticos a serem empregados no estudo, e ainda o treinamento de pessoal, para evitar classificações erradas.

Variações de Tipos de Estudos

Caso-Controle aninhado e Coorte retrospectiva e os Estudos Ecológicos -

Os tipos básicos de estudos observacionais são os já mencionados, porém podemos ter variações especialmente nos estudos de coorte. Podemos por exemplo dentro de um estudo de coorte formar um caso-controle, que é então chamado de **caso-controle aninhado numa coorte,** ou simplesmente **caso-controle aninhado.** Este tipo de estudo surgiu para facilitar o estudo de determinadas doenças raras (em geral câncer) em coortes númerosas como a coorte do estudo de Woman´s Health Study, que é uma coorte de cerca de 40 mil mulheres enfermeiras nos EUA. Desde o inicio do estudo amostras de sangue eram guardadas nos acompanhamentos que aconteciam mais ou menos a cada 2 anos. O sangue das 40 mil mulheres nem sempre era analisado toda vez, porque o custo seria muito alto. Assim, as amostras são guardadas em nitrogênio, e quando necessário essas amostras são analisadas. Para compor o caso-controle aninhado, quando surgia um caso de cancer de mama, um ou mais controles eram sorteado de dentro da coorte, compondo assim um caso-controle aninhado. A

vantagem deste tipo de caso-controle é a disponibilidade de muitas observações coletadas nos anos anteriores pelos pesquisadores diminuindo a possibilidade de vies de informação.

Retrospectivo e prospectivo relacao cronologica do início do estudo e a ocorrência do fenomeno de estudo (exposição e desfecho). Essas sao medidas que Mietnem chama de timing. Num estudo completamente prospectivo o pesquisador observa os dois fenomenos (exposição e registra a ocorrência do desfecho) em termos de direcionalidade ele é pra frente (foward). Completamente retrospectivo tanto o exposicao como desfecho já aconteceram, ai ele pode ter qualquer direcionalidade (foward ou backwards).

Os estudos de coorte clássicos são **prospectivos** partem da exposição e seguem ao longo do tempo futuro até acontecer o desfecho. No caso-controle tudo aconteceu no passado em relação a quando o estudo começa logo é **restrospectivo**. Existe, no entanto, um estudo chamado coorte retrospectiva. Como a base da coorte é que o estudo começa identificando expostos e não expostos e neste momento ninguém pode ter o desfecho, essa característica permanece na coorte retrospectiva. Ha necessidade então de identificar no passado grupos de expostos e não expostos, e rastrea-los até que desenvolvam a doença. Um ótimo exemplo, é se decidissemos em 2015 estudar o que aconteceu, especificamente desenvolvimento de câncer, nas pessoas que foram expostas ao césio em Goiana no ano de 1987. Para quem não tem noção do que foi o acidente existe um site do governo de Goiania (http://www.cesio137goiania.go.gov.br/).

Naquele ano, em Goiana, uma família desmontou sem saber um aparelho que continha césio (substância radioativa) contaminando a vizinhança inteira. Se hoje fossemos fazer um estudo poderiamos identificar estas pessoas expostas no passado, por meio do endereço e talvez de contas de luz que nos forneceria o nome dos chefes de familia que moravam ao redor da casa onde o césio foi manipulado. Provávelmente teriamos que demarcar alguns quarteiroes ao redor da casa com césio e denominar todos estas familias como expostas ao césio. Em alguma parte de Goiania com nível socioeconomico semelhante e que teriamos certeza que não foram contaminados pelo césio, marcariamos alguns outros quarteiroes e conseguiríamos os nomes dos chefes de

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

familia que habitavam tais casas em 1987. Apartir destes registros, tentariamos achar todos os moradores expostos e não expostos para saber como estão hoje em dia ou se já havia falescido e do que. Temos que seguir cada familia independente de onde estejam morando atualmente. Se nós estamos estudando leucemia, por exemplo, então registrariamos como desfecho todos os indivíduos que tiveral leucemia morrendo ou se recuperando entre expostos e não expostos. Neste estudo tudo tanto a exposição como o desfecho aconteceram no passado. A vantagem deste tipo de estudo é a possibilidade de se estudar algo que já aconteceu no passado e economizar tempo. No entanto, é claro que a qualidade das informações talvez não seja tão boa quanto as informações da coorte prospectiva, principalmente as informações sobre fatores de confusão.

A coorte retrospectiva também tem vantagem para o estudo de doenças com período de latência muito grande. Período de latência seria o período de tempo desde a exposição até o inicio de desenvolvimento de uma doença. Um exemplo é justamente a exposição à radiação e seus efeitos tardios. Por vezes pode ser que quando acontece uma exposição marcante a radiação o DNA não seja totalmente alterado para começar uma doença, mas com o envelhecimento outras alterações se somam e resultam em doenças décadas mais tarde. Desta forma as coortes retrospectivas são vantajosas embora o seguimento das pessoas ao logo de toda a vida fosse ideal.

Antes de começar a descrever os estudos Ecológicos vamos resumir os tipos de estudos até o momento estudados e fazer o primeiro exercício.

Tipos de estudos básicos

Experimentais

- Experimentos (estudos clínicos randomizados ou aleatorizados)
- Quasi-experimentos

Observacionais

- Estudos de Casos ou Série de Casos
- Estudos transversais

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

- Estudos de Coorte
- Estudos de Caso-controle
- Estudos Ecológicos

Exercício I. Tente agora descrever as vantagens e desvantagens de cada tipo de estudo.

Caso-controle	**Coorte**	**Transversal**
Vantagens	Vantagens	Vantagens
Desvantagens	Desvantagens	Desvantagens

Exercício I I

Tente classificar os estudos abaixo!

_____a. Sujeitos do estudo: soldados americanos que participaram da Guerra do Vietnan (1969 – 1971), e soldados americanos que permaneceram na Europa no mesmo período. Na decada de 1980 investigadores comparam a mortalidade nos dois grupos.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

_____ b. Pacientes terminais de câncer, um novo tratamento é oferecido, e a sobrevivência é observada até o período de 2 anos.

_____ c. Pacientes com tricnose confirmada por laboratório e um controle sadio. Todos os participantes respondem questionário sobre consumo de carne de porco, e outras carnes.

_____ d. Crianças de um convênio de saúde aos 18 meses são aleatóriamente designadas a dois tipos de vacinas contra a gripe. Os efeitos colaterais são registrados nas duas semanas seguintes.

_____ e. A história de consumo de cigarro de pacientes entrando em hospitais com câncer de pulmão são comparados com a história de fumo de outros pacientes que foram admitidos com outras condições que precisavam de cirurgia.

_____ f. Aspirantes a soldados do exército respondem a um questionário sobre a história de hábitos de uso de cigarros. Os fumantes e não fumantes são subsequentemente seguidos em relação ao desenvolvimento de câncer e outras doenças crõnicas.

_____ g. Num estudo da relação entre anormalidades reprodutivas e exposição *in utero* ao dietilestilbestriol (DES), a taxa de incidência de anomalias reprodutivas em indivíduos cujas mães foram expostas ao DES quando estavam grávidas 20 a 30 anos atrás são comparados a taxa de incidência de anomalias reprodutivas em indivíduos cujas mães não foram expostas ao DES.

_____ h. O nível de estrógeno é medido no sangue numa amostra de mulheres de 50 a 69 anos de idade. Ao mesmo tempo, a densidade mineral óssea também foi

medida. A proporção de mulheres com baixa densidade mineral óssea e comparada em relação ao nível de estrógeno

Estudos Ecológicos e estudos com variáveis ecológicas

Até o momento foram descritos estudos em que indivíduos eram organizados em grupos para que fossem estudados. Os fatores de confusão e modificadores eram atributos de cada indivíduo. Por exemplo, no estudo de associação entre álcool e câncer coleta-se informação sobre exposição e desfecho de cada indivíduo, e também informações sobre os fatores de confusão. Dizemos que a unidade do estudo é o indivíduo.

Alguns estudos, no entanto, são realizados não com informações sobre o indivíduo, mas sobre um grupo de indivíduos. Por exemplo, posso fazer um estudo para verificar se a prevalência de cárie dentária numa cidade está associada à fluoretação de água na cidade. Para fazer este estudo, por exemplo, no Estado de São Paulo, temos que levantar qual a prevalência de cárie em cada cidade e obter a informação sobre fluoretação das águas. Se os dados não existirem, o pesquisador deverá examinar todas as cidades do estado de SP e estabelecer a prevalência de cárie para cada uma das cidades. Note ele irá utilizar apenas uma informação resumo (prevalência) de cada cidade. Em relação a água fluoretada ele pode usar apenas a informação sobre se é adicionado ou não fluoreto à água ou ainda coletar várias amostras de água em cada cidade e avaliar a média de dosagem de fluoreto em cada cidade. Assim, as medidas que serão utilizadas no estudo são medidas “resumo” (summary measures) que chamamos de medidas ecológicas ou variáveis ecológicas. O sentido de ecológico é que são características gerais médias que descrevem um determinado grupo populacional, no caso cada cidade do Estado de São Paulo.

Se por um acaso, o governo tem como prática fazer estudos de tempos em tempos sobre cárie dentária no Estado inteiro, e estes dados estão disponíveis nas Secretarias de Saúde ou outro órgão qualquer, podemos utilizar estas informações e não precisamos coletar.

O exemplo descrito a seguir é de um estudo ecológico publicado em sobre associação de consumo de cerveja e câncer de intestino nos Estados Unidos publicado

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

em 1977 no British Journal of Câncer. Informações sobre mortalidade por câncer de intestino e consumo per capita de cerveja para cada estado americano.

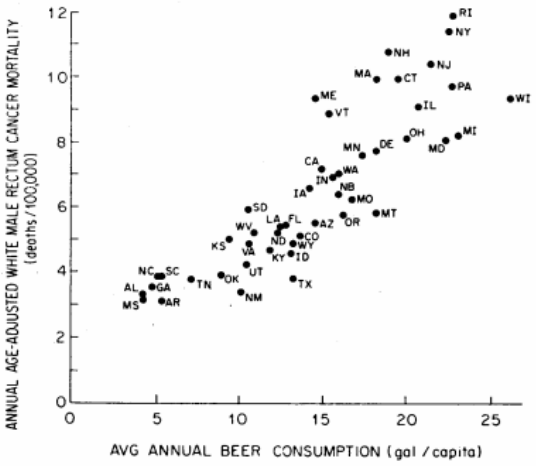


FIG. 1.—Scattergram showing relationship between 1941–60 average annual *per capita* beer consumption and the 1950–67 average annual age-adjusted mortality for white male rectal cancer in 47 states of the United States.

Outro exemplo do estudo sobre dureza da água e mortalidade por doença cardiovascular na Inglaterra (Crawford, Proc Nutri Soc, 1972), porém com associação negativa, quanto maior a dureza menor a mortalidade.

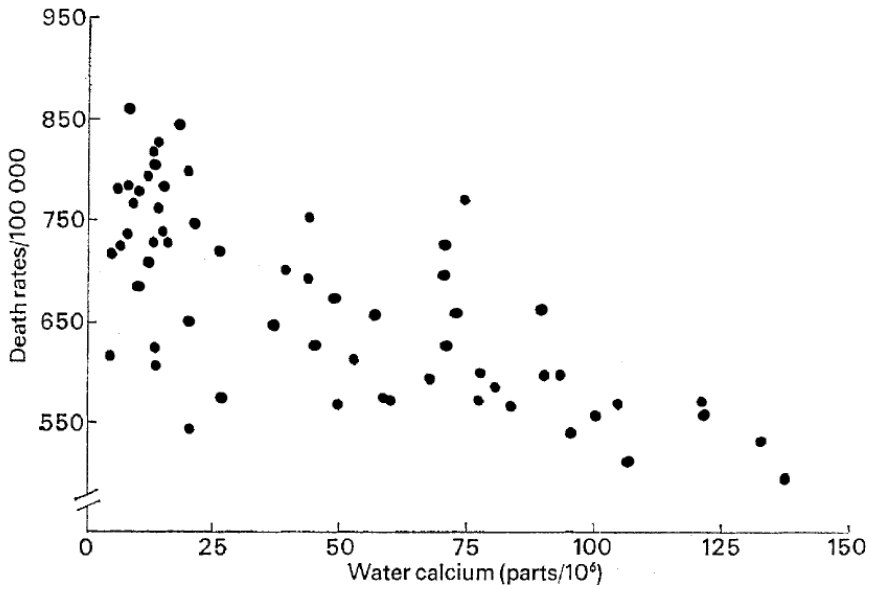


Fig. 1. Cardiovascular mortality (1958–64) in males aged 45–64 and water calcium in sixty-one large county boroughs in England and Wales.

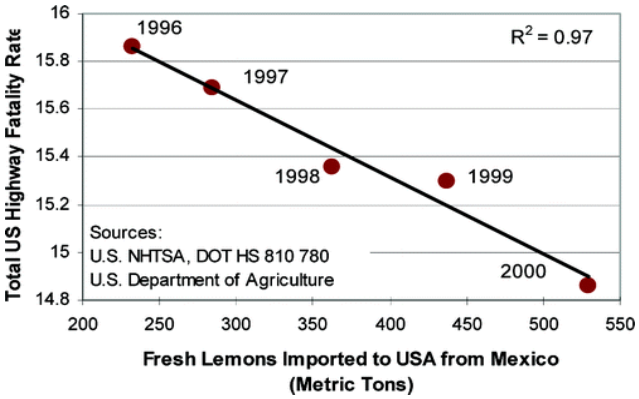
mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Esses dois estudos são exemplos de estudos ecológicos, não sabemos se realmente as pessoas que morreram de câncer de intestino no estudo sobre cerveja realmente tomaram cerveja. Pelo gráfico observamos que quanto maior o consumo de cerveja no estado maior a mortalidade por câncer de intestino, mas será realmente a cerveja ou o consumo de carne e gorduras que deve acompanhar o consumo de cerveja, além do excesso de alimentação? Será também que outros fatores poderiam explicar a associação como o fato de fumar? Falta levar em consideração outros fatores de confusão, e mesmo assim, fica a dúvida se realmente aquele que morreu da doença realmente ingeria cerveja com constância e em quantidade. Um estudo ecológico é aquele em que tanto a variável de exposição e quanto a variável de desfecho são variáveis ecológicas, isto é, representam uma média ou medida de resumo. Neste exemplo, média de consumo de cada estado americano e taxa de mortalidade do estado.

No estudo sobre dureza da água (alta concentração de cálcio e outros minerais) a associação com morte por doença cardiovascular é negativa, isto é parece proteger de morte por doença cardiovascular. Como em todo estudo ecológico não se sabe se realmente o que cada indivíduo realmente foi exposto. Será que quem estava protegido da doença cardiovascular realmente bebia água sem dureza? Ou talvez uma melhor pergunta fosse: será que as regiões em que existe água com dureza alta existem algum outro fator associado com a proteção de doença cardiovascular? Poderiam ser cidades mais desenvolvidas em que as pessoas se alimentam melhor, fazem mais exercício físico etc. Além dos fatores de confusão possíveis, que seriam comuns a todos os estudos observacionais, os estudos ecológicos trazem a questão da dificuldade de se saber ser realmente a pessoa exposta/ou não exposta é aquela que sofreu o desfecho. A falha de se poder extrapolar a associação observada a nível ecológico para o individual é chamada de **falácia ecológica**. Alguns autores dizem que a falácia ecológica é um viés, porém de forma alguma caracteriza um erro de coleta e execução do estudo, mas um erro de interpretação. Portanto falácia ecológica não é viés de forma alguma. Estas correlações também podem ser completamente sem sentido, e um exemplo clássico é o número de mortes em autoestradas nos estados unidos e a quantidade de limões

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

importados do México. O que será que estaria acontecendo, os limões mexicanos diminui as mortes em acidentes de carro? Porque seria? Claro que não tem sentido algum. Correlação não significa causalidade.



Este gráfico copiei do site <https://blogs.oregonstate.edu/econ439/2014/02/03/distinguishing-correlation-causation-key-critical-thinking/>

Embora estudos ecológicos tenham algumas limitações, eles são úteis em vários momentos, principalmente para explorar associações e levantar hipóteses como consumo de carne e câncer de intestino. Estudos ecológicos ao longo dos anos foram capazes de levantar hipóteses sobre causa de vários canceres desvendando que o ambiente em geral é muito mais importante do que a genética. Claro que existem alguns genes específicos como BRCA1 que aumentam em muito a chance de desenvolver câncer de mama e ovário em idades bem precoce. Comparando incidência de câncer e estilos de vida entre japoneses e chineses e americanos pode-se levantar hipóteses sobre alimentação. No Japão o câncer de estômago era dominante enquanto nos EUA o câncer de intestino era consideravelmente maior, e o consumo de carne foi a primeira hipótese levantada. A confirmação da influência do meio ambiente se deu em seguida com os estudos de migrantes japoneses que migraram para os EUA. Comparando as gerações de japoneses observou-se que as primeiras gerações possuíam basicamente a mesma incidência de câncer do país original, mas as gerações seguintes a incidência se assemelhava a dos americanos. O mesmo foi observado em relação ao câncer de mama entre chinesas que migraram para os EUA, com aumento da incidência até se assemelhar ao país de migração. Assim, os estudos ecológicos podem ser muito úteis desde que interpretados com cautela e de forma apropriada.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Portanto, os estudos ecológicos tem seu papel na ciência, o problema não está com o estudo, mas nas interpretações erradas que as pessoas fazem. O estudo é realizado no nível ecológico e neste nível devem ser mantidas suas conclusões. Se o estudo foi realizado testando-se associação entre consumo anual de carne e incidência de câncer, a conclusão tem que ser de que "foi observado correlação positiva entre consumo anual de carne e incidência de câncer de intestino". Isso levanta suspeitas de que o consumo de carne aumente o risco de câncer de intestino, mas esta não é a conclusão do estudo.

Outro fato importante, é que mesmo nos estudos ecológicos temos os problemas de fatores de confusão e devemos sim levá-los em consideração na análise estatística. Por exemplo, o gráfico apresentado anteriormente sobre incidência de morte por doenças cardiovasculares é simples sem levar em consideração fatores de confusão. O estudo não levou em consideração fatores de confusão, mas deveria ter levado em consideração fatores como idade da população em cada área, atividade física, migração de idosos etc. Num estudo recente sobre associação entre fluoretos na água e hipotireoidismo, além de varias outras falhas metodológicas, os autores não levaram em consideração fatores de confusão importantes como a concentração de cálcio e selênio na água nem mesmo o consumo de iodo que está diretamente associado ao hipotireoidismo.

Relembrando, o que define um estudo ecológico é que tanto exposição como desfecho são variáveis ecológicas. Estes estudos podem ser realizados de forma transversal, como por exemplo, associar consumo de carne num determinado ano com a incidência de câncer em vários países. Mas podem ser realizados de forma longitudinal, avaliando ao longo de uma década o consumo anual de carne numa população e observando a incidência de câncer anual. Se a incidência de câncer subir proporcionalmente pode indicar que possa haver associação. No entanto, se pensarmos bem, o consumo de carne não leva ao aumento de incidência de câncer de um ano para o outro. Portanto, precisamos pensar bem se é plausível nesta situação a associação proposta. Talvez com cárie que acontece mais rapidamente que o câncer seja mais fácil observar mudanças na quantidade de cárie com mudanças anuais de consumo de açúcar. Note que mesmo sendo conhecida a associação causal entre açúcar e cárie, é

interessante notar que nos EUA entre 1960 e 1990 o consumo de açúcar aumentou muito, mas a incidência anual de cárie diminuiu. Uma das explicações para esta associação é a falta de consideração de fatores de confusão, pois conforme houve aumento de consumo de açúcar aumentou também a utilização de flúor na água, na pasta de dentes e também outras medidas preventivas como selantes.

Outra informação importante é que existem estudos observacionais que coletam informações de indivíduos como coortes, caso-controle e transversais nos quais uma ou mais variáveis de exposição (logo independentes) podem ser ecológicas, sem que o estudo seja ecológico. Lembre-se que para ser ecológico desfecho e exposição precisam ser ecológicos. Por vezes, pode-se estar interessado em sabe qual o efeito de se morar num bairro de nível socioeconômico baixo independente do nível socioeconômico do individuo terá na hipertensão. O nível socioeconômico do bairro deverá ser construído como uma variável ecológica, ou a partir de estimativas que existam na prefeitura ou IBGE, ou fazendo-se uma média do nível socioeconômico existente no bairro. Nesta situação, lembrem-se, o estudo não é ecológico, apenas a variável de exposição que é ecológica.

De volta aos critérios de causalidade

Como acabamos de ressaltar a discussão sobre causalidade nos estudos ecológicos este é o momento de retornar aos critérios de causalidade que foram mencionados no inicio da apostila. Os critérios clássicos de causalidade nos estudos em saúde foram compilados por Austin Bradford Hill em 1965, num artigo intitulado " The environment and Disease: association or causation?" relatado no Proceedings of Royal Society of Medicine 1965 (pag: 295-300). Não era intensão de Hill ter um check-list de critérios, ele apenas menciona "quais aspectos da associação devemos especialmente considerar antes de decidir que a interpretação mais provável seja causalidade?"[tradução minha].

Relembrando, precisamos seguir alguns critérios para concluir sobre causalidade. Mencionamos anteriormente a **plausibilidade**, por exemplo, não existe

plausibilidade entre associação de limões importados do México e taxa de fatalidade em acidentes de carro. Não esquecendo que nem sempre agente conhece a plausibilidade de uma associação, como por exemplo, na associação entre *H. pylori* e úlcera, pois não se achava que bactérias poderiam viver no estômago. Também, levou se bastante tempo até que se pudessem ter evidências de plausibilidade biológica da ação do estresse no sistema imunológico.

Além da plausibilidade biológica há necessidade de haver vários trabalhos já publicados que tenham validade interna **(retornaremos a este termo).** A este critério damos o nome de **consistência**, que significa que vários estudos apontam para a mesma associação. É importante que esta consistência seja de estudos com metodologia diferentes que tentam refutar pequenas falhas dos demais.

Ainda sabendo-se da possibilidade da existência de fatores de confusão é importante verificar se a **força de associação** é grande ou não. Força de associação é medida pelas medidas de associação (risco relativo, odds ratio, razão de prevalência). Uma força de associação pequena é considerada aquela > 1 e mais ou menos menor que 2,5. Acima de 2.5 a 4 é uma força de associação moderada, e acima disso forte. Essas forças de associação somente devem ser resultantes de estudos válidos, logo em que os fatores de confusão foram controlados. Esta avaliação do tamanho da força de associação é bem relativa e tem que ser combinada com outras evidencias. A princípio se temos uma associação fraca (> 1 ate por volta de 2.2) podemos imaginar que talvez algum fator de confusão não muito bem medido possa ser responsável por este valor, e se levado em consideração resultaria no valor 1 (nulo). Mas se a associação final do estudo é de 3 ou 4, fica mais difícil que fatores outros desconhecidos de confusão possam ser responsáveis por este valor. Embora o valor de 1,7 seja considerado não grande, na verdade representa 70% a mais de indivíduos com o desfecho em comparação com quem não tem o desfecho. Um exemplo é a associação entre colesterol e doenças cardiovasculares que foi estabelecido como causal, e tem risco relativo de certa de 1,7.

Além de verificar a existência na literatura das informações acima, antes de se aceitar que um fator é causal para um desfecho, deve-se também verificar se existe evidencia de temporalidade, isto é, se foram realizados estudos de coorte ou

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

experimental. Mesmo para doenças raras como câncer que levam tempo para desenvolver, há sim necessidade destes estudos. Por exemplo, as indústrias de tabaco apenas começaram a aceitar a associação causal com câncer de pulmão, depois que estudos de coorte foram realizados, demonstrando que quem era exposto e sem sinais da doença. Os estudos são demorados sim, e se envolverem doenças raras necessitam de amostras enormes, mas tem que ser feitos. Portanto, o critério de causalidade que acabamos de discutir é a **temporalidade**.

Além de temporalidade, existem outros indícios que contribuem para evidências de associação causal, e uma delas é **o gradiente de dose e efeito**. Se nós estamos estudando algo como exposição ao benzeno, e observamos que quanto maior a exposição, maior o risco de a pessoa desenvolver certo câncer isso da ideia de que talvez a associação realmente exista. No entanto, nem toda exposição tem relação de gradiente com o desfecho, por vezes algumas exposições tem limites, por exemplo, até uma determinada concentração não leva ao desfecho e a partir de um valor passa a causar o desfecho. Por exemplo, a radiação solar até determinados níveis não provoca câncer, causará câncer apenas quando passar de certo nível.

Portanto para concluir sobre causalidade precisamos primeiro buscar na literatura tudo que foi publicado sobre o assunto (tudo mesmo e existem sites especializados para isso), depois ler atentamente cada trabalho avaliando se tem validade interna ou não. Apenas os trabalhos que tem validade interna podem participar da avaliação final. Uma vez separados todos os estudos com validade interna devemos nos certificar que existe plausibilidade biológica, que os resultados são consistentes nos vários estudos publicados (vários estudos mostrando a associação) e que existem estudos demonstrando temporalidade. Ainda, devemos avaliar a força de associação e somente depois de todas estas avaliações poderemos concluir se aceitamos a associação causal ou não.

Nos parágrafos anteriores, eu menciono os mais importantes criteriso de Hill como sendo plausibilidade, consistência, força de associação, temporalidade e gradiente dose-efeito. Ainda em seu artigo original Hill menciona **coerência**, **evidências obtidas por experimentos ou quasi experimentos**, **especificidade** e ainda **analogia**.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Esses critérios têm sido utilizados por alguns pesquisadores como check-list para causalidade. No entanto, minha visão é que estes critérios são uteis para estimular a reflexão sobre o processo causal de forma simples, e pelo visto era esta a intensão de Hill e não uma lista para um check-list para se estabelecer causa. No artigo do Hill ele menciona

> *" .... here then are nine diferente viewpoints from all of which we chould study association before we cry causation. What I do not believe […] is that we can usefully lay down some hard-and-fast rules of evidence that must be ogeyerd before we accept cause and effect.* ***None of my nine viewpoints can bring indisputable evidence for or against dhe cause-and-effect*** *[grifo meu] hypothesis and none can be required as sine qua non.* ***What they can do, with great of less strength, is to help us to make up our minds on fundamental questions*** *[grifo meu] – is there any other way of explaning the set of facts before us, is there any other answer equally, or more likely then cause and effect?*

Ainda Hill comenta sobre Teste de significancia

> *"No formal tests of significance can answer those questions. Such testes can, and should, remind us of the effects that they play of chance can creaste, and they will instruct us in the likely magnitude of those effects. Beyond that they contribute nothing to the proof of our hypothesis.*

Validade interna

Validade interna é a capacidade do estudo em avaliar corretamente a associação que se propôs avaliar ou fazer a inferência que foi proposta pelo estudo. Abordaremos o que é inferência posteriormente. Por enquanto, inferir significa a partir de um estudo infere-se associação estatística que **talvez** seja causal. Um estudo com validade interna é de forma prática aquele é realizado corretamente dentro dos conhecimentos atuais sobre métodos de pesquisa. Por exemplo, se vamos testar a hipótese de associação entre chumbo e cárie dentária, temos primeiro que escolher o tipo de estudo, e a partir dai temos que pensar em manter a validade interna do mesmo isto é evitar viés de seleção, viés de informação, pensar em todos os fatores de confusão e modificadores e coletá-los de forma adequada para serem utilizados na análise estatística, planejar e

executar a análise estatística de forma adequada. Ainda calcular o número adequado de indivíduos permitindo que posteriormente os fatores de confusão sejam controlados na análise estatística.

Um estudo que não controla fatores de confusão de forma adequada ou suficiente não terá validade interna. Um estudo de caso-controle que coleta controles de conveniência da própria clínica ou hospital terá viés e, portanto não será válido. Existem exceções quando os controles de um próprio serviço são considerados aceitáveis, mas de forma geral, talvez em 99% dos estudos, isso não seja aplicável. Um experimento em que a aleatorização não for apropriada não terá validade interna. Lembre-se ameaçam a validade interna de um estudo os vieses (de seleção e informação) e as terceiras variáveis (fatores de confusão e fatores modificadores) e também a análise estatística inapropriada.

É muito difícil ter um estudo perfeito, algumas falhas pequenas sempre podem acontecer, mas não podemos ter erros gritantes principalmente quando já o conhecemos e sabemos como controlá-los. Se já sabemos que temos que escolher controles oriundos da mesma população que os casos, para que teimar em escolhê-los do próprio hospital ou clínica? Quem irá acreditar no estudo?

Como evitar viés de seleção? Em cada estudo a seleção deve ser realizada de forma adequada. Na coorte, ao estabelecer o grupo de expostos deve-se sempre pensar quem seriam os melhores não expostos para servir de comparação. Além disso, como podemos ter perdas de indivíduos na coorte temos que evitar esta perda. Ainda lembrar-se de evitar o viés do trabalhador sadio quando for realizar estudos ocupacionais ou com populações em fábricas e locais de trabalho.

No caso-controle a melhor maneira de se evitar viés de seleção é escolher os controles da mesma população de onde veio caso.

No estudo de transversal deve-se selecionar a amostra de forma a representar a população que se pretende estudar de forma adequada. Lembrando que se o objetivo do estudo é de estimar uma doença em crianças numa cidade, não podemos apenas selecionar duas escolas que ficam mais próximas da faculdade ou onde tenha dentista, para facilitar a vida. Com certeza terá viés de seleção. Ainda, deve-se tomar cuidado para que o recrutamento seja uniforme e representativo da população. Por exemplo, no

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

estudo transversal do SB2003 (Saúde Bucal 2003) realizado no Brasil, houve participação maior das pessoas mais pobres, mulheres e negras, portanto, houve viés de seleção (na verdade de participação). No final do estudo sem levar a perda de brancos, ricos e homens, a prevalência de cárie relatada para o país foi muito maior do que realmente era.

Com relação aos fatores de confusão, que são terceiras variáveis que interferem na associação estudada, podemos coletar as informações e levar estas informações em consideração na análise estatística. Note que para tanto, devemos montar sempre um **diagrama causal** para visualizar todos os fatores e não se esquecer de nenhum fator de confusão. Outra opção é restringir as pessoas no estudo, por exemplo, se sabemos que sexo pode ser um fator de confusão podemos realizar o estudo somente com mulheres ou somente com homens. Restrição pode ser utilizada, mas tendo vários fatores de confusão, pode ficar impossível de se achar um possível participante. Por exemplo, ao realizar um estudo sobre modificação de marcadores biológicos de inflamação antes e depois do tratamento periodontal existem vários fatores de confusão, porque vários marcadores estão associados a outras doenças como as doenças cardiovasculares. Assim, para encontrar indivíduos com perda periodontal restringindo os fatores de confusão precisaríamos de pessoas que estivessem entre 40 e 60 anos de idade (idade em que se observa quantidade grande de diagnósticos e antes de perda de dentes por outros motivos), mas que não sejam hipertensas, que não tenham nenhum outro problema inflamatório, que tenham colesterol controlado, não sejam fumantes, que não bebam muito, etc. Fica muito difícil encontrar pessoas assim. Em geral se formos fazer um estudo de coorte, caso-controle ou experimento restringindo todos os fatores de confusões possíveis vamos terminar talvez com 10 pessoas, e a triagem seria absurdamente exaustiva depois de examinar centenas ou milhares de pessoas.

Além de restrição, no estudo de caso-controle podemos utilizar o pareamento, isto é ao selecionar um caso que seja do sexo feminino escolhemos um controle do sexo feminino também. Isso faz com que o sexo não seja mais considerado fator de confusão no estudo. Mas novamente, é difícil parear por vários fatores de confusão. O pareamento tem também suas desvantagens porque se parearmos por muitos fatores pode ser que não encontremos nenhuma diferença entre casos e controles. Assim,

tende-se a parear as pessoas apenas por sexo e nível socioeconômico. A decisão é sempre modulada de acordo com o desfecho e exposição que estão sendo avaliados.

Resumindo, o controle de fator dos fatores de confusão pode ser realizado no planejamento do desenho do estudo por meio de restrição ou pareamento, ou coletando-se informações sobre todos os fatores de confusão e ajustando (levando em consideração na analise estatística).

De forma geral os fatores de confusão podem ser controlados nas análises estatísticas desde que tenham sido coletados. Por outro lado, os vieses uma vez existentes não podem ser concertados. Existem algumas exceções como no exemplo de um estudo transversal em que pessoas sorteadas não possam participar o que caracteriza um viés de participação (que não deixa de ter o mesmo efeito de um viés de seleção), no entanto, desde que a amostra tenha sido probabilística, com a informação da probabilidade do recusou participar é possível calcular pesos adequados para levar em consideração a representatividade da amostra. Um exemplo real, foi o viés de observação detectado no estudo transversal NHANES III em que um examinador (dentre os 8 examinadores) de forma sistemática considerou negros com mais doença do que deveriam ter, provávelmente enviesado pelo conhecimento de negros em geral teriam mais perda de inserção que brancos. Este viés somente foi detectado porque os pacientes eram aleatoriamente distribuídos aos examinadores, e, portanto, a presença de algum viés poderia ser detectada. Sendo detectada de forma adequada, a solução, embora não corrija o viés foi de levar em consideração na analise estatística a presença de 8 examinadores.

Conflito entre termos prospectivo e retrospectivo.

Cada livro que lemos tem uma interpretação diferente sobre o que é retrospectivo e prospectivo. Vamos começar com o termo definido por Miettinen que está descrita no livro de Kleibaum et al 1982. Segundo ele, temos duas características no estudo direcionalidade e tempo (timing). Estas descrições a seguir são traduzidas e copiadas direto do livro do Kleimbaum et al 1982.

Direcionalidade é descrito como a dimensão chave dos estudos observacionais, referindo-se a relação temporal entre NOSSA observação da exposição e NOSSA

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

observação do desfecho. A direcionalidade pode ser forward (pra frente), backward (para trás) ou sem direção. Um estudo forward, o investigador começa observando a exposição e acompanha verificando a incidência (casos novos) ou alterações do desfecho. Assim todos os experimentos envolvem a direção forward.

Um estudo backward , o investigador começa com a classificação do desfecho e depois obtém a informação sobre a exposição. Porém, sempre que envolver casos incidentes é possível em teoria determinar se a exposição veio antes da doença.

No estudo sem direção (nondirectional) o investigador observa simultaneamente a exposição e o desfecho e não dá para dizer quem veio antes. Exemplo dado pelo Kleimbaum et al 1982, seria num estudo onde se verificaria simultaneamente a ocorrência de infarto e hipertensão, não da para saber quem está ocorrendo antes. Explicação minha agora, seria se ao elevar momentaneamente a pressão sanguínea acarreta no momento um infarto. A intensão não é verificar se um é causa do outro, mas se acontecem ao mesmo tempo.

A caracterização do tempo (prospectivo e retrospectivo) do estudo refere-se à relação cronológica entre o ***estabelecimento do estudo e ocorrência do fenômeno sob estudo****. "Num estudo completamente prospectivo, o pesquisador observa diretamente DIRETAMENTE observa após o início do estudo tanto a exposição como o desfecho após o inicio do estudo. Num estudo retrospectivo tanto a doença como a exposição já aconteceram. Um estudo completamente retrospectivo pode ter qualquer direcionalidade*

Estas definições não são muito fáceis de entender porque são muito próximas, e os livros fazem muitas misturas sobre tudo isso. Por exemplo, no livro de Bonita et al, um caso-controle seria tanto prospectivo como retrospectivo e também pode ser longitudinal. Segunto Bonita et al (Basic Epidemiology) " os termos retrospectivo e prospective também são usados para descrever o timing da coleta de dados em relação a data corrente. Neste sentido um estudo de caso-controle pode ser retrospectivo quanto todos os dados se referem ao passado, e prospectivo no qual os dados continuam com a passagem do tempo". Esta explicação é confusa, porque mesmo que inclua casos incidentes no estudo de caso-controle, o caso somente entre para o estudo quando vira caso. Note que de acordo com a referência de Kleibaum, prospectivo e

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

retrospectivo tem a ver com o inicio do estudo onde ou o desfecho já aconteceu (retrospectivo) ou ainda não aconteceu como nos estudos prospectivos.

O Rothman argumenta que o caso-controle é retrospectivo, mas quando a intensão for estudar a associação entre uma exposição que já esteja registrada num prontuário médico então se teria certeza que a pessoa tomou o remédio antes do desfecho, e se teria um bom registro da exposição e, portanto poderia ser prospectivo. Novamente segundo Kleimbaum o termo se refere a quando começou o estudo. Se o registro da exposição é bom e está num prontuário, para mim não interessa, já aconteceu no passado quando o pesquisador nem sequer pensava em fazer o estudo.

Sklo e colaboradores em seu livro utilizam o termo concorrente (concurrent), assim ele descreve a coorte prospectiva como sendo aquela que pesquisador começa observando a exposição e segue até o desfecho. E a coorte retrospectiva como não-concorrente. No livro não encontrei os termos retrospectivo e prospectivo.

O que é mais importante é entender cada estudo em particular e suas limitações. No caso controle tudo aconteceu no passado mesmo quando a exposição está registrada num prontuário, tudo bem que se tem certeza que a receita foi realizada antes, mas os outros fatores de confusão serão coletados quando o estudo começar. Portanto, falar que este caso-controle seria prospectivo dando uma conotação de melhor é bastante apelativo e confunde qualquer pessoa. Um autor chamado Vandenbroucke (BMJ 1991; 302:249-50), tem um ótimo comentário que é abandonar estes nomes prospectivos e retrospectivos de forma geral e não dar muita importância a estes termos. No livro do Sklo apenas o termo concorrente e não concorrente é utilizado.

No livro do Medronho et al, o termo concorrente é também utilizado. Segundo os autores um estudo transversal pode ser concorrente e não concorrente. Se a exposição for, por exemplo, peso ao nascer e como este aconteceu no passado este é chamado de não concorrente.

No nosso curso vamos adotar apenas os termos de forma clássica coorte retrospectiva e prospectiva, já o caso-controle será sempre retrospectivo, assim como o estudo transversal. Desta forma, quando lerem que caso-controle pode ser tanto longitudinal, como prospectivo ou retrospectivo esqueçam essas definições, faz muita confusão e em nada ajuda! Se o caso-controle foi feito de casos incidentes, então, na descrição do estudo isso e especificado “caso-controle de casos incidentes” e se prevalentes “caso-controle de casos prevalentes”.

Antes de prosseguir com as medidas de associação tenha certeza que entende a diferença entre os tipos de estudo, e o significado de viés, fator de confusão e fatores

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

modificadores. Façam os exercícios no Anexo sem olhar as respostas e tente entender. Lembre-se sempre que epidemiologia não é questão de decorar definições, mas entender os processos.

Se ainda não conseguiu entender a diferença entre os tipos de estudo, lembre-se sempre que o estudo tem duas etapas: a primeira é a seleção de indivíduos para compor o estudo e a segunda é a obtenção de informações. Estas fases são bem distintas.

As confusões frequentes que observamos na literatura acontecem entre caracterização de caso-controle e coorte retrospectiva, entre caso-controle e estudos transversal, e também entre coorte e quasi-experimentos ou ainda entre estes dois últimos e série de casos.

Um artigo de 2012, Dekker et al, esclarecem as diferenças entre série de casos e coorte. No entanto, em vários exemplos que eles relatam como coorte na verdade caracterizam quasi-experimentos.

Vejamos os exemplos que ele descreve no artigo. Os autores apresentam uma condição inicial no exemplo e oferecem duas alternativas de publicações (a e b), tente atribuir o nome de estudo corretamente a cada alternativa de publicação.

Exemplo 1. Um cirurgião realiza um novo procedimento em 20 pacientes com uma condição severa de uma doença que leva a morte. Dez pacientes sobrevivem.

a) Descrição de todos os pacientes e seguimento com cálculo de risco de mortalidade
b) Além da descrição o pesquisador compara com um grupo histórico da mesma instituição e compara com a mortalidade.

Respostas:
Na situação a) Se o cirurgião realiza um procedimento e segue estes pacientes temos um quasi-experimento do tipo antes-depois.
Na situação b) o pesquisador compara o grupo com dados históricos de outros tratamentos. Portanto, seria um quasi-experimento do tipo antes-depois com grupo controle externo.

Comentários: Alguns podem ter denominado este tipo de estudo como sendo coorte, mas lembre-se que uma vez que o pesquisador faz a intervenção caracterizamos melhor como sendo um quasi-experimento. Este exemplo assim como os outros 3 abaixo foram tirados do artigo do Dekker et al. Ele no entanto não considera a terminologia quasi-experimento. Para Dekker et al seria apenas um estudo de coorte, ele argumenta que seria melhor designar como coorte do que serie de casos. Concordo, mas ser um quasi-

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

experimento é melhor do que ser uma coorte, porque tivemos a intervenção realizada pelo cirurgião.
Em minha opinião se ele recrutasse pacientes que haviam realizado o tratamento X sem saber quem havia realizado ai sim poderia ser chamado de coorte, mas a intervenção fez parte do estudo.

Exemplo 2. Dados são coletados de pacientes que tiveram depressão de medula óssea em um hospital. Potenciais fatores de risco incluindo uso de drogas que poderiam causar depressão da medula óssea são levantados. Após um ano os pacientes são avaliados e verificados se a depressão ainda estava presente dependendo das drogas utilizadas inicialmente para resolver o problema de depressão medular.

a) Descrição de todos os pacientes com depressão medular e frequência de potenciais fatores de risco para a depressão celular.
b) Comparação do risco de depressão medular persistente após um ano de tratamento.

Respostas:

Na situação A seria um estudo de coorte transversal (amostra de conveniência) com descrição dos fatores associados coletados naquele momento. Então seria mais bem caracterizado como uma serie de casos.

Na situação B, o risco seria observado após um ano nos grupos expostos a diferentes drogas para o tratamento da doença. Como não e relatado como a droga foi aplicada se de forma aleatória ou não, então, torna-se novamente um exemplo de quasi-experimento antes-depois com controle interno antes-depois.
Para os autores do artigo, a situação A seria considerada serie de casos, e a segundo como um estudo de coorte, o que novamente discordo, é mais bem caracterizado como quasi-experimento.

Exemplo 3: em um hospital, um grupo de pacientes hospitalizados com Escherichia coli que induziu síndrome uremia hemolítica (SUH) desenvolvem sintomas neurológicos durante a hospitalização. Característica clínica e demográficas dos pacientes com HUS são coletadas.

a) Descrição dos pacientes que tiveram SUH e sintomas neurológicos
b) Comparação para ver se o risco de sintomas neurológicos foi maior entre homens ou mulheres.

Respostas:

Na situação A teríamos novamente uma serie de casos, mas nesta situação vemos que os pacientes foram seguidos, então é um seguimento de casos.

Na situação B, ok, pode-se dizer que poderia ser uma coorte considerando que homens e mulheres foram seguidos e o risco de desenvolver problemas neurológicos foi registrado. Como a exposição estudada é sexo, e não uma intervenção, ok de considerar como um estudo de coorte.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

	Validade de testes diagnósticos **Sensibilidade, Especificidade, Valores preditivos e máxima verossimilhança**

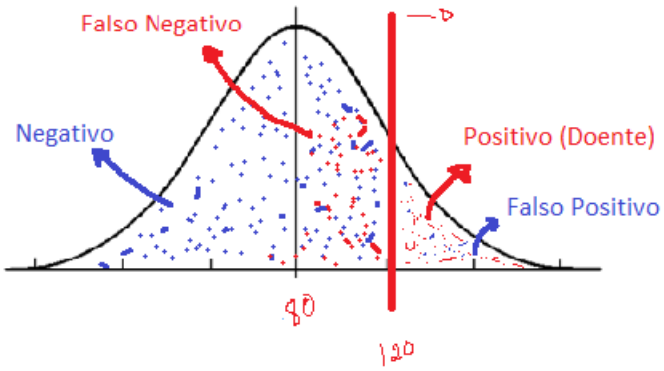
Quando lidamos com doenças ou medidas de condições de saúde como por exemplo depressão, hiperatividade, diagnóstico de cárie, de bruxismo, de doença cardiovascular, câncer precisamos de instrumentos para medir tais condições. Instrumentos podem ser questionários com grupos de perguntas previamente elaboradas e testadas, ou mesmo algum teste de detecção de alguma substância no sangue, ou sinais observados em radiografias ou tomografias entre outros, e mesmo sinais e sintomas clínicos observados e medidos por um profissional de saúde. Seja qual for o instrumento para medir uma determinada condição sempre existe a possibilidade de erros: é possível que o instrumento detecte uma condição, mas que seja um resultado falso (falso positivo), ou que não detecte e seja um resultado também falso (falso negativo).

Porque existe essa possibilidade de resultados falsos? Em geral pensamos em diagnósticos apenas como existindo ou não uma condição. Um dente é ou não cariado, um indivíduo tem ou não tem diabetes, no entanto chegar ao diagnóstico em geral não é tão fácil a nao ser por vezes em que os sinais sejam muito definidos com doença avançada. Claro que se você estiver diante de um dente com um buraco visível ao olho nú, que atinge a dentina e esta está amolecida e tem até uma semente de goiaba dentro, certamente é um dente cariado. No entanto, quando uma doença está no início nem sempre ela é tão evidente para nós humanos. A cárie dentária no início apresenta-se apenas como mancha branca (processo de desminarilização do esmalte) com bordas meio amolecidas é por vezes de difícil diagnóstico.

Para compreender o processo de um diagnóstico precisamos entender que independente da doença, o processo de origem de uma doença tende a ser contínuo embora em geral queiramos dicotomizar o diagnóstico em sim e não. Por exemplo, a cárie dentária começa com a desmineralização do esmalte que aumenta até o ponto de se ter uma cavidade, já diabetes provávelmente não é de um minuto para outro que o

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

pâncreas passa a ter disfunção na produção de insulina. Vamos utilizar o exemplo de insulina para explicar como o diagnóstico de uma condição funciona. Imaginem um teste rápido para diagnosticar diabetes por meio de nível de glicose no sangue, quando vamos considerar que é diabetes e quando não é diabetes.Vamos imaginar (inventar) que numa população de uma cidade com 5 mil pessoas entre 20 e 25 anos e coletamos amostras de sangue para medir a glicose em jejum. Certamente vamos encontrar níveis bem baixos de glicose em algumas pessoas, e níveis bem altos em outros e a maioria num nível intermediário. Se cada pessoa fosse representada por um palitinho e se empilhassemos esses palitinhos sobre uma régua com os possíveis níveis de glicose , vamos supor de 0 a 800, o resultado seria uma figura como a representada abaixo.



Esse formato é de uma curva de Gauss ou também denominado de curva Normal. Detalhes sobre a curva normal pertence a estatística, por enquanto sabemos apenas que ela tem esse formato de morro simétrico e perfeito. Vamos imaginar que estabelecemos o nível de glicose de > 120µg/dl como ponto de corte para considerar que um indivíduo é diabético. Entre as pessoas que tiveram acima de 120 vamos encontrar verdadeiramente diabéticos isto é **positivos** para diabetes (no teste), mas podem passar alguns que não são diabéticos, portanto indivíduos **falso positivos**. Por outro lado, quem está abaixo de 120µg/dl terá o resultado do teste como negativo e será considerado sem diabetes. Porém, entre estes teremos alguns que serão diabéticos, mas no dia estavam com nível de glicose abaixo do limite considerado.

Mas você deve estar pensando, como se decide qual o ponto de corte ideal num exame de glicose para considerar que um indivíduo tem diabetes? A resposta é, um

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

pesquisador (grupo de pesquisadores) deve ter dedicido isso por meio de pesquisas e depois vamos falar sobre isso. Agora vamos voltar como deve funcionar este ponto de corte. Pense bem se movemos o ponto de corte para > 140 µg/dl o que aconteceria? Teremos mais certeza que quem foi considerado positivo seja realmente diabético, e a porcentagem de **falso positivo** deve cair. No entanto, para o outro lado mais pessoas serão consideradas sem a doença, mas a porcentaem de falso negativos será maior.

A decisão do ponto de coorte pode ser estabelecida no nível de glicose que maximiza a identificação tanto de positivos como negativos. Vamos discutir isso tecnicamente e de forma objetiva por meio de algo chamado curva ROC (Receiver operator curve) em outro momento. No entanto, além de considerações objetivos existem decisões subjetivas, mas que são necessárias ao se estabelecer ponto de cortes para melhor **eficiência** do teste. Eficiência se refere a como o teste (ou algo) deve se comportar **na prática**. Como assim? Não queremos sempre o ideal científicamente? Nem sempre, e portanto para você entender vamos abrir um parenteses para explicar a diferença entre eficiência e eficácia e já voltamos aos testes. Um medicamento pode ser eficaz mas não ser eficiente. Eficaz seria nas condições ideais, mas nem sempre na prática vai funcionar. Imagine que se uma pessoa nunca comer açúcar e carboidratos e escovar os dentes 6 vezes ao dia não vai ter cárie (estou inventando esta condição). Essa medida pode ser eficaz (funciona nessas condicões), mas não é nada eficiente pois como vamos fazer as pessoas não comerem de forma alguma açúcar e escovar o dente bem escovado 6 vezes ao dia.

Voltando ao testes a aplicação desta subjetividade aos testes e vamos dar como exemplo o teste de diagnóstico para HIV. Imagine quando a HIV e AIDS foram descobertos lá na dácada de 80. Nesta época, a doença era basicamente uma sentença de morte rápida, basicamente sem tratamento, e com muito preconceito em torno de quem adquiria o vírus. Muitas pessoas que receberam o diagnóstico na época cometeram suicídio. Assim, não se poderia arriscar dando um diagnóstico errado para uma pessoa. Por outro lado, não se podia deixar ninguém sem receber o diagnóstico para começar a se previnir e evitar a infecção de outras pessoas. Todo sangue que chegava em laboratórios para exames rotineiros deveria passar por exames de HIV, até mesmo para serem manipulados com mais cuidado no laboratório. Desta forma, havia

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

necessidade de um teste rápido que abrangesse a maioria das pessoas com possibilidade de doença, mesmo que a pessoa não tivesse o HIV. Era apenas o primeiro teste de rastreamento. Imaginem na curva normal apresentada, nela iriamos colocar o ponto de corte bem baixo, não importando em termos muitos falsos positivos. Posteriormente, as amostras de sangue consideradas positivas neste teste eram submetidas a um teste melhor e mais caro utilizando a técnica de Western Blot que daria a confirmação da presença de HIV.

No teste inicial para HIV queriamos que identificasse o máximo de pessoas com HIV, isso é a probabilidade do teste ser positivo entre os doentes deveria ser máxima, logo dizemos com **Sensibilidade** alta. A definição de sensibilidade é a probabilidade do teste ser positivo dado que o indivíduo é doente. Não importando portando nos falso positivos (positivos que eram falsos) porque a pessoa ainda não receberia este resultado. Por outro lado, se estamos diante de uma doença como diabetes, se por um acaso o indivíduo não for diagnosticado no primeiro exame, ele não irá passar a doença para outras pessoas (porque diabetes não é transmitida por vírus ou bactérias), e nem vai morrer rapidamente (proque diabetes é doença crônica e não mata de um dia para o outro). Assim, para diabetes não precisamos usar um teste tão sensível como para o HIV. Desta forma, a decisão do ponto de corte depende da doença que o teste irá identificar.

Um outro aspecto do teste se refere ao resultado negativo. A probabilidade do teste dar negativo dado que o indivíduo é sadio é chamado de **Especificidade**. Um teste muito especifico irá identificar a maioria dos indivíduos que não são doentes, restando poucos falso positivos. **Sensibilidade** e **Especifidade** de um teste são estabelecidos na elaboração do teste diagnóstico. Para isso, é necessário identificar indivíduos conhecimendamente doentes e indivíduos não doentes, isto é não se tem dúvidas que são ou não doentes. Vamos imaginar que para um tesde de diabetes, utizamos 100 indivíduos com diabetes diagnosticada por vários exames por vários medicos, sem nenhuma dúvida que são diabéticos. Por outro lado precisamos de 100 indivíduos que não tem diabetes diagnosticada, condição esta certificada por outros exames e médicos. O número 100 é apenas para exemplo. No nosso exemplo os 200 indivíduos são submetidos ao teste e o resultado é o que vemos na tabela abaixo. Neste exemplo, o

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

teste positivo conseguiu identificar 80 dos 100 indivíduos doentes, logo sensibilidade de 80%. Entre os não doentes 95% foram identificados. Assim, temos 5 (falso positivo) indivíduos não doentes que foram considerados doentes, e 20 doentes que foram identificados como não doentes (falso negativos)

	Doença		
	Sim	Não	
Teste +	80	5	**85**
Teste -	20	95	**115**
	100	**100**	

Esse resultado de sensibilidade e especificidade interessa a empresa que vende o teste ou ao pesquisador que desenvolveu o teste. Em termos de fórmula a sensibilidade é P(t+/D) e especificidade é P (t-/ND), aqui ND é não doente.

Imagine para o paciente ou mesmo para o médico/dentista que está interessado na saúde do paciente, ele quer saber na verdade a resposta a pergunta: “Então o teste deu positivo, para este paciente, portanto, qual a proabilidade dele ser realmente doente?” Em termos de fórmula temos : P (D/t+). Se o teste deu negativo ele quer saber qual a probilidade de realmente não ter a doença dado que o teste deu negativo : P (ND/t-). Para compreender melhor tente pensar em coisas drasticas. Imagine com um tese positivo o paciente precise amputar uma perna. Desta forma não me interessa a sensibilidade do teste, o paciente vai pensar “o médico pediu para eu amputar a perna porque o teste deu positivo, qual a probabilidade de realmente eu precisar amputar a perna já que o teste deu positivo”. Esta resposta se refere ao **Valor Preditivo Positivo** (VPP), que é P (D/t+). Se o tesde der negativo o paciente quer saber o quanto pode confiar no teste e ficar tranquilo , logo P (ND/t-) que é chamado de **Valor Preditivo Negativo (VPN)**.

No exemplo, o VPP seria 80/85 e o VPN 95/115, portanto 94,1% e 82,6%. Embora as sensibilidade seja de 80%, o valor VPP é de 94,1%, isto é se o tesde der positivo posso acreditar nele pois 94,1% das vezes será mesmo positivo. Já comparado a especificidade que era de 95% se o teste der negativo, apenas 82,6% será negativo mesmo. O valores

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

preditivos indicam a eficiência do teste na prática. Porém ainda não explicamos todas as facetas dos testes de diagnósticos. Faça o exercício abaixo e calcule os valores preditivos. Ao terminar os cálculos reflita sobre os resultados antes de continuar a ler o texto explicativo. Este momento em que você se esforcar para entender o processo fará com que você jamais esqueça do aprendizado. Portanto, faça as contas sem preguiça e reflita sobre os resultados que encontrar.

Exercício III

Considerando um teste com Sensibilidade de 80% e Especificidade de 95% calcule os valores preditivo positivo e negativo para as tabelas abaixo. Após calcular todos os valores, o que você consegue concluir?

A	Doença			VP +	VP-
	Sim	Não			
Teste +					
Teste -					
	1500	**1500**			

B	Doença			VP +	VP-
	Sim	Não			
Teste +					
Teste -					
	1125	**1875**			

C	Doença			VP +	VP-
	Sim	Não			
Teste +					
Teste -					
	1000	**2000**			

D	Doença			VP +	VP-
	Sim	Não			
Teste +					
Teste -					
	800	**2200**			

E	Doença			VP +	VP-
	Sim	Não			
Teste +					
Teste -					

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

	600	**2400**			

F	Doença			VP +	VP-
	Sim	Não			
Teste +					
Teste -					
	400	**2400**			

Depois de calcular todos os valores preditivos positivos e negativos o que você conclui? Responda a esta pergunta, e escreva o que concluiu antes de continuar a ler o texto.

Bom, você deve ter notado que conforme a prevalência da doença no grupo de indivíduos submetidos ao teste cai, o VPP é menor e o VPN é maior. Portanto, quanto mais rara uma doença numa população pior o desempenho de um mesmo teste com mesma sensibilidade e especificidade. Por isso, que quando a pandemia começou em 2020, muitos testes para diagnóstico da Covid foram questinados quanto a eficiência além é claro da baixa sensibilidade e especificidade. Na época, em algumas entrevistas epidemiologistas alertaram que os testes de farmácias eram pouco informativos pois a Sensibilidade e Especificidade eram baixas, e associando com a baixa prevalências da doença na população os resultados não eram confiáveis. Por outro lado, epidemiologistas falavam que os testes deveriam ser reservados para pessoas que trabalhavam em hospitais onde seriam mais confiáveis. Essas recomendações não ficaram claras para o público, e algumas pessoas achava que estavam negando teste à

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

população porque talvez não tivesse testes suficientes. Para a população que não conhece como funcionam os testes diagnósticos era estranho: como um teste pode servir para testar a doença entre profissionais de saúde, mas não para a população em geral? Agora você sabe o porque, pois a época os profissionais de saúde eram os que mais se contaminavam ao tratar os pacientes com Covid. Com certeza não tinhamos muitos testes, mas a explicação de se utilizar estes testes relativamente fracos (com baixa sensibilidade e especificidade) para profissionais da saúde tinha uma explicação. Agora você já sabe.

Agora vamos refletir mais um pouco sobre como funcionam testes de diagnóstico. A sensibilidade muito alta de um teste leva a diagnosticar os verdadeiramente positivos em grande porcentagem, mas também arrasta muitos negativos junto (falso-positivos). Logo por outro lado, neste teste sensível de mais quem é descartado como doente (teste negativo) fica mais seguro que realmente não deve ter a doença. Lembre-se do exemplo do teste de HIV de Elisa que tem sensibilidade enorme. Assim, quando qualquer sangue chega a um laboratório de análises clínicas todas as amostras são submetidas ao ELISA e a confiança de que os negativos realmente não tenham doença é bastante grande. Reforçando, quando um teste tem alta sensibilidade é fácil descartar a doença em quem é negativo para o teste.

Por outro lado, quando um teste altamente específico indica resultado negativo, os falso-positivos serão poucos. Então como no caso de cárie dentária temos testes mais específicos do que sensíveis, portanto, quando um dentista diagnostica uma cárie (teste positivo) provávelmente é porque tem mesmo cárie. Por outro lado, como a sensibilidade é baixa vamos ter muitos falso-negativos e mandamos os pacientes embora com cárie, que em geral são as lesões mais difíceis de diagnosticar. Por isso mesmo, para explorar melhor a existência de cárie em superfícies interproximais, o dentista utiliza radiografias, e por vezes detecta lesões que não estavam visíveis. Ao se fazer o exame radiográfico o dentista está fazendo um segundo **teste em paralelo**. Chamamos de teste em paralelo, quando os mesmos indivíduos são submetidos a dois testes com critérios diferentes ao mesmo tempo. Com isso, **aumenta a sensibilidade** do

teste, diminuindo os **falso-negativos**. No caso de diagnóstico de HIV, o teste Elisa é utilizado primeiro, e somente as amostras positivas para o ELISA são levados para o teste Western Blot. Nesta situação, depois de usar um teste muito sensível os falso positivos são eliminados **aumentando a especificidade** geral. Este tipo de utilização de um segundo teste é chamao de **teste em série**.

É preciso ficar atento aos testes em paralelo e em série, e a ocorrência do evento para que se leve em consideração a probabilidade do seu julgamento estar errado. Às vezes é bem difícil entender o porque dos falsos positivos e negativos e somente fazendo as contas agente consegue compreender. É preciso compreender porque apenas decorar vai levar você a acreditar de novo em crendices quando for para a clínica. A clínica nos engana a cada dia, pois o paciente que você diagnosticou direito retorna ao consultório, e o outro desaparece e você não contabiliza como erro. Ainda dentro de um consultório você não percebe que aquele paciente é um membro da comunidade. Precisamos ficar atentos. Recentemente tivemos uma polêmica sobre frenctomia em recém-nascidos para ajudar na amamentação (escrever sobre isso depois em detalhes e com referencias).

Antes de prosseguir, vamos falar de um termo que em geral é utilizado por alguns autores **a acurácia** de um teste. Bom, definição de acurácia na física é ausência de erro sistemático e de erro aleatório. O mais correto ao falar sobre testes diagnósticos é validade, que é o título que colocamos nesta sessão. O quão válido é um teste para diagnosticar uma condição. Deixamos o termo acurácia para a parte da estatística, se bem que como vamos discutir posteriormente muitas pessoas utilizam o termo acurácia como sinônimo de ausência de erro sistemático o que não é o mais correto.

Razão de Verossimilhança, probabilidade Pré e pós-teste

Não se assuste com o termo acima que vamos explicar, pois na verdade o termo em inglês é bem menos assustador, pois é *likelihood ratio.* Bom para alguns ficou mais assustador, mas vamos explicar aos poucos, por enquanto Ignore o termo por que vamos entender a informação que queremos saber. Quando fizemos o exercício

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

verificando que um teste de mesma sensibilidade e especificidade em populações com diferentes níveis de doença resultam em diferentes valores preditivos positivos e negativos, estamos mostrando que a *performance* do teste varia de acordo com prevalência da doença na população.

Bom, foi fácil de calcular os valores preditivos no exercício em que já estabelecemos a ocorrência da doença. No entanto, na prática do dia a dia, como posso ter a idea de valor preditivo e desempenho do teste se eu não tenho o número total de doentes e de não doentes e de falso positivos? Uma alternativa é você planejar uma tabelinha como fizemos no exercício com a possível prevalência da doença ou fazer continhas na hora que vamos explicar agora. Estas continhas são no fundo semelhantes, mas mais diretas, porém nem sempre tão intuitivo. O que fizemos antes foi verificar qual seria o VP+ e o VP- dado certa prevalência da doença e um conjunto de sensibilidade e especificidade específicos.

Bom, lançamos mão da noção de possível **probabilidade de ter a doença *à priori*** e o que queremos saber é a probabilidade ***a posteriori*** do teste, o que seria equivalente aos valores preditivos. A probabilidade a priori seria referente ao que se espera na população em geral como fizemos no exercício. Mas para calcular esta probabilidade a posteriori precisamos do que se chama de razão de verossimilhança positiva (RV+) e negativa (RV-). Você vai encontrar em alguns livros essa noção mais formalmente explicada como aplicação de teoremas bayesiana ou de Bayes. Vamos tentar simplificar isso aqui.

Repetimos aqui a tabela do primeiro exemplo para que possa seguir melhor o raciocínio.

A	Doença			VP +	VP-
	Sim	Não			
Teste +	80	5	**85**	0,94	0,83
Teste -	20	95	**115**		
	100	**100**			

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Para a *performance* do teste de diagnóstico seria interessante sabermos qual **a razão entre doentes verdadeiros e falso-positivos entre aqueles com teste positivo**? Na tabelinha com 50% de doentes e 50% de não doentes seria de 80/5. Essa conta é uma chance equivalente à porcentagem 80/85 que era nosso valor preditivo positivo. Como assim? Lembra que se num grupo temos 1 menina e 4 meninos, podemos representar em porcentagem 1/5 ou em termos de chance 1:4. Entre as pessoas com teste positivo a chance de ser positivo é de 80 para 5. Quanto menor for o número de falso- positivos melhor o teste. Esta fórmula seria: **sensibilidade /(1 – especificidade).** Esta razão é chamada de Razão de Verossimilhança positiva (RV+). Assim, 80/5 = 16. E o que significa isso? Nessa situação existem poucos falso-positivos, logo RV ou LR (likelihood ratio) é alta. Uma LR+ baixa significa que existe muito erro ao se detectar alguém positivo para um teste. Assim uma LR de 16 significa que o teste positivo é 16 vezes mais provável de acontecer num paciente com a doença do que num paciente sem a doença.

Aplicando isso ao conhecimento prévio de que a prevalência pré-teste seria de 10% , porque seria a prevalência na população da qual veio esta pessoa. Porém uma vez que o teste deu positivo e como é pouco provável que seja errado, qual é a probabilidade dela agora ter a doença? Temos primeiro que transformar 10% em chance que seria 1:9. Para transformar Probabilidade em chance, dividimos a probabilidade / pelo complemento da probabilidade. Por exemplo, odds de 10%, seria 0.10 /(1 – 0.10) -> 0.10/0.90 -> 0.11. Agora vamos multiplicar esta chance que se refere à probabilidade pré-teste pela LR que era de 16. Assim 0.11 x 16 = 1,77. Agora temos que trasformar este valor em forma de chance para probabilidade. O retorno será 1.77 * (1 –+ 1.77) que é igual a 0.6389, isto é, 63,89%. Muito próximo do VP+ que calculamos de forma direta nesta tabela abaixo com prevalência de doença de 10%. A diferença de 63,89 para 64 é que apenas de aproximação.

B	Doença			VP +	VP-
	Sim	Não			
Teste +	8	4,5	**12,5**	0,64	0,98
Teste -	2	85,5	**87,5**		
	10	**90**			

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Existem tabelas/gráficos que se chamam nomogramas para o cálculo rápido da probabilidade pós-teste se um teste deu positivo ou negativo. Aqui ficamos apenas com este conhecimento que pode ser mais bem desenvolvido em literatura específica, mas de forma intuitiva é o que apresentamos aqui, a utilização do conceito do teorema de bayes para nos ajudar avaliar a probabilidade de que um teste positivo devido à probabilidade esperada da doença seja estimado, ou da mesma forma se o teste for negativo.

Curvas ROC

A curva ROC é foi primeiro utilizada com este nome por XX. O problema que tentavam resolver com sua aplicação era a seguinte: suponha que um observador receba uma voltage variando com o tempo durante um intervalo de observação e é perguntado a decidir se sua fonte é noise ou sinal mais ruído . Qual o método ele deve usr para . Curva para detecção de decisões.

“ the theory of detection de sinais na engenharia eletrica, mas a teoria é aplicável a qualquer processo de informação envolviendo both observação e subsequente decisão binária. Já comenta sobre diagnostico medico por maquinas. L2 function eh chamado de ROC character, tha tis one to one with graphical presentation .

The problem of signal detectability treated in this paper is the following: Suppose an observer is given a voltage varying with time during a prescribed observation interval and is asked to decide whether its source is noise or is signal plus noise. What method should the observer use to make this decision, and what receiver is a realization of that method? After giving a discussion of theoretical aspects of this problem, the paper presents specific derivations of the optimum receiver for a number of cases of practical interest. The receiver whose output is the value of the likelihood ratio of the input voltage over the observation interval is the answer to the second question no matter which of the various optimum methods current in the literature is employed including the Neyman - Pearson observer, Siegert's ideal observer, and.Woodward and Davies' tlobserver.lV An optimum observer required to give a yes or no answer simply chooses an operating level and concludes that the receiver input arose from signal plus noise only when this level is

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

exceeded by the output of his likelihood ratio receiver. Associated with each such operating level are conditional probabilities that the answer is a false alarm and the conditional probability of detection. Graphs of these quantities, called receiver operating characteristic, or ROC, curves are convenient for evaluating a receiver. If the detection problem is changed by varying, for example, the signal power, then a family of ROC curves is generated. Such things as betting curves can easily be obtained from such a family. The operating level to be used in a particular situation must be chosen by the observer. His choice will depend on such factors as the permissable false alarm rate, a priori probabilities, and relative importance of errors. With these theoretical aspects serving as sn introduction, attention is devoted to the derivation of explicit formulas for likelihood ratio, and for probability of detection and probability of false alarm, for a number of particular cases. Stationary, band-limited, white Gaussian noise is assumed.. The seven special cases which sre presented were chosen from the simplest problems in signal detection which closely represent practical situations. Two of the cases form a basis for the best available approximation to the important problem of finding probability of detection when the starting time of the signal, signal frequency, or both, are ll&TlOUn. Furthermore, in these two cases uncertainty in the signal can be varied, and a quantitative relationship between uncertainty and ability to detect signals is presented for these two rather general cases. The variety of examples presented should serve to suggest methods for attacking other simple signal detection problems and to give insight into problems too complicated to allow a direct solution. 1. INTRODUCTION The probl

Ate agora apresentamos o termo validade em termos de teste diagnostico. Existem também outras medidas de validade em geral relacionadas a questionários ou instrumentos de medida de algo, como por exemplo validade de face, validade de conteúdo, validade de construto. A validade de face é a primeira impressão de validade de um instrumento que pode ser uma serie de perguntas que alguém agrupa para avaliar algo como dor ou tensão econômica numa família. Não existe uma maneira de se avaliar realmente a validade de face a não ser “achando” coerência no instrumento. A validade de conteúdo se o instrumento inclui todos os itens que compõem aquele suposto instrumento. Imagine que um questionário sobre estresse econômico pergunte apenas se alguém perdeu emprego na família, mas teria q ter algo mais como falta de dinheiro para se fazer o que rotineiramente a família faria, ou ainda se deixou de fazer algo que faria caso tivesse dinheiro regularmente. Esse critério também é subjetivo. Validade de construto é o grau que seu instrumento realmente mede o construto comparado com outras coisas fora do construto (não entendi)

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Reprodutibilidade de um exame (Kappa de Coehn)

Além de um teste ou uma medida ser válida para se avaliar uma doença ou condição de saúde, ao se aplicar um determinado teste, esta aplicação precisa ser igual, isso é feita da mesma maneira. Quando falamos de teste com sensibilidade e especificidade pode ser, por exemplo, um conjunto de sinais e características clínicas que definem uma doença como a cárie dentária (superfície de esmalte esbranquiçada de aspecto amolecido) que ao ser confrontado com evidencias mais contundentes por meio de radiografias e evidências histológicas, ou exames prospectivos. No entanto, durante o desenvolvimento de um estudo observacional ou experimental em que existe a necessidade de se diagnosticar um determinado estado, há necessidade de que o examinador apresentente consistência nos exames. Aqui novamente chamamos de consistência de reprodutibilidade, isto é, se um examinador detecta uma doença num exame e se ele realizar o mesmo em outro dia, ele seria capaz de reproduzir o mesmo diagnóstico. Chamamos esta consistência de reprodutibilidade intra-examinador. Se o estudo utiliza mais de um examinador, o que é comum em estudos com muito participantes, há necessidade de que esta reprodutibilidade seja alta, também entre os examinadores, o que é denominado de reprodutibilidade entre-examinadores.

Quando testamos a reprodutibilidade, examinamos duas vezes um mesmo paciente, e verificamos se o diagnóstico foi o mesmo, isso é se houve concordância. Porém temos que levar em consideração o acaso. Quem nunca chutou respostas num teste e acertou? Portanto, parte de concordâncias podem acontecer ao acaso. Pense em uma sala com 100 adultos em que sabemos que 70 teem asma, e estes estejam situados na sala de forma aleatória. Se pedirmos a uma pessoa leiga, uma criança de 9 anos ir a sala, de olhos vendados escolher, 4 adultos é possível que os 4 adultos tenham asma. Diríamos que isso teria acontecido ao acaso, e talvez não seja tão difícil, uma vez que 70% dos adultos teem asma. No entanto, se na sala de 100

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

pessoas existirem apenas 5 adultos com asma, e pedirmos para que a criança nos traga ao acaso 4 pessoas, será pouco provável que a criança traga 4 dos 5 adultos com asma. Portanto acertar diagnósticos pode acontecer ao acaso, e quando avaliamos reprodutibilidade temos que considerar apenas a concordância além do acaso.

Como avaliamos isso na prática. Existe uma medida de reprodutibilidade chamada de Kappa de Coehn que mede justamente a proporção de condordância além do acaso. Imagine que duas pessoas avaliem um conjunto de 200 lâminas histológicas sendo que um dos avaliadores é considerado padrão ouro, porque tem experiência na leitura de lâminas para diagnóstico de câncer inicial. A outra pessoa a ler as lamina seria um técnico que está sendo treinado. Os resultados dos dois examinadores se encontam na tabela abaixo.

	Técnico		
Pesquisador	Positivo	Negativo	
Positivo	140	60	200
Negativo	10	90	100
	150	150	300

Nesta tabela sabemos que a probabilidade de ter laminas com sinais de câncer de acordo com o pesquisador é de 0.667, ou seja 66,7%. Já o técnico encontrou 50% de laminas com indícios de câncer. Em temos de concordância, pesquisador e técnico concordaram que 140 lâminas eram positivas, e que 90 eram negativas, portanto o total de concordância foi de 230 entre as 300 lâminas isso representa concordância geral de 76,7%. Se eles tivessem concordado em tudo, seriam 100% de concordância equivalente a 300 laminas em concordância.

Como calculamos o acaso? Primeiro fixamos alguém como sendo a verdade e neste acaso diríamos que a verdade é do pesquisador experiente e que reconheceu câncer com a probabilidade de 0.667. Qual seria a capacidade do técnico ao acaso reconhecer lâmina com câncer se ele fizesse isso ao acaso, como por exemplo, se usasse “une-duni-te”? Vamos considerar como fixo que ele reconhecedia 50% como tendo câncer. Assim, como você aprendeu em probabilidade no fundamental que a probabilidade de dois eventos independentes é a multiplicação simples destas probabilidaes, então Pp = 0.667, e a do técnico de 0.50, assim 0.667*0.50 = 0,334. Isso significa que a proabilidade de se sortear uma pessoa, em que esta tenha sido reconhecida tanto pelo professor como pelo técnico como positivo para câncer é de 0.334. Agora podemos proceder de duas fomas, continuar a trabalhar com as probabilidades ou calcular uma tabela de números esperados ao acaso. Para isso, vamos copiar a tabela acima, mas sem os valores internos, vamos manter apenas os números marginais.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

	Técnico		
Pesquisador	Positivo	Negativo	
Positivo			200
Negativo			100
	150	150	300

E vamos preencher os vazios com números esperados dado o acaso. Como temos na tabela 300 laminas, então sabemos que 0.334 * 300 = 100,2 , será o número esperado ao acaso de concordância positiva.

	Técnico		
Pesquisador	Positivo	Negativo	
Positivo	100,2		200
Negativo			100
	150	150	300

Uma vez, que temos um dos números na casela de positivo/positivo os demais núermos são calculados.

	Técnico		
Pesquisador	Positivo	Negativo	
Positivo	100,2	99,8	200
Negativo	49,8	50,2	100
	150	150	300

Agora, nesta tabela de esperados temos 100,2 + 50,2 = 150,4 concordancias. Não tem problema dos números não serem inteiros. Portanto a porcentagem de concordância ao acaso seria de 0,501 ou seja 50,1%.

Com estas estimativas, podemos agora pensar em responder qual seria a proporção de aconcordância além do acaso entre o pesquisador e o técnico. A concordância geral observada havia sido de 66,7%, sendo então que 50,1% seriam ao acaso. Podemos trabalhar tanto com porcentagem como com os números. Vamos aos números primeiro. Sabemos que o máximo de concordância seria de 300, sendo que 150,4 delas seriam ao acaso. Removo o acaso do total de concordância possível 300 – 150,4 e temos 149,6 possibilidades de concordância além do acaso. E o técnico e professor haviam concordado em 230 laminas, removendo destas as 150,4 que seriam ao acaso, nos resta 79,6. Agora vamos calcular qual a porcentagem de concordância além do acaso que será 79.6/149,6 = 0,532. Assim, embora a concordância global tenha sido de 76,7 %, a concordância alem do acaso foi de 53,2%. Assim, o kappa é de 53,2%. Essa concordância é bastante fraca, mas vamos ver o que é considerao fraco ou forte.

Você pode trabalhar apenas com as probabilidades e pensar que a probabilidade de acertos seria a probabilidade de acertos positivos ao acaso que seria o já calculado (0.667*0.5) mais os acertos negativos [(100/300)* (150/300)] ou seja 0.337*0.50. Assim (0.667*0.5) + (0.337*0.50) = 0,334 + 0,169 = 0,503. Assim, agora subtraindo esta porcentagem de

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

concordância da porcentagem de concordância geral e dividindo pela proporção possível de concordância além do acaso teremos (66,7 – 0,503)/ (1 – 0,503) = que será o mesmo kappa que calculamos anterioremente. Pequenas diferenças de casas decimais sempre acontecerão.

O importante é compreender o conceito de Kappa e não ter que decorar para fazer os cálculos. Além do Kappa existe o Kappa ponderado que é recomendando quando se tem variáveis ordinais e não somente dicotômica. Ordinal é uma variável que tem mais de dois níveis de forma ordenada, como por exemplo, uma doença ou condição que se caracteriza como ausente, leve, moderada e grave. Pode ter maior concordância entre leve e ausente, por exemplo, e assim o Kappa ponderado (weighted Kappa) pode ser útil. Aqui não vamos abordar a fórmula do kappa ponderado, mas se tiverem necessidade de calcular esta estatística é facilmente encontrada na internet.

Ainda é importante ressaltar que aqui apresetamos soluções para quando as variáveis forem categóricas dicotômicas ou ordinais (em níveis), porém podemos ter situações em que nossas medidas são chamadas de contínuas como, por exemplo, medida de profundidade de bolsa em milímetros. Assim, nós temos eu verificar se duas ou mais pessoas conseguem medir com precisão além do acaso. Nesta situação, calculamos o que se chama de correlação intraclasse, que seria um valor de 0 a 100% que representa o quanto as medidas de duas pessoas se correlacional bem além do acaso. Essas fórmulas também são mais complexas e não serão aqui abordadas, em geral são encontradas em livros de estatística com abordagem de análise de variância.

Existem casos especiais de Kappa como quando temos avaliações de condições que são ordinais e não simplesmente dicotômicas como doentes/não doentes. Nesta situação utilizamos o que se chama de weighed Kappa, mas não iremos dar detalhes aqui.

É importante contextualizar a necessidade de se verificar concordância intra e interexaminadores num estudo que tem como objetivo reduzir erros sistemáticos que podem levar a vieses.

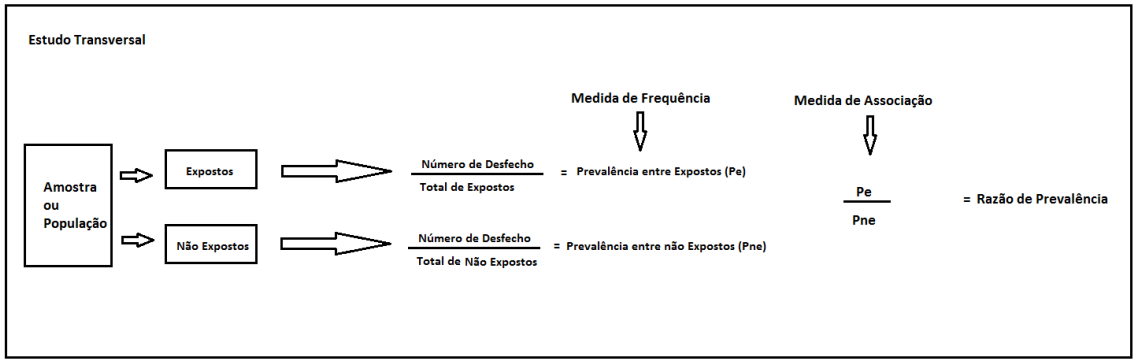
Medidas de Frequência e de Associação

Em todos os estudos que fazemos vamos resumir seus resultados por meio de números. Numa série de relatos de casos como o Gregg temos que resumir os resultados e dizer que dos 78 casos de crianças com catarata congênita 87,18% suas mães foram expostas a catarata congênita. Da maneira como o estudo é montado, já sabemos quais as contas que devemos fazer. Note os quadros abaixo. Em geral nos livros primeiro se

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

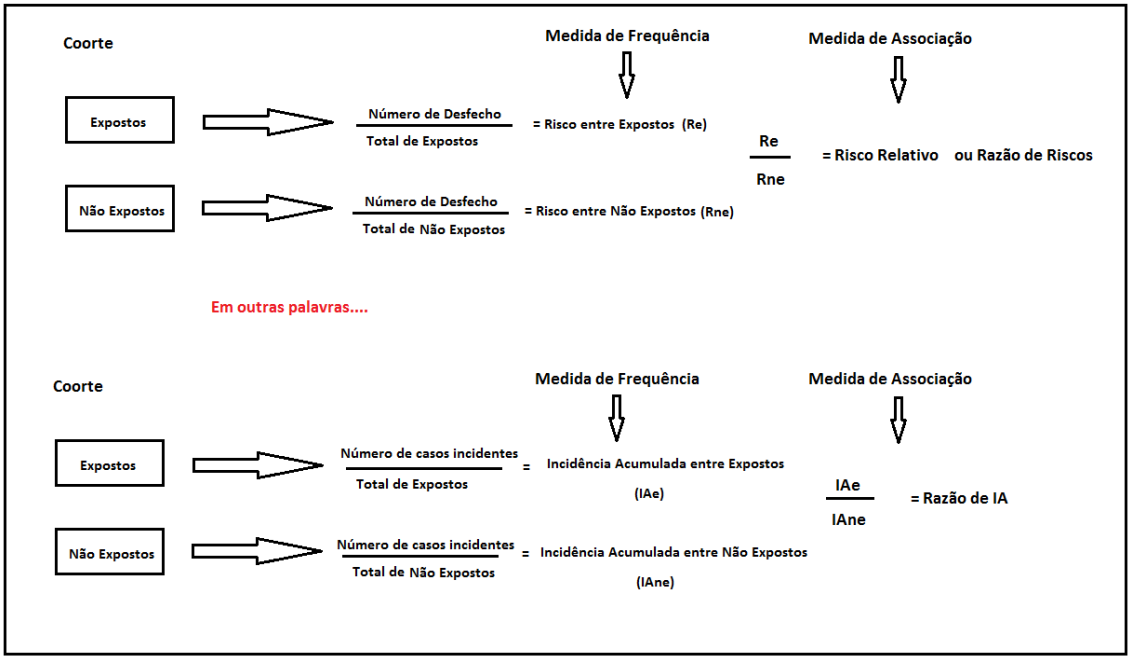
comenta sobre as medidas de frequência e depois de associação. O principio básico é que para cada estudo damos um nome ao que calculamos, e quando estamos diante de um grupo simples tipo quantidade de pessoas com desfecho entre expostos, chamamos de medida de frequência. É medida de frequência porque descreve quantas pessoas (porcentagem) com o desfecho existe naquele grupo. Quando comparamos a frequência de um desfecho em um grupo com certa característica com a de outro grupo, estamos diante de uma medida de associação. Isso porque ao falarmos que um grupo tem tantas vezes mais desfecho que o outro, o que estamos descrevendo é a se existe ou não associação entre exposição e doença.

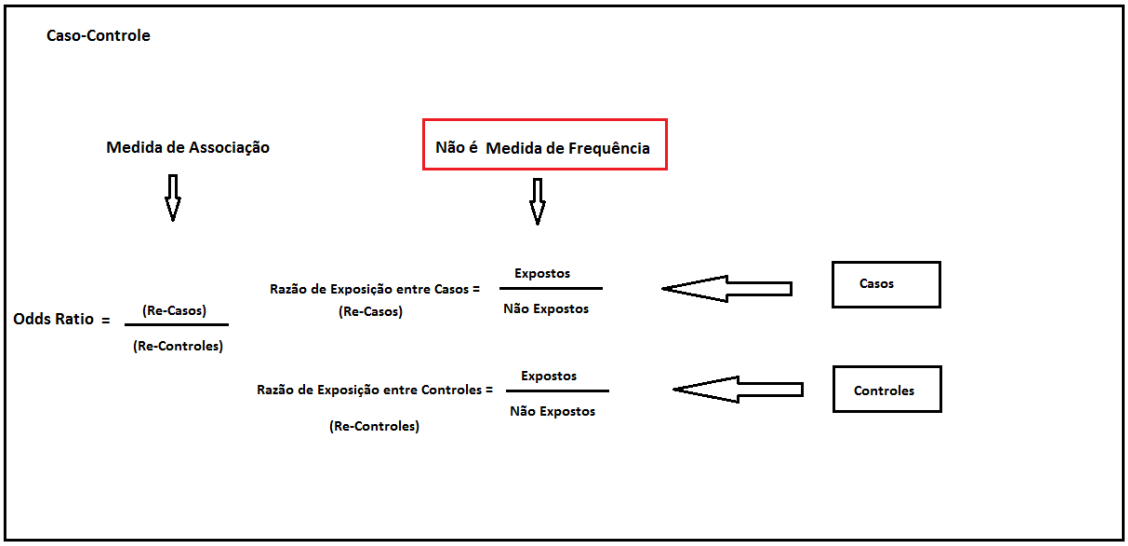
Preste atenção nas figuras que resumem as medidas para estudo de coorte, caso-controle e estudos transversais. Note como é fácil entender e calcular as medidas de frequência e de associação.



mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva





Prevalência e Incidência

Os termos Prevalência e Incidência são utilizados rotineiramente de maneira incorreta em revistas/jornais não científicos e até mesmo em jornais científicos e, o que é mais grave, em algumas revistas de saúde pública. Há necessidade de se diferenciar

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

estes termos porque eles representam **medidas** bem definidos, e cada um destas é medida de forma totalmente diferente. A melhor forma de entender as medidas de frequência de doenças como a prevalência, incidência e taxa é esquecer completamente todos os exemplos que você já ouviu anteriormente.

Existem termos que às vezes são utilizados na vida comum de forma solta e que na ciência têm significados próprios, e assim, devem ser empregados por nós. Prevalência no dicionário do Aurélio significa "caraterística do que prevalece, superioridade, supremacia". Mas para nós prevalência é uma medida de frequência de doenças ou estados de saúde e se refere à porcentagem de casos de doenças (ou qualquer agravo) num determinado momento, enquanto a incidência se refere aos casos incidentes, isto é, novos casos. Logo incidência se refere a novos casos. Se bem que às vezes usamos também a expressão "esta doença é muito prevalente na população", significando que sua prevalência é alta. Tudo bem de falar assim, mas se lembre de que prevalência é também uma medida e como uma medida deve ser calculada da forma que discutiremos a seguir.

A prevalência é dada em porcentagem, não em número. Ela é calculada em estudos transversais apenas, ou em determinado momento mesmo dentro de um estudo de coorte. O fundamental é que prevalência é a porcentagem de indivíduos com o desfecho que está sendo estudado. Por exemplo, podemos fazer um estudo de prevalência de asma em crianças de Ribeirão Preto de 7 a 12 anos de idade no ano de 2014, e descobrir que (números inventados) entre as 5mil crianças examinadas 250 possuiam asma logo a prevalência (medida de frequência) foi de 5%. Concluímos, portanto, *que 5% das crianças de 7 a 12 anos de Ribeirão Preto em 2014 possuiam asma.* A conclusão correta do estudo tem que expressar a faixa etária envolvida, a localidade e o ano em que foi realizado o estudo. Se o estudo fosse realizado em 2013 talvez a prevalência tivesse sido outra, e se envolvesse crianças de 7 a 15 anos também talvez fosse outra. Se apenas meninos tivessem sido observados em 2014 a interpretação correta deveria ser que em *5% dos meninos de 7 a 12 anos de Ribeirão Preto em 2014 possuiam asma.* Se o estudo for realizado com uma amostra, e não com todas as crianças de Ribeirão Preto de 7 a 12 anos de idade no ano de 2014, acrescentamos a

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

interpretação a palavra o termo *"em média"* isso se deve a possível variação amostral que abordaremos posteriomente.

A incidência, no entanto, precisa de estudo de coorte em que se seguem indivíduos expostos por um periodo de tempo. Vamos supor que no início de 2014 examinassemos todas as 5 mil crianças de Ribeirão Preto de 7 a 12 anos, e excluíssimos os casos prevalentes, iriamos ficar com 4750 crianças sem asma. Assim poderíamos seguir essas crianças (sem asma) por um ano e no inicio de 2015 verificariamos quem desenvolveu asma durante o ano. Esses novos casos são casos incidentes. Vamos supor que 100 novos casos tenham surgido, então, a **Incidência Acumulada** neste periodo foi de 100 casos entre 4750 que em porcentagem significa 2,1%. Imaginando manter as mesmas crianças do ano anterior que agora já tem de 8 a 13 anos, a prevalência neste grupo seria de 250 +100 casos de asma, em 5mil crianças, assumindo que nenhuma criança foi perdida. Entao, teríamos a prevalência de 350/5000 = 7%.

Note que a prevalência tem relação direta com a incidência. Se a incidência de novos casos aumenta a prevalência tende a aumentar dependendo se a doença tem ou não cura ou se leva a morte. Como exemplo, nos temos o casao da AIDS. Quando a doença apareceu, não havia nenhum tratamento e as pessoas morriam rapidademente. Hoje em dia diferenciamos infectado pelo HIV de inviduos com sintomas que seriam os que têm AIDS. No inicio, uma vez infectado logo apareceiam os sintomas. Como a doença era transmitida facilmente e as pessoas não sabiam que ela existia, a incidência aumentou muito, porém as pessoas morriam rapidamente. Quanto mais rapido as pessoas morrem menor a prevalência, mesmo que a incidência continue constante. Se as pessoas comecam a viver mais, como aconteceu com os infectados por HIV, a prevalência aumenta mesmo sem aumentar a incidência. Dependendo, mesmo diminuindo a incidência, às vezes com o aumento de longevidade das pessoas com a doença, a prevalência aumenta. Imagine um tanque com entrada e saída de água, se a entrada tiver a mesma velocidade (agua nova) da saída, então a água que vai estar dentro do tanque será sempre a mesma. Se abrimos mais a entrada de água (aumento de incidência) e ao mesmo tempo fecharmos mais a saída, a àgua vai acumular mais (prevalência). E assim, a prevalência vai aumentar. Esta situação seria a situação de aumento de contaminação do virus HIV e inicio de tratamento que começou a manter

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

mais pessoas infectadas na população. Naquele momento, a prevalência aumentou muito. Em seguida mesmo diminuindo a contaminação pelo HIV (diminuição da incidência) por causa da prevenção, com o aumento da sobrevida a prevalência não parou de aumentar. Se olharmos apenas a prevalência podemos pensar que as pessoas continuam a se infectar com velocidade, mas não foi isso que aconteceu. Com o tempo, as pessoas pararam de se preocupar com a prevenção, e a incidência voltou a aumentar nos ultimos anos.

Note então que a prevalência está diretamente associada à incidência e também ao tempo que as pessoas ficam com a doença. A prevalência seria identica à incidência caso a pessoa se infecte com algo e morra em seguida. Nesta situação, mesmo fazendo um estudo transversal teriamos Idea da incidência, mas isso aconteceria apenas teoricamente. Nos livros vamos encontrar de forma simplória a fórmula Prevalência = incidência versus o tempo de sobrevida da doença (P = I x T). Isso quer dizer apenas que prevalencia está diretamente relacionada à incidência e ao tempo da doenca. Esta fórmula não é tão simples assim. O livro do Kleimbaum et al 1982, traz uma dedução bem mais realista desta relação, mas esta relação simples serve para entendermos o processo.

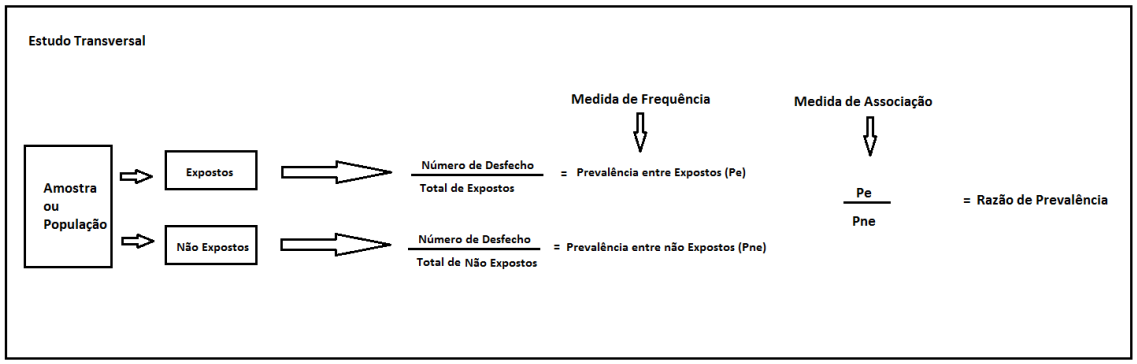
Razão de Prevalência – medida de associação

Num estudo de transversal em que se quer saber se meninos tem mais asma do que meninas, ou se crianças pretermo term mais asma do que crianças atermo (não pretermo), podemos então calcular duas medidas de frequência, uma prevalência para pretermos (expostos) e uma prevalência para atermos (não expostos) e dividir uma pela outra. Esta razão entre duas prevalências de expostos sobre não expostos é chamada de **Razão de Prevalência.** A razão das duas prevalências é uma medida de associação, associação entre pretermo (exposição) e asma (desfecho). A razão de duas coisas obviamente não pode ser uma porcentagem porque cada prevalência é medida num grupo específico. Quando a razão de prevalência for igual a 1 quer dizer que a prevalências entre os dois grupos de expostos e não expostos são iguais. Se for maior

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

que um é que existe associação positiva entre ser exposto à prematuridade e ter maior prevalência de asma. Se for menor do que um, quer dizer que ser exposto à prematuridade tem associação negativa com a prevalência de asma.

Vamos supor que a razão de prevalência obtida no estudo para estudar prematuridade e asma fosse de 3,7. Como interpretariamos corretamente este valor? Considerando que o estudo foi realizado entre crianças de Ribeirão Preto de 7 a 12 anos de idade no ano de 2014, a interpretação correta é " *a prevalência de asma entre prematuros foi 3,7 vezes que a prevalência de asma entre não prematuros em crianças de 7 a 12 anos em Ribeirão Preto no ano de 2014*". Se for resultado de uma amostra acrescenta-se o termo "....em média 3,7 vezes...."

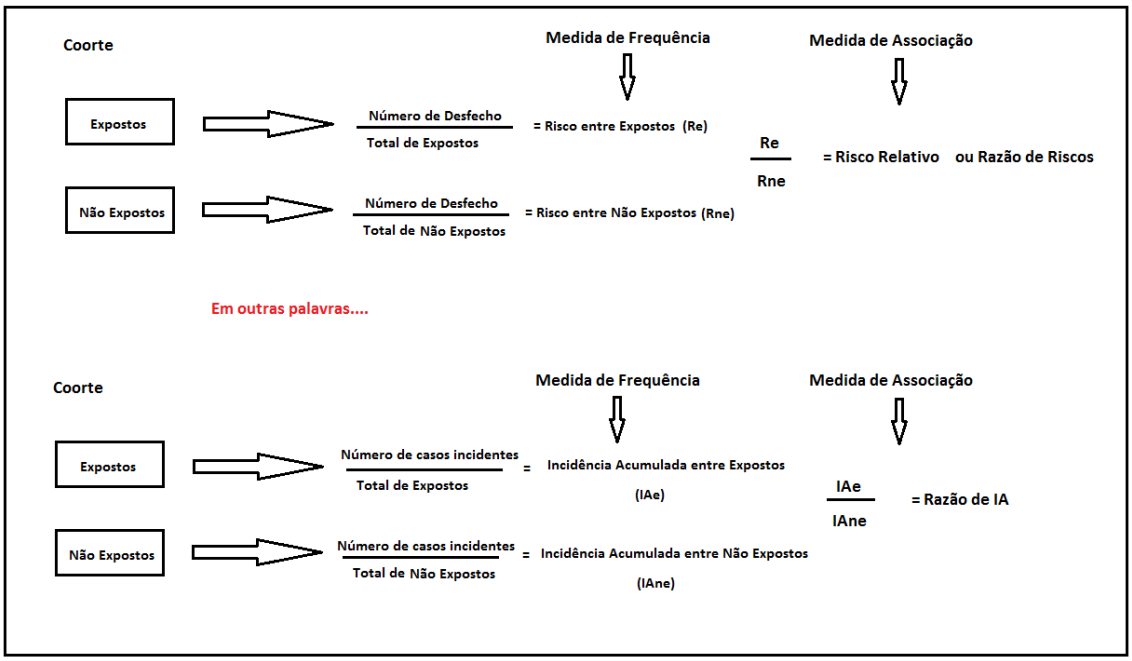


Num estudo de coorte começamos com pessoas expostas e não expostas e nenhum indivíduo tem no tempo zero a doença. Com o passar do tempo os casos incidentes são anotados e ao final do estudo temos a **incidência acumulada** no período do estudo. Vamos a um exemplo, imaginemos uma coorte que acompanhou 1000 crianças pretermo e 3000 atermo desde o nascimento até os 5 anos de idade, e foi constatado que entre as pretermo 150 desenvolveram asma e entre as atermo 150 desenvolveram asma. A incidência acumulada entre os pretermo seria de 150/1000 = 15%. Dizemos então que a "*incidência acumulada de asma em crianças pretermo foi em média de 15% no período de 5 anos apartir do nascimento".* Utilizamos ém média assumindo que temos uma amostra. Ainda podemos dizer que "*a probabilidade média de uma criança pretermo desenvolver asma no periodo desde o nascimento até os 5 anos de idade foi de 15%".* Ainda, podemos utilizar a palavra RISCO. O "*risco médio de uma*

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

criança pretermo desenvolver asma no periodo desde seu nascimento até os 5 anos de idade foi de 15%. A palavra risco quer dizer **probabilidade média de vir a desenvolver um evento num determinado periodo de tempo.** Portanto, risco somente pode ser calculado em estudos de coorte! Em nenhum outro estudo podemos estimar risco. Um estudo de prevalência não tem estimativa de risco, nem mesmo num caso controle. Então sempre que ver um estudo publicado dizendo que cálculou o risco, veja se é um estudo de coorte, caso contrário, está errado!

Resumindo risco (também chamado de incidência acumulada) é a medida de frequência que se calcula num estudo de coorte. Existe outra medida de frequência nos estudos de coorte que é a taxa de incidência, porém somente vamos falar sobre ela depois. Quando se faz a razão entre duas incidências acumuladas (ou dois riscos , que eh sinonimo) temos a **razão de incidência acumulada ou também chamado razão de riscos ou ainda risco relativo.**



A interpretação do Risco Relativo, ou Razão de Incidência Acumulada, segue o mesmo padrão. Se entre os expostos o risco era de 15 %, entre os não-expostos 150/3000 era de 5 %. Assim, o risco relativo foi igual a 3 e a interpretação correta é que "*em méia o risco de vir a desenvovler asma em uma criança pretermo desde o*

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

nascimento até os cinco anos de idade é 3 vezes o risco de desenvolver asma entre crianças atermo". Note que sempre precisamos especificar o tempo, porque se fosse um periodo mais longo a incidência acumulada seria diferente e o risco relativo também poderia ser muito diferente.

A noção de risco e risco relativo é sempre empregada em medicina e na prática clínica de odontologia. Embora se fale muito na clínica sobre risco de cárie dentária, e existam questionários para se estimar o risco do indivíduo ter cárie, é interessante notar que pouquissimos estudos de coorte foram realizados até hoje. Isso acontece porque até o ano de 1989, a odontologia não tinha ideia do que era estudo de coorte e muito menos ideia do que era risco apropriadamente. Assim, o conceito até hoje que muitos dentistas têm de risco não é o correto.

Apenas relembrando risco (ou incidência acumulada) é medida de frequência, e risco relativo (ou razão de incidência acumulada) é medida de associação.

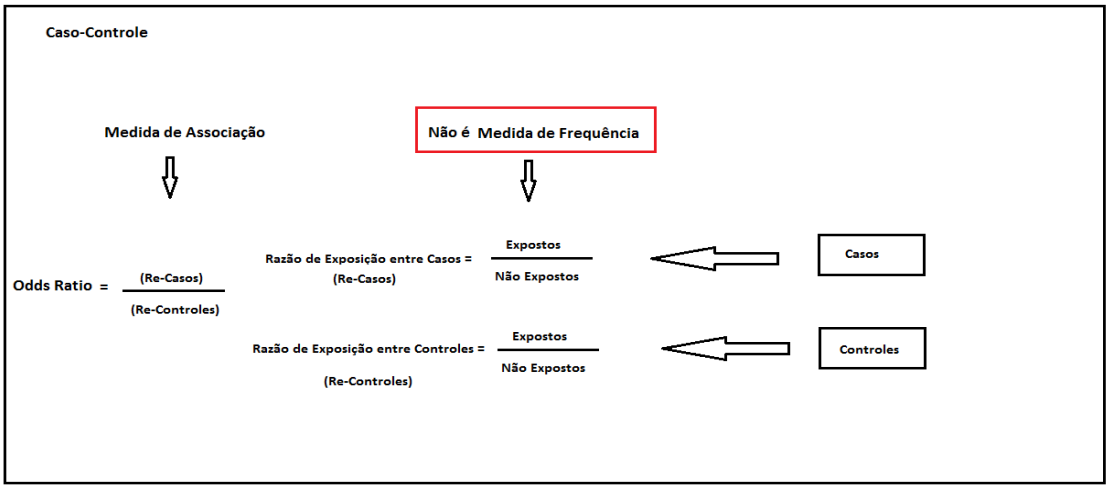
Uma questão importante é a interpretação da magnitude do Risco Relativo. Embora 1,7 de risco relativo pareca ser pouco, significa que o grupo de expostos tem 70% a mais de pessoas doentes do que no grupo de não expostos. Interessante quando o RR é abaixo de 1. Um risco de 0.90 significa que o risco com uma certa exposição foi reduzido em 10% e um risco de 0,25 significa redução comparativa de 75%. Mas agora não estamos interpretando em % de individuos a mais apenas redução relativa do risco. Nesse exempo de 0,25 se tivessmos o inverso estariamos diante do equivalente a 1/0,20 = 5, que seria um aumento de 4 vez ou seja 400% a mais de pessoas afetadas. No entanto, quando temos proteção não podemos ter mais do que 100% de proteção. Assim, um RR de 0.001, significa uma redução comparativa de 0.999 %.

Odds Ratio- medida de associação

Nos estudos de caso-controle temos uma diferença crucial com os estudos de prevalencia (transversal) e de coorte. Tanto no transversal como na coorte conseguimos reunir grupos homogeneos de indivíduos expostos e não expostos, mas no caso-controle não. No caso controle, escolhemos indivíduos com desfecho que são os casos, e buscamos indivíduos que servirão de comparação. Não temo todos os casos da população e todos os não casos. Desta forma, não podemos calcular medida de

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

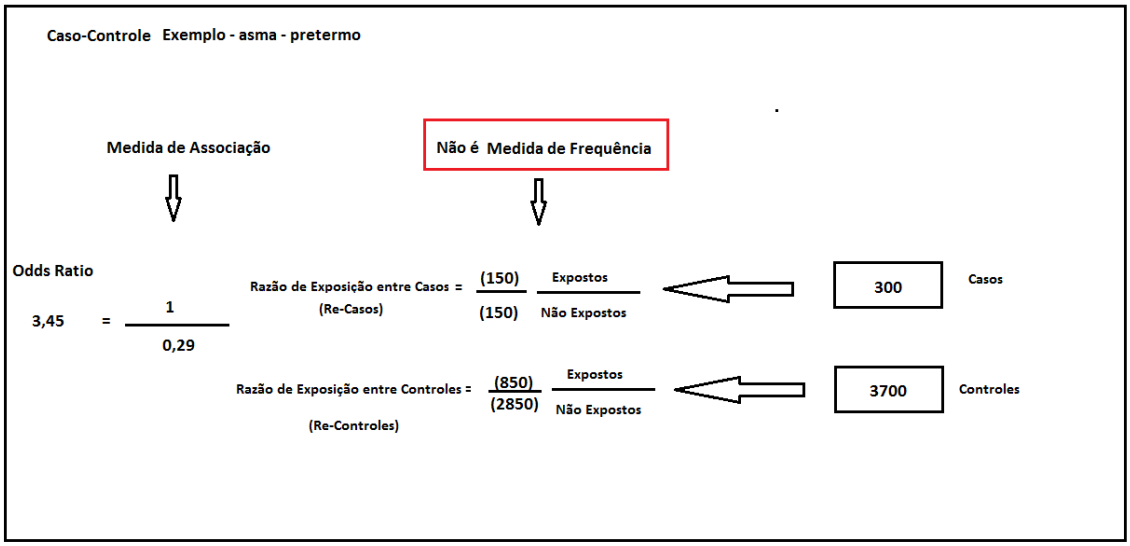
frequência. Mesmo num caso-controle aninhado numa coorte, vamos incluir no caso-controle os casos que desenvolverem a doenças e uma amostra de controles especificamente para os casos. Logo temos "um punhado de gente" sem ter um grupo homogeneo. Assim, não podemos calcular uma medida de frequência (que seria uma porcentagem) calculamos no máximo uma razão que é a razão de exposição dos casos e a razão de exposição entre os controles. Veja o próximo quadro sobre caso-controle.



No caso controle o que podemos calcular então é apenas a razão de expostos sobre não expostos para casos e razão de expostos sobre não expostos. Se transformassemos a o estudo de coorte hipotético descrito acima em um caso controle teríamos 150 casos dos expostos e mais 150 casos dos não expostos, portanto teríamos 300 crianças com asma aos 5 anos de idade e logo 3700 crianças sem asma que seriam nossos controles. No exemplo a razão de exposição entre casos foi de 1, e a razão de exposição entre controles de 0,29. A divisão das duas razões resulta no que nós chamamos de Odds Ratio (razão de razões de chance) que foi igual a 3,45.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva



Como interpretamos corretamente esta Odds Ratio? *A razão de exposição de pretermos e atermos entre casos é 3,45 vezes a razão de exposição de pretermos e atermos entre os controles entre crianças de cinco anos de idade.* A resposta é simples, é só ler o que a figura do quadro está mostrando, nada mais. Não precisamos de tempo nem nada, a nao ser especificar a população onde foi realizada que no caso incluiu crianças de 5 anos de idade.

O erro mais comum é a interpretação errada da odds ratio num caso controle assumindo, por exemplo, que os prematuros são 3,45 vezes mais prováveis de terem asma. Isso não é verdade! O estudo caso-controle parte dos casos e vai em direção a exposição.

Quando interpretamos o RR dissemos por exemplo que um RR de 1,7 significa que temos 70% a mais de pessoas doentes no grupo de expostos. No entanto, quando interpretamos uma OR de 1,7, o excesso de 70% está em relação não a pessoas, mas em relação a razão de exposição. Assim, 1,7 significa que a razão de exposição entre casos é 70% maior do que a razão de exposição entre os controles. Assim preste atenção às interpretações.

Podemos calcular odds ratio nos estudos de coorte e transversais? Calcular podemos, mas não é a melhor escolha, se pudermos calcular razão de prevalência nos estudos transversais ou risco relativo para os estudos de coorte para que vou calcular odds ratio? Note que com os mesmos números quando calculamos o risco relativo ele

foi igual a 3 , no entanto, ao calcular odds ratio chegamos a um número maior de 3,45. A odds ratio em geral fornece um número "inflado" em relação ao risco relativo e a razão de prevalencia que se baseiam em porcentagens. Por que isso acontece? Imagine um grupo de 5 crianças sendo um menino e 4 meninas, temos então 1/5 = 0,20 ou seja 20% de meninos. Se usarmos a razão de 1 menino para 4 meninas e dividirmos estes números teremos o resultado de 0,25. Na verdade este 0,25 não é uma porcentagem, é apenas o resultado da razão 1:4. No entanto o número 0,25 é maior que 0,20. No final, ao calcular razão de razões o número resultante será em geral maior do que em razões de porcentagens. Você pode pensar em odds também em termos de razão de duas probabilidades 1:4 também é igual a probabilidade de ser menino (0.20), sobre a probabilidade de ser menina (0,8), sendo que 0,20/0,80 = 0,25. Pode-se dizer que odds é a razão de probabilidade complementares mas diferentes.

Portanto, embora possamos calcular odds ratio em outros estudos não existe nenhuma vantagem, a não ser que alguma estatística complexa não possa ser feita para o cálculo de razão de prevalência ou risco relativo o que vai exigir utilizar a odds ratio. Nestes casos, é bom especificar que para o estudo transversal foi utilizado a **odds ratio prevalência** e para o estudo de coorte a **odds ratio de desfecho.** Mas para efeito de aprendizagem inicial jamais vamos utilizar estes termos em provas. Qual a diferença de odds ratio de desfecho e odds ratio de exposição? No caso controle utilizamos e interpretamos a odds ratio de exposição (razão de exposição entre casos e controles), já no estudo coorte seria a odds ratio de desfecho, isso é a razão de desfecho entre expostos e razão de desfecho entre não expostos.

Odds ratio para estudos caso-controle pareados

Um detalhe, a odds ratio que calculamos acima é a odds ratio simples, considerando que temos um grupo de casos e um grupo de controles, mas não pareados. Por exemplo, posso ter escolhido um caso do hospital, e procuro um controle em geral sem parear por sexo ou idade, o que é hoje em dia bastante incomum. Mas imaginando esta situação de não pareamento é como se calcula a odds ratio de forma geral. Na verdade quando pareamos cada caso com um controle, o cálculo da odds ratio

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

não é tão simples assim. A tabela é montada de maneira diferente, mas não vamos entrar em detalhes aqui sobre isso. A Idea da apostila é simplificar. Vou apenas mencionar para que os curiosos entendam. Se nos pareamos casos e controles o que queremos ver é quantos pares são discordantes ou concordantes. Quantos pares de casos e controles são concordantes e quantos discordantes. Vamos imaginar 500 pares de casos e controles (note que na tabela não temos os mil indivíduos descriminados como expostos e não expostos temos os pares).

		Controles	
		Expostos	Não Expostos
Casos	Expostos	100	250
	Não Expostos	50	200

Nesta situação de pares, o que nos importa são os pares discordantes e não os concordantes. Quer dizer se tanto casos e controles são expostos, isso não contribui em nada para a associação, quer dizer não haveria associacao se todos os casos e controles concordarem em relação a exposição. Assim os discordantes nos dizem algo. No exemplo cima notamos que em 250 pares os casos são expostos e os controles não. E em apenas 50 pares os casos não sao expostos enquanto os controles sao. Logo 250/50. A odds ratio em estudos pareados é então calculada desta forma 250/50 que resulta é OR de 5. Assim a continha vira discordância de casos expostos e controles não expostos sobre discordância de casos não expostos com controles expostos.

Preste atenção na composição da tabela.

		Controles	
		Expostos	Não Expostos
Casos	Expostos	+ +	+ -
	Não Expostos	- +	- -

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

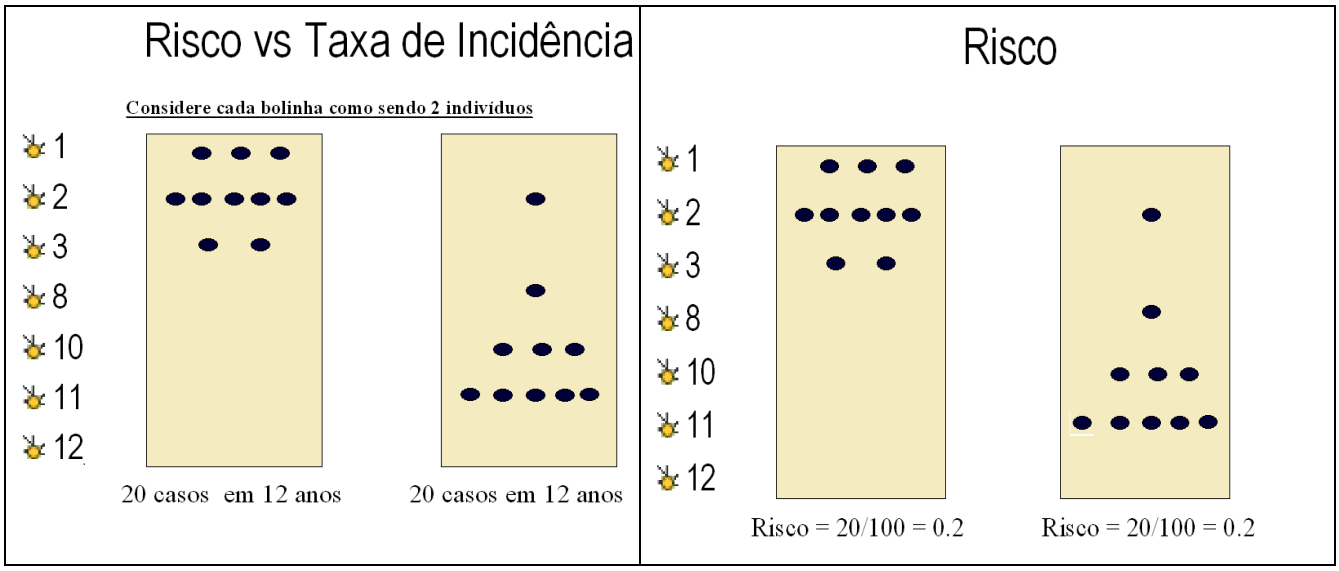
Outro detalhe é que se o pareamento foi realizado em relação a idade e sexo, não temos como avaliar o efeito da idade e do sexo na associação entre exposição e desfecho. Isso porque idade e sexo já que os casos e controles foram pareados não pode ser considerado mais um fator de confusão!

Antes de prosseguir para o estudo da medida chamada Taxa, faça os exercícios do Anexo B.

Taxa de Incidência ou Densidade de Incidência

A medida de risco é a mais utilizada na literatura para estudos de coorte. No entanto, por vezes o risco não nos serve tirar conclusões sobre o verdadeiro efeito de um fator etiológico numa doença.

Vejamos o exemplo abaixo.



mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Imagine que as colunas sejam duas populações, uma que consome suco de laranja e outra que não consome. Imagine também que temos 100 indivíduos com uma doença tipo HIV/AiDS. Os números 1 a 12 se referem aos anos de estudo. Ao final dos 12 anos, observamos que 20 indivíduos morrem da doença em cada uma das populações. O risco, portanto, é o mesmo para as duas populações.

No entanto, ao observar as duas figuras você preferiria estar na população da direita ou da esquerda? Bom na população A que não recebeu a droga em estudo, as pessoas morreram mais rápido. Este efeito de velocidade em que a morte aconteceu na população não pode ser observado por meio do risco. Se estivessemos avaliando a droga num estudo, utilizando risco, concluiríamos que a droga não foi eficaz, mas na verdade retardou a morte dos indivíduos. Portanto, nem sempre o risco é uma medida correta para se avaliar um fator etiológico. Risco serve para dar uma ideia de probabilidade de algo acontecer num certo período de tempo.

Portanto, calcular o risco não foi adequado para o estudo. Precisamos de uma medida que expresse esta velocidade em que o desfecho, morte no nosso exemplo, atinge a população. Note velocidade média é denominado de taxa na física e assim empregamos este termo neste momento, taxa da ideia de velocidade de fluxo. Esta medida também é chamada de densidade de incidência, porque se prestar atenção existe uma diferença de densidade de mortalidade na figura A em relação a B. Na figura A as mortes estão mais concentradas nos primeiros anos e na figura B mais dispersas.

Para montar a figura acima foi necessário saber quando exatamente cada morte aconteceu, caso contrário não poderiamos fazer os desenhos acima e verificar que a velocidade foi diferente. Por exemplo, se estamos estudando incidência de cárie num estudo de coorte, mas somente examinamos as crianças depois de 10 anos não podemos estimar esta velocidade. Para tanto, deveriamos examinar as crianças todos os anos ou no máximo a cada dois anos. Nesta situação da cárie é difícil precisar exatamente o momento de aparecimento da mesma, porque em geral procura-se tratamento quando a lesão já está visivel. Se fizermos exames anuais, utilizaremos como unidade de tempo o ano, se for bianual podemos dizer que a cárie desenvolver no meio dos dois anos.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Observando as figuras acima podemos dizer que os indivíduos da população A morram mais rápido do que na população B, ou ainda podemos dizer que a sobrevida na população B foi maior. Assim, podemos pensar em comparar a sobrevida dos dois grupos. Para tanto, somaremos a quantidade de anos que cada indivíduo sobreviveu em cada população. O tempo contado na sobrevida é o quanto o indivíduo realmente estava vivo, assim se um indivíduo morre no último dia do ano, eu tenho certeza que ele estava vivo 364 dias. Para simplificar da figura acima que está em anos, vamos considerar que indivíduo morto na primeira linha referente a um ano, tenha morrido um dia depois de um ano e assim por diante. Portanto, contamos o ano inteiro para ele.

Assim, a sobrevida na população A (998) será menor do que na população B (1095). O cálculo da sobrevida é a soma da sobrevia de cada indivíduo. Temos 100 pessoas e sabemos que 80 sobreviveram até o final do estudo tanto para a população A como para a B, logo 80 x 12 anos = 960 anos em cada grupo. Na população A seis pessoas (cada pontinho vale duas pessoas) sobreviveram apenas 1 ano (6 x 1), 10 pessoas 2 anos (10 x 2), e quatro pessoas três anos (4 x 3). Somando a sobrevida dos 80 (960 anos) mais a sobrevida dos demais (6 + 20 + 12 = 38) temos 960 + 38 = 998. A mesma conta de sobrevida deve ser feita para a população B cuja sobrevida foi maior totalizando 1095 anos.

A taxa é calculada dividindo o número de desfechos pelo tempo calculado para todos. Logo a taxa de incidência para a população A é igual a 20/998 = 0,020 e para a população B 20/1095 = 0,018. Note que isso não pode ser uma porcentagem porque no numerador temos o número total de casos incidentes, ou seja, a incidência acumulada, mas no denominador temos outra medida que é tempo de sobrevida de todos. Assim a interpretação correta para da taxa de incidência *é que para a população A temos 0.020 mortes por pessoas-ano no período de 12 anos*, *enquanto na população B termos 0.018 mortes por pessoa-ano*. O termo pessoa-tempo (no caso pessoa-ano) é sempre utilizado para se referir a sobreviva total da população, o termo parece estranho, mas é isso que ele representa são os anos por pessoa. A taxa tem a ver com a velocidade que os desfechos ocorretam, notem que na população A, a velocidade foi maior. Se utizassemos o risco, não conseguiriamos demonstrar esta diferença que é facil de ser observada. Assim taxa é uma medida de frequência, quando fazemos a razao de duas

taxas, temos a **Razão de Taxas**, que pode tambem ser chamada de **Taxa Relativa, ou Razão de Densidade de Incidência.** Neste exemplo 0,018/0,020 = 0,9. A interpretação correta deste valor é *a taxa de mortalidade na população B que recebeu o tratamento foi em média 0,9 vezes a taxa de mortalidade da população A num periodo de 12 anos.*

Ainda temos mais uma forma de interpretar a taxa que é como força de mortalidade ou de morbidade. Se estudamos morte dizemos força de mortalidade e se for uma doença, dizemos força de morbidade. Assim podemos dizer *que força de mortalidade na população B foi em média 0,9 vezes a força de mortalidade na população A num período de 12 anos*. Ainda alguns chamam esta medida de pontencial de impacto, pois descreve o impacto que a doença teria na população. No nosso exemplo o impacto na população A foi maior do que na população B.

Note que pessoa-tempo tem a ver com sobrevida do indivíduo e não com tempo de exposição. A exposição é definida no início do estudo. Se as pessoas do grupo de expostos têm exposições diferentes, então, temos que calcular a taxa para cada grupo com exposição diferente.

Vamos a outro exemplo de taxa de cárie dentária em um grupo de 10 crianças acompanhadas por 12 meses sendo que uma desenvolveu cárie no terceiro mês, e duas outras no oitavo mês o que foi observado quando foram examinadas por um dentista. As demais crianças não desenvolveram cárie durante todo o estudo. Desta forma, a criança que foi diagnosticada com cárie no terceiro mês, passou 2 meses com certeza sem cárie. As duas que desenvolveram cárie no oitavo mês, passaram certamente 7 meses sem cárie. As demais crianças passaram certamente os 12 meses livres de cárie. Assim, computamos como pessoa-tempo 2 + 7 + 7 + (7 x 12) = 100 que neste caso é pessoa-mês. Como o número de casos incidentes de cárie foi igual a 3, então a taxa de incidência é a divisão entre o número de casos incidentes e pessoas tempo , isto é, 3/100 = 0,03. Dizemos, então que a força de morbidade da doença, ou a *taxa de incidência média da doença é de 0,03 pessoas com cárie por pessoa-mês durante o período de 12 meses.*

Algo importante é que quando consideramos os indivíduos expostos a alguma coisa, imaginamos que eles sejam expostos de forma semelhante. Ao classificar um grupo de fumantes e não fumantes em uma coorte nos temos que ter certeza que os

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

fumantes tenham exposições semelhantes. Por exemplo, se os indivíduos classificados como fumantes incluem tanto pessoas expostas há 30 anos fumando 20 cigarros por dia como fumantes expostos há 3 anos consumindo 1 cigarro por dia, teremos viés de informação. Se isso acontecer teremos que classificar adequadamente os expostos em grupos de diferentes padroes de consumo de cigarro.

Até o momento os exemplos incluíram coortes imaginárias em que todos os indivíduos foram incluídos ao mesmo tempo, e todos permanecem no estudo até o final, exceto os que desenvolveram o desfecho. Lembre-se ao desenvolver o desfecho o indivíduo não é mais seguido porque já contribuiu com a informação sobre incidência.

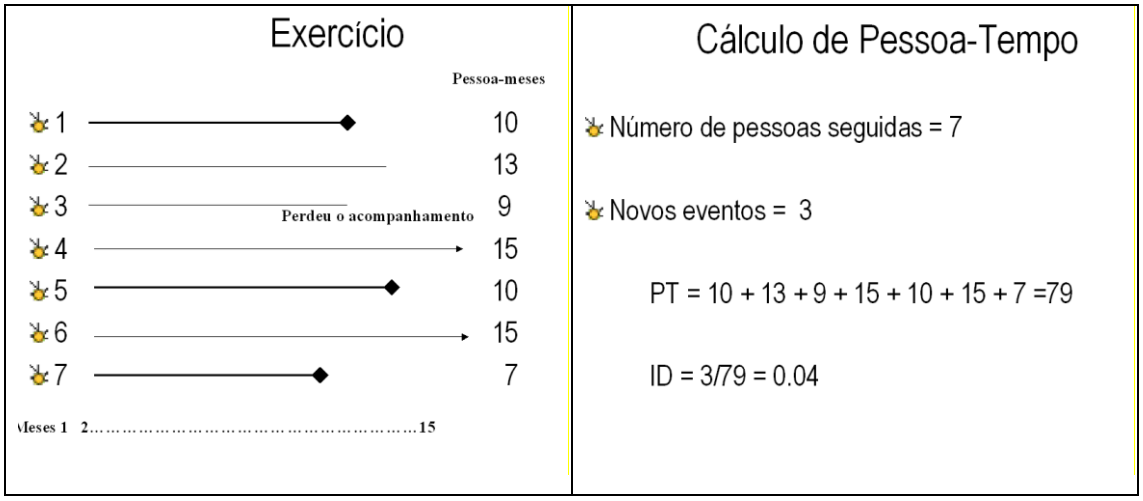
No entanto, nem sempre as pessoas permanecem no estudo por todo o tempo até desenvolverem o desfecho, às vezes as pessoas desistem do estudo ou saem porque se mudam para outros lugares ou morrem de outras causas. Nesta situação de desistência seja por que motivo for, podemos aproveitar na contagem da taxa de incidência a sobrevida até o momento que a pessoa participou do estudo. Por exemplo, se uma pessoa desiste de uma coorte e a última vez que foi examinada foi no terceiro ano e ela não possuia o desfecho, então ela contribui com três anos para o pessoa-tempo. Assim, aproveitamos ao máximo a informação dada pela pessoa que desistiu. Quando calculamos risco, se a pessoa desiste apenas a eliminamos do cálculo. Se o estudo começou com 20 pessoas e 5 desistiram, o cálculo do risco vai ter 15 pessoas no denominador e não vinte. A perda de indivíduos tem potencial de introduzir viéses num estudo, se apenas as pessoas mais sadias sairam do estudo porque estavam cansadas de participar, a taxa de incidência para aquele grupo será maior do que deveria ser. Isso será um problema especial principalmente quando a perda acontecer de forma diferente entre os grupos de expostos e não expostos.

Além da desistência nem sempre as pessoas começam na coorte no mesmo momento. È possível que as pessoas sejam inseridas num estudo em tempos diferentes e seu tempo no estudo passa a ser contado apartir daquele dia. As coortes cujos participantes começam no mesmo momento são chamadas de coortes fixas, já aquelas que permitem entradas em momentos diferentes são chamadas de coortes dinâmicas.

Na figura abaixo mais um exemplo de cálculo de taxa.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva



Embora estejamos utilizando exemplos de epidemiologia esta noção de medidas de frequência e de associação são utilizadas em todas as áreas. Vamos supor que você esteja estudando a ação de uma substância na regeração de tecidos, com técnicas *in vitro*, isto é estudando proliferação celular. Poderemos determinar que nosso evento final de observação seja o número de células que se formaram num período de 3 semanas, isto é, estaríamos medidando qual a porcentagem de células finais que poderiamos traduzir na probabilidade de geração de novas células ou risco de formação de novas células. Ao final de 2 semanas pode ser que nossa substância teste não seja diferente da outra, mas talvez a formação inicial destas células seja maior no grupo teste do que no controle. Neste caso estamos falando de velocidade de formação das células. Podemos então expressar esta característica através da taxa de formação celular. Se pesquisador calcular apenas a formação celular no final das duas semanas, vai perder a informação da velocidade de proliferação que pode ser importante para definir qual a melhor produto.

Relação entre incidência e prevalência

Sempre se pergunta como a incidência se relaciona com a prevalência. Todas as suposições entre estas medidas são feitas considerando-se o que se chama de *steady state,* ou estado de equilibrio, isto é que tudo seja constante, que os novos casos sejam

gerados na mesma velocidade por ano, e que a mortalidade ou saída de pessoas da população também seja constante. Caso isso não aconteça fica mais difícil falarmos sobre a relação destas medidas de associação. De maneira bem simples, vamos imaginar que uma população esteja sem nenhum caso doente e é mês de Janeiro, mais precisamente primeiro de Janeiro. Certa doença começa a acometer a população numa taxa de 1% ao dia. Considerando-se a população inicial de 1000 sadios vamos ver o que acontece se a taxa permanece constante. Note que por vezes teremos número de doentes fracionados como 9.1. É claro que não existe "0.1 de uma pessoa doente", mas para efeito de cálculos assim apresentaremos.

Data	Total de Indivíduos	Sadios	Doentes no dia	Total de Doentes	Prevalência
1 Janeiro	1000	1000	10	10	1%
2	1000	990	9,9	19.9	1.99%
3	1000	980,1	9,801	29,701	2.9%

Note que a prevalência vai aumentando com o tempo, caso nenhum indivíduo se cure ou morra. Desta forma a duração do indivíduo doente na população é um fator importante para a prevalência da doença. A prevalência então depende da taxa de incidência e da duração da doença na população. Esta relação em geral é expressa de maneira simplidicada como Prevalência = incidência x a duração da doença. Esta relação não é tão simples assim, mas o mais importante desta fórmula que encontramos na maioria dos livros é dizer que a prevalência depende diretamente da incidência da doença e duração da mesma. Ela também vai ser inversamente proporcional à taxa de mortalidade geral na população ou evasão de pessoas.

Na tabela anterior, podemos notar também que com a diminuição das pessoas sadias o número total de pessoas que ficam doentes a cada dia diminui. Começamos com 10 pessoas, e no quarto dia tinhamos 9.801 pessoas doentes novas. Mas se lembre de que a taxa é a mesma. No entanto, esta situação é de equilíbrio da população, e ainda o doente permanece com a doença. O cenário seria completamente diferente em face

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

de uma doença em que as pessoas vão se curando e ou adquirindo imunidade. Se a pessoa se cura, e a doença não leva a imunidade, logo que ela se cura, ela volta para o grupo de sadios passível de se infectar de novo. Se ela passa a ser imune então o cenário será diferente, embora a pessoa imune entre para a composição de prevalência ela não irá mais entrar para a composição da incidência acumulada. De forma simples, essa é a relação entre incidência e prevalência, discutiremos esta relação com mais detalhes nos próximos capítulos. Se quiser entender melhor e brincar continue a fazer a tabela acima por mais 30 dias.

Diferenças Fundamentais entre Risco e Taxa

- Risco
  - Refere-se a indivíduo
  - Útil para decisões clínicas
  - É uma probabilidade logo varia de 0 a 1
  - Unidade é o tempo
- Taxa
  - Refere-se a uma população
  - Útil para comparar populações
  - Varia de 0 a infinidade
  - Unidade é pessoa-tempo

Regras Gerais para escolha de Medidas de Frequência

- se vc quer relatar proporção de indivíduos que tem um determinado estado de saúde use PREVALÊNCIA
- se ve quer relatar a probabilidade de uma pessoa na média ter um evento num período de tempo use o INCIDÊNCIA ACUMULADA - RISCO- (predição individual)
- se vc quer relatar o impacto da doença numa população , ou grupo de pessoas então use a DENSIDADE DE INCIDÊNCIA (inferência etiológica).

Posso interpretar Prevalência como se fosse Risco?

- Lembre-se que risco necessita ser medido por um período de tempo. Portanto, prevalência NÃO pode ser interpretada como risco.
- Entretanto, para dados de uma idade específica, nos podemos derivar aproximações de risco quando a doença for irreverssível & a doença não afetar a probabilidade de morrer & assumindo-se ao mesmo tempo uma condição de "steady-state" (ex: glaucoma)

$$R_j = (P_{(j+1)} - P_j) / (1 - P_j)$$

Ex:

$$Risk_{30yrs\text{-}old} = (P_{31yrs\text{-}old} - P_{30yrs\text{-}old}) / (1 - P_{30yrs\text{-}old})$$

$$Risk_{30yrs\text{-}old} = (0.2 - 0.15) / (1 - 0.15)$$

$$= (0.05) / (0.85) = \mathbf{0.0588}$$

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Quando a Prevalência é proxima da Densidade de Incidência? Apenas quando a duração da doença for muito pequena & a prevalência também for muito pequena.	

Taxa de Mortalidade e Taxa de Morbidade

Existe uma medida chamada de taxa de mortalidade que é explicada na maioria dos livros e recebe este nome de taxa. Baseado no que explicamos, taxa de mortalidade deveria refletir a velocidade com que a população morre, e, portanto para seu cálculo um estudo de coorte deveria ser realizado. No entanto, a taxa de mortalidade que em geral é reportado em livros e estudos, na verdade não é uma taxa, apenas recebe o nome erradamente por vício de se falar "taxa de mortalidade". Em geral a taxa de mortalidade é medida registrando-se número de mortes em um período, em geral ano, que é dividido pela estimativa do número de pessoas vivas no mês de julho de acordo com órgãos oficiais como o IBGE que estimam estes números. Logo a taxa de mortalidade é apenas uma proporção cujo denominador é aproximado por meio das estimativas de algum órgão de pesquisa do governo. No livro do Rothman et al, a taxa de mortalidade já é denominada de proporção de mortalidade. Eu considero mais apropriado utilizar o termo proporção de mortalidade, mas o vício de dizer taxa de mortalidade é grande. Apenas saiba que quando ler o termo taxa de mortalidade ou morbidade na verdade não é uma taxa, é apenas uma proporção!

Por vezes você verá o termo taxa de incidência relacionada à algumas doenças em Boletins Epidemiológicos como o Boletin de Sífilis do Governo. Como dissemos anteriormente, taxa é uma medida de velocidade de eventos que acontecem num período de tempo. Leia o parágrafo em itálico que foi copiado do Boletin Epidemiologico de Sífilis de 2020.

"Neste novo Boletim Epidemiológico, pode-se observar que a sífilis adquirida, agravo de notificação compulsória desde 2010, teve uma taxa de detecção de 72,8 casos por 100.000 habitantes, em

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

2019. Também em 2019, a taxa de detecção de sífilis em gestantes foi de 20,8/1.000 nascidos vivos; a taxa de incidência de sífilis congênita, de 8,2/1.000 nascidos vivos; e a taxa de mortalidade por sífilis congênita, de 5,9/100.000 nascidos vivos. Assim como no ano anterior, nenhuma Unidade da Federação (UF) apresentou taxa de incidência de sífilis congênita mais elevada que a taxa de detecção de sífilis em gestantes, o que pode refletir a melhora da notificação dos casos de sífilis em gestantes no país."

Em geral utilizam o termo taxa por vezes com o mesmo tipo de cálculo que discutimos sobre mortalidade. Aqui se comenta sobre "taxa de detecção", é calculada com as notificações compulsórias de sífilis num certo ano e é expressa em casos por 100.000 habitantes. Certamente o número de notificações foi dividido pelo número de habitantes estimados para Julho daquele ano, e expresso na razão por 100 mil habitantes. Note que ao comentar sobre a taxa de sífilis em nascidos vivos, possívelmente são os casos detectados em nascidos vivos. A estimativa de nascidos vivos pode ser que seja a do final do ano, pois existem dados reais no SINASC, Sistema Nacional de Nascidos Vivos. Neste caso se o denominador for o número de nascidos vivos naquele ano, estamos diante de incidência acumulada em nascidos vivos. Uma proporção (que ao ser reportado por 100 mil nascidos) é denominada de taxa embora não seja uma taxa verdadeira.

As estimativas realizadas pelo governo por vezes obedecem a metodologias específicas. Veja particularidades do cálculo de Taxa de incidência no site do Ministerio da Saúde < HTTP://idsus.saúde.gov.br/ficha16s.html> acesso em 22/06/2021 sobre "taxa de Incidência de sífilis congênita em residentes menores de 1 ano".

Conceituação: número de casos novos de sífilis em menores de um ano residentes em determinado município por nascidos vivos de mães residentes do mesmo município, no período considerado.

Interpretação: "expressa a qualidade do pré-natal, uma vez que sífilis pode ser diagnosticada e tratada ao longo do período de gestação". Fonte Utiliza o SINAN (Sistema Nacional de Agravos de Notificação) e o SINASC.

Limitações "Depende das condições técnico-operacionais do sistema de vigilância epidemiológica, em cada área geográfica, para detectar, notificar, investigar e realizar testes laboratoriais específicos para a confirmação diagnóstica da sífilis em gestantes e recém-nascidos;

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Demanda cautela na análise de séries temporais, pois deve considerar o processo de implantação do sistema de notificação na rede de serviços, a evolução dos recursos de diagnóstico (sensibilidade e a especificidade das técnicas laboratoriais utilizadas) e o rigor na aplicação dos critérios de definição de caso de sífilis congênita."

$$= \frac{Número\ de\ casos\ de\ sífilis\ congênita\ em\ menores\ de\ 1\ anos, residentes\ municipio\ no\ período}{Número\ de\ nascidos\ vivos\ mães\ residentes\ no\ municipio, no\ período\ considerado}$$

Nesta situação se assemelha mais ao risco do que a Taxa embora o MS se refira a taxa.

Medidas de Impacto em Potencial

Alem das medidas de associação existem medidas que são denominadas de medida de potencial de impacto que especificamente são denominadas de fração etiológica e fração previnível.

Além de saber quantas vezes mais um fator pode estar associado a um desfecho, podemos calcular o quanto um fator por se só poderia levar a mais doenças na população.

Alerta! Esta parte está incompleta e não pertence aos cursos de graduação e pós graduação da FORP. Depois irei acrescentar explicação detalhada. Estas medidas são descritas na maioria dos livros de Epidemiologia recomendados para este curso.

Sobre Fator de risco e marcador de risco

Um questionamento e confusão frequente é quando devemos utilizar o termo fator de risco e o que significa. Vimos que a medida Risco se calcula em estudos de coorte mostrando por exemplo a existência de temporalidade entre exposição e desfecho. Fator de risco não implica necessariamente causalidade, mas deve ser um fator que aconteça antes do desfecho que vá predizer no futuro a presença de um desfecho.

Veja como pode ser confuso. Uma vez vi em um post de instagram que defeitos de esmalte denominado hipomineralização de molar e incisivo (HMI) seria fator de risco para asma. Segundo a autora do post é muito comum observar o HMI e perguntar aos pais se a criança tem ou teve asma e a resposta é assertiva. Assim, crianças com HMI são mais prováveis de apresentarem asma, porém é impossível que o HMI aumente a probabilidade de o indivíduo ter asma. É possível que seja uma causa comum, ou o

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

oposto que é o mais provável, isto é que a criança com asma no primeiro ano de vida venha a desenvolver HMI nos dentes permanentes que estão se formando nesta época especialmente devido a medicamentos para asma. Assim, HMI seria um marcador de risco e não um fator de risco.

Outro elemento nesta discussão é sobre o que é causa. Causa na verdade como já mecionamos é uma discussão filosófica. No entanto, vamos pensar na seguinte situação de consumo de alimentos e obesidade. O alto consumo calórico sem ser gasto por meio de atividade física é a causa da obesidade. Pessoas ansiosas acabam ingerindo muitas calorias e assim podemos considerar também que ansiedade é parte da causa da obesidade. Em geral dietas para emagrecimento são mais efetivas se acompanhadas de aconselhamento ou acompanhamento com psicólogos para diminuir ansiedade. Nessa situação é fácil das pessoas aceitarem a ansiedade como parte da causa da obesidade, e até mesmo ela entra no tratamento.

Uma outra situação semelhante é com relação a cárie e consumo de açúcar. O açúcar como sabemos é a causa da cárie, mas porque não considerar os fatores antecedentes que como ansiedade ou maus cuidados de uma criança como também causa da cárie? Seriam também pobreza que leva ao consumo de açúcar que é um alimento barato assim, como conhecimento (educação) da mãe sobre os malefícios do açúcar.

Assim falamos de causas diretas e causas indiretas.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Tipos de Viéses

Embora tenhamos descrito vários viéses anteriormente de forma intuitiva, por vezes você irá encontrar estes vieses com determinados nomes específicos. Viéses podem ser introduzidos num estudo por diversas razões e em diferentes momentos por questões de seleção de amostra, outros por coletas de informação, e também existem vieses decorrentes de análises de dados inapropriados ou ainda de interpretação de dados de forma incorreta.

Mais importante do que saber nome de viés é reconhecer quando acontece algo que pode enviesar um estudo. Tentei não colocar nomes na maioria dos vieses para que o leitor não fique decorando nomes e descrição de vieses. No entanto, como a maioria dos livros falam de vieses específicos aqui exemplificamos alguns.

Por exemplo, a primeira descrição a seguir é sobre viés de sobrevivência, que é um termo utilizado quando um estudo não avalia o impacto da perda dos não sobreviventes, ou seja, ou se concentra nos sobreveviventes. Este tipo de viés na verdade pode acontecer em qualquer estudo, pois sempre vamos incluir indivíduos que são sobreviventes. O impacto e significado da sobrevivência será peculiar a cada estudo depedendo da exposição e desfecho que se estuda. Vamos supor que estejamos estudando o impacto da amamentação na obsedidade infantil aos 6 anos de idade, nesta situação a morte de crianças neste período não é tão grande, e talvez não esteja associada a amamentação. No entanto, se estamos estudando os efeitos de exposições durante a gravidez no nascimento de crianças de baixo peso, pode ser que não encontremos nenhuma associação e concluir que a exposição não tem efeitos deletérios, porque na verdade provocou o aborto dos mais susceptíveis. Um exemplo na odontologia seria estudar a saúde bucal incluindo doenças ligadas a dentes como

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

cárie e doença periodontal num estudo de adultos acima de 50 anos. Nesta situação, teremos no estudo apenas as pessoas que não perderam seus dentes anteriormente.

Existem inúmeros tipos de viéses e o site https://oxfordbrazilebm.com/index.php/catalogo-de-vieses/> contém varias definições. Aqui descrevo alguns destes viéses.

A seguir a descrição de alguns viéses, sendo que alguns como viés de colisão e de confusão já foram abordados anteriormente.

(1) **Viés de sobrevivência** é um tipo de viés de seleção que acontece ao não considerar não sobreviventes de um evento. Por exemplo, em estudos de fatores gestacionais como exposições a contaminações ou infecções em que o desfecho principal seja nascimento de prematuros. Nesta situação pode ser que um dos efeitos da exposição seja o aborto e ao nascimento aparentemente não é detectado aumento de prematuros porque na verdade a exposição resulta também em maior incidência de abortos.

Outro exemplo comum na literatura é em relação a aviões na segunda guerra mundial Abraham Waldo que analisou áreas nas quais os aviões que deveriam ser reforçadas avaliando as áreas onde havia mais buracos de tiros. A solução do matemático foi de reforçar onde não havia tiros, isso porque as aeronaves que retornavam sem serem abatidas haviam recebido tiros em áreas que ainda as possibilitavam regressar.

(2) **Lead time bias** é um viés que nos engana ao concluir que um tratatamento ou intervenção leva a sobrevida maior em indivíduos devido a antecipação de diagnóstico e não devido a um tratamento. Por exemplo, se num determinado momento decide-se realizar um programa de rastreamento de câncer precoce numa população, pode-se ter a impressão que o tratamento aumenta a sobrevida, mas esta sobrevida aumentada pode ser apenas devido a antecipação do diagnóstico.

(3) **Viés de atrição** se refere a viés em um estudo devido à perda de indivíduos em um estudo clínico de seguimemto.

(4) **Viés de identificação** (ascertainment bias) viés provocado quando membros de uma população alvo são menos prováveis de serem incluídos nos resultados finais.

(5) **Compliance bias** – viés de não aderência a protocolos terapêuticos. Por exemplo, se um grupo que está recebendo um medicamento num experimento e o não o utiliza da forma apropriada irá resultar em viés.

(6) **Viés de confirmação** (confirmation bias) se refere a avaliar de forma diferente evidencias que confirmam prévias expectativas. Este tipo de viés é considerado um viés cognitivo e foi demonstrado por um psicólogo Peter Cathcart Wason no artigo “; Este tipo de viés esta presente em nossa vida o tempo inteiro, mesmo de cientístas. Se achamos que não existe evidência suficiente em uma associação pelo que vem sendo

publicado na literatura, nossa tendência é ser super criterioso e procurar tudo que possa estar de errado nos estudos publicados. Por outro lado se achamos que uma associação é plausível ao sair mais um estudo que vá na direção que esperamos vamos dar maior valor a este.

Por exemplo, tenho uma amiga que acha que fluoretação de água faz mal, então a cada trabalho que sai publicado apontando algum malefício da fluoretação de águas, ela me envia para me provocar porque sabe que sou a favor do flúor. Sempre leio cuidadosamente os trabalhos e percebo erros consideráveis de metodologia e que invalidam as conclusões. Eu não posso me excluir deste viés completamente, porque ao ler um trabalho contra o flúor sou extremamente crítica. Porém, sempre ressalto que os trabalhos até então publicados tem problemas metodológicos graves, e, portanto não temos evidências de que fluoretação de águas nas doses recomendadas faça mal. Isso não quer dizer que no futuro alguma metodologia nova revele algum efeito indesejável da fluoretação, mas no momento temos mais razões para utilizar o fluoreto do que não.

(7) **Efeito de Howthorne** - se refere a mudanças de comportamento quando se está sob supervisão num estudo. Este nome vem do bairro em Chicago onde um experimento foi realizado para se verificar se melhorando a iluminação do ambiente resultaria em melhor eficiência dos empregados. Os resultados mostraram que a melhor iluminação estava associada ao aumento de eficiência, mas também houve melhoria de eficiência dos demais empregados.

(8) **Viés de alocação** - é o viés decorrente de alocação de indivíduos em experimentos de forma diferente.

(9) **Viés de publicação de resultados positivos.** Este viés é bem comum, se uma hipótese é plausível e for testado seu resultado é imprevisível. Por exemplo, recebemos como uma das explicações de recusa de um artigo o seguinte comentário " The research methods and theoretical basis of this paper are strong enough, but the results of this paper do not confirm the author's hypothesis". Ficamos pensando que se os resultados tivessem sido positivos, se o artigo estaria aceito apensar das outras limitações que foram apontadas. O que representa este comentário? Uma incapacidade do revisor em compreender o que é ciência. Provávelmente este revisor nunca teve aulas de Epidemiologia adequadas ou noções de filosofia de ciência que fazem parte da Epidemiologia.

(10) **Viés de semelhança –** num estudo deve-se ter certeza do diagnóstico que se está realizando e tomar cuidado que uma condição seja semelhante ao diagnóstico.. Um exemplo é um estudo de Morrison et al 1977 em que o diagnóstico de hepatite pode ter sido confundido com o diagnóstico de icterícia. Esse estudo investigava a associação causal entre contraceptivos e hepatite.

(11) **Viés de interpretação -** a interpretação de resultados de pesquisas mesmo admitindo que sejam bem realizadas e pouco sujeitas a vieses durante sua realização e análise estatística pode ser ainda susceptível a vieses de interpretação. Um recente artigo 2021 de Kaptchuk (Effect of interpretative bias on research evidence) especialmente

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

motivados pela negacionismo e falta de conhecimento de ciência, discute os vieses de interpretação. Se fornecermos resultados de uma mesma pesquisa e pedirmos para dois pesquisadores escreverem um artigo com certeza cada artigo será diferente e até mesmo em relação conclusão. Isso existe porque pesquisadores carregam experiências e conhecimentos diferentes e opniões diferentes sobre um mesmo assunto. Dependendo dos conhecimentos e mesmo interpretações de resultados de pesquisas um pesquisador pode realizar uma revisão sistemática da literatura e ser influenciado por suas impressões. Por exemplo, desde 1994 Offenbacher por evidências empíricas de suas pesquisas com animais chegou levantou a hipótese de que doenças periodontais fossem fatores de risco para baixo peso ao nascer em animais. Em 1996 este pesquisador realizou um estudo caso-controle que foi bem recebido pela comunidade de periodontia, associava uma doença bucal à um evento de saúde relevante. As limitações do estudo foram ignoradas pela comunidade de periodontistas, devido a em parte a falta de conhecimentos de estudos epidemiológicos. Na década de 90 vários estudos foram realizados demonstrando a associação apesar de claros problemas metodológicos, e foram compilados em revisões sistemáticas que também ignoravam as limitações metodológicas.

	Quasi-experimentos

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

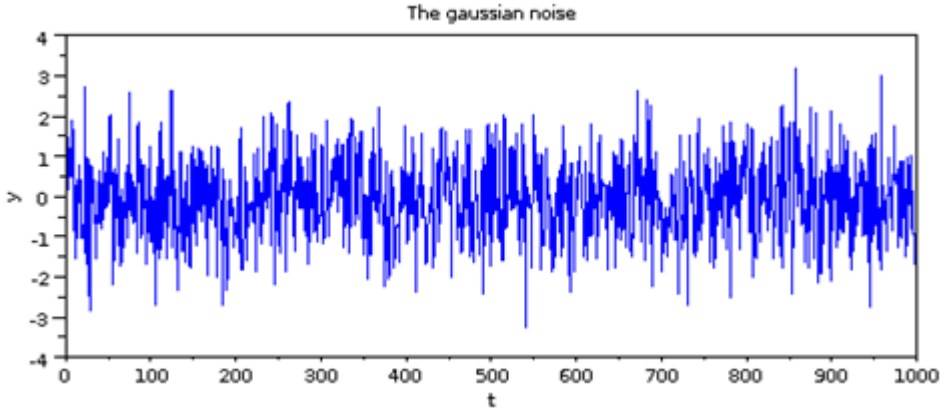
Amostragem e Erros

Até o momento descrevemos os desenhos dos estudos, viéses, fatores de confusão, além das medidas de frequência e associação. Os viéses levam a imprecisões nos estudos que na verdade são erros de medidas ou de seleção de indivíduos. Além dos viéses e imprecisões de informações e ainda além do fato de fazermos uma ciência empírica, os estudos tem outra fonte de imprecisão pelo simples fato de serem amostras e não populações inteiras. Estudos de laboratório com ratinhos têm amostras de ratinhos, estudos de laboratório com corpos de prova têm amostras de corpos de provas, e em geral nossos estudos em humanos são realizados com amostras. Estudos que incluem a população inteira são raros além do censo, mas se existirem eles apenas terão os problemas relacionados a qualidade das medições que forem realizadas. Podemos então dizer que podemos ter imprecisões e erros relacionados à coleta de informações (sejam elas exposição, desfecho ou terceiras variáveis) e também imprecisoes e erros relacionados à amostragem.

Erros e imprecisões, vamos esclarecer os signficado dos dois. Se formos medir a altura de uma população de crianças, não teremos erros relacionados a amostragem, apenas a medição. O metro utilizado pode estar com algum problema, pois foi feito na fabrica com os milimetros maiores que milímetros, e assim a altura das crianças será maior do que deveria ser. Esse erro é um viés, causado pelo metro que estava errado. Se o metro estiver correto, mas medirmos as crianças no final da tarde, o resultado final será uma altura média menor do que a verdadeira. Teremos de novo um viés. Vamos supor que decidimos realizar todos os exames de manhã com um metro perfeito para que seja logo após as crianças levantarem e assim todas terão medidas mais

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

comparáveis (comparáveis não iguais). Embora o metro seja perfeito alguem vai medir a criança, e temos sempe alguma variação que podemos chamar de aleatória ao medir alguma coisa. Não chega a ser um viés porque não altera muito a medida final, ainda mais que algumas crianças serão medidas para mais e outras para menos. Estamos nos referindo a uma oscilação constante e pequena que chamamos de aleatória, pois acaba tendo característica de ao acaso (aleatório). Em algumas áreas como economia e física, é comum chamar este erro aleatório ao redor de uma medida de barulho (*noise),* na área biológica este termo barulho não é comum. Note na figura abaixo que existe uma média que seria zero com um barulho ao redor.

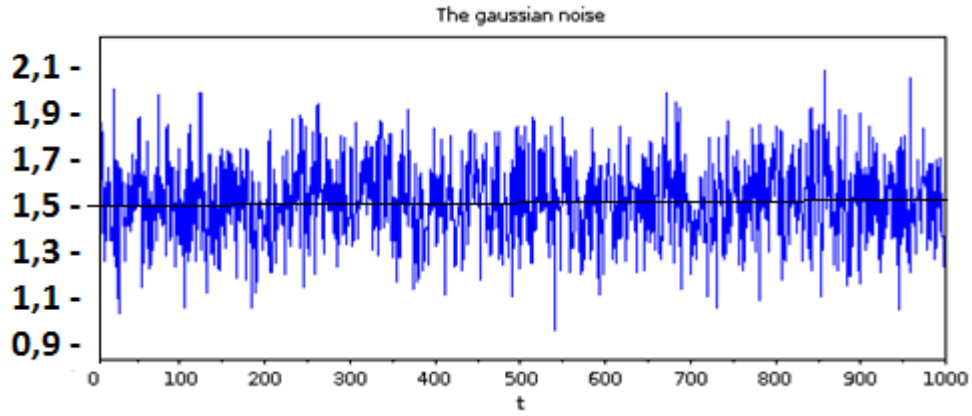


Independende de se aleatório ou não, são erros! Assim, dizemos que as medidas podem ter **erro aleatório** e também erro sistemático (que seria o viés). Logo, **viés é um erro sistemático**. O erro sistemático pode ser evitado, já o aleatório acontece, e é de difícil quantificação no caso das medidas.

A amostra também pode ser selecionada de forma errada para representar o que queremos sendo chamado de viés de seleção como já discutimos. O erro de seleção também é um erro sistemático e pode ser evitado planejando e realizando a amostra de forma adequada. Além do erro sistemático, a amostragem é sujeita também a um erro aleatório sempre que a amostra é realizada por meio de um processo aleatório. Num exemplo simples, se fizermos um sorteio aleatório para selecionar uma amostra de uma população, esta amostra é apenas uma das milhares possíveis amostras que poderiam sair daquela população. Imagine se realizarmos milhares de amostras de crianças com altura média total (imaginária) seria de 1,50 metros. A figura abaixo é uma montagem

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

da figura que utilizei para representar o erro aleatório na figura anterior, apenas adicionei hipoteticamente as medidas de altura com média de 1,50.



Isso quer dizer que ao fazermos um estudo com uma amostra, esta amostra é apenas uma das milhares possíveis que poderiam ter resultados diferentes. Certissimo, mas se fizermos esta amostra aleatória corretamente, podemos estimar este erro! Não podemos evitar completamente, mas como veremos posteriormente podemos diminuí-lo e podemos estimar sua magnitude. Para diminuí-lo completamente a única alternativa seria trabalhar com a população inteira.

Concluindo, temos dois tipos de erros nos estudos, o erro sistemático e o aleatório. O sistemático pode vir tanto da coleta de informações como da amostragem, e este podemos e devemos evitá-los. Temos o erro aleatório das medidas e o erro aleatório da amostragem. O erro aleatório das medidas em geral não pode fazer nada, ele acontece e pronto. O erro aleatório da amostragem, no entanto, acontece, porém pode ser reduzido, porém não eliminado, mas o erro aleatório de amostragem não eliminado pode ser estimado! Como você irá entender, ai entra a estatística!

Você pode imaginar se temos sempre erros qual é a verdade? Bom, a verdade não você saberá nunca, vamos dizer que é hipotética, ainda mais levando em consideração que nossa ciência é empírica. No entanto, se evitarmos os erros sistemáticos, e se estimarmos o erro aleatório da amostragem provávelmente nós vamos ter uma boa aproximação da “verdade”. Vamos chamar a *verdade desconhecida* de **parâmetro,** que é aquilo que queremos estimar por meio de nosso estudo.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Imagine que queremos saber a altura média de crianças com 15 anos em Ribeirão Preto. Se examinarmos todas as criânças e de forma correta com o metro sem viés, vamos ter uma média, mas que não é a verdadeira, pois nela estará embutido o erro aleatório de medida. Portanto esta média resultante de nosso estudo é apenas um **estimador** da verdade que seria o parâmetro. Se esta medida é resultante de uma amostra da população de crianças de 15 anos de Ribeirão Preto, então além do erro aleatório de medida teremos o erro aleatório de amostragem. Esta medida continua sendo o estimador do parâmetro que agora vai aprensetar um erro amostral aleatório, mas que pode ser estimado tambem. Veja a fórmula a seguir que representa o que acabamos de comentar.

$$\mu = \widehat{M} + \varepsilon$$

μ é a representação do parâmetro (a verdade desconhecida), o $\widehat{M}$ com chapeuzinho é a média que acabamos de estimar, logo estimador e o épson (ε) representa os erros tanto aleatórios como sistemáticos. No caso utilizamos a letra "m" para representar a média do exemplo de média de altura, no entanto, podemos imaginar a fórmula para qualquer coisa que queiramos estudar. Por exemplo, poderia ser o parâmetro verdadeiro do Risco Relativo de uma associação, sendo que por meio de nosso estudo conseguimos chegar a um estimador do neste $\widehat{RR}$, e, portanto precisamos colocar um "chapeuzinho" $\widehat{RR}$, e o éspson representaria todos os erros do estudo.

$$\mu = \widehat{M} + \varepsilon$$

Você deve estar pensando porque esta fórmula não tem sinal de mais ou menos "+ - ". Acontece é signfica que é mais o erro que pode ser ele próprio negativo ou positivo. A fórmula está correta! O erro pode vir ou de medição (o metro com qual está sendo medido pode se alterar com calor e humidade, os olhos do observador podem ler

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

de forma inadequada, as vezes para mais e as vezes para menos) e ou do processo amostral se for realizado com amostra. Assim esse "erro" representa todos os erros.

Vamos voltar à discussão do erro aleatório de amostragem depois de discutir detalhes de amostragem probabilística e não probabilistica que será discutido a seguir no capítulo de amostragem. Quando terminarmos a discussão dos tipos de amostragens nós retornaremos a discussão de como estimar o erro aleatório que nada mais é do que a estatística.

Amostragem Probabilística e Amostragem não Probabilística

Amostragem às vezes é ensinada como capítulo à parte dentro da estatística, mas na verdade não é um capítulo à parte e sim o fundamento de estatística. Embora os livros de estatística sempre comecem a ensinar estatística definindo variáveis, desvio padrão, eu considero mais importante começar com o propósito da mesma que é a inferência. Acho mais fácil entender estatística começando pelo processo amostral do que pela maneira clássica, especialmente porque em 99,9% dos estudos trabalhamos com amostras. Difícilmente vamos analisar dados da população inteira. Não confunda com dados populacionais com a população inteira. O termo **dados populacionais** é utilizado para expressar que a amostra veio da população seja ela inteira ou de uma amostra da mesma. Quando lidamos com uma população inteira não temos amostra, e logo aquele erro amostral que mencionamos não existe. Porém pode ter erros de observação.

Ignorando o erro de observação que podemos considerar "culpa" do pesquisador, o que resta é erro amostral. Este erro amostral pode ser sistemático que também é culpa do pesquisador, restando apenas o erro amostral aleatório. Vamos começar como na pratica se faz uma amostra de uma população, e retornaremos ao erro amostral posteriormente.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Amostra populacional na prática

Vamos começar dando um exemplo de amostragem e como é realizado este processo. Vamos supor que queremos saber a prevalência de asma em adultos de 30 a 40 anos em Ribeirão Preto. O ideal seria examinar todas as pessoas desta faixa etária perguntar se a pessoa tem asma. No entanto, a população é muito grande e levaria muito tempo, e também não saberíamos onde estas pessoas estão.

Definindo limite geográfico

A primeira coisa que temos que fazer é definir o que é Ribeirão Preto: vamos considerar a cidade ou o município, vamos incluir zona rural ou apenas zona urbana? Uma vez definido o limite geográfico vamos procurar as pessoas, mas como? Como não existe no google uma lista de moradores de 30 a 40 anos com endereço e atualizada, temos que encontrar algo melhor. Pode-se pensar nos títulos de eleitor, mas nem todo mundo que vive em Ribeirão Preto vota na cidade, e moradores de outras cidades votam em Ribeirão Preto. Quem sabe podemos usar carteiras de motorista, mas nem todos desta faixa etária têm carteira de motorista e existem moradores de outras cidades com carteira de Ribeirão Preto e vice-versa. Podemos ir às fábricas, mas nem todo mundo trabalha nas fábricas da cidade. Podemos ir aos escritórios, mas da mesma forma nem todo mundo trabalha em escritório. Se estimarmos asma em trabalhadores de fábricas não é o mesmo que estimar a prevalência real de asma na cidade que é o que queremos. Outra ideia é ir para as praças principais e chamar as pessoas que tem esta idade. No entanto, nem todo mundo passa pelas praças do centro da cidade; são pessoas específicas que passam pela praça que podem ter mais ou menos asma que o restante da cidade.

Podemos ainda restringir a um bairro, mas não sabemos o que o bairro representa para a cidade em termos de asma. Não é nada fácil. Que tal bater de porta em porta na cidade inteira, visitar todas as residências procurando pelas pessoas de 30 a 40 anos? Vai levar muito tempo. Assim, uma saída é sim estabelecer a residência (domicílio) como meio de se encontrar os indivíduos (unidade da amostra). Desta forma temos a oportunidade de incluir todos os indivíduos da cidade independente de trabalho (evitando viés de trabalhador sadio) ou de onde estejam momentaneamente. É claro

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

que escolhendo domicílios, estamos excluindo as pessoas que vivem em instituições, mas nada impede de incluí-los também.

Algumas outras dificuldades vão surgir ao se resolver sortear residências. Talvez conseguíssemos na prefeitura uma lista de todos os imóveis com IPTU, pois todos tem IPTU (supostamente) e poderíamos sortear os números das casas com IPTU. Nem todo imóvel é residencial, e talvez pudéssemos separar aqueles que estão registrados como comerciais. No entanto, às vezes temos famílias morando em imóveis comerciais, assim como temos comércios e escritórios em casas residenciais. Para sortear um domicilio, temos que eliminar comércios, indústrias, escolas, igrejas e também casas abandonadas. A única forma que conseguimos fazer isso apropriadamente é enviando pesquisadores à rua, e pedindo que façam um mapa para sorteio de cada quadra, apontando onde é residencial ou não.

Ainda o pesquisador precisa anotar as casas onde moram pessoas de 30 a 40 anos, pois somente estas casas podem ser sorteadas e as demais serão excluídas. É preciso ainda decidir o que fazer com residências que tem um número, mas onde moram várias famílias. Por vezes vamos à porta de uma casa, e descobrimos que tem mais três edículas no fundo e cada uma mora uma família diferente. Tudo isso pode ser identificado na construção do mapa para sorteio.

Vamos supor que tudo isso seja possível. Na verdade é possível e é assim que se monta uma pesquisa populacional, desde que a cidade não seja muito grande.

Uma vez resolvido e sorteada as casas apenas com pessoas de 30 a 40 anos, podemos enfrentar mais um obstáculo, pois algumas famílias têm mais de uma pessoa com esta faixa etária. Assim deve-se decidir se todos devem entrar para a amostra, ou se apenas um indivíduo será sorteado.

Mais um inconveniente, depois de sorteada a pessoa ela não é encontrada. Se isso acontecer tenta-se retornar pelo menos três vezes em horários diferentes. Se possível é aconselhável perguntar a vizinhos ou moradores da casa quando a pessoas estará disponível. Mesmo encontrando a pessoa, ela pode não querer participar do estudo. Se ela não quiser, não podemos fazer nada. Uma alternativa que não pode ser adotada é substituir o sorteado com outro morador ou com o vizinho que está disponível. A amostra é resultado de sorteio aleatório e não pode haver substituição.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Vamos supor que na investigação sobre asma, um indivíduo com asma foi sorteado e ele nunca está em casa porque está sempre em tratamento e fisioterapia. O seu vizinho sem asma está sempre em casa. Se isso acontecer a prevalência da doença vai ser menor do que deveria ser, pois eliminamos o indivíduo com asma. Portanto, uma vez sorteado, está sorteado e não podemos substituí-lo pelo vizinho nem outra pessoa qualquer.

Definindo a estrutura amostral....

Veja que pensamos em várias possibilidades de onde tirar informações sobre os indivíduos para que fossem sorteados para nossa amostra. Chamamos estas possibilidades de **estruturas amostrais.** Se fossemos estudar crianças de 7 a 18 anos, a estrutura amostral de escolha seria a escola, para o nosso estudo, domicílio foi a estrutura amostral de escolha. Caso decidíssemos estudar crianças de 0 a 6 anos, a creche poderia ser uma escolha, mas não adequada, porque muitas crianças não frequentam creches. Nesta faixa etária o domicílio seria a escolha mais certa. Óbvio que creche seria adequado numa cidade onde todas as crianças ou a quase totalidade frequenta creche. Da mesma forma, são exemplos de estruturas amostrais a lista telefônica, locais de trabalho etc. Em inglês estrutura amostral é chamado de "*sampling frame*".

Para identificar os domicílios comentamos em mapear a cidade, porém numa cidade grande fica inviável enviar pessoal para bater de porta em porta preparando o mapa de sorteio. Alternativas a estrutura amostral de domicílios poderia ser uma lista de telefone fixo pelo qual os domicílios seriam sorteados. Mas note que neste caso a estrutura não é de domicílios e sim de lista de números de telefones. As listas telefônicas podem ser uma opção para se sortear os domicílios, mas precisaria que a quase totalidade de domicílios tivessem telefones fixos. Mesmo assim, haveria casas com mais de dois telefones e tudo isso teria que ser resolvido antes de se sortear aleatoriamente os domicílios. A alternativa da lista telefônica seria útil se eliminássemos telefones comerciais e duplicidade de telefones residenciais. De fato, este tipo de amostragem tem sido utilizado em muitos estudos nos Estados Unidos e outros países. No Brasil é quase impossível, a não ser que a cidade tenha condições acima favoráveis.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Nos últimos anos com a substituição de telefones fixos por celulares ficou ainda mais difícil realizar estudos por meio de listas de assinantes de telefones.

Vamos começar de uma forma simples imaginando que a cidade tem uma lista de todos os domicílios já preparada limpa de comércios ou que já tenhamos nós preparado a lista e resolvemos fazer uma **amostra aleatória simples**. Este nome quer dizer que de uma lista de todos os domicílios sortearíamos uma quantidade de forma aleatória simples, isto é, de forma que todos os domicílios com pessoas de 30 a 40 anos têm a mesma probabilidade de ser escolhida. É possível que os mil domicílios sorteados estejam bem espalhados na cidade representando todos os estratos sociais, todas as cores de pele, e diferentes etnias se for o caso. Porém pode ser que neste sorteio aleatório simples a maioria seja de domicílios de alta renda não representando a população da cidade. Imaginem que tenhamos 10% de domicílios de alta renda na cidade e na amostra eles correspondem a 30%. Assim, se a asma é mais comum entre os mais ricos teríamos uma superestimação da ocorrência de asma.

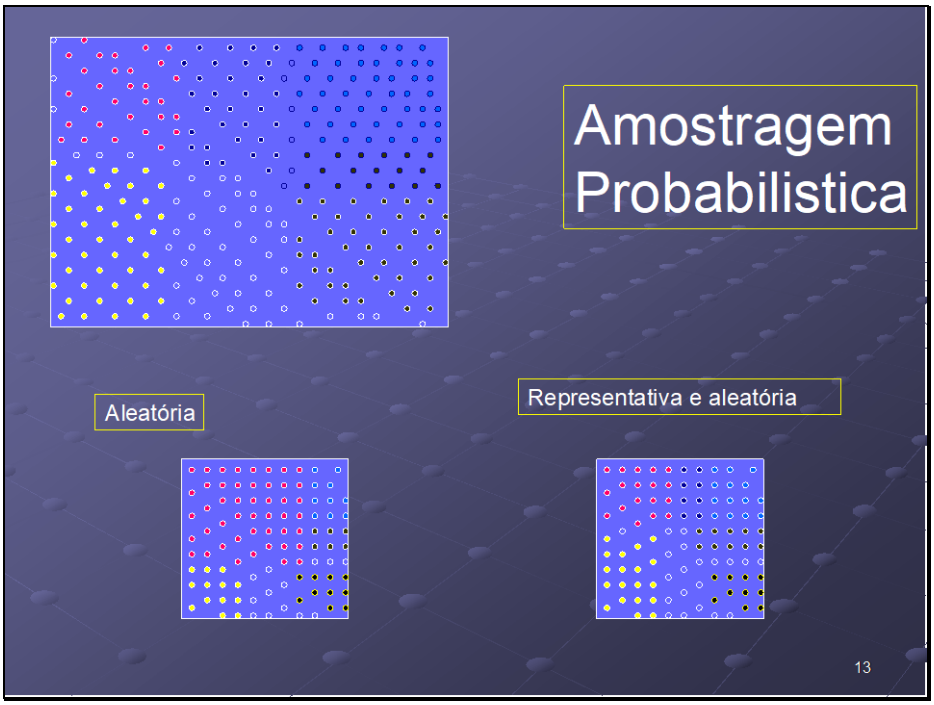
O que podemos concluir é que se fizermos uma **amostra aleatória simples** pode ser que não consigamos representatividade da população. Porém, para estimar a prevalência correta da asma na cidade, precisamos de uma amostra que seja representativa. Vamos dizer que a amostra tem que ser igualzinha a população no que se refere a distribuição dos principais fatores associados ao que vamos estudar, exceto se o que estamos estudando acontecer de forma aleatória na população. Aleatório significaria que o fator que vamos estudar não seria influenciado pelo sexo, etnia, cor de pele, escolaridade, nível socioeconômico, tipo de ocupação nem idade. Bom, é muito difícil ter alguma doença que não seja aleatória.

A amostra aleatória simples é um tipo entre outros de **amostra probabilística**, isto é, baseada na probabilidade (sorteio aleatório) onde *se conhece a probabilidade exata de cada membro participar da amostra*. Esta é a definição oficial de amostra probabilística, aquela em que se conhece a probabilidade exata de cada membro da população participar da amostra.

A amostra aleatória simples é uma amostra probabilística em que cada elemento da população **tem a mesma probabilidade de ser sorteado**. Note na figura abaixo onde a população esta representada pelo quadrado grande com seus elementos.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

A figura que tem escrito aleatório não é representativo, mas a outra que está abaixo do titulo “representativo e aleatório” é sim.



Portanto para uma amostra ser boa ela tem que ser além de aleatória representativa, se não ela não serve para estimar um parâmetro adequadamente. Entre aleatório e representativo, o representativo é mais importante!

Para garantir representatividade numa amostra, o procedimento aleatório simples não é suficiente. Se tivermos o conhecimento de que asma está associada ao nível socioeconômico, precisamos demarcar os bairros da cidade de acordo com esta característica e sortear aleatoriamente dentro de cada um deles. Por exemplo, dividir a cidade em grupos de bairros com diferentes estratos sociais de acordo com informações obtidas pelo IBGE. Se a princípio pensamos em amostrar mil domicílios e temos 4 níveis de estrato socioeconômico, a amostra tem que ter proporções equivalentes destes estratos. Não precisa prestar atenção nas fórmulas da figura abaixo, apenas atenha-se aos quadrados coloridos e a proporção dos mesmos na amostra. Note que a amostra na base da figura é uma copia proporcional da população.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Este procedimento de amostragem é chamado de estratificação. Dentro de cada estrato a amostra deve ser aleatória simples. Chama-se esta amostragem de **amostragem estratificada**. Vamos dizer que temos duas etapas ou estágios no processo de amostragem, estratificação seguida de uma amostra aleatória (two-stage sample). A amostragem estratificada resulta em uma amostra probabilística onde se sabe exatamente qual a probabilidade de cada elemento pertencer à amostra. Com a estratificação garantimos a representatividade de cada nível socioeconômico. Note que quem pertence ao quadrado branco, tem maior probabilidade de pertencer a amostra do que um indivíduo do quadrado preto. Se fosse a mesma probabilidade talvez o indivíduo do quadrado branco não fosse sorteado ao acaso para se montar uma amostra aleatória simples. Ao melhorar ou "garantir" representatividade do fator que se quer minimiza-se alguns fatores de confusão, mas não elimina.

Vamos imaginar que conseguimos estratificar Ribeirão Preto em 4 grandes áreas de acordo com o nível socioeconômico, e ai vamos aleatoriamente sortear os domicílios. Poderíamos partir para uma amostra aleatória simples se tivéssemos uma listagem dos domicílios em todos os estratos. Se não tivermos a lista até podemos elaborá-la como já mencionado. Ok, se a cidade fosse pequenina, mas se a cidade é grande e não temos uma lista de domicílios o que fazer? O que temos são quadras e mais quadras de casas e mais casas e para bater na porta de cada uma seria muito trabalhoso e levaria muito

tempo. Uma alternativa é verificar blocos de quadras chamadas de áreas censitárias já estabelecidas pelo IBGE para realizar o censo. Podemos dentro de cada estrato sortear algumas dessas áreas, e ai, podemos bater de porta em porta em todas as casas das áreas censitárias sorteadas e elaborar o mapa para sorteio.

Neste procedimento da amostragem incluímos um elemento novo que é o sorteio de toda uma área censitária. Esta área sorteada é que vai participar do estudo, deixando de fora os que moram nela. Chamamos esta área de **conglomerado (**em inglês é chamado de *cluster*). Porque conglomerado? Porque é um aglomerado de casas próximas. Não fizemos o sorteio aleatório das pessoas, sorteamos o conglomerado de forma aleatória e todos que estão nele entram para o estudo. Esta foi uma saída razoável para evitar que andássemos muito em cada estrato caso tivéssemos sorteado casas longe uma das outras.

Com certeza o conglomerado facilita e barateia o estudo, mas você pode morar em um estrato não sorteado e começar a indagar que sortearam bloco censitário onde justamente moravam vários indivíduos da mesma família e que esta família tem 90% de seus membros com asma. Sim, as pessoas que vivem num conglomerado tendem a ser mais semelhantes entre si do que entre os conglomerados. Desta forma, os conglomerados são úteis, mas temos que entender que estes indivíduos do conglomerado podem ter um peso grande na prevalência de uma doença. Porém, ao mesmo tempo em que um conglomerado pode ter uma prevalência alta de asma, outros conglomerados também podem ter prevalência de doença muito baixa pelo mesmo motivo de que os moradores são de uma mesma família. Estas concentrações diferentes de doença em geral se diluem, mas provocam como veremos posteriormente probleminhas, contornáveis na estatística.

Outro desafio pode aparecer se você resolver sortear algumas casas, e tiver mais do que um morador na casa com a idade que se quer amostrar. Nesta situação teremos de novo o mesmo problema do conglomerado inicial que era a área censitária. Isto é, o domicilio com mais de um morador elegível pode contribuir com vários adultos com asma. Se decidirmos que todos os moradores de cada casa entram para o estudo, a casa será outro conglomerado do estudo. Para evitar outro conglomerado podemos estabelecer regras para sortear aleatoriamente um dos moradores elegíveis. Este

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

morador terá que ser realmente sorteado, e não pode ser, por exemplo, aquele que atender o entrevistador por ser o mais fácil de encontrar. Se um dos moradores fica mais em casa, e isso pode ser devido a ter a doença, ele será mais provável de ser encontrado, e, portanto irá inflar a prevalência do estudo. Se decidirmos sortear um dos moradores, precisamos bater na porta, identificar quantos moradores tem, e sortear um deles, se estiver em casa ótimo, se não se deve voltar depois para tentar entrevistá-lo.

Até o momento vimos que uma amostra aleatória simples pode não garantir representatividade, assim podemos usar procedimentos como estratificação seguida de amostra aleatória simples (chamado de amostra estratificada), e se for difícil fazer a amostra aleatória simples dos indivíduos podemos utilizar o sorteio de conglomerados. Quando misturamos várias etapas na amostragem incluindo estratificação e conglomerados, nós denominamos a amostra de **amostra complexa**. Sempre que estabelecemos etapas com aleatorização envolvida temos o que chamamos de amostra probabilística cuja definição é aquela em que se sabe exatamente (pode-se calcular) a probabilidade de cada indivíduo em particular pertencer à amostra.

A **amostra sistemática** é outro tipo de amostra probabilística bastante útil e que pode ser realizado quando a amostra aleatória simples não pode ser empregada. Existe muito preconceito em relação à amostra sistemática entre as pessoas que não conhecem amostragem de forma adequada e somente ouve falar dos problemas que ela tem. O preconceito com a amostra sistemática me lembra do preconceito com os estudos ecológicos que são abominados por muitos que os utilizam e interpretam de forma não adequada. De maneira bastante simples a amostra sistemática é aquela que se faz quando não se tem uma forma de fazer um sorteio aleatório numa lista de alguma coisa. Se quisermos selecionar aleatoriamente 10 alunos da sala do primeiro ano de odontologia que tem 80 alunos. Podemos seguir a lista e chutar um número, por exemplo, o 17 e a cada 8 alunos a partir do número 17 será sorteado para participar da amostra. Esta é uma forma fácil de realizar um sorteio sem tabela de números aleatórios ou programa especifico para aleatorização. O problema que existe é se tiver alguma tendência na ordem alfabética dos nomes que coincidir com a listagem, ai a amostra não vai ser representativa. Outro exemplo seria de utilizar a amostra sistemática para sortear casas de uma rua resultando sempre no sorteio de casas que são de esquina. As

casas de esquina são mais valorizadas que as demais, e, portanto, estaríamos selecionando as pessoas mais ricas. Por outro lado se o processo nunca seleciona casas de esquina a amostra final será de moradores de nível socioeconômico mais baixo do que a verdade. O pesquisador deve refletir sobre o que está sendo amostrado e identificar possíveis empecilhos na utilização da amostra sistemática.

Em geral a amostra sistemática é útil para selecionar coisas quando não se tem uma lista dos possíveis elegíveis como no caso de se selecionar casos incidentes de uma doença conforme vão aparecendo num serviço médico. Se quisermos uma amostra de casos novos diagnosticados com câncer durante o ano de 2015, podemos decidir que entrarão para a amostra todo o quarto caso que aparecer no hospital. Nesta situação não tem problema algum. Já na situação de verificar estudar infartos que entram no hospital e o pesquisador optar por sistematicamente escolher os casos que chegam à segunda e a sexta-feira talvez tenhamos problemas. Em geral os casos da segunda feira são diferentes dos demais, e incluem pessoas que resolveram no final de semana fazer exercício físico em excesso. Já á sexta feira talvez cheguem principalmente os casos de pessoas que aguentaram sintomas durante toda a semana e resistem a ir ao médico, resolvendo ir apenas à sexta. Se houver possibilidade de algum viés na amostragem então devemos optar por outra solução, que pode ser ainda sistemática, mas que drible estes problemas.

Amostra sistemática corretamente foi utilizada no estudo de coorte de nascidos vivos realizada em São Luís, Maranhão. Em São Luís nascem cerca de 20 mil crianças por ano, e o estudo precisava apenas de 7 mil ao longo do ano. Para tanto a amostra sistemática foi utilizada para selecionar a cada terceiro bebe nascido vivo nos hospitais da cidade. A amostra sistemática foi utilizada porque em estudos preliminares não se observou nenhuma tendência atrelada á ordem de nascimento. Nesta situação não teria como fazer uma amostra aleatória, pois havia necessidade de examinar os bebes ao nascimento conforme iam nascendo. Portanto, corretamente foi utilizada uma amostragem sistemática.

Nos livros em geral encontra-se a seguinte regra para se realizar uma amostra sistemática quando se tem uma lista prévia de indivíduos para o sorteio. Vamos supor que queremos fazer uma amostra de 300 indivíduos de uma população de 8mil. Primeiro

dividimos 8000/300 e termos 27. Assim, selecionaremos a cada vigésimo sétimo indivíduo na lista. Também começaremos por sortear um número entre 1 e 27 como ponto de partida para a amostragem. Sempre começar com o número aleatório de K-1 , neste caso K = 27 logo entre 1 e 26 inclusive, mas não 27.

As amostras descritas acima são todas probabilísticas, pois saberemos a probabilidade de cada indivíduo participar da amostra. Elas são utilizadas em pesquisas científicas e em pesquisas governamentais como, por exemplo, realizadas pelo IBGE em estudos como a PNAD, POF etc. Em geral estas pesquisas populacionais para PNAD e POF incluem amostras complexas.

Em alguns momentos não há necessidade de saber qual a probabilidade do indivíduo participar de uma amostra por que não precisamos fazer inferência de porcentagem de algo ou média. Por vezes precisamos de amostras com motivos específicos para explorar ou entender melhor um assunto, e nestes casos podemos utilizar uma amostra não probabilística. A seguir vamos comentar os principais tipos de **amostras não probabilísticas.**

AMOSTRA NÃO PROBABILÍSTICA

As amostras não probabilísticas não são utilizadas para testar hipóteses por meio delas não há como se estimar a probabilidade de cada indivíduo pertencer a amostra e nem é esse o objetivo do estudo. Estas amostras têm aplicações específicas para explorar algum assunto como intensão de votos, para conhecer melhor determinados grupos de pessoas e em pesquisas qualitativas.

- **Amostra por Cota**

O nome de cota é porque cada entrevistador tem uma cota (isto e número exato) de pessoas a serem entrevistadas. Por exemplo, vamos supor que o IBOPE decida que mil pessoas seriam suficientes para uma pesquisa de intensão de votos em Ribeirão Preto. Se 10 entrevistadores são utilizados, cada um receberia uma missão de entrevistar uma cota, por exemplo, de 100 indivíduos. Cada um iria para uma área de

Ribeirão Preto, como por exemplo, na Praça XV de Novembro e ficaria lá até cumprir a meta de 100 pessoas. A pesquisa por cota é muito utilizada nas eleições para estimar a porcentagem de votos para os candidatos a precisão não é a de uma pesquisa, e a finalidade é apenas ter uma ideia das intensões de votos.

- **Pessoas Típicas**

Por vezes pode-se ter intensão apenas de conhecer a opinião de algumas pessoas numa companhia, instituição, ou grupo de pessoas. Por exemplo, se a intensão é explorar a percepção dos alunos quanto as aulas de Epidemiologia, poderíamos montar grupos de 8 alunos, que participariam de uma entrevista semiestruturada sobre a disciplina. O mais correto seria lançar mão de uma pesquisa chamada grupo focal em que os alunos seriam reunidos em grupo de 6 a 8 pessoas de acordo com suas notas. Os melhores alunos formariam um grupo e assim por diante. Estes indivíduos poderiam trocar ideias e opiniões sobre a disciplina. Os grupos focais são pesquisas científicas sim, que são realizadas para avaliação em geral seja de um produto em uma indústria ou um serviço médico, dentário, ou qualquer serviço. Os grupos tem que ser homogêneos para se explorar de forma mais adequada o que realmente é relevante. O objetivo deste estudo não é levantar a porcentagem de pessoas que gostam ou não de um serviço, mas para avalia-lo de forma correta.

- **Amostra por Conveniência**

Grande parte de estudos na odontologia, infelizmente, são feitos com amostras de conveniência de pacientes atendidos na faculdade. Como não sabemos a origem dos pacientes, não podemos, por exemplo, estudar fatores etiológicos. Já discutimos isso em relação aos estudos de caso-controle. No entanto, quando se tratar de experimentos para teste de drogas ou tratamentos, as amostras de conveniência são utilizadas. Para testar uma droga contra o câncer, não tem outra escolha a não ser recrutar estes pacientes diretamente num hospital de tratamento de câncer ou por meio de divulgação de recrutamento. A consequência de se utilizar amostra de conveniência e voluntários

num experimento é que pode existir alguma seleção de pacientes que não reflete os pacientes em geral. Já comentamos anteriormente que por isso, quando um medicamento é lançado no mercado, ele fica sob controle dos órgãos de vigilância de saúde por algum tempo e se algum efeito colateral aparece o medicamento é suspenso.

As amostras de conveniência também servem para se pretextar questionários antes de um estudo observacional. Para tanto, escolhe-se alguns voluntários da mesma população.

MAIS UM EXEMPLO DE AMOSTRAGEM E INTRODUÇÃO AO ERRO AMOSTRAL E INTERVALO DE CONFIANÇA

Agora que você já sabe formalmente quais são os tipos de amostras vamos os mais alguns detalhes necessários para compreender como estudos são realizados e suas limitações.

Vamos supor que um prefeito encomenda um estudo sobre a altura média das crianças com 15 anos de idade da cidade de Ribeirão Preto no ano 2015. O primeiro passo é refletir sobre o pedido que foi feito pelo prefeito, e, portanto, refletir sobre os objetivos do estudo. Assim o objetivo é claro “altura média das crianças de 15 anos de idade da cidade de Ribeirão Preto”. Definido o objetivo esclarecer se a pesquisa será realizada na cidade ou município de Ribeirão Preto, e também se área rural será incluída ou não. Portanto**, a demarcação geográfica** da cidade é fundamental para que a inferência seja realizada de maneira correta.

Demarcada a área geográfica, precisamos discutir onde encontrar os indivíduos para selecioná-los aleatoriamente, isto é definir a estrutura amostral. Crianças de 15 são facilmente encontradas no Brasil nas escolas, seria mais fácil examiná-las nas escolas do que em suas casas. Assim, escola será nossa estrutura amostral. Lembre-se que a unidade do nosso estudo é a criança, mas vamos selecioná-la apartir das escolas. Desta forma, escola é um conglomerado.

Sabemos que temos escolas públicas e privadas e que são bem diferentes em relação ao nível socioeconomico, ainda existem áreas de nível socioeconomico bem

diferentes na cidade. Deve ser estabelecido se haverá estratificação por áreas e por tipo de escola.

Ao sortear as escolas participantes deve-se decidir se todas as criançsa com 15 anos entrarão para o estudo, ou se serão sorteadas crianças dentro de cada escola. Caso sejam sorteadas, deve-se decidir se o sorteio será aleatório por meio de uma lista ou se salas de aula serão sorteadas e todas as crianças com 15 anos na mesma sala serão incluidas. Caso opte-se por sortear salas de aula, cada sala de aula será um novo conglomerado. Ao final teremos uma amostra complexa em que o amostrista irá calcular com base nas estratificações e conglomerados a probabilidade de cada indivíduo teve de participar do estudo.

Para simplificar e introduzir a noção de erro aleatório, nós vamos assumir que temos uma lista de todas as crianças da cidade e vamos utilizá-la. Vamos admitir também que não existirá erro sistemático de observação. Imaginando que temos 10 mil crianças de 15 anos na cidade, nós vamos realizar um sorteio aleatório de mil crianças, e vamos trabalhar com esta amostra de mil. No entanto, se pensarmos bem, esta é apenas uma, uma única amostra, das milhares que poderiam ser sorteadas, e vou ter que me contentar com este resultado. Com certeza vamos ter que nos contentar com uma amostra, porque seria impossível examinar 10 mil crianças em tempo hábil.

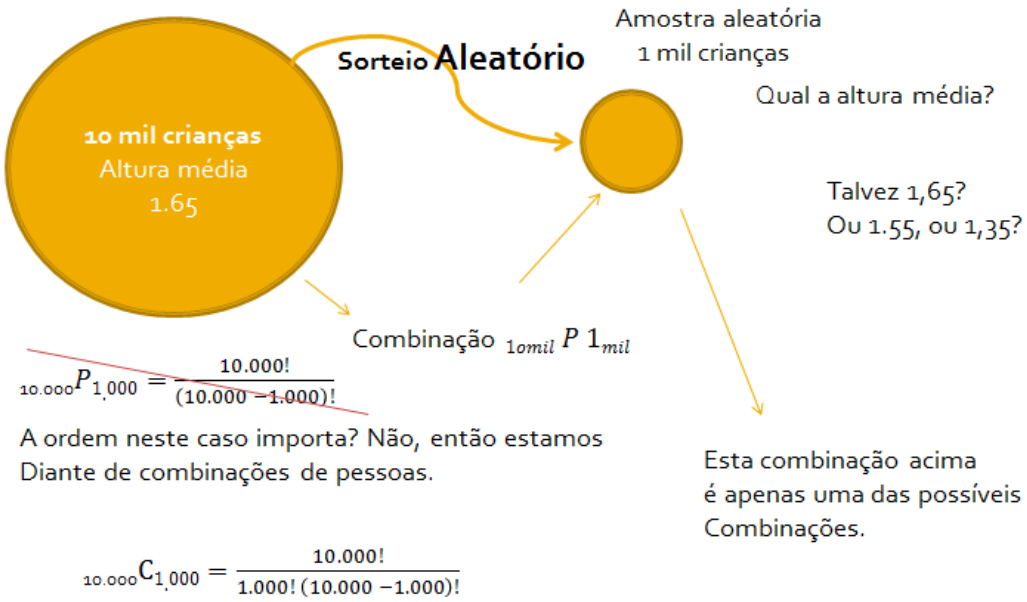
Quando fazemos uma amostra aleatória simples obtemos uma combinação, no exemplo, de mil crianças para compor a amostra. Nesta amostra vamos ter uma determinada média decorrente da combinação. Se devolvermos as mil crianças a população, e sortearmos de novo mil crianças (sorteio com reposição)[a], é possível que algumas voltem a segunda amostra combinadas com outras crianças. Pode ser também que outras mil crianças totalmente diferentes daquelas sorteadas na primeira amostra sejam sorteadas na segunda amostra. Possívelmente esta nova amostra terá uma outra média. Se continarmos a fazer isso, indefinidade, vamos selecionar milhares de amostras, e cada uma com sua combinação, vamos chegar a um ponto que as possíveis combinaçoes irão se esgotar e começaremos a repetir a mesma amostra. Veja a figura a seguir,

a- Um sorteio pode ser com reposição ou sem reposição. Se fizermos reposição, e o sorteio for aleatório simples, o novo sorteio manterá a mesma probabilidade de se sorteado para cada membro da população. O sorteio tem que ser feito de uma vez só. Se for sem reposição, ao remover

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

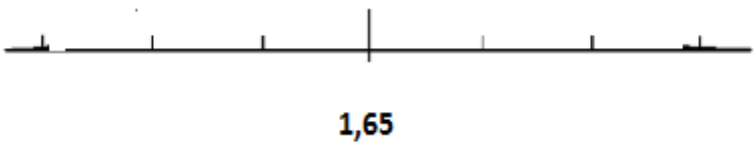
no primeiro sorteio, quem resta tem probabilidade diferente de ser escolhido. Sorteando a primeira pessoa a probabilidade dela seria de 1/n, ao sorteá-la, a probabilidade do restante é de 1/n-1, e assim por diante.



$$_{10.000}C_{1.000} = \frac{10.000!}{1.000!\,(10.000-1.000)!}$$

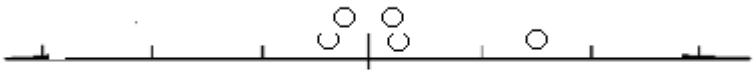
Esgotando todos os milhares de possíveis combinações vamos ter todas as possíveis médias que poderiam sair da população de 10 mil crianças, por meio de combinações de 10 mil crianças de mil em mil.

Imagine também se cada criança fosse um palitinho redondo, e se empilhassemos estes palitinhos sobre uma superfície em que a altura estivesse marcada, como se fosse uma regua.



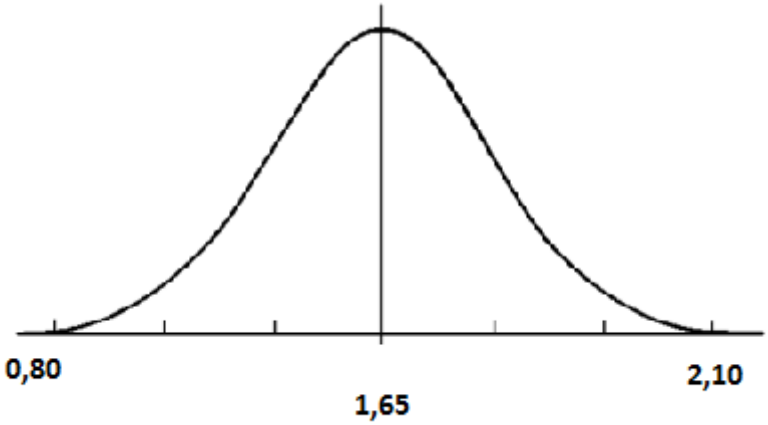
Vamos ter algo se formando como

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva



1,65

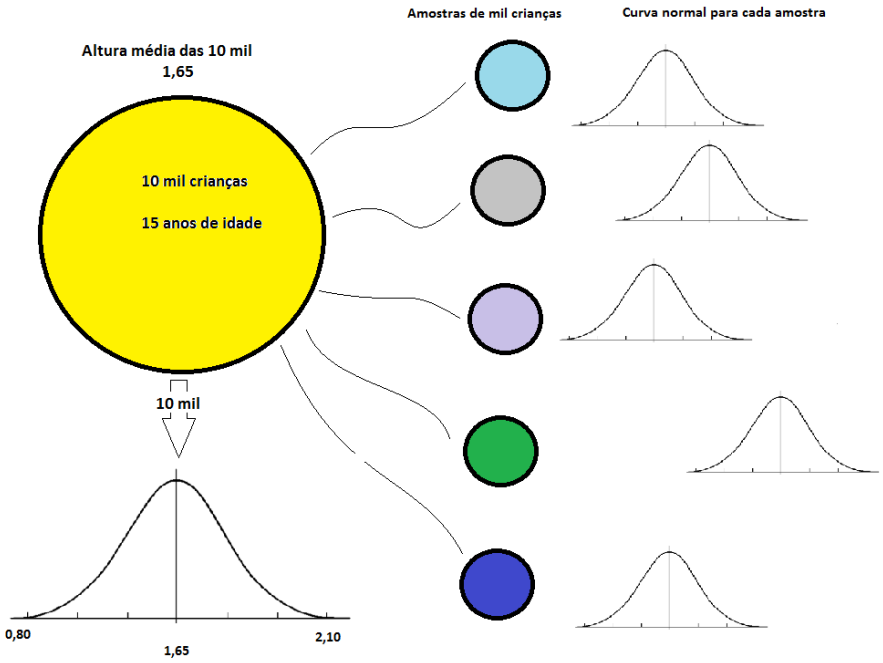
Se continuarmos a empilhar sabendo que a média é de 1,65, então a maioria das crianças deve ter esta altura, e bem poucos serão muito baixinhos e poucos serão muito altos. Ao se empilhar não sabemos a figura final, mas pode ser que a distribuição se assemelhe a uma distribuição normal, em que a média realmente reflita a médiana, isto é o ponto que divide 50% da população. Vamos supor que depois de empilhar os 10 mil palitinhos, cada um referente a altura de cada criança, tenhamos a figura abaixo, com a menor criança de 15 anos de 0,80 centímetros e a mais alta de 2,10. É pouco provável que tenhamos estas alturas aos 15 anos de idade, mas estamos inventando os valores apenas para poder explicar.



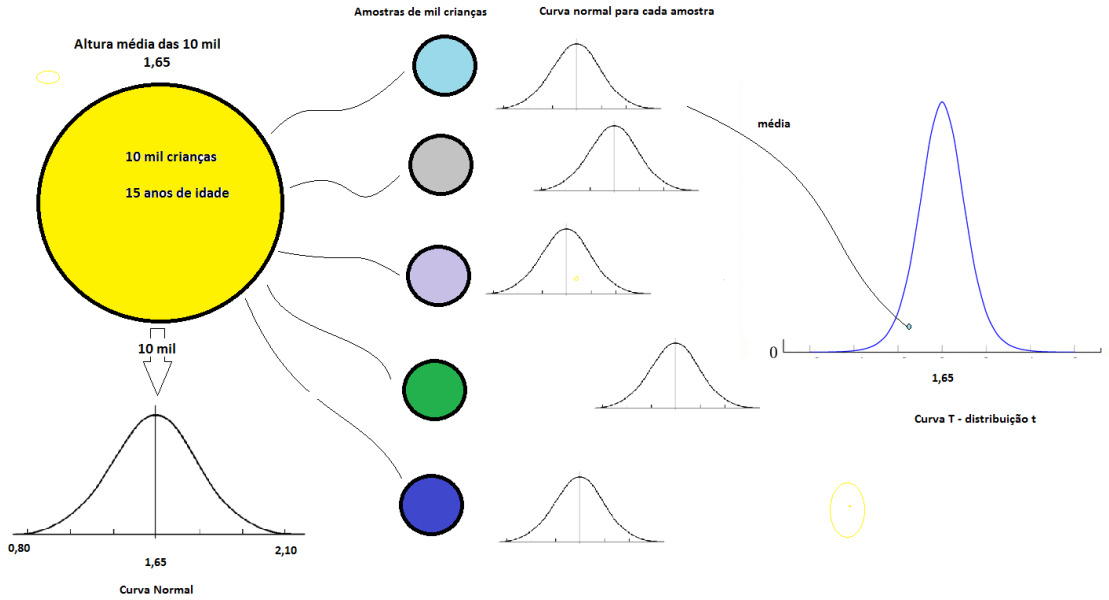
Para cada amostra de mil crianças que fizermos também podemos fazer uma curva de gaus. A curva de Gauss pode ser explicada de maneira grosseira como sendo esta curva em forma de sino, bem simétrica ao redor da média que deve coincidir com a mediana. Ao empilhar valores de alguma coisa, podemos ter vários formatos além de curva de Gauss, chamada também de curva normal, ou distribuição normal.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva



A figura acima representa uma simulação de inúmeras amostras. Se fizermos mais um exercício de empilhar os valores de todas as milhares médias geradas pelas amostras nós teremos ao final uma curva que se assemelha a curva normal, mas é mais magrinha e baixinha, embora a figura possa parecer que é mais alta.



Esta nova curva mais magrinha que a normal é composta por médias. A curva normal tanto da população total como as outras para cada amostra, era composta por

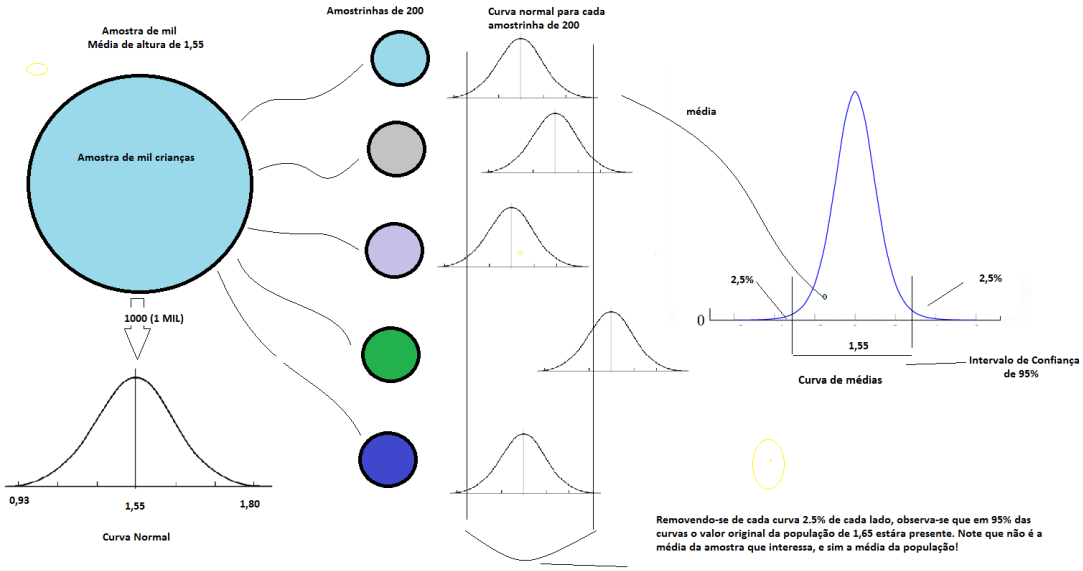
palitinhos que representavam os indivíduos empilhados. Na nova curva os palitinhos representam apenas as médias. É nesta nova curva de médias que vamos refletir. Ela é uma curva mais magrinha é mais magrinha porque nunca nesta curva vamos ter uma média de 0,80 porque o indivíduo com altura mais baixa de 0,80 vai estar sempre misturado com outros na amostra e, portanto a menor média terá um valor maior que 0,80. O mesmo acontece com o maior valor. Também esta curva será mais baixa do que a curva normal.

Quem bolou esta simulação observou um padrão. Vamos supor que tenhamos uma média X de uma população, vamos imaginar 1,65cm. Inumeras amostrinhas de mil pessoas são realizadas, e nem sempre o valor médio era de 1,65, pois era calculado apartir de combinação de mil pessoas dentre 10 mil possibilidades. Sendo assim, nestas mil amostras possuímos todas as possiveis combiações que poderiam vir da população. Na prática quando fazemos uma amostra para um trabalho podemos sair com qualquer uma destas possíveis mil combinaçoes. O mil aqui foi utilizado apenas para exemplificar, o número de possíveis amostras depende da variabilidade de altura dos indivíduos de uma população. Se na população temos pouca variábilidade as possibilidades de combinações serão poucas, mas se grande as possíveis combinações serão grandes. Como a média verdeira é de 1,65, é mais provável que as combinações sejam próximas de 1,65, mas não sabemos se a amostra que fizemos é proxima ou não, ainda mais que não sabemos a verdade. Existe uma regra na estatística em que se considera 5% um evento raro. Quem estabeleceu isso, foi meio num “chutometro” bem pensado (*educated guess*) porque precisavam de um número para dizer “ isto é pouco provável” assim decidiu-se que 5% seria o ponto de corte de pouco provável. Então assumindo que na curva de médias temos todas as possíveis médias que poderiam vir da população, porque não eliminar os 5% menos prováveis (2,5% de cada lado), e considerar o resto como provável?

Mas você deve estar pensando que não trabalhamos com milhares de amostras, se fosse assim, era melhor fazer logo uma só amostra com 10 mil do que milhares de mil. Mas ai existe outra simulação interessante. Se pegarmos uma amostra aleatória qualquer de mil indivíduos, e a partir dela fizermos milhares de pequenas amostrinhas, vamos imaginar de 200 indivíduos cada, vamos gerar inúmeras amostrinhas. O

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

interessante é que em cerca de 95% das amostrinhas a média origina da população de 1,65 estará no intervalo.



Isso quer dizer que ao fazer uma amostra, provávelmente a média não é a verdadeira, mas apartir da amostra podemos estimar um intervalo no qual possívelmente, ou melhor temos 95% de cofiança que deve conter a verdadeira média. Preste atensão que o correto é dizer que “temos 95% de confiança de que a o intervalo estimado deve conter a verdadeira média” Está errado dizer que temos 95% de confiança de que a verdadeira média está no intervalo. A confiança não está na verdadeira média, a confiança está no intervalo construido.

Desta forma sempre que calcularmos algo num estudo que foi feito com amostra, a média obtida (chamada de média pontual), não é o mais importante. O importante é o intervalo estimado com a amostra! Portanto, o valor pontual de uma prevalência estimada não é o importante, o importante é o intervalo de confiança ao redor da prevalência. Desta forma, você ja deve estar imaginando que o valor pontual de uma Odds Ratio não é o mais importante, nem o valor do RR, e que estes valores sozinhos não são tão importantes se não viérem com um intervalo de confiança.

O que me diz o intervalo de confiança? Se a média pontual de altura das crianças de 15 anos foi de 1,55 metros e o intervalod foi de 1,25 e 1,85, o que podemos concluir

é que " temos 95% de confiança de que o intervalo que vai de 1,25 a 1,85 contém a verdeira média". Talvez a verdadeira média seja 1,30 ou talvez seja 1,70, não sabemos.

O interevalo de confiança é muito importante também para fazer comparações. Se queremos comparar a altura de crianças de Ribeirão Preto e Sertaozinho, pois disconfiamos que a as crianças em Sertãozinho são mais altas, podemos fazer amostras nas duas cidades e calcular os intervalos de confiança e compará-los. Vamos supor se o intervalo de Ribeirão é o mencionado anteriormente de 1,25 e 1,85. Pode ser que a pontual média resultante em Sertãozinho seja de 1,70. Logo, bem maior do que 1,55 em Ribeirão Preto, mas ao verificar o intervalo para Sertaozinho observa-se que é de 1,40 a 2,00 metros. Comparando os intervalos notamos que os intervalos se interpoem, então temos possiveis valores reais comuns. Logo pelo menos neste estudo não podemos chegar a conclusão de que a média de altura de Sertãozinho é diferente da média de Ribeirão Preto. Isso não quer dizer que são iguais, eu não posso concluir que são iguais! Apenas não são diferentes. Isso porque em amostras infinitamente grandes, a tendência será de ter intervalos de confiança muito pequenos e, portanto, diferenças podem serão encontradas.

Algo importante do intervalo de confiança, é que ele pode variar conforme a quantidade de pessoas na amostra. Fizemos amostras de mil pessoas do total de 10 mil. Caso tivessemos feito amostras de 999 pessoas, qual seria o intervalo? Com certeza muitíssimo pequena, porque a amostra só deixou fora 1 pessoa. E se tivessemos amostrado a população toda? Neste caso, não temos erro amostral, não existe intervalo de confiança porque ele é 100% de confiança!

Algo que confunde as pessoas é pensar que quanto maior o número de pessoas o intervalo de confiança aumenta, no sentido de ter mais confiança. So que não é isso, o intervalo na verdade é de possíveis erros (possíveis médias) logo quanto maior é o intervalo maior possibilidades de médias, logo se tem na verdade incerteza, pois existirão muitas opçoes para a verdadeira média. A confiança é que este intervalo, seja ele largo ou estreito, contenha a verdadeira média.

Portanto, se tivessemos medido todas as crianças de Ribeirão Preto, e todas de Sertãozinho, mesmo que chegassemos a conclusão de que a média de Ribeirão Preto era de 1,65 e a de Sertãozinho 1,66 como utilizamos a população toda, a conclusão é

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

que sim, as médias são diferentes. Pode ser que você conclua que são diferentes, mas é pequena a diferença para ser relevante.

O que o intervalo de confiança de um Risco Relativo, ou qualquer razão que seja nos diz? Se a Odds Ratio pontual de um estudo foi de 1,7 e o intervalo for de 1,2 a 4,6, isso indica que temos 95% de confiança de que o intervalo entre 1,2 e 4,6 contém a verdadeira OR. Como este intervalo não inclui o valor 1 que seria nulo, sabemos que a força de associação parece ser positiva. Se o intervalo fosse de 0.8 a 3,5, isso significaria que uma das possíbilidades seria o valor 1 que significa nulidade, isto é não haver associação, e até mesmo pode ser que seja uma associação negativa (< 0,1). Logo neste caso, o estudo **não foi capaz** de mostrar associação estatisticamente significante.

Note que se temos um risco relativo, que é resultado da divisão do risco de expostos e não expostos, podemos também calcular o IC (intervalos de confiança) para o risco dos expostos e compará-lo com o IC dos não expostos, e tirar conclusão se os intervalos se interpoem. Isso é análogo a se calcular o IC para o risco relativo.

Quando temos diferenças de médias a lógica é a mesma, por exemplo, poderiamos calcular o IC para a altura média de Ribeirão e Sertãozinho, ou para a diferença das médias. Se Sertaozinho era de 1,70 e Ribeirão Preto 1,55, a diferença seria de 0,15, e então se calcularia o IC para a diferença. Uma vez que a diferença nula é zero, se o IC para o valor pontual de 0,15 incluir o zero, então zero seria uma possibilidade e, portanto, a conclusão seria de que não foi possível encontrar diferença entre a altura média de RP e S.

Portanto, comparando intervalos de confiança nos permitem comparar médias, medidas de frequência, interpretar razões etc. È importante ter em mente que com amostras grandes os intervalos são pequenos e, portanto é mais facil que ao comparar duas coisas elas se apresentem diferentes. Com amostras pequenas, por outro lado, pode até ser que o que se compara seja diferente, mas o intervalo vai ser tão grande que não mostrará o que se quer verificar. Assim, a quantidade de indivíduos que precisamos num estudo é importante! Existem maneiras para se calcular o número mínimo de indivíduos num estudo. Ainda pelo exposto, nota-se que o IC serve também para dar uma Idea se foi utilizado número de indivíduos suficiente no estudo. Intervalos de confiança muito largos nos dão a ideia de número insuficiente de pessoas no estudo.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

IC largos de mais dão ideia de imprecisão do estudo. IC pequeno dá ideia de estudo mais preciso.

Note que para as razões, o valor do intervalo de confiança não é simétrico, em geral é mais curto quando abaixo do valor 1. Inventando um valor é comum ver intervalos de confiança de uma razão como sendo por exemplo Risco Relativo de 2,7 sendo que o intervalo vai de 0,7 a 8.9. Vejam um exemplo de Intervalo de Confiança para a Odds Ratio num estudo de associação de um polimorfismo genetico com periodontite publicado por Kormam et al (1999).

To control the effect of age on disease severity, data were analyzed separately for non-smokers aged 40–60 years. In this age range, the composite genotype was present in 78% of severes (n=9), 26% of moderates (n=30), and 16% of milds (n=32) (odds ratio: severe versus mild=18.90 (1.04–343.05), P<0.002). The influence of genotype on disease severity is evident in the cumulative frequency distribution (Fig. 2) that shows	Para controlar o efeito de idade na .....[ ] Odds ratio : severo vs moderado = 18,90 com Intervalo de Confiança de 1,04 a 343,05. Valor de p < 0.002.

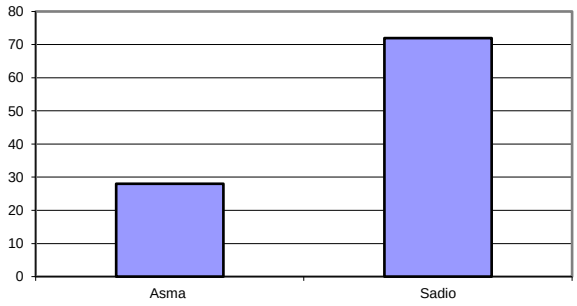
Note que o valor pontual da Odds Ratio é de 18,90 ,mas o intervalo vai de 1,04 a 343,05. Sim um intervalo muito largo demonstando que existe algo de errado no estudo, ou amostra muito pequena ou variabilidade muito grande dos tipos de indivíduos no estudo.

Intervalo de Confiança para Proporções

O exemplo utilizado para introduzir as noções de intervalo de confiança partiu da variável altura que tem distribuição normal na população. Depois mencionamos sobre intervalo de confiança para risco, que é uma frequência apartir de uma variável que não tem distribuição normal. De forma geral as variáveis podem ser quantitativas e qualitativas. Variáveis quantitativas quantificam coisas como peso, altura, pressão arterial, distâncias, número de dentes cariados, perdidos e obturados e etc. As variáveis qualitativas qualificam os indivíduos como, por exemplo, em doentes e não doentes,

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

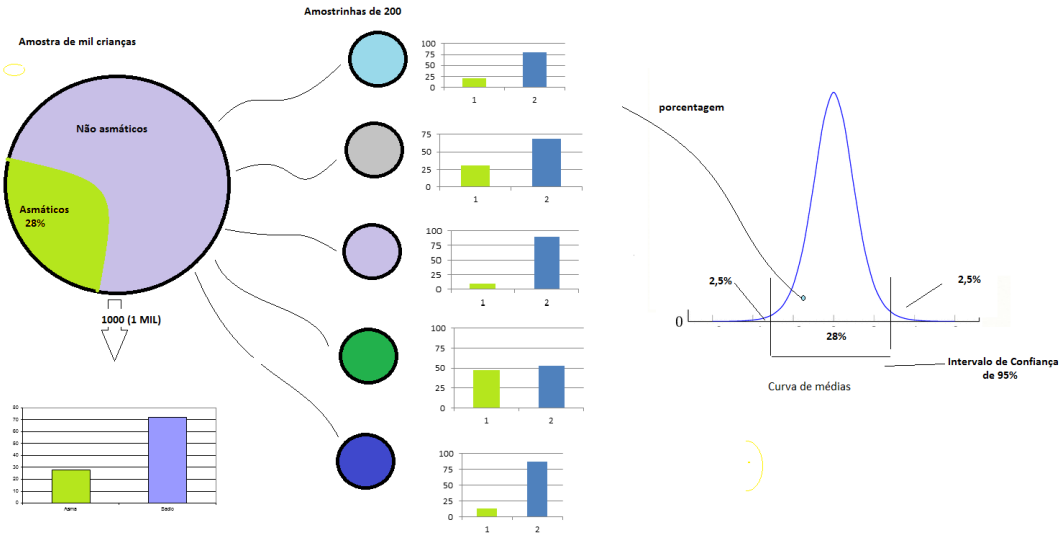
votantes do indivíduo X ou Y nas eleições. As variáveis qualitativas não tem distribuição normal, se você empilhar palitinhos que se referem a pessoas doentes e não doentes vai encontrar uma distribuição como a da figura a seguir.



Nesta figura temos 28% de indivíduos com asma. Como é que a partir desta variável qualitativa, chamada de categórica (com dois níveis: asmáticos e não asmáticos), vamos gerar um intervalo de confiança? A ideia é a mesma. Vamos partir da nossa amostra de 10 mil crianças de 15 anos e agora vamos verificar a prevalência de asma e não mais a altura das crianças. Sorteando mil crianças qual será a prevalência de asma na amostra? Como sempre não sabemos. Porém se a amostra foi aleatória simples de todas as crianças de 15 anos de Ribeirão Preto, podemos a apartir dela, sortear amostrinhas pequenas de 200 cadas, e calcular a prevalência. Se depositarmos estas prevalências (palitinhos) em uma base vamos notar que vai se formar uma curva de prevalências geradas das amostras que se assemelha a curva t. Esta curva será formada pelas prevalências das inúmeras amostrinhas de 200 pessoas. Nesta curva também removemos os extremos 2,5% e geramos assim, um intervalo de confiança!

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva



Legal não? Então quer dizer que mesmo com uma variável que é qualitativa eu consigo geral uma curva de possíveis porcentagens que tem um jeito de distribuição normal ou distribuição t? Sim, porque esta nova distribuição é distribuição de possíveis prevalências que podem ir de 0 a 100! Note que asma não é variável quantitativa e sim qualitativa, mas as várias prevalências em diversas amostras constituem valores numéricos que podem lembrar uma variável contínua (lembrar apenas, nao quer dizer que é).

Esta produção de uma curva com aspecto de distribuição entre normal e t- independe do tipo de distribuição inicial da variável é chamada de TEOREMA DO LIMITE CENTRAL. Isto é , o teorema diz que a distribuição de uma média (esperança) de qualquer que seja distribuição original de uma variável resulta numa distribuição normal ou aproximadamente normal. Foi exatamente isso que foi verificado, mesmo utilizando uma distribuição original que era de uma variável dicotomica (categórica) a distribuição de suas esperanças (porcentagem esperada de cada amostra) acabou virando uma distribuição aproximadamente normal. Note que cada valor desta nova distribuição não são indivíduos mas porcentagens. Ainda o teorema vale apenas para quando temos amostra é muito grande (mais de 30 ou 40 pessoas ou mais).

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Intervalo de Confiança para amostras não aleatórias simples (amostras estratificadas e complexas).

Até agora comentamos sobre amostras aleatórias simples, mas e se a amostra for complexa com estratificação e conglomerados, como será que calculamos o IC? Simples, obedecendo a amostragem. Se a amostra de mil crianças tivesse sido obtida por meio de estratificação em 4 áreas da cidade, quando fossemos fazer as amostrinhas de 200 crianças, nós devemos dividir a amostra de mil de acordo com as 4 áreas e compor esta amostrinha de 200 apartir da estratificação.

Preste atenção nos problemas acarretados por não considerar o tipo de amostra ao gerar o IC. Se por um acaso tivessemos conglomerados, a amostrinha de 200 deve obdecer os conglomerados também. Se não fizemos isto, vamos gerar um intervalo de confiança falso. Imagine que estratificamos as mil crianças em 4 estratos socioeconomicos, então garatimos representatividade dos 4 estratos, porém ao fazer a amostrinha eu desconsidero os estratos, e posso acabar incluindo apenas crianças mais altas, resultando numa média alta e consequentemente alargando o IC. Isso significa que ao estratificar uma amostra, não somente eu consigo representatividade mas reduzimos também a variabilidade da amostra e consequentemente do IC. Se ao produzir o IC eu não levo em consideração o estrato, ele será maior do que deveria, e perdemos em precisão. Se vamos utilizar o IC para comparar as duas cidades com o IC mais largo, temos maior chance dos IC se interporem e não encontrarmos diferenças estatisticamente significantes.

Quando usamos conglomerados a tendência do próprio conglomerado é de aumentar a variância final da amostra. Se não levarmos em consideração o conglomerado, o IC tende a diminuir e ai aumentamos a possibilidade de encontrar diferença quando a diferença não deveria existir. Por isso, sempre que usamos estratificações, e conglomerados devemos levá-los em consideração para calcular o IC de forma adequada. Portanto, o intervalo de confiança será diferente se fizermos uma amostra aleatória simples ou uma amostra complexa. A diferença do erro padrão quando a amostra é complexa em relação ao que seria se a amostra fosse aleatória simples chamamos de efeito de desenho de estudo (design effect).

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Um outro detalhe sobre amostragem é que por vezes precisamos sobre amostrar algum grupo que é subrepresentado na população. Por exemplo se queremos no nosso estudo sobre altura em Ribeirão Preto, comparar a altura de descendentes de japoneses com os demais, e estes representam apenas 2,5% da população, no total da amostra teremos cerca de 25 japoneses. Pode ser que este número seja relativamente pequeno, então decidimos dobrar o número de japoneses, para que a medida de altura entre eles seja mais confiável, ainda mais que queremos dizer que os japoneses de Ribeirão Preto são ou não semelhantes as demais etnias. No entanto vamos ter 1025 indivíduos, e agora eles representam mais do que 2,5% da amostra. Neste caso, podemos ponderar o cálculo da média de altura, e considerar no total cada japones com valor de meio. Isso é apenas para que a altura deles não pese muito no cálculo de altura geral para a cidade , uma vez que a proporção deles é maior na amostra do que na população. Se eles forem mais altos irão puxar a média da cidade para cima, e se forem mais baixos vão puxar a média para baixo.

Estamos admitindo que basicamente todos que foram chamados para a nossa amostra para a estimação da altura em Ribeirão Preto, participaram do estudo. Vamos supor que embora tanhamos estratificado e sorteado quantidades que iriam represenar a população, os indivíduos não participam. No final a amostra não vai ter a mesma distribuição percentual do planejado e passa a não refletir mais a população. O que podemos fazer é, uma vez que se tem conhecimento no planejamento da proporção real, pode-se calcular pesos e ajustar a contribuição de cada estrato para se manter a representação original. Obvio que o ideal seria que todos participassem, mas as vezes isso não acontece, e temos que usar pesos calculados meticulosamente pelo amostrista, para que possamos fornecer um valor representativo para a população.

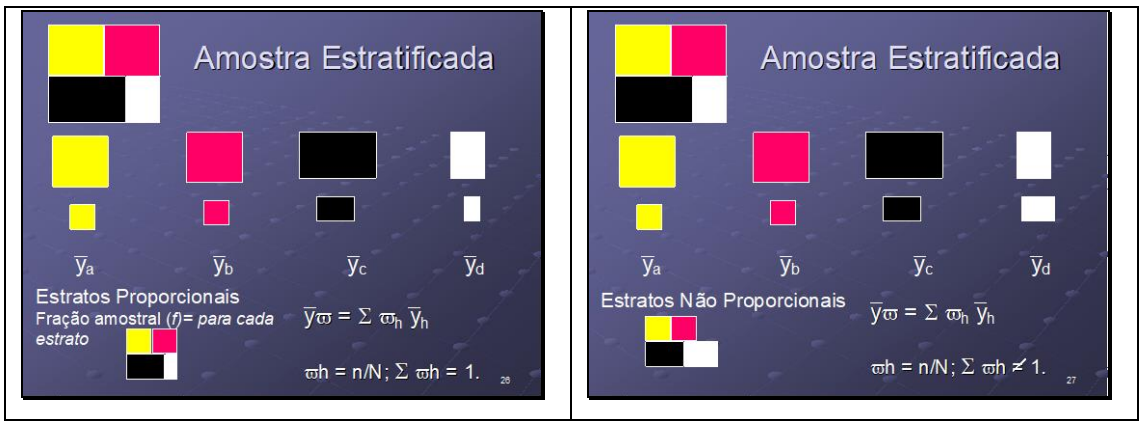
Como ressalta Valliant et al “sem a utilização de pesos amostrais, estimadores refletem apenas nuances de uma amostra particular e pode conter niveis significantes de viéses”.

Valliant, Richard; Dever, Jill A.; Kreuter, Frauke. Practical Tools for Designing and Weighting Survey Samples: 51 (Statistics for Social and Behavioral Sciences) (p. 307). Springer New York. Edição do Kindle.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

No estudo do levantamento Nacional de Saúde Bucal em 2003, embora tenham planejado o estudo de forma razoavelmente adequado, maior quantidade de pessoas brancas, mais ricas e homens se recusaram a participar do estudo, predominando portanto uma amostra com sobreamostragem de mulheres, negros e pobres. Sem levar em consderação os pesos, que não foram calculados, o estudo apresentou prevalência de cárie muito maior do que deveria ser. Não somente teve viés de seleção (introduzido pela não participação) como também ao não levar em consideração pesos e o desenho do estudo, os intervalos de confiança foram calculados erradamente. Assim, em todos os estudos realizados com o banco de dados deste estudo não foram apropriados.



Nas duas figuras acima, temos amostra estratificada proporcional e do lado direito a amostra estratificada não proporcional (onde um estrato foi sobreamostrado). Note que a única diferença entre eles é que wh é diferente de 1, isso significa o peso de cada elemento vai ser diferente de 1 para a amostra estratificada não proporcional. Quando se tem amostra estratificada simples, faz se amostra de cada estrato e simplesmente soma-se tudo. O que vai acontecer é que garantimos a representatividade (se todo mundo responder propocionalmente), e o IC tende a ser menor do que fosse realizada uma amostra aleatória simples.

A seguir um pequeno resumo de alguns detalhes sobre amostra probabilistica que não foram mencionados em ordem.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Amostragem – Resumo

Agora que você já entendeu o que é uma amostra, para que serve a estatística, e o que é o intervalo de confiança, será mais fácil entender alguns outros detalhes da amostragem.

- Amostragem é utilizada por razões de economia (de tempo e de recursos) e de acurácia, isto é reduzir erro sistemático, ao máximo. Diferentes processos de amostragem resultam em diferentes erros de amostragem. Todo esforço deve ser feito para escolher um desenho (processo de amostragem) que resulte no menor erro de amostragem possível.
- Para ser útil a amostra deve satisfazer os seguintes critérios: (1) deve ser representativo da população, (2) ser economicamente eficiente, (3) resultar em estimadores não enviesados, precisos e capazes de serem testados pela confiabilidade.

Problemas relacionados à amostragem

Em relação à estrutura amostral (*sampling frame*) os principais problemas são: (1) elementos faltantes, (2) elementos estranhos, (3) listas com duplicatas, (4) aglomeração de elementos (conglomerados). Por exemplo, se usarmos uma lista de telefone algumas pessoas podem não estar na lista porque não querem seus nomes divulgados nas listas (1), além de residêntes temos telefones cormeciais (2), algumas pessoas podem ter dois telefones (3), pessoas de uma mesma residência podem ter vários telefones (4).

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Resolvendo os problemas de estrutura amostral. (1) Se o problema for **muito pequeno** ignore. Mas a relevância pode ser diferente em diferentes populações. Nos EUA quando mais de uma família mora numa casa isso é ignorado, pois raramente encontraremos duas famílias dividindo um mesmo teto. Porem, no Brasil isso pode ser relevante dependendo da cidade. (2) Se necessário escolha outra estrutura amostral. (3) se possível corrija a lista da população, como por exemplo, fazendo um censo antes de selecionar os domicílios.

Tipos de Amostragem

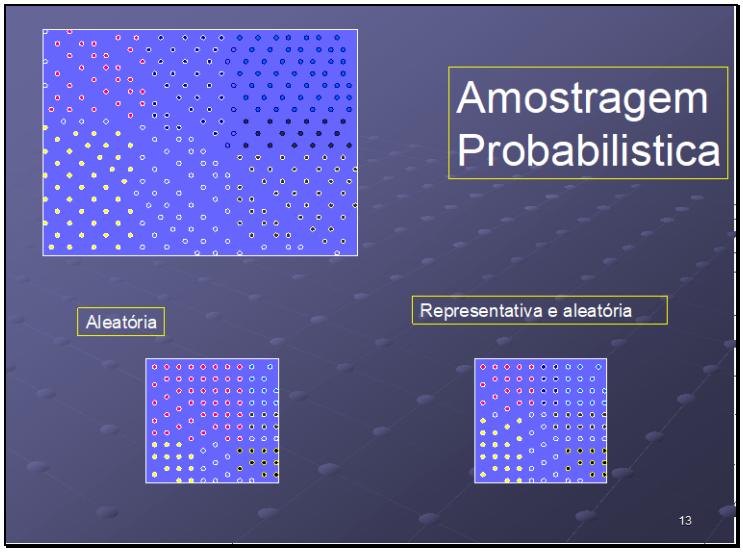
Tipos de Amostragens

- Não Ciêntifica ou não probabilística
  - amostra proposital
  - cota
  - pessoas típicas
  - amostra por conveniência

- Amostra probabilística
  - amostra simples aleatória
  - sistemática
  - estratificada
  - conglomerado

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva



- **Amostra Simples**

A amostragem simples é o processo mais básico de amostragem, todos os outros são variações da amostragem simples. Nesta amostra cada pessoa tem chance igual de ser selecionado para o estudo. Mas isso, não significa que cada grupo de indivíduos dentro da população a ser amostrada terá a mesma chance. Por exemplo, se um grupo tem 30% de mulheres e 70% de homens a chance de um homen ser selecionado é maior do que a chance de uma mulher. Mas em termos de indivíduos (independente do sexo), a chance é igual para todos.

A amostra aleatória serve para não envisarmos o estudo escolhendo pessoas por algum motivo especifico. Vamos supor que fizéssemos uma amostra olhando para o individuo e escolhendo sem critério. Embora pensassem que não existe critério, podemos nos simpatizar mais com uns do que com outros, e assim, cada indivíduo não vai ter a mesma chance de participar da amostra. Assim, dizemos que uma amostra sendo aleatória é mais um passo importante para mostrarmos que nossa pesquisa tem validade interna.

Vantagens:

- Relativamente simples

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

- Erro da amostragem pode ser calculado facilmente

Desavantagens :

- Não garante representatividade
- Não garante a representação de pequenos grupos o que pode dificultar comparabilidade
- Necessita de uma lista de todos os elementos da população para o sorteio

- **Amostra Sistemática**

Vantagens

- fácil de ser realizada
- investigador não precisa saber a estrutura amostral, a estrutura amostral pode ser construída conforme o estudo estiver em andamento. Por exemplo, num estudo clínico randomizado, os pacientes vão sendo alocados a cada grupo de tratamento conforme eles vão aparecendo. Por exemplo, seria impossível reunir 300 casos de cirurgia cardíaca para depois fazer a cirurgia de todos de uma vez só.

Desvantagens

- se a lista tiver alguma periodicidade , o processo não será mais aleatório
- é apenas possível para pequenas populações desde que a lista de todas as pessoas esteja disponível
- Não se pode estimar variância para grupos de indivíduos dentro do estudo

- **Amostra Estratificada**

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

A população é dividida em subgrupos e então aleatoriamente os indivíduos são sorteados dentro de cada grupo. A alocação então poderá ser proporcional ou não proporcional. Por exemplo, uma população com 30% de mulheres e 70% de homens, podem sortear numa amostra de 100, 70 homens e 30 mulheres, ou se for necessário por alguma razão podemos sortear proporções diferentes. Portanto, podemos ter amostra estratificada com **alocação proporcional** ou **desproporcional**.

Vantagens

- Reduz o erro amostral - isto porque garante que diferentes grupos sejam representados proporcionalmente.
- Garante representatividade dos grupos que constituem a amostra e que foram usados para a estratificação

Desvantagem

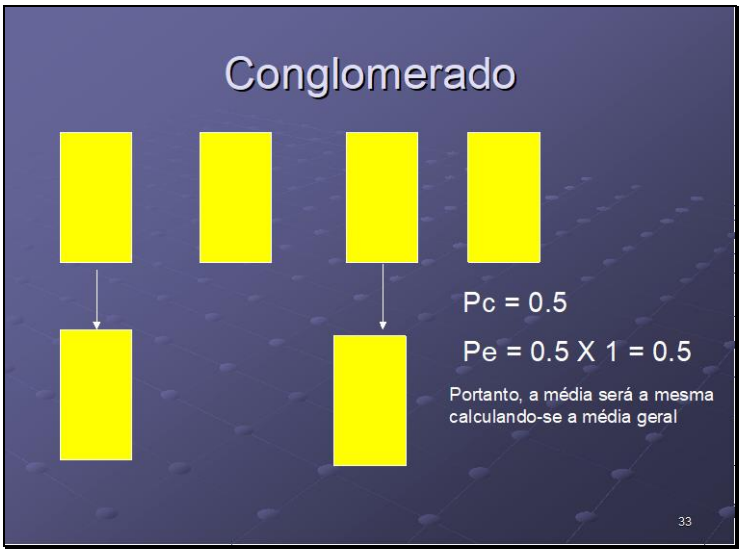
- Precisa de lista da população
- Precisa de um certo conhecimento da população para identificá-los

❑ **Amostra por conglomerado**

Para economizar dinheiro e tempo, o conglomerado é frequentemente utilizado em amostragem. Podemos amostrar quarteirões, escolas, salas de aula, e dentro de cada conglomerado ou fazer um novo sorteio aleatório dos indivíduos ou incluir todo mundo. A vantagem e que podemos aumentar o tamanho da amostra aumentando a acurácia do estudo, mas isso tem que ser balanceado com a distribuição de diferentes grupos na população. Em geral, a amostragem por conglomerado **diminui** a acurácia quando comparado com a amostra simples aleatorizada. Porque o intervalo de confiança aumenta? Tente acompanhar na figura abaixo. Imagine uma população formada de 4 conglomerados, e você escolhe dois deles, isto é com uma probabilidade de 0.5 (Pc). Como todos os elementos do conglomerado entram para a amostra, não haverá erro amostral dentro dos conglomerados, e, portanto, a prevalência da doença será a prevalência simples. Porém o erro amostral será calculado entre os

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

conglomerados. Isto é podemos ter diferenças muito grandes entre os conglomerados e portanto, o IC tende a ser maior.



Para o cálculo de IC com amostra por conglomerados precisamos também de pacotes estatísticos especiais, porque há necessidade de se levar em consideração que todos os elementos vêm de um mesmo conglomerado, pois eles têm algo em comum.

Vantagens

- menor custo
- menos trabalho

Desvantagem

- menor acurácia (IC maior)

- **Amostragem de multi-estágios (complexas)**

As amostras complexas misturam várias das possibilidades sitadas acima: amostras estratificadas, seguidas de amostragem simples, conglomerados e etc. Como a amostra estratificada diminui o IC e o cluster aumenta, as amostras complexas fazem arranjos entre estratificação, conglomerados e amostras aleatórias simples para que se cheque ao melhor custo e facilidade de amostragem.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Vantagem

- não é necessária lista de toda população

Desvantagem

- o erro da amostragem será maior do que na amostragem simples mas isso é compensado por não requerer uma lista completa de toda população. Além disso, garante representatividade.

Intervalo de Confiança nas pesquisas de intensão de votos em eleições

Nas eleições várias agências de pesquisas de campo como IBOP, Data-Folha entre outras, fazem pesquisas e divulgam seus resultados de intensão de votos por meio de estimadores pontuais e o intervalo de confiança que eles chamam de "margem de erro" que vai em geral de 3% para mais e para menos do valor pontual. O que eles chamam de margem de erro, nós chamamos de Intervalo de Confiança. Nas eleições de 2014 a imprensa melhorou a forma de explicar o IC. Antes eles apenas falavam que a margem de erro era de 3% para cima e para baixo. Em 2014 começaram a explicar (de forma melhor mas não completamente correta) que o intervalo significava que se fizessem 100 pesquisas em 95% delas os valores estariam contidos no intervalo. Esta explicação não é a correta, mas é melhor do que a que davam antes.

Quando os jornalistas se referem a empate técnico (quando os intervalos se sobrepoem) é o mesmo que nós dissemos que "não foi possível encontrar diferenças entre os candidatos com aquela pesquisa".

Acurácia, precisão e validade de um estudo.

Chegou um bom momento para se falar sobre a diferença entre precisão e validade de um estudo. Já discutimos validade que seria o estudo bem realizado em termos de ausência de viéses, e fatores de confusão (ou que pelo menos esteja

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

controlado). A precisão é outro lado. Podemos ao final de um estudo ter uma medida bem precisa com intervaldo de confiança bem pequeno, mas cuja validade interna é questionável devido à viéses e outros problemas. Da mesma forma podemos ter um estudo válido com precisão duvidosa.

Assim, controlar viéses e fatores de confusão não afeta a precisão do estudo, mas a validade interna do mesmo. Um número grande de indivíduos em nada melhora a validade interna do estudo, apenas a precisão do mesmo!. Portanto, dizer que erros de amostragem (viés de seleção) num estudo de 300 mil pessoas serão minimizados pela quantidade enorme de pessoas é completamente insano! Existe uma tendencia de se dizer que erros se diluem num estudo com número grande de indivíduos, isso não acontece! Precisamos sim de amostras com números adequados, mas isso não suprime viéses.

Essas duas virtudes de um estudo (validade e precisão) por vezes competem entre si, e o investigador vai ter que decidir para não sacrificar muito nem um nem outro. Vamos supor que gostariamos de fazer um estudo para analisar cárie dentária na população, e o ideal seria examinar cuidadosamente durante 1 hora, se certificar com mais dois dentistas, e tirar radiografias dos dentes posteriores para ter certeza na avaliação de cáries interproximais. Seria um método com ótima validade, porém não coseguiriamos examinar muitas crianças para se conduzir um estudo de prevalência de cárie na população. A decisão então é utilizar um exame sem tanta validade, e aumentar o número de pacientes. A validade menor, não seguinifica mal feito sem validade, significa que talvez possamos perder algumas lesões de cárie, e isso é aceitavel até um determinado ponto.

Se a virtude de um estudo é ser válido e preciso, isso quer dizer que seria um estudo sem erros sistemáticos e sem erros aleatórios tanto de observação como de amostragem. Este estudo sem nenhum erro é apenas um desejo, e é chamado de acurácia de um estudo. **Acurácia** significa ausência de erro sistemático e também de erro aleatório. Este é o significado do termo na física de onde ele surgiu. Em alguns livros de epidemiologia e odontologia você irá encontrar que acurácia é ausência de erro sistemático apenas, mas esta errado! Acurácia é ausência dos dois erros tanto sistemático como aleatório. Ausência de erro sistemático significa validade.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Um adendo do livro de Campbell e Stantley interessante é quando comentam sobre ameaças a validade interna de um estudo. Basicamente já mencionamos em toda a apostila sobre os itens mencionados por estes autores. Irei fazer um resumo por achar que historicamente é muito importante. Para quem tem curiosidade aconselho ler o primeiro livro destes autores de 1963. Oito diferentes classes de variáveis estranhas ou ameaça a validade interna de um estudo.

1. Historia – eventos específicos que ocorram entre a primeira medida e a segunda num experimento além da intervenção. No decorrer do texto falamos que se eventos acontecem ao longo do estudo mesmo que seja um experimento esses eventos podem desequilibrar os fatores de confusão ou mesmo alterar o efeito da intervenção.
2. Maturação – processos que acontecem entre os participantes do estudo operando como função de passagem do tempo per se incluindo crescimento, ficar mais velho, tornar-se mais irritado, mais cansado.
3. Testagem – efeitos de um teste sobre outro teste. Por isso em alguns experimentos toma-se cuidado em ter um período de washout.
4. Instrumentação – em calibração de uma medida ou alteração dos observadores ;
5. Regressão estatística – ocorre quando grupos são selecionados com base em seus scores estremos, o que se chama de *regression towards the mean*
6. Viés – resultante de escolha diferencial entre grupo experimental e controle, aqui seria especialmente relacionado ao viés de seleção.
7. Mortalidade experimetnal ou perda diferencial de respondentes entre grupo experimental e controle. Comentamos a importância do viés de sobrevivencia, que podem acontecer em estudos de coorte ou experimentos com perda de participantes seja por morte ou mesmo abandono do estudo.
8. Interação seleção-maturação – especialmente quando se tem multiplos grupos

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Anexo A – Tipos de Estudos e Causalidade

Questão 4. Marque V pra verdadeiro ou F para falso e se falso justifique.

( ) Num estudo de caso-controle para verificar associação entre prematuridade e asma , o pesquisador observa a perda de indivíduos que decidem sair do estudo durante os primeiros anos de observação. Esta perda de indivíduos pode levar a um viés de seleção chamado de viés de aderência que é muito comum neste tipo de estudo.

( ) Vários fatores ameaçam a validade de um estudo como por exemplo os vieses que precisam ser ajustados posteriormente durante a análise estatística.

( ) Um estudo observacional para ser válido não precisa levar em consideração as variáveis de confusão porque elas são controladas durante a aleatorização dos indivíduos.

Anexo B – Medidas de Associação e Frequência

Utilize os dados da tabela abaixo para responder as questões a seguir.

Questão 1.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

		BRONQUITE		TOTAL
		Sim	Não	
Álcool	Sim	273	753	1026
	Não	84	1046	1130

Questão 1. Considerando os dados acima calcule, e interprete corretamente e de forma completa as medidas abaixo. Se precisar de informações extras utilize quando necessário o ano de 2000, tempo de 10 anos e indivíduos de 40 a 45 anos. Lembre-se ou está correto ou completamente errado.

a) Medida de associação que pode ser calculada considerando que é um estudo de coorte.

b) Medida de associação que pode ser calculada se o estudo for um caso-controle

c) Medida de associação que pode ser calculada se o estudo for um estudo transversal

Questão 2. Nas tabelas abaixo, foi realizada estratificação pela condição de fumante.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Fumante

		Bronquite Sim	Bronquite Não	TOTAL
ÁLCOOL	+	260	556	816
	-	30	82	112

Não Fumante

		Bronquite Sim	Bronquite Não	TOTAL
ÁLCOOL	+	13	197	210
	-	54	964	1018

158

Questão 3. Comparando a medida que você calculou na Questão 1 para o estudo transversal com os resultados da Questão 2, o que como você classificaria o papel do fumo na associação entre álcool e bronquite. Justifique.

Questão 4. No estudo abaixo sobre hipertensão e derrame do tipo caso-controle aninhado numa coorte prospectiva responda:

Hipertensão	Derrame	Não Derrame	
Sim	500	4500	5000
Não	20	980	1000
Total	520	5480	6000

a) Qual a medida de frequência que você pode calcular

b) Calcule esta medida e interprete

c) Qual a medida de associação que você pode calcular

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

d) Calcule esta medida e interprete

Bioestatística

Embora seja uma continuidade dos assuntos anteriores, vamos chamar esta parte da apostila de Bioestatística. Começamos tentando desvendar causalidade por meio de desenhos de estudos, aprendemos o que estes estudos estimam. Discutimos possíveis problemas que ameaçam a validade interna na montagem e condução destes estudos (fatores de confusão, modificadores e vieses). Vimos também que estes estudos são conduzidos com amostras que devem ser apropriadamente realizadas, mas que, no entanto, geram possíveis erros aleatórios. Esses erros são levados em consideração nas análises gerando os intervalos de confiança que são utilizados para ajudar a descrever um evento fazendo inferência para a população como também testar hipóteses de associações.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Quando falamos de erro aleatório de amostragem já estamos falando de estatística propriamente dita, no caso inferêncial, isto é com a finalidade de fazer inferência a população. O principal objetivo desta apostila é que o leitor consiga entender princípios básicos de estatística para que consiga analisar artigos científicos, e entender qual a estatística apropriada quando for realizar sua iniciação científica ou trabalho de conclusão de curso. Em geral o que mais utilizamos é inferência estatística na prática e por isso é nosso foco.

Um livro de estatística em geral começa com revisão de probabilidade, segue pela estatística descritiva, medidas de tendência central e medidas de dispersão incluindo desvio padrão, teste de hipóteses, comparação de médias etc. Nossa estrutura será diferente, como já vem sendo diferente dos livros de epidemiologia. Continuamos abordando desafios e resolvendo os mesmos. Já vimos o conceito de intervalo de confiança e vamos expandi-lo. Em nenhum livro de estatística você irá ver intervalo de confiança e erro padrão (que ainda não foi abordado) antes de desvio padrão, mas faço questão de manter esta ordem estressando que são conceitos diferentes, com propósitos diferentes. É uma duvida frequente dos alunos e pesquisadores quando usar desvio padrão e erro padrão, e a duvida sure por não saber o que significam e que são conceitos completamente diferentes.

Outra grande diferença com livros será a escassez de fórmulas, vamos tentar da mesma forma que apresentamos epidemiologia ressaltar sua capacidade intuitiva para entender conceitos básicos de estatística. No entanto, várias fórmulas serão apresentadas mostrando que o raciocínio pode ser resumido com fórmulas, assim se você decidir estudar com um livro de estatística poderá comparar com o assunto da apostila.

Além da parte teórica o texto contém como fazer análises no programa estatístico SAS e temos uma apostilinha à parte com princípios do **Rstudio**. Embora muitas pessoas que não conhecem o programa SAS têm preconceito dizendo ser um programa difícil e complicado, na verdade ele é muito intuitivo e mais fácil de usar do que um programa de clicar. Além disso, ele ajuda a desenvolver raciocínio lógico. Outro motivo de utilizar o SAS é que a Universidade de São Paulo tem este programa disponível

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

que pode ser utilizado na sala de aula. Você não precisa aprender a usar o SAS, mas ele será utilizado na sala de aula apenas ilustrar as estatísticas.

Adicionando no ano de 2021 algumas noções do R, e também vamos falar da existência de dois programas chamados Jamovi e Jasp que estão populares nas redes, porém são limitados. Qual o grande problema de programas limitados para se fazer algo? O problema é que nem sempre são dotados de ferramentas de diagnóstico adequado, e isso leva as pessoas a fazerem as coisas erradas. Por outro lado, fazem coisas erradas, porque não apredenram estatística básica, aprenderam apenas apertar os botões de programa. Podemos usar um programa limitado, sabendo das limitações, e se necessário vamos pedir ajuda.

Um exemplo prático aconteceu há dois anos. Alguém fez uma análise de variância de duas entradas num programa que usam em laboratório. A revista retornou um artigo alertando que os gráficos mostravam que as variâncias entre os grupos eram diferentes, isso invalidava a utilização de anova-two way. Um pesquisador me pediu para verificar, e eu sem tempo disse para testar se as variâncias eram mesmo diferentes e resolver. Resultado, o tal programa tem nos manuais uma nota dizendo que se forem diferentes as variâncias para o pesquisador procurar outro programa. Quando fui avaliar os dados para ajudar, notei que nem mesmo a distribuição era normal, e as variâncias completamente diferentes. Solução foi mudar a estatística e utilizar um pacote importado dentro do SAS e que poderia ter sido feito no R também. Usar um pacote estatístico adequado é importante, mas nada mais importante do que aprender a teoria com base sólida. Apertar botão qualquer criança faz.

A estatística que estamos vendo aqui é bem básica e lida com modelos probabilísticos chamado de estocástico também. Contrapondo ao estocástico temos processos que são determinísticos, como por exemplo, quando queremos estimar qual seria o local que uma pessoa está dado que ela dirige a uma determinada velocidade. Não temos nessa predição o componente aleatório é determinístico. Se a uma pessoa saiu de Ribeirão Preto em direção a Araraquara numa velocidade de 100km por hora constante e o percurso tem 100km, quando der meia hora deve estar no km 50 da distância e não vamos ter dúvida sobre a estimativa. Modelos determinísticos são muito comuns em física. Aqui sempre lidaremos com

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Teste de Hipótese

Retornando ao intervalo de confiança vimos que por meio dele fazemos muitas coisas, como por exemplo, atribuir incertezas criadas pela amostragem às medidas pontuais de frequência ou de associação calculadas em vários estudos. Assim, se quisermos comparar duas médias, vamos verificar se os intervalos se sobrepõem ou não, se quisermos comparar três ou dez médias simplesmente comparamos os intervalos de confiança de várias médias. Se quisermos comparar proporções também faremos o mesmo. Vamos dizer que resolvemos basicamente todos os nossos problemas comparando intervalos de confiança. Intuitivamente é assim mesmo. No entanto, nos livros clássicos, estatística é apresentada de forma diferente, sempre cheio de fórmulas e compartimentado, e em geral os aprendizes se perdem não entendendo o princípio básico e lógico que existe por traz. Eu comparo à dificuldade de entendimento das fórmulas para se calcular odds ratio, e medidas de associação, pois não adianta apenas decorá-las. Tudo é sempre um pouco mais cheio de detalhes e talvez complicado do que estamos expondo, mas para isso existem livros mais detalhados. O intuito nesta apostila é compreender os princípios.

Teste de Hipótese

Quando comparamos duas médias, utilizamos o intervalo de confiança para verificar se eles se interpõem ou não. Se eles não se interpuserem, quer dizer que nenhuma das possíveis médias de um grupo foi igual ao do outro grupo, logo são diferentes estatisticamente. Se houver interposição dizemos que não foi possível encontrar diferenças.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Se estivermos comparando duas médias, isso quer dizer que cada uma deve vir com uma característica diferente, por exemplo, ao comparar as médias de altura das cidades de Ribeirão Preto e Sertãozinho é o mesmo que testar a hipótese de que as médias de alturas são iguais entre as duas cidades, ou ainda se existe associação entre cidade e altura.

Note que se vamos fazer um estudo para comparar estas médias, significa que temos dúvidas se elas são ou não semelhantes. Logo, montamos o estudo para testar a hipótese de que as médias nos grupos são iguais, ou podemos dizer que montamos o estudo para verificar se existe associação entre cidade e altura.

Só fazemos um estudo se houver necessidade de esclarecer algo, isto é para testar uma hipótese. O teste de uma hipótese começa com o desenho do estudo bem montado (como já discutimos) e termina na estatística na análise de resultados. A parte da estatística assume que o estudo foi muito bem montado, e que não existem viéses, apenas erro aleatório de amostragem.

Ao planejar um estudo, em geral temos uma pergunta que não foi respondida com as evidências empíricas que existem. O primeiro passo então é escrever a pergunta.

Pergunta: existe diferença entre a altura de adolescentes de 15 anos de Ribeirão Preto e de Sertãozinho?

Montar como será o teste de hipótese é preparatório do estudo, deve estar estabelecido no protocolo de pesquisa, porque ele vai guiar como será realizado a teste estatístico. Não confunsa teste de hipótese (teórico) com o teste estatístico que vai realizar depois.

Já falamos que na prática depois de coletar os dados construiremos uma curva com as possíveis médias que saírem da amostra assumindo que irá representar aquelas médias que deveriam sair da população como um todo. Fazemos isso primeiro para uma das cidades e depois para a outra e assim verificamos se as curvas se interpõem. Se tomarmos como base Ribeirão Preto e estivermos comparando com Sertãozinho, vamos contrastar a curva de Ribeirão.

Para efetivamente comprarar (teste estatístico) devemos obedecer ao que estabelecemos como teste de hipótese no protocolo de pesquisa. Para isso, no

protocolo devemos estabelecer o que é a Hipotese que estamos testando (Ho), e a hipótese alternativa. Em geral as pessoas confundem os dois.

Dizemos que temos sempre duas hipóteses, sendo que a primeira representa a pergunta que o teste estatístico vai responder. Esta primeira hipótese é chamada de hipótese de nulidade (H zero = H0).

A hipótese de nulidade (H0) então seria:

H0 = a média de altura de Sertãozinho é igual à média de altura de Ribeirão Preto.

Note que a hipótese de nulidade não é uma pergunta é uma "afirmação" uma hipótese que deve ser testada. Assim após concluir o teste estatístico (para se testar a hipótese) teremos a resposta sobre "qual a probabilidade da média de altura de Sertãozinho ser igual a média de altura das crianças de Ribeirão Preto? ". O teste estatístico vai nos dar um valor desta probabilidade, o famoso "valor de p" que você encontra nos artigos, e todo mundo pergunta, mas não sabe o que significa. Esta probabilide das médias serem iguais pode ser grande ou pequena. Lembre-se que podem ser médias, proporções, razões, variâncias e etc, aqui simplificamos para média, porque estamos dando exemplo de média de altura.

Como eu vou executar este teste estatístico vai depender da **hipótese alternativa** (Ha = hipótese alternativa).

A **hipótese alternativa serve para guiar como vamos executar** o teste estatístico, isso é como iremos gerar o valor da probabilidade das médias serem iguals, que foi a hipótese a ser testada. Lembre-se a hipótese testada não é a alternativa, é a de nulidade.

Se o pesquisador não tem ideia ao escrever o protocolo de pesquisa, se a média de altura das crianças de Sertãozinho é maior ou menos do que a altura média em Ribeirão Preto, então a hipótese alternativa (Ha)

Ha = Media de Sertãozinho é diferente de Ribeirão Preto

Se a intuição antes de coletar os dados do estudo é de que a média de altura de Sertãozinho era maior do que a média de altura de Ribeirão preto a Ha seria:

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Ha = Média de altura de Sertãozinho é maior do que a média de altura de Ribeirão Preto.

Se a intuição é que a média de altura de Sertãozinho seria menor do que a média de altura de Ribeirão Preto, então temos:

Há = Média de altura de Sertãozinho é menor do que a média de altura de Ribeirão Preto.

Como procedemos ao teste de hipótese? Fazendo a estatística apropriada de acordo com a amostra que foi realizada. Lembre-se que se para levantar a média de altura em Ribeirão Preto e Sertãozinho utilizamos amostras probabilísticas complexas, temos que construir as curvas de possíveis médias de acordo com o plano amostral. Uma vez feito isso, vamos colocar as curvas uma do lado da outra. Se eu não tenho ideia se altura dos adolescentes de Sertãozinho é maior ou menor do que Ribeirão Preto, vamos proceder ao teste de hipótese da seguinte maneira. Construir a curva para Ribeirão Preto, e depois para Sertãozinho e colocar uma do lado da outra. Sertãozinho pode tanto ir do lado direito como esquerdo de Ribeirão Preto. Vamos supor que a média observada de Sertãozinho tenha sido maior que a de Ribeirão, então, colocamos a curva de Sertãozinho á direita de Ribeirão.

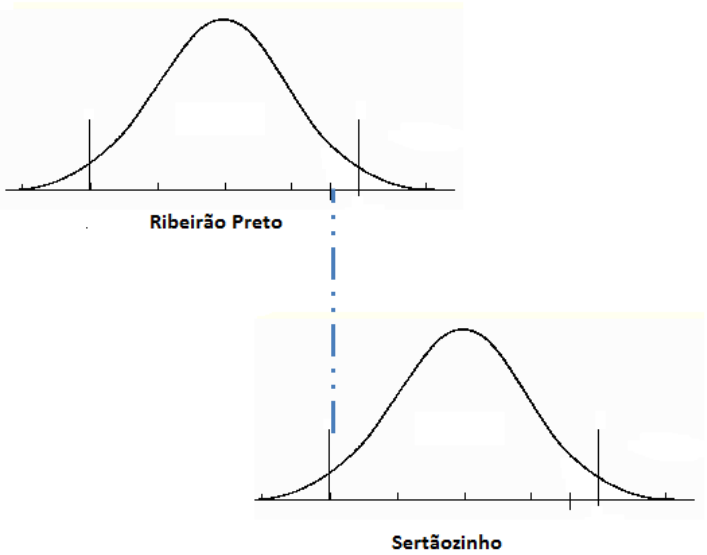


Fig.1 da página 136

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Nesta figura nota-se que o menor valor de Sertãozinho pertence a um valor provável de Ribeirão Preto, portanto não posso dizer que as possíveis médias são diferentes. Nesta outra figura abaixo o resultado é diferente.

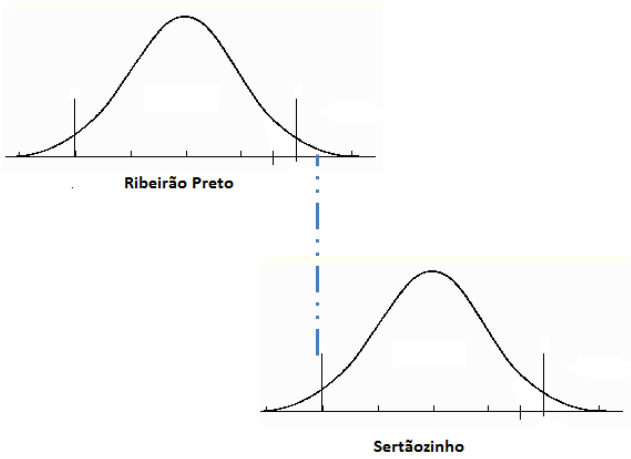
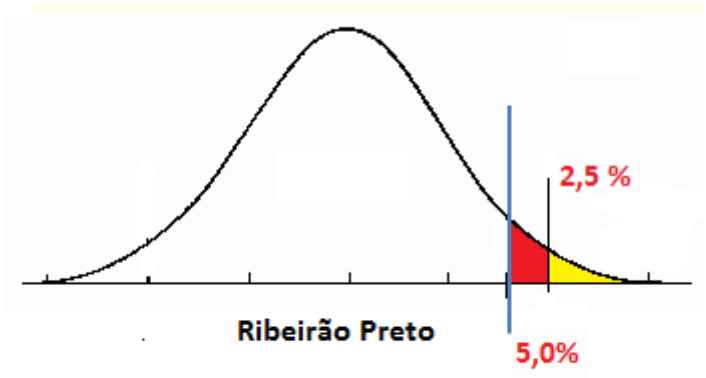


Fig.2 da pgina 136

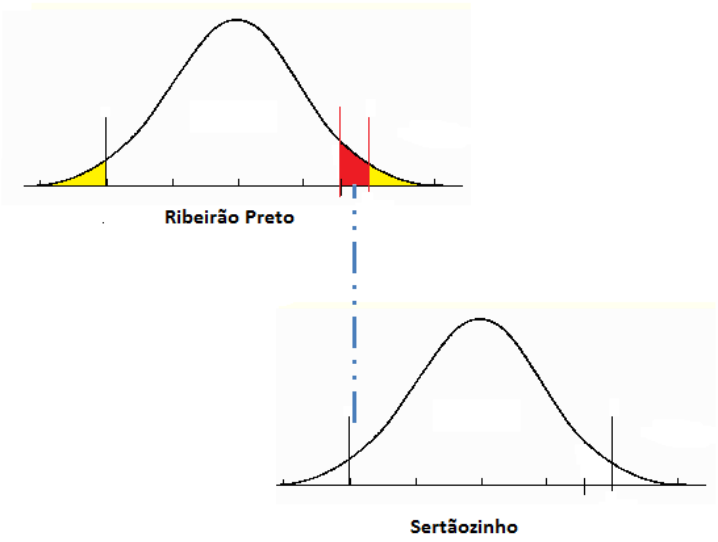
Nesta figura, observa-se que o menor valor de Sertãozinho é maior do que aqueles considerados prováveis para Ribeirão Preto, por isso, concluem que as médias são diferentes, sendo que a média de altura de Sertãozinho é maior do que a média de altura de Ribeirão Preto. O teste estatístico que acabei de fazer apenas me diz que as médias são diferentes e observando onde elas estão concluímos que a média de Sertãozinho é maior do que a média de Ribeirão Preto.

Note que nestes dois testes acima, o que era considerado raro (5%) foi dividido nos dois lados da curva, utilizando o clássico intervalo de confiança de 95% para fazer o teste. No entanto, quando temos a intuição de que Sertãozinho tem altura maior, durante o teste de hipótese, ao invés de dividir o erro de 5% dos dois lados posso jogar todo o erro na curva de um só lado. Mas note que precisamos ter realmente uma intuição grande de que a altura de Sertãozinho é maior do que Ribeirão Preto.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva



Assim, se fossemos fazer a mesma comparação com os 5% de um lado apenas, teríamos a figura a seguir. Note que a conclusão anterior era que não foi possível encontrar diferença estatisticamente significante. No entanto, agora que joguei os 5% de um lado da curva, a menor média de Sertãozinho não é mais uma das possíveis médias de Ribeirão Preto, e, portanto concluímos que as médias são diferentes. Para explicar de forma mais fácil e visualizar melhor estabelecemos que na curva de referência os 5% estará de um lado, mas na curva que será testada mantemos o intervalo de confiança com 2,5% de cada lado.



Logo, o que se observa é que se montamos um estudo desde o início com a hipótese alternativa de que a média de Sertãozinho é maior do que Ribeirão, nós temos que jogar toda a incerteza de Ribeirão Preto para a direita (apenas para fazer o teste), e

com isso será mais provável encontrar diferenças do que na situação anterior. O teste em que a princípio apenas assume que as médias são diferentes é chamado de **teste bicaudal** (as incertezas são colocadas nos dois lados da curva referência). Já o teste em que se tem a priori a ideia de que a média é maior ou menor do que a referência chamamos de teste **monocaudal**.

O efeito de usar teste bicaudal ao invés do teste monocaudal é como se déssemos um "empurrãozinho" para achar diferenças já que a impressão que se tem é que ela deve estar naquela direção. Assim o teste bicaudal é mais conservador do que o monocaudal, sendo assim, devemos ser cautelosos ao usar o teste monocaudal, ele deve ser usado sim, mas tem que se ter uma boa "certeza ou intuição" sobre a hipótese alternativa.

Embora o correto seja empregar o teste monocaudal sempre que se tiver a ideia de que a associação existe numa direção, a maioria dos programas estatísticos fazem testes bicaudais. Para realizar um teste nomocaudal, é necessário alterar as estatísticas automáticas dos pacotes estatísticos para que façam um teste monocaudal.

Agora que você já entendeu o que significa o teste de hipótese, vamos acrescentar mais alguns detalhes. Uma vez que sempre se tem uma referência para comparação (no exemplo dado Ribeirão Preto é a base da comparação), e nesta base temos a probabilidade das possíveis médias acontecerem, podemos responder a seguinte pergunta: qual a probabilidade das médias de Ribeirão Preto e Sertãozinho serem iguais?

A determinação do valor do p vai depender se o teste é monocaudal ou bicaudal. Apenas para questão didática vamos imaginar que o menor valor do Intervalo de Confiança da curva de Sertãozinho seja fixo, pois você elimina aqueles valores que poderiam ser raros para Sertãozinho. Você então vai procurar onde fica este valor na curva de médias de Ribeirão Preto. O ponto em que ele se situa se refere a uma probabilidade dentro da curva de Ribeirão Preto, certo? Isso porque dentro de toda a curva de Ribeirão Preto, você tem 100% das possíveis médias que você pode gerar com amostra que você coletou. Qual é essa probabilidade?

Se você tem uma hipótese alternativa estabelecida desde o protocolo de pesquisa que a altura de Sertãozinho seria bem maior do que a de Ribeirão, você vai

procurar este valor nos valores maiores de Ribeirão Preto do lado direito da curva. Sendo o teste, portanto monocaudal, os 5% que consideramos raros para Ribeirão vão ser colocados apenas entre os valores da direita (maiores valores). Os 100% da curva de Ribeirão passam a ser contados apartir do menor valor da curva de Ribeirão, esquecendo-se do intervalo de confiança. Com o valor menor sendo o mais provável e os maiores os menos prováveis.

Vamos supor que o menor valor de Sertaozinho se encontra no ponto 0.08 da curva monocaudal de Ribeirão Preto. Neste caso, uma vez que 0.05% era o alfa escolhido, podemos dizer que a probabilidade de se encontar um valor de Sertaozinho na curva de Ribeirão Preto é de 8% ou menos. Mas 8% é um valor grande, portanto, não foi possível encontrar diferença estatisticamente significante entre altura de Ribeirão Preto e Sertaozinho.

Outra forma de explicar, vamos partir da figura do teste monocaudal, nele verificamos que o menor valor de Sertãozinho se encontra entre os valores de 2,5 e 5%, vamos imaginar que seja 3,5%. Então a probabilidade da menor média de Sertãozinho pertencer a Ribeirão Preto, é de apenas 3,5% o que é menor do que 5%, portanto, é uma probabilidade tão pequena (em relação aos 5%), que dizemos que a probabilidade das médias serem iguais (ou seja, de uma das possíveis médias de Sertãozinho pertencer a Ribeirão Preto) é tão pequena que, portanto, elas são diferentes.

Num teste bicaudal é mais ou menos o mesmo raciocínio, no entanto, como a o alfa de 5% foi dividido e não sabemos se é do lado direito ou esquerdo, o valor encontrado no caso de 3,5% tem que ser multiplicado por dois resultando em 7%. Assim, consideramos que a probabilidade das médias serem iguais é de 7% que é uma probabilidade muito grande e, portanto, as médias não são diferentes. Em outras palavras, não foi possível encontrar diferença estatisticamente significante entre as médias de altura de Ribeirão Preto e Sertãozinho. Aquele valor é o valor exato da probabilidade que chamam nos livros de valor de “p” (p minúsculo).

O exemplo abaixo é um teste de comparação de duas médias realizado no SAS. O exemplo se refere à diferença de notas em uma prova do meio do semestre entre alunos que acertaram ou erraram uma questão no teste de admissão do curso. Note que

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

com o teste monocaudal, o valor da probabilidade das médias serem iguais é exatamente a metade do valor bicaudal.

Teste Monocaudal X Bicaudal

Teste Bicaudal

Q5	Method	Mean	95% CL Mean		Std Dev	95% CL Std De
0		4.1112	3.4195	4.8028	2.2475	1.8532 2.856
1		5.3472	4.5497	6.1446	2.1357	1.7009 2.87
Diff (1-2)	Pooled	-1.2360	-2.2807	-0.1913	2.2025	1.8923 2.63
Diff (1-2)	Satterthwaite	-1.2360	-2.2729	-0.1991		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	71	-2.36	0.0211
Satterthwaite	Unequal	64.528	-2.38	0.0202

Teste Monocaudal

Q5	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev
0		4.1112	3.4195	4.8028	2.2475	1.8532 2.856
1		5.3472	4.5497	6.1446	2.1357	1.7009 2.871
Diff (1-2)	Pooled	-1.2360	-Infty	-0.3628	2.2025	1.8923 2.635
Diff (1-2)	Satterthwaite	-1.2360	-Infty	-0.3697		

Method	Variances	DF	t Value	Pr < t
Pooled	Equal	71	-2.36	0.0105
Satterthwaite	Unequal	64.528	-2.38	0.0101

Note no exemplo acima, que o programa fornece os Intervalos de Confiança de 95%, mas ele faz o teste estatístico de maneira diferente. No bicaudal o valor do p foi de 0.0211, e para o teste bicaudal foi de 0.0105. Nos dois casos, continuou a ser significante independente do teste.

No próximo exemplo, quando a média final de três provas foram comparadas entre meninas que acertaram ou erraram a questão 5 no teste de admissão vemos que os resultados em termos de significância são diferentes. Com o teste bicaudal não foi possível encontrar diferenças estatisticamente significantes, enquanto que o teste monocaudal revela que as médias são diferentes. Note que para o teste bicaudal podemos comparar os intervalos de confiança de 95% para chegar a conclusão se as medias foram diferentes ou não. No entanto, no teste bicaudal a comparação dos intervalos e confiança de 95% não funciona. Fiquem atentos!

Teste Bicaudal

Q5	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev
0		5.3527	4.7032	6.0021	1.8013	1.4441 2.3948
1		6.2289	5.6354	6.8224	1.3724	1.0614 1.9424
Diff (1-2)	Pooled	-0.8763	-1.7738	0.0213	1.6370	1.3761 2.0209
Diff (1-2)	Satterthwaite	-0.8763	-1.7350	-0.0175		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	53	-1.96	0.0555
Satterthwaite	Unequal	52.782	-2.05	0.0457

Teste monocaudal

Q5	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev
0		5.3527	4.7032	6.0021	1.8013	1.4441 2.3948
1		6.2289	5.6354	6.8224	1.3724	1.0614 1.9424
Diff (1-2)	Pooled	-0.8763	-Infty	-0.1271	1.6370	1.3761 2.0209
Diff (1-2)	Satterthwaite	-0.8763	-Infty	-0.1595		

Method	Variances	DF	t Value	Pr < t
Pooled	Equal	53	-1.96	0.0277
Satterthwaite	Unequal	52.782	-2.05	0.0228

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

A seguir uma cópia de uma parte de uma tabela do estudo sobre ingestão de vitamina D por gestante e posterior desenvolvimento de cárie em crianças (Tanaka et al, Annals of Epidemiol,2015).

Table 1
Distribution of characteristics in 1210 mother-child pairs, KOMCHS, Japan

Variable	Overall	Caries status		*P* value
	Number (%) or mean ± SD	Caries (*n* = 267)	Caries free (*n* = 943)	
		Number (%) or mean ± SD	Number (%) or mean ± SD	
Baseline characteristics				
Maternal age, y, mean ± SD	31.6 ± 4.0	31.9 ± 4.0	31.5 ± 4.0	.14
Region of residence				.0004
Fukuoka Prefecture	693 (57.3)	127 (47.6)	566 (60.0)	
Other than Fukuoka Prefecture in Kyushu	403 (33.3)	110 (41.2)	293 (31.1)	
Okinawa Prefecture	114 (9.4)	30 (11.2)	84 (8.9)	
Household income, yen/y				.03
<4,000,000	387 (32.0)	96 (36.0)	291 (30.9)	
4,000,000–5,999,999	457 (37.8)	104 (39.0)	353 (37.4)	
≥6,000,000	366 (30.3)	67 (25.1)	299 (31.7)	
Maternal educational level, y				<.0001
<13	249 (20.6)	82 (30.7)	167 (17.7)	
13–14	406 (33.6)	87 (32.6)	319 (33.8)	
≥15	555 (45.9)	98 (36.7)	457 (48.5)	
Paternal educational level, y				<.0001
<13	350 (28.9)	104 (39.0)	246 (26.1)	
13–14	177 (14.6)	40 (15.0)	137 (14.5)	

Esta tabela tem a descrição, por exemplo, da média de idade materna geral na primeira coluna (chamada overall) e a média de idade para as mães das crianças que tiveram cárie e para as que não tiveram cárie. Note que na última coluna está escrito "*p value*" isto é o **valor de p**. A média de idade das mães de crianças com cárie foi de 31,9 e das mães de crianças livres de cárie 31,5. O valor de p para esta comparação foi de 0,14, que é um valor alto comparado aos 0.05 logo vamos interpretar que a ***probabilidade da média de idade entre as mães de crianças com cárie ser igual a média de idade das mães de crianças sem cárie é de 14%, logo, uma probabilidade muito alta, e portanto, não podemos dizer que as médias são diferentes***. AO lado das médias de idade existe uma informação chamada "sd" que é de "*standard deviation", ou seja,* desvio padrão, mas que vamos abordar posteriormente. Por enquanto, o desvio padrão não nos interessa.

Você deve estar sentindo falta do intervalo de confiança destas médias. Se a tabela tivesse os intervalos, você poderia compara-los e verificar se iriam sobrepor. Na verdade devemos sempre reportar o intervalo de confiança, porque além de servir para comparar, o intervalo de confiança nos dá uma ideia de precisão da amostra. Se o

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

intervalo for muito grande significa que talvez o número de participantes na amostra tenha sido pequeno.

Note esta descrição do trabalho pioneiro de Kormam KS de 1999, sobre associação de um polimorfismo de interleucina e periodontite.

To control the effect of age on disease severity, data were analyzed separately for non-smokers aged 40–60 years. In this age range, the composite genotype was present in 78% of severes (n=9), 26% of moderates (n=30), and 16% of milds (n=32) (odds ratio: severe versus mild=18.90 (1.04–343.05), P<0.002). The influence of genotype on disease severity is evident in the cumulative frequency distribution (Fig. 2) that shows	Para controlar o efeito de idade na .....[ ] Odds ratio : severo vs moderado = 18,90 com Intervalo de Confiança de 1,04 a 343,05. Valor de p < 0.002.

Veja que Korman relata uma Odds Ratio de 18.90 que é um valor considerável. Ela significa que a razão de expostos ao polimorfismo e não expostos ao polimorfismo para quem tem periodontite severa é 18.90 vezes a razão de exposição entre os controles que aqui são os indivíduos com periodontite moderada. Se você olhar somente o valor de p, notará que é muito pequeno. Isto significa que a probabilidade da razão de exposição entre casos ser igual a razão de exposição dos controles é menor que 0.002 (isto é 0,2%), que é uma probabilidade muito pequena de serem iguais e, portanto, são diferentes. O pesquisador olhando somente para o valor de p deve ficar muito contente, mas ao olharmos para a imprecisão deste valor representado pelo Intervalo de Confiança que vai de 1,04 a 342,05, vemos que é um intervalo muito largo. A conclusão é de uma imprecisão gigantesca e que pode na verdade revelar algo de errado, talvez um número pequeno de participantes e uma heterogeneidade muito grande. Simplesmente tem algo estranho que nos leva a acreditar na conclusão de associação entre o polimorfismo e a periodontite severa. Bom, a associação apontada pelo Korman nunca mais foi reproduzida em nenhum estudo.

Esta é outra tabela de um estudo sobre associação entre haplotipo de anamelina e cárie dentária.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Table 3. Crude and Adjusted Odds Ratios (OR) Associated with the Caries Phenotype Using Haplotype Analysis

Haplotype	SNP Order*	OR_{crude}		*p* Value	$OR_{adjusted}$**		*p* Value
G A T T C G	------	Reference haplotype		–		–	–
G A T **C** C **A**	--- 17 - 21	2.04	[0.98 - 4.26]	.056	2.66	[0.99 - 7.20]	0.05
G A T **C** T G	--- 17,20 -	2.58	[0.97 - 6.86]	.057	0.80	[0.28 - 2.27]	0.67
G A **C** T C G	-- 19 ---	4.22	[1.57 - 1.33]	.004	1.17	[0.33 - 4.12]	0.80
A A T **C** C **A**	5 - - 17 - 21	NA		NA	NA		NA
A G T T C G	5,12 ----	NA		NA	NA		NA
A G T **C** C **A**	5,12 - 17 - 21	1.03	[0.49 - 2.14]	.94	1.24	[0.49 - 3.13]	0.65

*SNP5, rs144929717; SNP12, rs139228330; SNP17, rs2609428; SNP19, rs7671281; SNP20, rs36064169; SNP21, rs3796704.
**Adjusted for consumption of soft drinks, toothbrushing frequency, and parental occupational status.

Independente do que realmente esteja representando, vemos duas colunas com Odds Ratio. uma chamada “crude” que significa bruta , isto é sem ajuste de fatores de confusão, e outra chamada OR ajustada. Note que quando o valor do Intervalo de confiança que está entre braquetes inclui o valor 1, os valores de p são em geral maiores ou iguais a 0.05. Assim, concluímos que por meio do intervalo de confiança conseguimos ler se existe ou não associação, e ainda o intervalo nos indica se a amostra tem muita variabilidade ou não.

Para evitar equivocos comuns em relação a comparação de intervalos de confiança aqui está um resumo:

1- Ao comparar dois intervalos de confiança apenas queremos saber se ele se interpõe ou não. Neste tipo de comparação incluir zero ou 1 não importa.

2- Se formos analisar o intervalo **para diferenças de médias** teremos apenas um intervalo de confiança para analisar, e ai verificará se eles contem o valor zero ou não. O zero importa porque uma diferença nula é igual a zero.

3- Se formos analisar o intervalo para razão de alguma coisa (Razão de Prevalência, Odds ratio, razão de risco e etc) ai verificaremos se o intervalo contem ou não o valor 1. Porque a razão nula de duas coisas inclui o 1. O zero aqui nem vai existir.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Possíveis erros ao se testar uma hipótese

Agora que você consegue entender o que significa o valor do p, e como se faz um teste de hipóteses, vamos avançar em alguns conhecimentos. Ao fazer o teste de hipóteses consideramos o valor de 0.05 como ponto de corte para decidir se algo será diferente estatisticamente. Lembre-se que este ponto de corte de 0.05 foi inventado, ou melhor, sugerido para ser utilizado. Se o assumimos, pode ser que tenhamos eliminado nos 5% alguma possível média que seria a verdadeira, e estamos descartando o verdadeiro. Porém não sabemos da verdade e corremos este risco. Chamamos o ponto de corte de 0,05 (ou qualquer outro valor que escolhermos para um trabalho) de **alfa** e não valor de p. Valor de p é a probabilidade exata do que você está comparando ser igual. O ponto de corte que se decidiu usar é chamado simplesmente de alfa. Num projeto de pesquisa devemos escrever na parte de planejamento da estatística qual o alfa que vamos utilizar (e não o valor de p). O valor de p é calculado ao se realizar a estatística, ou seja, o teste de hipótese depois dos dados coletados.

Vamos falar de novo de possibilidade de erros, e não confundir aqui com o erro sistemático e aleatório. Agora estamos começando a falar de possibilidade de erro ao fazer o teste estatístico. É possível que façamos um teste, que encontremos diferenças, logo os intervalos de confiança não vão se sobrepor, porém, não exista diferenças. Um exemplo é o estudo do Korman que até hoje não foi replicado. Kormam encontrou diferenças, se bem que suspeito pelo grande intervalo de confiança que algo não estava certo no estudo dele. Acredito que a montagem do estudo, o desenho, e algum viés de seleção tenha levado a encontrar associação quando esta na verdade não existe. Chamamos de erro tipo I, quando encontramos associação e esta associação na verdade não existe.

Em geral, este tipo de erro, é resultado de algum viés do estudo, mas também pode ter simplesmente acontecido ao acaso porque resolvemos remover os 5% das prováveis médias de valores mais extremos. Independente de vieses, contando apenas com a possibilidade de erro aleatório, corremos o risco de ter este tipo de erro de 5% (ou o valor de alfa que estabelecermos).

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Além desta possibilidade, o que mais resulta no erro de encontrar associação quando ela não existe são os vieses de forma geral. Acredito que no estudo do Korman ou foi a estatística não bem realizada ou a montagem do estudo que acabou ou os dois, que acabou resultando em diferenças entre os grupos quando na verdade não existia. Ainda é comum encontrar associação quando na verdade não existe quando não se ajusta por fatores de confusão. Note que na tabela de enamelina para o SPN 19, o intervalo de confiança não incluía o valor 1 e o valor de p correspondente era significante, no entanto depois de ajustado a significância sumiu. Ainda fatores de confusão não conhecidos podem também levar a este erro. Como nunca sabemos a verdade para saber se erramos ou não como podemos chegar a conclusão de erro tipo 1? No caso do trabalho do Kormam, não tínhamos certeza na época, a critica feita ao trabalho era que não havia sido bem estruturado, faltava controle de outros fatores de confusão e ainda com o intervalo muito impreciso ressaltava muita heterogeneidade no estudo. Com o tempo, outros estudos realizados não conseguiram replicar os achados do Korman, acredita-se então que aquele foi um típico **erro tipo I** , não devido apenas ao fator aleatório mas a vieses.

Outro tipo de erro seria não encontrar associação quando na verdade a associação existe. Este erro é chamado de **erro tipo II**. Como dissemos anteriormente, o teste de hipótese depende do tamanho do intervalo de confiança, e este tamanho depende do número de indivíduos na amostra. Se a amostra é pequena, o intervalo será grande e maior será a possibilidade de que intervalos que estão sendo comparados se interponham. Assim, o motivo mais comum de erro tipo II é o número pequeno de indivíduos num estudo. É comum sempre que não se consegue encontrar associação estatisticamente significante num estudo, que se comente na discussão do mesmo que uma das razões tenha sido o pequeno número de indivíduos. Isso virou quase um mantra em estudos quando não se encontra associação. Ainda, também em estudos com poucos indivíduos, porque o intervalo de confiança tende a ser grande, os autores sempre escrevem que apensar do pequeno número uma vez que diferenças foram observadas estas realmente existem. Porém nem sempre é assim, quando temos amostras muito pequenas, pode ser que aquele pequeno número de indivíduos em cada grupo sejam completamente diferentes ou por causa de viés de seleção ou mesmo ao

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

acaso. Portanto, ter amostra pequena e encontrar diferenças pode representar não é 100% de garantia de que a diferenças exista. Temos que ficar atentos.

Resumindo ao realizar um teste de hipótese podemos ter dois tipos de erros: o erro tipo I e o erro tipo II. Não confundir com erros do desenvolvimento do estudo que incluem erros sistemáticos e erro aleatório que ameaçam a acurácia do estudo (lembre-se acurácia é ausência de erro sistemático e aleatório).

Relembrando para se montar um estudo primeiro temos que estabelecer uma pergunta bem fundamentada de razões para que seja necessário realizar o estudo. Segundo, nós temos que escrever claramente quais serão as hipóteses nula e hipóteses alternativas. Terceiro temos que desenhar o diagrama causal onde vai constar exposição desfecho e todos os fatores de confusão e modificadores necessários para se testar a hipótese de associação entre exposição e desfecho independente dos fatores de confusão e modificadores.

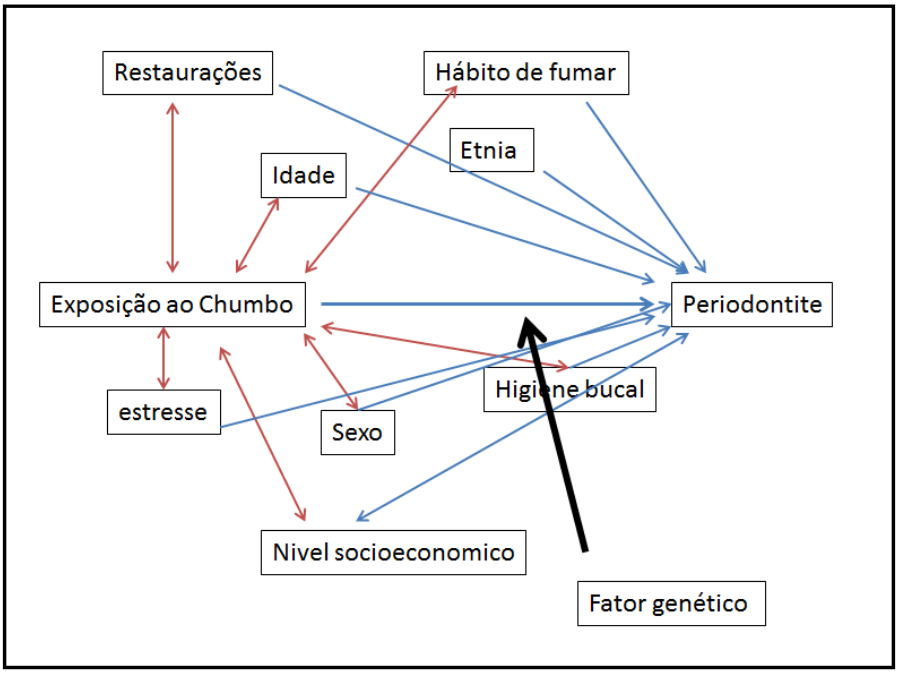
Pergunta: existe associação entre exposição ao chumbo e periodontite no adulto?

H0 : a média de perda de inserção entre indivíduos expostos ao chumbo é igual a média de perda de inserção em indivíduos não expostos ao chumbo.

Ha: a média de perda de inserção entre indivíduos expostos ao chumbo é maior do que a média de perda de inserção entre não expostos

Diagrama causal

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva



Note que neste diagrama causal apenas os fatores de confusão foram adicionados. As setas entre fatores de confusão e exposição devem ser em dupla direção (setas vermelhas), já as setas dos fatores de confusão para o desfecho são unidirecionais. Note que existe uma seta preta mais grossa que incide diretamente entre a exposição e o desfecho, ela representa um fator modificador que está aqui hipoteticamente descrito como "fator genéticos". Os fatores genéticos em geral modificam a associação entre exposição e desfecho.

O diagrama causal acima é simplificado. Na verdade as relações entre as variáveis pode ser muito mais complicadas. Neste diagrama causal simplificado é apenas para servir de lembrete no planejamento de um estudo. Falta neste diagrama simples, a conexão entre os fatores de confusão entre si, por exemplo, higiene bucal também está associada ao sexo e ao nível socioeconômico entre outros.

Os diagramas causais são fundamentais para não se esquecer dos fatores de confusão que devem ser coletados. Posteriormente, os diagramas são úteis e fundamentais durante a análise estatística para lembrar o pesquisador das relações entre as variáveis.

As setinhas do diagrama causal indicam a direção da associação, por exemplo, se a doença altera a exposição, a relação causal será muito complicada, e montar um

estudo apropriado também será complicado. Por exemplo, se estou estudando fumo e câncer de pulmão, o fumo tem que vir antes do câncer para ser causal. Se a doença puder afetar o uso do fumo, a seta tem que ser bidirecional. Em geral registramos o fumo passado e o câncer. Mas imagine que alguém está realizando um estudo transversal e pergunta apenas se o indivíduo no momento está fumando, e ele responde que não. Mas esta não exposição se deve ao fato de não estar mais fumando devido a doença. Vamos inventar um exemplo, inventado mesmo, isso não é verdade, é apenas para dar um possível exemplo. Alguém pode alegar que a susceptibilidade de câncer (antes de ter câncer) está associada à depressão nata do indivíduo, e, portanto, quem tem o gene susceptível para o câncer é deprimido por natureza e por isso fuma. Neste caso a associação seria meio que inversa, mas sendo o cigarro também nocivo aos tecidos do pulmão os dois eventos se somariam e poderiam resultar em interação. Para estudar um fenômeno assim, é bem mais complicado, embora existam formas e estatísticas apropriadas denominadas de complexidade. Aqui vamos simplificar e lidar apenas com situações que achamos que não exista esta relação inversa. Mas se lembre de que ela é possível em algumas situações, e se for assim, tem que ser estudada com outra abordagem de sistemas complexos. Este texto é apenas introdutório e motivacional, existem muitos detalhes para serem aprofundados o que não é nosso objetivo.

Uma vez estabelecidos os objetivos do estudo e montado o diagrama causal, passamos a outra etapa que seria estabelecer como exposição, desfecho e fatores de confusão vão ser medidos e posteriormente devermos decidir como analisar estas informações.

Já fizemos vários comentários sobre como montar estudos, viéses, fatores de confusão, e já temos noção básica do que seria o intervalo de confiança. Vamos supor que alguém leve a um estatístico um banco de dados para ser analisado posteriormente a coleta de dados. O estatístico ou quem entende de estatística deve ser parte do time que irá desenvolver o estudo, e o investigador principal que vai montar o estudo tem que ter noções básicas de estatística para montar adequadamente o mesmo. Impossível como irá notar a possibilidade de um pesquisador não entender nada de estatística,

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

porque para ler artigos e saber se são validos os argumentos e achados o pesquisador precisa saber estatística.

Se houver necessidade de se buscar um estatístico posteriormente, no mínimo o investigador deve levar ao estatístico um relato contendo a pergunta do estudo, o H0 e a Há escritas e um diagrama causal detalhado. Ainda, precisa detalhar para o estatístico o tipo exato de estudo que realizou e de que forma foi realizado. Detalhadamente precisa saber como foi realizada a amostra, se simples, ou estratificada, como foi realizado a alocação aleatória caso tenha sido um experimento.

Até o momento você deve ter aprendido:

1- **Fazer o teste de hipótese**
   a- **Definer as hipoteses nulas e alternativas**
   b- **Utilizar adequadamente os testes mono e bicaudais e reconhecer os efeitos de cada um deles**
   c- **Interpreter e diferenciar os valores de alfa e p**
2- **Definir os erros tipo I e tipo II e compreender as causas destes erros**
3- **Comparar médias e proporções utilizando o interval de confiança.**
4- **Montar um diagram causal**

Antes de continuar com tópicos clássicos sobre unidade de estudo e distribuição devemos adicionar mais um conhecimento pequeno e intuitivo de grande valia. Se pegarmos uma curva de médias (possíveis médias) geradas por simulação e cortarmos em 3 partes a partir do meio teremos a curva de médias com 6 pedaços iguais na base.

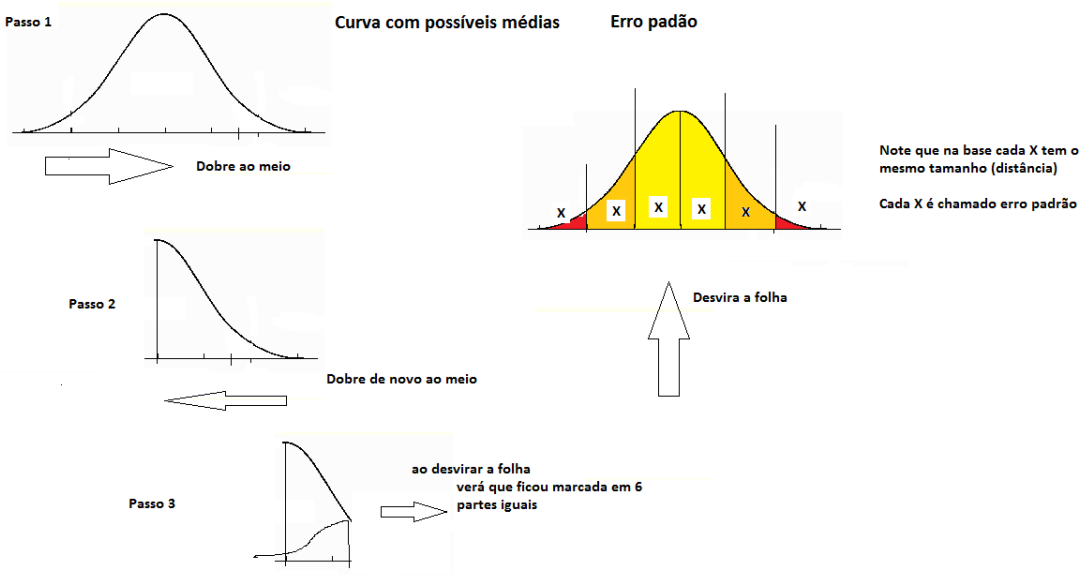
Desenhe uma curva tipo sino numa folha inteira de papel, se possível uma curva bem simétrica. Dobre a folha ao meio, e depois ao meio de novo. Você terá uma curva dividida em 6 partes iguais na base (veja quadro a seguir). Cada parte, que tem distâncias iguais, é chamada de **erro padrão**. Na verdade, é uma aproximação do erro padrão como veremos. O espaço marcado de amarelo na figura que compreende um erro padrão para cada lado da média representa aproximadamente 68% das possíveis médias. O espaço marcado de amarelo mais espaço laranja, que representam dois erros padrão de cada lado da média corresponde aproximadamente a 95¨% das médias. Logo, aquilo que chamamos de intervalo de confiança corresponde a dois erros padrão de cada lado. Os três erros de cada lado incluiriam 100% das possíveis médias. Porque erro? Porque os que temos na curva de médias são possíveis médias que representariam erro aleatório.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Na verdade, não são dois erros inteiros para cada lado para montar o intervalo, mas é aproximadamente, na verdade são 1,96 erros de cada lado. Mas grosseiramente assumimos que são dois erros. Nos livros você vai encontrar 1,96. Isso é, é um erro inteiro mais 96% de outro erro.

O que é importante do erro aleatório amostral é que se em uma tabela tivermos a média de 150 cm (referente à altura) e se sabemos que um erro padrão tem distância de 3cm, podemos calcular o intervalo de confiança, considerando dois erros para cada lado. Logo o intervalo de confiança de 95% será de 144 a 156. Óbvio se quiser um intervalo de 90% então seria um erro mais uma parte de outro erro.

A figura abaixo descreve como chegamos ao erro padrão. O erro padrão é importante para calcular o intervalo de confiança, para fazer inferência estatística.



Unidade de Estudo

Além dessas informações para decidir qual estatística utilizar, o pesquisador vai precisar saber alguns detalhes como qual é a unidade do estudo.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

A unidade do estudo pode ser um indivíduo, uma família, uma cidade, um país. Num caso-controle de câncer a unidade do estudo deve ser um indivíduo. Um estudo recente sobre associação de concentração de flúor na água e prevalência de hipotireoidismo na Inglaterra é um bom exemplo para se explicar a unidade de um estudo. No caso, pelo titulo pensaríamos que a cidade seria a unidade de estudo. No entanto, como os dados de prevalência foram coletados de cada Unidade de Saúde Básica, os pesquisadores resolveram verificar a concentração de flúor em cada área correspondente a unidade básica de saúde. Assim, uma cidade poderia ter dezenas de unidades consideradas no estudo. Naquele estudo, portanto, a unidade de pesquisa foi cada unidade básica de saúde e os dados são do tipo ecológico.

Na odontologia os estudos de cárie e periodontite são bastante interessantes como exemplos de unidade de estudo. Em geral fatores são associados à periodontite como genes, fumo, chumbo entre outros. A doença periodontite é de difícil definição, o que se sabe é que leva a perda de inserção periodontal (perda de osso). Assim, a doença é diagnosticada medindo vários sítios, em geral 6, por dente. Assim, somam-se todos os valores e divide-se pelo total de superfícies observadas, resultando na perda média de inserção por sítio para aquele indivíduo. Se a perda média for maior do que certa medida, em geral 2,5mm, considera-se que o indivíduo tem periodontite. Existem outras definições baseadas na perda de inserção como, por exemplo, considerando-se doente aquele indivíduo que tem mais do que 3 dentes com sítios com mais de 3mm entre outros. De qualquer forma, todas estas definições se baseiam em medidas ecológicas onde se reúne informações sobre perda de inserção dos 28 dentes possíveis (em geral eliminam-se os terceiros molares dos cálculos). Se a pergunta da pesquisa é se um fator de exposição leva a maior perda de inserção em um indivíduo, então a unidade é o indivíduo. No entanto, pode ser que queiramos responder se a presença de restaurações nas proximais leva a maior perda de inserção nas próximas. Para tanto, realiza-se um estudo transversal e examinam-se todos os sítios de um indivíduo. Se um indivíduo contribuir com mais do que uma superfície com restauração interproximal quem será a unidade do meu estudo? A unidade será a superfície dentária, porém o indivíduo serve de conglomerado.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Uma vez definido a unidade do estudo é necessário classificar tanto desfecho como exposição com relação ao tipo de variável que representam. Podemos ter variáveis quantitativas e variáveis qualitativas. As variáveis quantitativas quantificam coisas, e as qualitativas estabelecem qualidades a coisas.

Exemplos de variáveis quantitativas são peso, altura, número de dentes cariados perdidos e obturados (Indice CPOD – o D é de dente), número de dentes com mais de 3 mm de perda de inserção, nível sérico de chumbo, concentração de glicose no sangue entre outros. Dentre estas variáveis, peso e altura, por exemplo, podem assumir inúmeras possibilidades fracionadas enquanto número de dentes com cárie não é fracionado, é apenas um número. Quando podemos fracionar a variável quantitativa temos o que chamamos de **variável quantitativa contínua**. Já a variável número de CPOD é um exemplo de **variável quantitativa discreta** assim como o número de dentes com mais de 3mm de perda de inserção.

Exemplos de variáveis qualitativas temos câncer (sim ou não), presença de cárie (sim ou não) , cor de pele (branca, negra, parda), etnia (europeia, indígena, japonesa). Estas variáveis atribuem qualidade as pessoas e não existe ordem entre elas mesmo que se atribuam números para identificar se o branco é número 1, negro seria 2 e pardo igual a número 3. Mesmo atribuindo número não existe uma ordem e recebe o nome de **variável qualitativa categórica (nominal)**. No entanto, para algumas variáveis qualitativas como severidade de doença periodontal classificada como severa, moderada e leve a ordem é importante. Se atribuirmos valores como leve =1, moderada = 2 e severa = 3 , sabemos que 1 é menor do que 2 que é menor do que 3. Quando a ordem for importante chamamos de **variável qualitativa ordinal.**

Distinguir o tipo de variável é importante para se decidir como se pode analisá-las. Com uma variável quantitativa continua podemos calcular medidas como média (se a distribuição for normal) e mediana. Já com uma variável qualitativa em duas categorias como doente/não doente podemos apenas fazer proporções.

Note que por vezes podemos categorizar uma variável quantitativa em categórica. Se a variável original é altura, podemos categoriza-la se necessário em altura baixa, altura média e altura alta estabelecendo pontos de corte.

Tipos de Variáveis:

- Quantitativa :
  - continua
  - discreta
- Qualitativa:
  - Categórica (nominal)

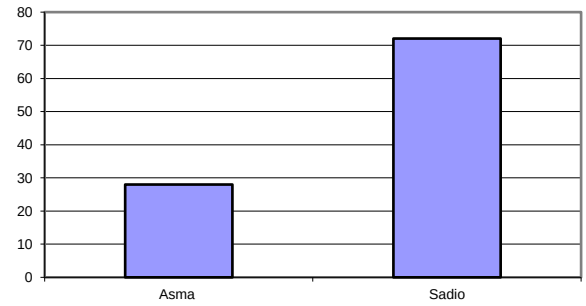
mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Existem duas terminologias que por vezes são utilizadas em relação às variáveis quantitativas que são classificadas em escalas intervalares ou razão. Essa terminologia foi proposta por um psicólogo Steven (1946) como sendo níveis de mensurações, e é muito utilizada especialmente em livros de estatística ligados a área de psicologia e psicometria. Nem sempre esta terminologia é utilizada e tem mais finalidade de se interpretar ou dar significado ao que se mede e como se pode comparar. Em relação às variáveis qualitativas os nomes são os mesmos, mas Stenvens ressalta que uma classificação nominal pode-se concluir que uma categoria é igual ou diferente de outra, já numa classificação ordinal pode-se assumir que uma categoria é menor ou maior do que a outra. Quanto a intervalar e razão temos que esplicar menor. Um bom exemplo é fornecido por Kerb Shedden (The Introduction to Data Science e-book is copyright 2020-2021) que exemplifica, por exemplo, a temperature em Fahrenheit que é uma escala onde não se tem o zero absoluto, e a temperatura considerada de congelamento se refere a 32 graus. Nesta escala a temperatura de 50 não significa que seja 2 vezes mais quente do que 25 graus o que seria uma razão, razão aqui não faz sentido e portanto é uma medida intervalar. Ao comparar as duas temperaturas nós dizemos apenas que existe uma diferença de uma para outra de 25graus. Por outro lado, se a temperatura é medida em graus Celcius cujo ponto de congelamento seria zero grau podemos dizer que uma temperatura de 50 é duas vezes mais quente do que 25 graus. Nesta situação razão faz sentido e seria considerada uma mensuração em escala de razão. Shedden argumenta que na prática essa tipologia não é muito útil e por vezes dúbia, pois mesmo no caso de Fahreinheit, poderia-se dizer que as distancias do zero absoluto poderiam ser consideradas como razão. O interessante desta classificação é poder raciocinar sobre o significado de uma medida. Esta tipologia não elimina a classificação em continua e discreta. Portanto, é bom saber, é bom refletir sobre o assunto, mas não estressem muito com esta classificação. Claro que psicometristas pensam mais sobre o assunto, pois frequentemente desenvolvem instrumentos para mensuração.

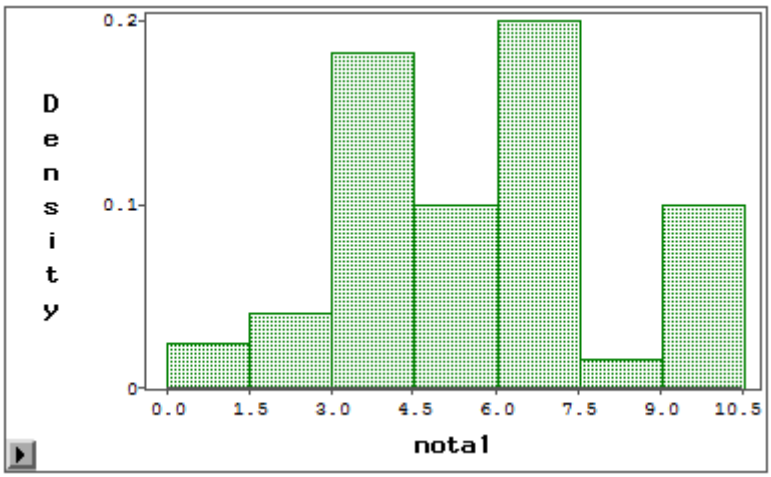
mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

❑ Visualização gráfica das variáveis

Se a variável é categórica sabemos que temos que analisá-la utilizando porcentagens. Para representar graficamente uma variável categórica usamos gráficos chamados de barra ou colunas como, por exemplo, no gráfico abaixo representando indivíduos asmáticos e saídos num estudo.



Se uma variável é quantitativa como, por exemplo, as notas de provas de uma determinada turma, a melhor forma de visualizar e representá-la graficamente é empilhando os valores das notas o que resulta num gráfico chamado **histograma.**



mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Existem regras e especificações de como devemos agrupar os valores de uma variável quantitativa contínua para montar a mão o histograma. No entanto, hoje em dia com os programas estatísticos é muito fácil pedir para o programa realizar o gráfico e não vamos entrar em detalhes de como fazê-lo a mão. O que basicamente foi feito pelo computador foi agrupar os valores (nestas colunas que você observa acima) de acordo com seu valor. Neste histograma notamos que tivemos notas desde zero até o valor de dez, e existem dois picos de acumulo de pessoas com as mesmas notas entre 3 e 4, e posteriormente de 6 a 7,5, e outro tanto conseguiu tirar 9 ou 10. Embora existam pessoas com nota dez, a tendência geral é de valores menores. De fato se pedirmos para o computador calcular a média ele nos dará os seguintes quadros:

Moments

N	80.0000	Sum Wgts	80.0000
Mean	5.5331	Sum	442.6500
Std Dev	2.5701	Variance	6.6052
Skewness	0.1189	Kurtosis	-0.5124
USS	2971.0475	CSS	521.8097
CV	46.4485	Std Mean	0.2873

Quantiles

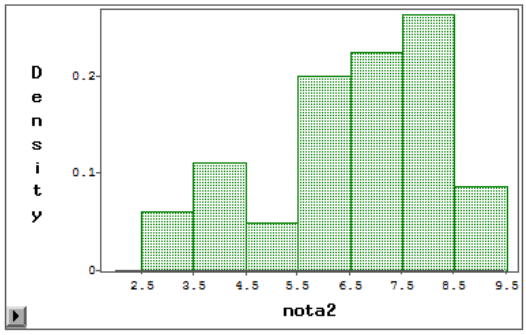
100% Max	10.0000	99.0%	10.0000
75% Q3	7.0000	97.5%	10.0000
50% Med	5.0000	95.0%	10.0000
25% Q1	4.0000	90.0%	10.0000
0% Min	0	10.0%	2.7500
Range	10.0000	5.0%	2.0000
Q3-Q1	3.0000	2.5%	0
Mode	7.0000	1.0%	0

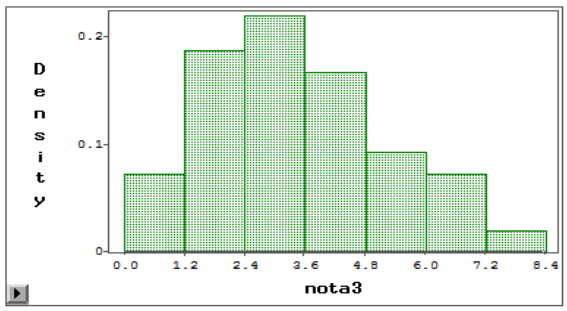
No primeiro quadro está escrito “momentos” (em inglês “*moments*”), isso significa que são algumas medidas únicas como média, mediana, curtose (kurtosis), simetria (skewness), entre outras que definem uma distribuição normal. O N significa o número de indivíduos que tem na amostra, no caso 80. “*Mean*” é a média que foi de 5,53, isto é um pouquinho acima do valor médio (de 0 a 10).

Vamos agora a um segundo histograma referente ao exemplo da segunda e a terceira prova que nos ajudará a explicar os medidas descritas no quadro de “moments”.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva





Note que a segunda prova (nota 2) os valores estão deslocados para a direita, isto é para os valores maiores, enquanto a terceira nota os valores estão deslocados para a esquerda, para os valores menores. Nenhuma destas curvas se assemelha muito a uma curva normal que seria simétrica (caudas iguais para os dois lados). A cauda da notas2 está para esquerda, e a cauda das notas3 está para a direita. Os alunos foram melhores na segunda prova, e suas notas caíram na terceira prova. No histograma notamos que a cauda está deslocada para a esquerda, em inglês é chamado de *skewed to the left* que seria assimetria à esquerda ou negativa. No histograma da terceira prova o deslocamento da cauda é para a direita dizemos então que tem assimetria positiva (*skewed to the right)*. Se calcularmos a média das provas 2 e 3, os valores pequenos vão puxar a média para baixo na prova 2, e os valores altos vão puxar a média para o alto. No quadro de resultados abaixo, você encontra uma palavra com uma medida chamada "*skewness" (= a simetria)* , pois é, é uma medida que revela se a assimetria estária mais a direita (positivo) ou a esquerda (negativo). Note que o valor da Prova 2 é negativo e da prova 3 positivo.

Moments

N	80.0000	Sum Wgts	80.0000
Mean	6.5450	Sum	523.6000
Std Dev	1.7619	Variance	3.1043
Skewness	-0.5284	Kurtosis	-0.5519
USS	3672.2000	CSS	245.2380
CV	26.9197	Std Mean	0.1970

Prova 2

Moments

N	80.0000	Sum Wgts	80.0000
Mean	3.5154	Sum	281.2333
Std Dev	1.7799	Variance	3.1681
Skewness	0.3426	Kurtosis	-0.4453
USS	1238.9344	CSS	250.2821
CV	50.6319	Std Mean	0.1990

Prova 3

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

A prova2 teve média de 6,54 e a prova 3 teve média de 3,51. Como as distribuições não são perfeitamente normais, as médias podem não refletir exatamente onde está a maioria dos indivíduos. Quando a distribuição não é normal, a média não dará a noção de "maioria/no meio". Média e mediana são **medidas de tendência central** (onde está a maioria), que será explorado com mais detalhes na próxima sessão. O que passa a representar melhor a característica da maioria é o que chamamos mediana, que representa o valor em ordem crescente que deixa para traz exatamente 50% da população. Este valor para a prova dois foi de 6,8, isto é um pouco maior do que a média. No caso da prova 3, a mediana foi de 3,4, isto um pouco menor do que a média. Nestes exemplos as diferenças entre médias e medianas não foram muito grandes, mas refletem as caudas que puxam os valores para onde estão mais acentuadas.

Pelas comparações acima, fica claro que quando temos uma variável continua precisamos avaliar o histograma identificando qual o formato da distribuição da mesma, para saber se a melhor medida para a representar de forma resumida seria média ou mediana. Se a distribuição é normal ou aproximadamente normal, a média é uma medida excelente, mas se a distribuição não for normal, a mediana deve ser a medida de escolha.

Ainda para compreender se a distribuição é normal precisamos olhar outro aspecto que é chamado de **curtose** (Kurtosis) que representa o quanto caudas são pesadas em relação à distribuição normal. Quando as caudas da curva são mais pesadas, contribuindo mais para a distribuição, a curva é chamada de **leptocurtica** (tem aspecto de cúpula do sino alto e mais fino, a cauda é longa, embora fina ela "pesa" e contribui bem para a distribuição embora longa) e quando as caudas participam pouco da distribuição à curva é recebe o nome de **platicurtica** (a cauda é curtinha, meio bracinhos de horácio). **Mesocurtica** seria uma distribuição normal em que as caudas contribuem harmoniozamente para a distribuição. Existem 3 maneiras de calcular a curtose e cada pacote estatístico tem suas preferências. A curtose calculada no SAS é a proposta por Snedecor e Cochran (1967) dois estatísticos bem famosos. Sendo que a curtose perfeita (mesocúrtica) seria igual à zero. Assim maior que zero seria leptocurtica, e menor que zero seria a platicurtica. No nosso exemplo, as notas de provas 2 e 3 são negativas logo platicurticas.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

O programa Stata, utiliza outra fórmula de curtose proposta por Bock(1975), porém a interpretação é que a curva normal teria curtose de valor 3 (mesocúrtica). Logo abaixo de 3 seria platicurtica e acima de 3 leptocurtica. Ainda uma terceira fórmula é utilizada pelo SPSS que foi proposta por Sheskin (2000), e utiliza como valor equivalente a mesocurtica de zero. Vamos voltar a comentar sobre curtose e simetria quando discutirmos melhor como avaliar uma distribuição normal.

Ainda nos "momentos" do SAS você encontra os termos USS que se refere à soma dos quadrados que equivale à soma dos valores da variável ao quadrado. Por vezes precisamos deste cálculo para fazer outros cálculos. E o CSS que se refere à soma dos desvios de cada valor em relação à média elevado ao quadrado. Essa informação representa qual a variabilidade total da amostra em relação à média elevado ao quadrado. Outra informação que pode parecer estranha são os pesos (weights) note que a soma dos pesos (Sum Wgts) é igual a 80 que seria o mesmo que o número de participantes. Isso significa que cada elemento aqui tem o peso de 1. Por vezes em amostras os pesos de cada indivíduo são diferentes se a amostra não for aleatória simples.

Calculando a mediana

Existem fórmulas para calcular a mediana, e regrinhas de como encontrar o valor, mas vamos aqui descrever de forma prática como se encontra a mediana. Definindo mediana como o valor que deixa 50% da amostra a sua esquerda depois da amostra organizada de forma crescente. Por exemplo:

B	C	D
Prova 1	Prova 2	Prova 3
6	5,25	2,5
3	7,2	1
4	9,2	2,5
7	6,9	5,75
10	8,3	3,5
3	6,5	4,5
0	5,75	2,5
4,5	7,6	4,25
3,5	8,4	3,5
7	8,4	4,5
4	6,5	4,25
5	3,7	3,5
0	3,8	5,75
7	6	4,5
6	7,9	2
2,5	5,3	3
2	7,65	2,25
4,5	7,7	2,75
4	5,8	3,25
4	7,9	5,75
4	5,1	2,5

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Estes valores acima estão organizados de acordo com nome dos alunos, mas para calcular a mediana precisamos organizar estes valores de acordo com a ordem crescente, do menor para o maior valor. Quando se pede a um programa estatístico para calcular a frequência acumulada, ele já organiza os valores. Se não tivéssemos computador faríamos isso a mão. Na figura abaixo temos a frequência acumulada para a terceira nota dos 81 alunos da sala.

The FREQ Procedure

nota3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1	1.23	1	1.23
0.5	2	2.47	3	3.70
1	4	4.94	7	8.64
1.4	1	1.23	8	9.88
1.5	4	4.94	12	14.81
2	10	12.35	22	27.16
2.25	3	3.70	25	30.86
2.5	6	7.41	31	38.27
2.75	2	2.47	33	40.74
3	3	3.70	36	44.44
3.25	5	6.17	41	50.62
3.45	1	1.23	42	51.85
3.5	5	6.17	47	58.02
3.75	1	1.23	48	59.26
4	3	3.70	51	62.96
4.25	2	2.47	53	65.43
4.5	8	9.88	61	75.31
4.75	2	2.47	63	77.78
5	4	4.94	67	82.72
5.5	1	1.23	68	83.95
5.75	4	4.94	72	88.89
6	4	4.94	76	93.83
6.5	2	2.47	78	96.30
7	1	1.23	79	97.53
7.75	1	1.23	80	98.77
8	1	1.23	81	100.00

Algumas notas foram iguais para vários alunos enquanto outras não. Note que 1 aluno tirou nota 0, 2 alunos tiraram nota 0,5, 4 alunos tiraram nota 1 e assim por diante. Na coluna nota temos os valores das notas, na coluna frequência temos quantos alunos tiraram aquela frequência. Um aluno nesta amostra de 80 representa 1,23% dos mesmos (1/81), sendo assim, na coluna porcentagem encontramos 1,23 quando se refere a apenas 1 aluno, 2,47 % quando se refere a dois alunos, 4,94% quando se refere a quatro alunos e assim por diante. Na frequência acumulada, note que a primeira linha tem o valor de 1 (que foi aquele que tirou zero), na segunda linha tem o valor de 3 , porque 3 alunos tiraram nota igual ou menor que 0.5. A próxima coluna (a ultima) se refere à porcentagem acumulada. É nesta coluna que vamos encontrar a mediana observando em que momento se deixa 50% para traz. Neste caso o 50% está justamente no valor 3,25 referente à nota de 5 alunos. Este é o valor que tenho certeza que deixa

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

50% dos valores para baixo. Se eu precisar dividir a amostra de acordo com a mediana vou ter que considerar os valores menores ou iguais a 3,25. Este valor representa 50,62% da população e não exatos 50%.

Além da mediana, podemos estabelecer os valores de referência de quartis (que divide a amostra em quatro). O primeiro quartil acontece quando se deixa para traz 25%. Assim é só procurar o valor de média que deixa para traz 25%. De acordo com a porcentagem acumulada vemos que o valor de nota 2 que teve 10 indivíduos com esta nota se refere a 27,16%, e este é o chamado primeiro quartil. O segundo quartil vai ser igual a mediana (inclui os valores acima do primeiro quartil ate incluir o valor da mediana), e o terceiro quartil se refere aos valores acima da mediana até o valor referente a 75% que é de 4,5. Podemos quebrar a amostra em decis (de 10 em 10) em percentis (1 em 1 porcento).

Junto com a noção de mediana acabamos de descrever o significado de quartil. Nos livros existem fórmulas e regrinhas que sempre que o número de indivíduos for par ai somamos e dividimos por dois, mas esqueçam disso, pois o que se precisa é ter noção do que a mediana significa. Na prática se precisamos calcular mediana ou quartil para categorizar uma variável continua sempre olhamos a frequência acumulada e pronto!

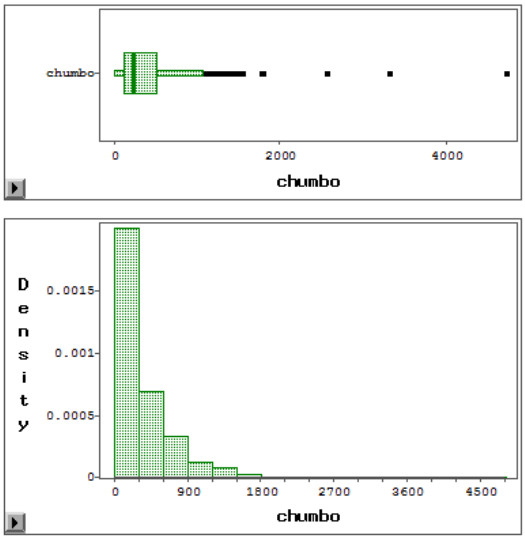
Ainda, a determinação dos quartis e mediana são necessários para ser construir o corpo (caixa) do box-plot.

Voltando ao significado da mediana em relação à média, note que nesta distribuição de notas da terceira prova, a média foi maior do que a mediana porque a cauda tinha tendência para a direita, causada por algumas notas melhores, mas a maioria das pessoas estava com notas mais baixas.

O intuito de se observar um histograma é compreender o qual a distribuição da variável que estamos trabalhando. Esta análise é importante porque é partir dela que decidiremos se à medida que mostra onde está à maioria seria a média ou a mediana. Dizemos que estas duas medidas são **medidas de tendência central** no sentido de informar onde está a maioria. No caso das notas a diferença entre média e mediana não foi tão grande, mas para algumas distribuições como exponenciais esta diferença pode ser muito grande, a ponto de não poder usar a média como medida de tendência central

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

aproximada de forma alguma. Veja a distribuição de chumbo em biopsia de dentes de crianças.



Note que neste caso a média deve ficar muito distante da mediana. Nesta figura existe um gráfico diferente acima do histograma, que é chamado de **box-plot**. O box-plot é reflexo organizado do histograma. O box-plot é formado principalmente por uma caixa, cujo inicio (da esquerda para a direita) representa o primeiro quartil. O risco no meio da caixa se refere à posição da mediana (segundo quartil), e a parede final da caixa se refere ao terceiro quartil (75%). Existe uma aste antes e uma depois mais longa que depois explicaremos como calcular que representam valores possíveis. Note que uma parte da aste da direita está mais escura e logo depois aparecem pontos. A parte escura foi formada pelo aglomerado de pontos pretos. Estes pontos pretos são valores extremos, que destoaram dos demais, e são chamados de **outiliers ,** ou simplesmente **valores extremos.** Podemos montar um box-plot a mão mas os computadores também os fazem, e bem mais bonitinhos que a mão. Portanto, você precisa é interpretar os dados.

Quando pedimos ao programa informações mais detalhadas verificamos que a amostra tem 277 indivíduos, com média de 389,72 que é bem maior do que onde se encontra a maioria que seria representada melhor pela mediana de 235,26. Estas duas caixinhas ainda trazem várias informações que por enquanto não nos interessa. No entanto, já é possível entender o que seria Q3 que é o terceiro quartil, e o Q1 que é o primeiro quartil que estão na caixa chamada "*quantiles*" ou seja quantis. Esta caixa ainda

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

traz informações importantes como valore máximo, valor mínimo, variação do menor para o maior valor (“range”), além dos 4 maiores valores e dos 4 menores valores. Note e compare o valor de assimetria que existia nas distribuições de notas de provas e no caso do chumbo. O valor de “skewness” é de 4.28 para o chumbo comparado a valores menores que 1 observados nas provas.

Moments

N	277.0000	Sum Wgts	277.0000
Mean	389.7213	Sum	107952.788
Std Dev	482.5248	Variance	232830.191
Skewness	4.2835	Kurtosis	28.6751
USS	106332628	CSS	64261132.6
CV	123.8128	Std Mean	28.9921

Quantiles

100% Max	4711.5025	99.0%	2548.9402
75% Q3	499.9299	97.5%	1486.7824
50% Med	235.2623	95.0%	1228.9963
25% Q1	121.1488	90.0%	857.8281
0% Min	0	10.0%	70.9414
Range	4711.5025	5.0%	29.4317
Q3-Q1	378.7811	2.5%	16.4232
Mode	.	1.0%	5.7049

Notamos que esta distribuição de chumbo com certeza não pode ser a média como seu representante de medida de tendência central, a mediana no caso é mais apropriada. Assim, se tivermos que compara altura entre duas cidades em que a distribuição é bem representada pela média vamos comparar as médias entre as duas cidades. No entanto, se formos comparar as concentrações de chumbo entre as duas cidades não se pode comparar médias e sim as medianas, isto é testar a hipótese de se as medianas são iguais. Isso é importante para guiar a escolha da correta da estatística a ser utilizada.

O que acabamos de ver é que média é uma medida boa para representar a medida de tendência central quando se tem uma distribuição normal. A distribuição do chumbo apresentada não é normal, ela é uma distribuição que podemos chamar exponencial. Existem vários tipos de distribuições são mais de 42 tipos e antes de prosseguir numa estatística, primeiro temos que definir qual é o tipo de distribuição. O livro de Evans, Hastings and Peacock (Statistical Distributions, Wiley Series in Probability and Statistics, 2000) identifica 42 tipos diferentes de distribuições.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Nesta apostila, vamos nos ater principalmente a distribuição normal em que a média é a referência quando trabalharmos com variáveis quantitativas contínuas. A análise de variáveis categóricas será abordada posteriormente embora já tenha visto que se derivarmos intervalos de confiança podemos comparar proporções.

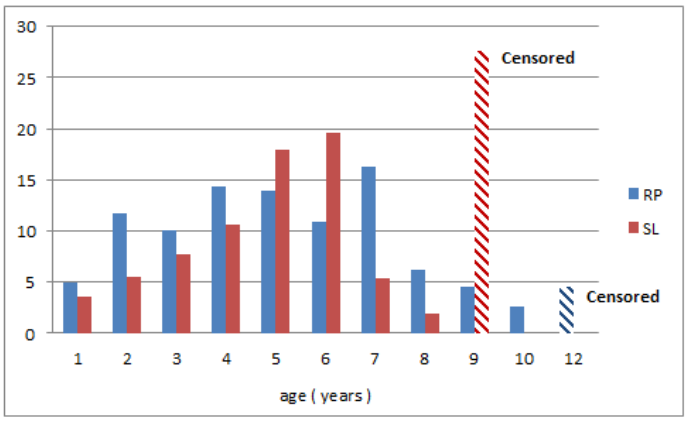
É bom se familiarizar com os termos **testes paramétricos** e testes **não paramétricos**. Estes termos são frequentemente citados em livros e ouvidos de professores de estatística. De forma simples, já vimos que se a variável é contínua e tem distribuição normal, a média pode servir de referência de tendência central. Assim se a distribuição é normal e se eu tiver que fazer algum teste estatístico com ela, eu sei que usarei como base a média além de outros parâmetros conhecidos como variância (que ainda não sabemos o que é, mas veremos posteriormente). Da mesma forma sempre que eu conseguir identificar uma distribuição eu consigo identificar parâmetros pertinentes à distribuição e consigo usar nos testes estatísticos apropriados. Chamamos de teste paramétrico aqueles que se baseiam numa distribuição conhecida (com parâmetros conhecidos). Quando não conseguimos identificar a distribuição correta de uma variável, ai é melhor usar o que chamam de testes não paramétricos. É comum que em testes não paramétricos a medida central de escolha seja a mediana e não a média.

Agora você já sabe que primeiro deve-se identificar a distribuição para saber como apresentar os dados por meio de medidas de tendência central e para escolher o tipo de estatística adequada que vai depender também do desenho do seu estudo. Para identificar a distribuição de variáveis quantitativa devemos sempre fazer gráficos incluindo histogramas e box-plots para visualizar os dados e decidir qual a melhor distribuição que representa os mesmos.

Embora seja sempre ressaltado que se devam examinar as distribuições de cada variável e refletir sobre a natureza das mesmas, frequentemente esquece-se deste passo. A figura abaixo é um destes exemplos, é um histograma com a apresentação da variável idade da primeira visita ao dentista para crianças de 11/12 anos idade da coorte de Nascidos Vivos de Ribeirão Preto de 1994 e de crianças de 7/8 anos de idade da coorte de Nascidos Vivos de São Luis (Maranhão) coletados em 2004. Esta variável é quantitativa discreta, porque é medida em anos inteiros. Num primeiro momento nota-se que a variável tem uma distribuição aproximadamente normal para as duas cidades.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

No entanto, embora tenha o jeito de sino, a distribuição normal é definida como distribuição de variável contínua, não discreta, e que supostamente vai de menos infinito à mais infinito. A variável idade da primeira visita ao dentista neste estudo tem outra particularidade além de ser uma variável discreta sua distribuição é censurada, isto é os indivíduos foram avaliados até os 8 anos em São Luiz, e vários deles ainda não tinham ido ao dentista e não sabemos se irão e quando irão. É como se algo fosse oculto na variável. Chamamos estes dados de dados censurados, pois a idade de visita ainda não existe para esta parcela de indivíduos. O mesmo acontece para as crianças cesuradas de Ribeirão Preto. Assim, calcular média para descrever esta distribuição não vai ser adequado, a não ser que falemos que exceto para aqueles que ainda não foram ao dentista a média de idade foi de X. Não é simplesmente uma distribuição normal, que se possa com os dados fazer uma média simples. Discutir o que fazer com estes dados não é propósito desta apostila. Este exemplo foi apresentado apenas como ilustração de que informações mais complexas exigem tratamento mais complexo. Se tivéssemos estas informações de adultos de 20 anos e nesta época todos já tivessem ido ao dentista, a média de idade da primeira visita ao dentista poderia ser calculada e analisada sem nenhum problema.



Até o momento falamos de distribuição do tipo gaussiana, mas existem outras. Por exemplo, o indice CPOD , indica o número de dentes Cariados, Perdidos ou Obturados. Desta forma, este índice quantifica o número de dentes. Portanto, é uma

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

variável quantitativa, mas não é contínua como no casao de altura que cujo valor pode assumir inúmeros posições. O CPOD pode ser no mínimo 0 e no máximo 32. Portanto, estamos diante de uma distribuição de uma variável que é **quantitativa discreta** e com limites superior e inferior, logo se aproxima mais de uma distribuição do tipo beta-binomial. A diferença é que a distribuição normal não tem limites de valores, embora você achar que altura tem um limite, pois ninguém é menor do que X centímetros, mas vamos dizer que teoricamente seria possível. Mas dentes temos no minimo 0 e no máximo 32, embora possam existir dentes extranumerários e anomalias.

Algo importante é que as distribuições podem mudar com o tempo. Por exemplo, no passado a distribuição de cárie dentária, ou melhor, do CPOD, tinha um jeito de normalidade, embora já tenhamos comentado que seria mais bem descrita pela distribuição beta-binomial. No entanto, com a diminuição de cárie na população a quantidade de pessoas livres de cárie aumentou e hoje a distribuição tende a ser exponencial. Ainda em populações com alta prevalência de cárie a distribuição permanece com formato de sino. A importancia disso é que a média serve como medida de tendência central quando a distribuição é em forma de sino, mas não serve para distribuição exponencial para qual a mediana é a melhor representante.

Neste novo seguimento você deve ser capaz de definir e identificar:

1- a unidade de estudo
2- e intepretar um histograma e um box-plot
3- interpretar e atribuir a medida de tendência central correta
4- e interpretar o que é desvio a direita ou a esquerda (skweness) e outilers
5- e calcular e interpretar o que é frequência acumulativa, e porcentagem acumulativa
6- quantis, quartis, mediana numa distribuição
7- os termos testes paramétricos e teses não paramétricos
8- entender a importância da distribuição

Reta

Análise descrtiviva

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Uma vez observada e classificada uma distribuição, passamos a ter condições de descrever a distribuição com medidas descritivas adequadas, tais como **medida**s que nos dêm noção **de tendência central** (onde encontramos a maioria da amostra/população)**,** de variabilidade ou dispersão dos dados (distâncias interquartis ou desvio padrão) e de valores extremos, que são chamados de ***outliers***. Vamos fazer revisão conceito de medida de tendencia central e adicionaremos outras medidas descritivas importantes além das medidas de tendência central.

Medidas de Tendência Central e o Medidas de Variabilidade (o desvio padrão)

Dentre as medidas de tendência central temos a média e a mediana. A média é calculada somando todos os valores e dividindo pelo número de indivíduos da amostra. Quando dizemos que numa sala temos 100 indivíduos e que a média de altura é de 1.68 cm, sem ter visto a distribuição real, imaginamos que metade dos indivíduos tem menos que 1.68 cm e outra metade acima de 1.68 cm e ainda imaginamos que a maioria dos indivíduos tem altura ao redor de 1.68 cm. Imagine agora a média sendo calculada para a distribuição exponencial apresentada acima. A maioria dos valores estão próximos de zero, e poucos indivíduos tem valores altos. Se calcularmos a média, veremos que ela não representará onde está a maioria dos indivíduos, mas um valor maior. No caso da distribuição do chumbo a média é de 305,7μg e a mediana é de 205,7μg. Neste caso, a medida de tendência central ideal será a mediana. A mediana é o ponto onde se divide 50% da amostra/população. A maneira mais fácil de entender como se calcula mediana é organizando a distribuição de frequência acumulativa da variável em questão. Imagine os seguintes 9 números: 7 27 13 2 14 10 11 2 4. O primeiro passo é organizar os números em ordem crescente, e montar uma tabela de porcentagem acumulada como na tabela a seguir.

Podemos notar que o valor de 50% corresponde ao valor 10. Desta forma, podemos calcular a mediana de maneira bastante simples, apenas observando a posição do valor na distribuição da variável. Existem fórmulas para calcular o valor referente a mediana e estas fórmulas são na verdade para se achar a posição do valor referente a mediana. Se temos uma população de 100 indivíduos o percentil 50 será localizado na

posição (n + 1)/ 2, logo, 101/2 , que seria igual a posição 50,5. Como não vai existir a posição 50,5 aproxima-se para a posição seguinte, onde ser pode garantir que abaixo daquele valor temos pelo menos 50% da amostra. Insisto em que não decorem a fórmula para se avaliar a posição central, apenas monte a porcentagem acumulada e veja qual o número que garante que 50% dos indivíduos estão a esquerda (ou para cima na tabela em ordem crescente).

Indivíduos	Frequência	Porcentagem	Porcentagem Acumulada
2	2	22,22	22,22
4	1	11,11	33,33
7	1	11,11	44,44
10	1	11,11	55,55
11	1	11,11	66,66
13	1	11,11	77,77
14	1	11,11	88,88
27	1	11,11	99,99

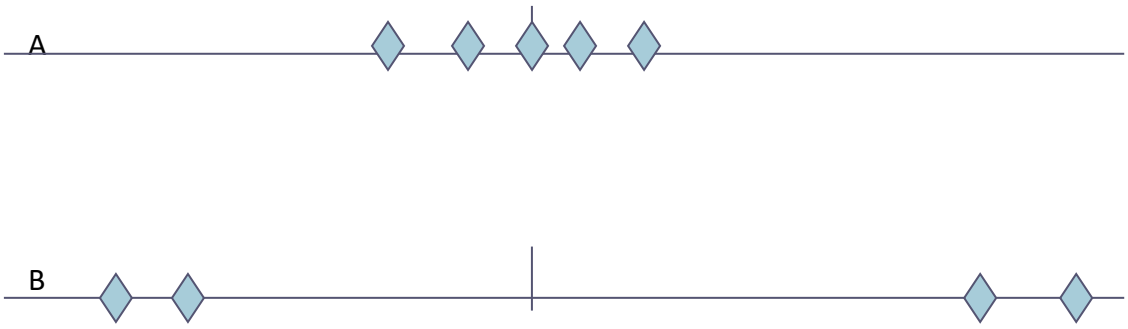
No exemplo acima, o valor da mediana foi igual ao valor da média. Na distribuição normal isso realmente acontece, mas em distribuições não normais, a média em geral é diferente. No exemplo da distribuição do chumbo a média foi bem maior.

Note que o maior problema nas provas é que o aluno esquece de organizar os dados em ordem crescente para se achar a mediana.

O desvio padrão

Além das medidas de tendência central, precisamos também descrever como a distribuição esta dispersa.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva



Vejamos a figura acima. Nesta figura, temos duas distribuições com médias iguais. No entanto, é notório que as distribuições não são semelhantes. Na distribuição A os elementos estão bem juntos a média , mas na distribuição eles estão dispersos.

Voltando ao nosso exemplo de altura é possível que tenhamos dois grupos de indivíduos com mesma média (1,68m) porém dispersões distintas. Em um grupo a variação de altura pode ir de 1,65 a 1,70 enquanto no outro de 1,45 a 2,05. No primeiro, as alturas teriam menor variabilidade e no segundo maior variabilidade. Essa variabilidade pode ser expressa em variação apartir da média, uma vez que a média seja uma medida válida para tal distribuição. Assumindo que isso seja verdadeiro, vamos calcular esta medida com os dados da tabela acima.

Indivíduos	**Média**	**Xi - M**	**(Xi-M)2**
2	10	-8	64
2	10	-8	64
4	10	-6	36
7	10	-3	9
10	10	0	0
11	10	1	1
13	10	3	9
14	10	4	16
27	10	17	289
Soma		0	488

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Vemos que quando somamos as diferenças dos valores em relação à média para ter uma noção de quanta variabilidade teríamos, o resultado é zero. Isto é, esperado, pois a média se encontra num ponto equidistante de todos os pontos. Assim, esta soma não resolve a necessidade de uma medida que demonstre a variabilidade dos dados. Uma forma de resolver é eliminando os valores negativos elevando-se ao quadrado. Com isso podemos somar as distâncias ao quadrado e calcular o desvio médio ao quadrado. Esta medida de desvios médios ao quadrado é chamada de **variância.** Sempre pode-se fazer a pergunta do porque não utilizar desvios absolutos, e a questão é que por outros motivos além de expressar a variabilidade ao redor da média , elevando ao quadrado tem propriedades desejáveis em outras estatísticas como regressão linear (diferenciabilidade quando x = 0).

A fórmula a seguir se refere à variância que cabamos de compor. É apartir desta idea de variância que muito das estatisticas de análise de variância e de modelos generalizados se baseiam como veremos depois. De forma simples, a variância nos dá ideia do quanto de variabiliade existe ao redor da média em média. O porque do valor "menos 1" será discutido mais adiante**.**

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

No exemplo acima, a soma dos desvios é de 488 e a média da soma dos desvios é de 54,2, que é a variância. Porém, a variância esta no mundo dos quadrados, e fica difícil situar os valores ao quadrado em relação à medida original. Assim, temos que retornar estes números ao nosso "mundo" dos números originais. Isso pode ser feito tirando-se a raiz quadrada da variância. O resultado desta operação é o que chamamos de **desvio padrão**. Neste caso, o desvio padrão é de 7,36. Costumo brincar que existem mundos diferentes mundo do quadrado, onde quando algo entra é transformado ao quadrado, da mesma forma, mundo do log etc. Em alguns momentos precisamos trabalhar nestes mundos porque fica mais fácil de trabalhar, porém mais difícil de interpretar, desta forma precisamos sair do mundo dos quadrados e retornar para a

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Terra, extraindo a raiz quadrada. Saimos do mundo do log e ao entrar na Terra ele sofre exponenciação e retorna a escala original para que possa ser interpretado.

A fórmula do desvio padrão é desta forma :

$$\sqrt{\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}}$$

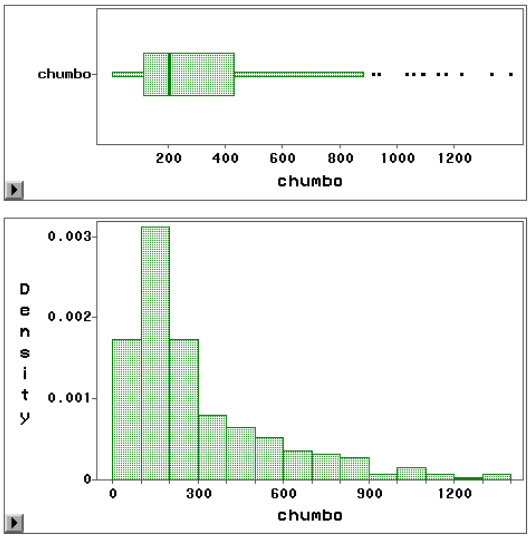
Note que nas duas fórmulas acima o n esta subtraído do valor 1. Isso acontece porque em geral tratamos de desvio padrão de amostras e não de populações totais. Na variância ou desvio padrão de uma população não teremos esta subtração. Note que você não deve decorar a fórmula, se você compreendeu o conceito sempre irá se lembrar como se calcula a variância e o desvio padrão. Se decorar hoje, irá esquecer amanha. Em geral os livros começam sempre explicando apenas variância e desvio padrão de uma população, e somente depois explicam o que seria estes valores para amostras. O mais importante, no entanto é entender o que são estas medidas. Ainda as anotações encontradas em livros diferenciam se estas medidas são populacionais ou de amostras. A variância populacional em geral é representada pela letrão σ^2 e a variância amostral pela letra s^2, consequentemente o desvio padrão populacional representado por σ e o amostral s.

O desvio padrão partindo de uma distribuição normal, vai nos dar uma ideia de variabilidade ao redor da média. De forma geral, um desvio padrão de cada lado da média nos revela que cerca de 68% dos dados estarão contidos neste intervalo; dois desvios padrão ao redor da média representam cerca de 95% dos dados, e 3 desvios padrão teriam aproximadamente 99.7% dos dados. Estas estimativas são chamadas de **regra empírica** (*empírical rule*).

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

O desvio padrão é muito útil para termos noção de variabilidade que pode ser importante para avaliarmos de certa forma a validade dos dados. Na década de 70/80 era muito comum vermos trabalhos de regeneração óssea, e membranas para recuperar osso periodontal com desvios padrão altíssimos. Por exemplo, uma determinada técnica levava ao ganho de inserção de 3mm em média com desvio padrão de 5mm. Isso queria dizer que cerca de 68% dos dados deviam estar entre – 2 e 8. Logo, era possível que varios dentes tivessem perdido inserção periodontal e que alguns ganharam. Isso nos revelava que a técnica talvez não fosse muito vantajosa clínicamente, muito estável, a não ser que se saiba predizer quem vai ganhar inserção periodontal.

O desvio padrão é uma medida de dispersão adequada para distribuição normal, mas não é adequada para outras distribuições. Quando a distribuição não é normal, devemos utilizar outra medida de dispersão. Já dissemos anteriormente que a mediana seria uma medida apropriada para descrever a tendência central de uma distribuição não normal. Junto com a mediana podemos utilizar a **distância interquartil** para demonstrar a variabilidade de dados. Esta medida seria a distância entre o quartil 1 (25%) e o quartil 3 (75%). Lembre-se que mediana se refere ao quartil 2. Graficamente esta dispersão pode ser demonstrada pelo box-plot.



mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Note nesta figura que o box-plot tem um risco marcado no meio, que representa a mediana. A caixa representa a distância interquartil, que contém 50% dos indivívuos de 25% a 75%). As astes que ficam de cada lado da caixa são calculadas como sendo 1.5 vezes a distância interquartil. E os pontinhos soltos são as medidas extremas que chamamos também de *outliers.*

Para o gráfico de chumbo os valores foram assim calculados

Distância interquartil = 432.25 –115.67 = 316.58

Limite inferior = 115.67 – 1.5 (316.58) = - 359, 20 (como não tem chumbo negativo o menor valor será mesmo zero, por isso a distancia da aste é bem curtinha)

Limite superior = 432.25 + 1.5 (316.58) = 907,12

O box-plot é um ótimo gráfico para ajudar a entender uma distribuição seja ela normal ou não. Quando a distribuição for normal, o box-plot será simétrico. Note que o box-plot da distribuição de chumbo é mais concentrado á esquerda (no lado dos valores mais baixos). Note também que os valores entre Q1 e Q2 tem pouca variação, o que aumenta apartir de Q2. O box-plot é complementar ao histograma pois ele nos fornece de forma mais nítica a posição da mediana e também dos outliers.

Ao explorar dados por meio de gráficos, principalmente quando queremos fazer comparações devemos ter o cuidado de colocá-los numa mesma escala. Escalas diferentes podem dar impressão que os valores são diferentes.

Um exemplo, muito comum de efeito de escala que pode nos surpreender, é a variação do dolar dia a dia mostrado nos jornais ou televisões. As escalas são muito pequenas e uma queda de 2 ou 3 centavos de dolar durante o dia, pode parecer uma queda enorme, porque as escalas são graduadas em décimos de centavo. Por isso, muitos dizem que podemos “mentir” com a estatistica. Na verdade a estatistica pode ser utilizada para ludibriar as pessoas despreparadas que ignoram noções básicas de estatística. É o mesmo que enganar uma pessoa humilde que não sabe ler ou lidar com dinheiro, podemos oferecer 10 notas de 1 real, em troca de uma nota de 100 reais. Portanto, cuidado com escalas de gráficos.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Significado prático dos valores extremos e o que fazer com eles

Acima falamos de valores extremos chamados de *outliers.* O que são esses valores e como lidamos com eles? Ter um outlier num estudo é algo a ser bem analisado antes de tomar uma atitude e prosseguir com a análise. Quando encontramos um outlier, devemos avaliar o que ele representa, devemos investigar o porquê de sua existência. Devemos voltar aos registros e nos certificar de que o valor não foi digitado errado, ou anotado errado durante a coleta de dados. Se for um erro de anotação, devemos corrigi-lo, e se não for possível corrigir devemos eliminá-lo. Se não for um erro, então devemos analisar bem antes de decidir o que fazer. A princípio se não for erro devemos ficar com os valores. No entanto, durante a análise devemos avaliar o quanto este valor altera os resultados das estatísticas, e devemos relatar em nossos resultados o efeito do outlier nos resultados. Jamais exclua outliers simplesmente porque o gráfico ia ficar mais bonitinho! Isso é mentir em ciência, é trapacear! Não há necessidade de trapacear em ciência ninguem quer provar nada de um jeito ou outro, pelo menos esta não deve ser a intensão do verdadeiro pesquisador. O verdadeiro pesquisador esta apenas buscando a verdade, e procurando melhorar o conhecimento.

Erro padrão vs desvio padrão

Pesquisadores sempre ficam em duvida sobre apresentar numa tabela o erro padrão ou desvio padrão. Se o objetivo é apenas demonstrar a variabilidade dos dados que foram coletados, então o desvio padrão é a medida a ser escolhida. No entanto, se a finalidade é representar o teste de hipótese do estudo em que uma medida pontual é igual a outra então o erro padrão é a medida que deve ser apresentada. Ainda o intervalo de confiança (que é composta de erros padrão) deve ser sempre a medida de escolha. Por economia de espaço, se o erro padrão é apresentado, é fácil para o leitor fazer as contas e calcular o intervalo de confiança.

Ainda como já comentamos em termos de conta o erro padrão é definido em alguns livros como o desvio padrão sobre a raiz de n que é o mesmo que dizer que seria a variância sobre n. Aí lembramos como deve ser a equivalência de fórmula para quando tivermos tratando de calculo de erro padrão para uma proporção. Proporção também

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

tem variância que depende da probabilidade de um evento vezes a probabilidade de não ter um evento dependendo do número de indivíduos isto é n*p*q.

A esperança de de uma distribuição binomial é n*p. Por exemplo, se temos probabilidade de 0.2 de uma doença e uma amostra de 50 pessoas, vamos ter como esperança o valor 10, isto é espera-se que 10 pessoas sejam doentes. A variância é dada por p*q e numa dada amostras a multiplicação pelo n. Se esta é minha variância o desvio padrão é de raiz de n*p*q.

Bom, a ideia é a mesma, mas não temos desvio padrão para proporção, porém temos estimativa de variabilidade que é dado pela p (1-q) e o erro padrão para proporção é a raiz quadrada de p (1-p) dividido por n.

N*p*q é a

Coeficiente de variação

Um coeficiente muito utilizado para expressar a variabilidade é o coeficiente de variação. Ele é uma razão entre o desvio padrão e a média. Assim, ele nos informa quão maior é o desvio em relação a média. Quanto mais próximo a zero menor a variabilidade dos dados e isso é considerado bom, pois a precisão é maior. Quanto mais próximo ou superior a 1 o que significa que a variabilidade é muito grande. Exemplo, uma média de ganho de osso periodontal de 7mm com desvio padrão de 8, resulta em um coeficiente de variação de 8/7 = 1,14 o que é muito grande. Se lembrarmos das regrinhas um desvio de cada lado da média significa que aproximadamente 68% dos indivíduos estão nesta variação. Neste caso, diríamos que a média de ganho de osso foi de 7mm mas que cerca de 68% dos indivíduos estavam entre -1 a 15mm, isto é uma variação muito grande.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Comparação de médias e proporções

Já vimos anteriormente que todos os testes de comparação de médias ou mesmo proporções podem ser realizados calculando-se o intervalo de confiança por meio de simulações. Esta é uma possibilidade, mas comumente você encontrará referência ao teste T (para comparação de duas médias) e ao teste do qui-quadrado (para comparação duas variaveis categóricas).

É preciso entender que não usaremos as anotaçoes de livros por enquanto como sigma e outras letras gregas. Vamos apenas utilizar “nossas anotações” para entender as estatísticas. Num capítulo posterior adicionaremos as diferenças destas terminologias.

Primeiro, é preciso entender que quando testamos se duas médias são iguais, como no exemplo entre média de altura de crianças de Ribeirão Preto e Sertãozinho, é o mesmo que dizer que estamos testando se existe associação entre cidade e altura. O que é o mesmo de se perguntar se a variação da variável altura é independente da variável cidade. Testar se crianças com asma tem a mesma proporção de cárie dentária do que as crianças sem a asma é o mesmo que testar se existe associação entre asma e cárie dentária.

Primeiro vamos ver as comparações de médias, seguida de análise de correlação e regressão linear e o teste formal de análise de variância.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

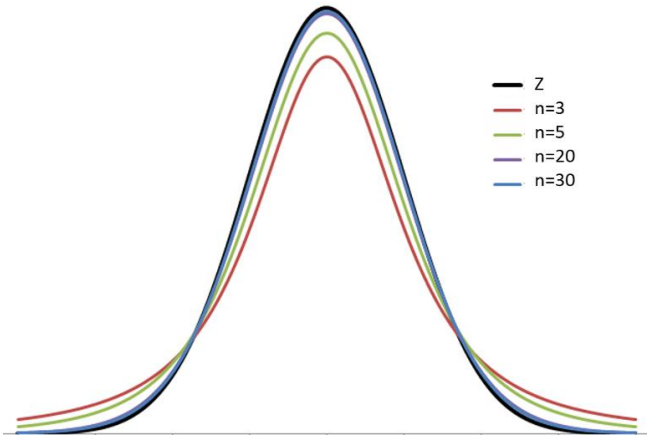
Comparação de médias

Agora que já vimos como se faz para comparar proporções fica fácil entender como o teste de comparação de médias funciona. Lembre-se que você pode criar intervalos de comparação por meio de simulações o que seria aceitável. Ainda temos um teste por meio de cálculo que gera uma estatística, como no caso da aproximação que vimos com a comparação de proporções.

O teste de comparação de médias mais utilizado é chamado de teste t de student, mas apenas para duas médias, quando quisermos comparar mais de duas médias, utilizaremos o teste chamado analise de variância, também conhecido como ANOVA. O teste tem esse nome porque considera comparação de médias de amostras e amostras especialmente pequenas seguem a uma distribuição que se chama t. O que significa a distribuição t (amostral). Bom, a distribuição normal assume que estamos diante de toda a população, o que na prática de estudos empíricos raramente acontece, a nao ser em censos e poucos casos. Se nós tivessemos uma população inteira em nossas mãos, a variância e desvio-padrão desta população seria conhecida. Acontece que em geral trabalhamos com amostras desta população, e portanto, nós não sabemos nada da variabilidade da população. O que sabemos é a variabilidade da amostra que é um estimador da variância verdadeira. Se tivessemos que comparar as médias populacionais de altura de meninos e menina por exemplo, como faríamos já que não temos aquela curva de amostras? Bom, tendo a população inteira, poderiamos calcular a variância que seria a verdadeira, e usar neste caso o desvio padrão como base de comparação à semelhança que fizemos com o erro-padrão. No entanto, em geral não trabalhamos com população inteira e sim amostras. Já é conhecido então que o estimador do desvio padrão verdadeiro que não conhecemos é o desvio-padrao da amostra dividido pela raiz quadrada de n (n = número de indivíduos na amostra). Se tivermos uma amostra grande a distribuição que teremos na amostra dado que a população tem distribuição normal, será proxima da normal. No entanto, se nossa

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

amostra é pequena a tendencia é que esta distribuição não seja tão proxima, embora tenha o mesmo jeitão da normal. Imaginem se eu tenho uma amostra pequena é possível como que tenhamos mais variabilidade de valores sorteados, e no final a distribuição tende a ser mais baixa e com caudas mais gordinhas. Se tivermos uma amostra grande é possível que se assemelhe a normal. Essa distribuição amostral menos númerosa é dita seguir uma distribuição chamada de "t" (distribuição t de student). Note na figura abaixo a diferença relativamente sutíl entre a normal e as distribuiçãoes t que dependem do número da amostra.



Se tivemos uma amostra grande, podemos utilizar a distribuição normal como parametro para analisar as diferenças de duas médias, mas se não utilizamos a distribuição t. Se utilizamos a distribuição normal vamos chamar de teste Z (estatística Z), e se utilizarmos asmostras pequenas o test t (estatística t). Nos computadores os programas em geral lançam o teste-t como o mais utilizado, porque em geral trabalhamos com amostras, e caso nossa amostra seja grande o test-t se aproxima do normal e o resulado é basicamente o mesmo. Em geral nos livros, começa-se apresentando todos os testes com populações e depois com amostras. Aqui começamos pelas amostras porque é o mais utilizado.

O teste t, comparação entre duas médias, tem alguns pressupostos como, por exemplo, que a variável (desfecho) tem uma distribuição normal, que os valores sejam independentes e ainda que as variâncias entre dois grupos sejam iguais. Como no caso

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

da comparação de proporções precisamos calcular a variância e erro padrão para a amostra que vamos estudar. Quando comparamos médias entre dois grupos é a mesma coisa que testar a hipótese de associação entre o que caracteriza cada grupo e a variável em questão. De outra forma, testamos a associação entre a variàvel indepentende e desfecho. Por exemplo, quando comparamos se a altura dos meninos é maior do que a altura das meninas é o mesmo que testar se existe associação entre sexo e altura. Lembre-se estudar os fatores que afetam a altura é o objetivo, então altura é também chamada de desfecho, ou variável dependente, e o sexo é considerado variável independente ou fator de exposição.

O primeiro passo para testar a hipótese de associação entre duas médias, é ter certeza de que os pressupostos podêm ser garantidos, isto é distribuição normal, independência dos dados, e ainda a variância igual para os dois grupos estudados. A distribuição normal pode ser verificada pelo gráfico de distribuição, tipo histograma, e ou ainda pela informação de que a medida seja conhecidamente normal. Mesmo que a amostra não pareça ser muito normal, o mais importante é que a medida seja conhecidamente normal. Existem testes estatísticos que testam a hipotese de que a amostram se assemelha a distribuição normal como por exemplo Shapiro-Wilk, Kolmogorov-Smirnov, entre outros, mas em geral não levamos estes testes muito a sério porque dependem muito do tamanho da amostra, e na verdade combinamos varioas informações para concluir sobre normalidade, incluindo curtose (kurtosis) e simetria (skewness) que veremos posteriomente.

A independência dos dados, ou melhor dizendo, que significa a independência dos indivíduos também tem que ser garantida. Por exemplo, não podemos ter conglomerados em nosso estudo, pois se não o teste t simples não pode ser aplicado. Podemos pensar então que não podemos comparar médias quando temos conglomerados? Não é bem assim, não podemos utilizar o test t comum que encontramos nos livros, mas podemos utilizar comparação de médias proprias que vão levar em consideração o conglomerado. Neste caso a grande diferença que vamos encontrar é em relação ao cálculo do erro padrão.

Veja a fórmula do test-t abaixo assumindo variâncias iguais para os dois grupos, temos que o valor do teste t é dado pela seguinte fórmula.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Vamos tentar entender esta fórmula. Baseado no que já foi exposto, podemos comparar os intervalos de confiança entre duas médias e isso seria uma forma de testar se as médias são iguais. Outra forma seria intuitivamente que poderíamos construir uma curva de diferença de médias e, portanto construir um intervalo de confiança para esta diferença e verificar se o intervalo de confiança incluiria o zero. Isto é, da mesma forma que fariamos hipotéticamente uma amostra de uma população calculariamos a diferença de média de altura para meninos e meninas e depois teríamos na nossa curva de "diferença de médias". O que a fórmula acima sugere é mais ou menos isso, ela mostra a divisão entre a diferença de médias encontradas na amostra de nosso estudo, dividida pelo erro padrão. Erro padrão? Sim, o S significa variância (é um sinal normalmente utilizado na estatística para representar variância de amostra), e multiplicado pela raiz do inverso do n (número de indivíduos) é o mesmo que dividir o desvio padrão pela soma da raiz de n. Como dissemos anteriormente o erro padrão em termos de fórmula (não conceito) é o desvio padrão dividido pela raiz de n. Desta forma, a fórmula acima seria a diferença de médias dividida pelo erro padrão. O quanto a média é maior que um erro padrão. Ainda parece estranho não? Vamos exemplificar. Imaginemos a comparação de dois grupos com média 7 e outro com média 4 e ambos com erro padrão de 1,5. A diferença seria de 3 dividido pelo erro padrão de 1.5 termiamos que 3 é duas vezes maior que 1.5. Logo um valor de 2. Este valor temos que levar para uma tabela pré-estabelecida que vai corresponder a um valor da probabilidade desta diferença ser igual a zero. Intuitivamente podemos imaginar que quando o valor da diferença for menor que um desvio padrão, então a probabilidade do intervalode confiança incluir o zero será grande. No exemplo se a diferença é de 3, então 1.96 erros para cada lado, teremos 2.94 (= 1.5 x 1.96). Assim o intervalo teria o limite

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

inferior de 3 menos 2.94 (0.06) e limite superior de 5.94. Nota-se que este intervalo não inclui o zero, portanto, podemos concluir que as médias são diferentes.

A fórmula, no entanto, nos da o valor exato de 2 que deve ser levado na tabela t, e la deve-se verificar qual a probabilidade das médias serem iguais para o valor de 2 dado os graus de liberdade. Depois explicarei o que é grau de liberdade, por enquanto fica que os graus de liberdade correspondem ao número de indivíduos. O importante é saber que o valor será correspondente a uma probabilidade que chamaremos de "p" ou valor de p (p de probabilidade). O computador faz isso automaticamente para nós, portanto hoje em dia o importante é entender o processo.

Na fórmula acima o "S" , que é utilizado para denominar o desvio padrão geral da amostra, assumindo que as variâncias de cada uma das médias são iguais. Lembre-se que em termos de fórmula, o erro padrão é o desvio padrão dividido pela raiz quadrada de n. Mas aqui temos dois grupos de indivíduos que embora tenhamos assumido que eles tem variâncias iguais, cada grupo pode ter um tamanho de amostra diferente. Por isso, que o desvio padrão é dividido pela raiz da soma do inverso de cada um dos n.

A variância (S) é a média ponderada de duas variâncias amostrais, onde os pesos são o número de graus de liberadade de cada amostra (isto é n de cada amostra menos 1) . Para fazer esta ponderação é calculado a variância para cada grupo a ser comparado separadamente. Utiliza-se então as duas variâncias caculadas para gerar uma variância única que será poderada. A poderação é conseguida pela fórmula:

$$s^2 = \frac{(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2}}{n_1 + n_2 - 2}$$

O que a fórmula do teste t (mostrada anteriormente) nos dá é qual a diferença entre as médias pontuais, em relação ao erro padrão. Se a diferença é muito pequena em relação ao erro padrão e formos reconstituir o intervalo de confiança, veremos que este intervalo deve incluir o valor zero, que seria ausência de diferenças. Se a diferença é grande em relação ao desvio, possívelmente o intervalo de confiança (usando o erro

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

padrão) não incluirá o zero demonstrando que a diferença é diferente de zero, e, portanto, as médias são diferentes. Mas não precisamos montar os intervalos de confiança para concluir se as médias são diferentes. O valor resultante do test pela fórmula acima, é utilizado da mesma forma que o valor do teste qui-quadrado, observando-se uma tabela de distribuição de valores correspondentes ao valor de p específico para distribuição t.

Você não será cobrado na prova sobre a fórmula acima. O que você precisa saber é que o teste t, serve para comparação de duas médias. Este teste tem, como os demais testes em estatística, pressupostos que devem ser analisados. O primeiro é que os dados precisam ser independentes, isso é cada indivíduo é independente do outro. Segundo, que a variável a ser estudada, o desfecho tem que ter distribuição normal ou pelo menos que tenha saído de uma distribuição conhecidamente normal, caso contrário a média não faz sentido e ai não se pode aplicar este teste. Além disso, a variância dos grupos a serem comparados tem que ser iguais. Fica difícil comparar dois grupos onde as variâncias são muito diferentes. Imagina comparar altura de meninos e de meninas onde os meninos teriam uma variância muito grande, isso é uma variabilidade grande, com um grupo de meninas onde a variabilidade é basicamente nula. Isso vai gerar uma instabilidade ao se compara os dois grupos. Assim, quando isso acontecer não poderemos utilizar o teste t formal, teremos que ponderar as variâncias. Isso seria mais ou menos assim, se as variâncias são iguais podemos fazer o teste t diretamente mas se não for, teremos que ponderar essas variâncias.

Algo importante para ficar claro que embora tenhamos apresentado as comparações de médias e proporções apenas com intervalo de confiança, por vezes, esta regra não vai bater com o test-t formal. Porque? Apenas porque a comparação é feita ponderando as variâncias. Veja um exemplo. A seguir você verá o resultado de um test-t comparando médias de provas para alunos que acertaram uma questão 5 numa prova de avaliação preliminar. A questão 5 era uma questão tirada do ENEM sobre proporções, e a prova deste exemplo era a terceira prova do curso de Epidemiologia e Bioestatística.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Q5	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	43	5.0558	1.9667	0.2999	0	9.2500
1	30	6.1117	2.0042	0.3659	2.2000	9.7500
Diff (1-2)		-1.0559	1.9821	0.4715		

Q5	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		5.0558	4.4505	5.6611	1.9667	1.6216	2.4997
1		6.1117	5.3633	6.8600	2.0042	1.5962	2.6943
Diff (1-2)	Pooled	-1.0559	-1.9960	-0.1157	1.9821	1.7029	2.3717
Diff (1-2)	Satterthwaite	-1.0559	-2.0017	-0.1100			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	71	-2.24	0.0283
Satterthwaite	Unequal	61.796	-2.23	0.0293

Note que os intervalos de confiança das médias separadamente se interpoem. O limite superior da média da prova para quem tirou zero na questão 5 foi de 5,6611 enquanto o valor inferior da média para quem acertou a questão cinco foi de 3,3633. Logo, se comparassemos apenas o intervalo concluiriamos que não houve diferenças estatisticamente significantes. No entanto, se você observar o intervalo de confiança para a diferença entre as médias verá que é completamente abaixo de zero, isto é a diferença é significante. Isso aconteceu porque o erro padrão calculado para a diferença entre as médias, é a soma das variâncias ponderada. Note que o erro padrão da média para quem tirou zero foi de 1,9667 e para quem acertou (nota 1) foi de 2,0042, e a soma dos erros deu menor que o maior erro isso é 1.9821. Por isso essa diferença.

Embora a ideia de comparação de intervalos seja intuitiva, e serve para compreendermos os processos, existem estas peculiaridades. Com certeza se encontrar dois intervalos de confiança que não se sobrepoem significa que as médias são diferentes. Mas se se interpoem por pouco, quando o erro padrao comum (somado) for calculado pode ficar menor e dar signigicante. Essa diferença é devido tanto a variabilidade diferente de cada grupo, como também a quantidade de pessoas em cada grupo. Fiquem atentos e não se desesperem quando isso acontecer.

O teste para verificar se as variâncias são semelhantes é o teste F (Fisher) que é um teste que verifica o quanto uma variância é maior do que a outra (razão de variâncias) levando-se em consideração os graus de liberdade de cada uma. O resultado, é verificado em comparação a distribuição F.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

$$F = s_1^2/s_2^2$$

Esse teste de igualdade de variâncias é apresentado no SAS como **Folded F**. Esse folded significa que SAS estabelece como regra sempre colocar a maior variância no numerador e a menor no denominador. Assim, o resultado do teste será sempre positivo. Se a probabilidade das variâncias serem iguais for baixa ($p < 0.05$) o teste-t não pode ser aplicado de forma usual, será necessário ponderar esta variância na comparação de médias, e isso levará a maior perda de graus de liberdade. Essa ponderação realizada pelo estimador de variância Satterwaite para ser utilizado no teste-t. Precisa ficar atento se utilizar outro programa, pois o SAS já nos fornece todas estas informações como veremos a segui com um comando apenas de *proc ttest*. Mas em outros é o pesquisador que tem que verificar se as variâncias são iguais e achar o estimador ponderado para utilizar se este for o caso.

Teste t no SAS

No sas o teste t é realizado com os comandos proc ttest. Os comandos são fáceis, mas é necessário que você se lembre dos pressupostos: a distribuição deve ser normal, os dados devem ser independentes, e as variâncias dos dois grupos estudados devem ser semelhantes. Lembrando que avaliar se que avaliar se existe dependencia dos dados é realizado por meio de compreensão do desenho do estudo e a normalidade é avaliada por gráficos e conhecimento da varíavel, e por ventuda testes estatísticos. Para verificar se as variâncias dos dois grupos são semelhantes precisamos fazer um teste chamado teste de igualdade das variâncias. No output do SAS este teste é feito automaticamente quando se pede o procedimento de test t (proc ttest).

Vamos ao exemplo sobre altura dos alunos, e vamos testar se a altura entre meninos e meninas é igual, ou seja se não existe associação entre altura e sexo.

Os comandos do SAS para as variáveis que demos o nome de altaluno e de sexo são:

```
proc ttest;
class sexo;
var altaluno;
        run;
```

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Note que o output do SAS libera vários números que podem parecer confuso a primeira vista, mas temos que olhar devagar passo a passo.

The TTEST Procedure

Variable: altaluno (altaluno)

Sexo	N	Mean	Std Dev	Std Err	Minimum	Maximum

Parte1

1	30	1.7780	0.0517	0.00944	1.6900	1.8600
2	42	1.6469	0.0542	0.00836	1.5400	1.7500
Diff (1-2)		0.1311	0.0532	0.0127		

Parte 2

Sexo	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		1.7780	1.7587	1.7973	0.0517	0.0412	0.0695
2		1.6469	1.6300	1.6638	0.0542	0.0446	0.0691
Diff (1-2)	Pooled	0.1311	0.1058	0.1564	0.0532	0.0456	0.0637
Diff (1-2)	Satterthwaite	0.1311	0.1059	0.1563			

Method	Variances	DF	t Value	Pr > |t|

Parte 3

Pooled	Equal	70	10.32	<.0001
Satterthwaite	Unequal	64.339	10.40	<.0001

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	41	29	1.10	0.8022

Primeiro o output é composto de uma primeira parte chamada "The ttest procedure" abaixo deste título você observa que existe o n (número de indivíduos), mean (média de altura), std dev (standard deviation = desvio padrão), std err (erro padrão), minimum (valor mínimo), maximun (valor máximo). As duas linhas inferiores a estas colunas tem os valores de médias para sexo 1 (meninos) e sexo 2 (meninas). Se você quiser com estas informações realizar sozinho o test t, você consegue, como? Você

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

tem nestas linhas as médias e também o erro padrão, então é só montar o intervalo de confiança para cada uma das médias e compará-las. Como montar o intervalo? pegue o erro padrão dado, multiplique por 1.96 e terá o comprimento que deve adicionar ou subtrair da média. Tente fazer isso no exemplo acima com a sua calculadora. Se os intervalos baterem em algum valor isso, significará que não foi possível encontrar diferença estatísticamente significante entre meninos e meninas.

Abaixo destes valores existe uma linha chamada Diff (diferença) que significa a diferença entre as médias de altura de meninos e meninas que no caso é de 0.13, este valor é seguido na linha pelo desvio padrão (0,0532), o erro padrão (0,0127) . Note estes desvios são em relação a diferença. Note que se você quiser pode calcular o Intervalo de Confiança de 95% da diferença entre as médias. Para tanto, é só pegar a diferenca de 0.1311 e adicionar e diminuir dela o valor do erro padrão da diferença multiplicado por 1.96. Se o intervalo da diferença contiver o valor zero, significa que zero é uma possível diferença, e portanto não terá sido possível encontrar diferenças estatísticamente significantes entre altura de meninos e meninas. No etanto, se o intervalo não contiver o valor zero, isso significa que a diferença existe entre altura de meninos e meninas.

Mas você não precisa na verdade calcular o intervalo de confiança, porque o SAS fornece os intervalos na parte 2. Na parte 2, repetem-se as médias, com seus respectivos intervalos de confiança de 95% (95% CL mean, o SAS chama de CL confidence limites, isto é limites de confiança). Note que se você quiser testar as hipoteses de diferença entre as médias poderá fazer apenas olhando os Intervalos de confiança fornecido. O IC para diferença das médias também se encontra aqui.

Note que nesta parte é dado ambém os desvios-padrão, e o intervalo dos desvios. Para que serve isso? Se você quiser pode testar a hipótese de que as variâncias são iguais por meio da comparação dos IC dos desvios-padrão de meninos e meninas. Note que o intervalo do desvio para meninos e meninas se interpoem, logo podemos assumir que as variâncias não são diferentes. Logo, o pressuposto de variâncias iguais ou semelhantes é aceito, e, portanto o test será valido.

A parte 3 contém o resultado final do test-t. Note que existem dois métodos, um chamado pooled e o outros Satterthwaite (na verdade Welch- Satterthwaite) o primeiro será utilizado quando o pressuposto das variâncias iguais for aceito. Este é o teste t,

digamos que original. Se o pressuposto de variâncias iguais for violado, devemos utilizar o outro método para variâncias não iguais (unequal). Quanto aos graus de liberdade vamos tentar explicar depois em inglês é o degrees of freedom (DF). O valor do teste já foi explicado anterioremente. O que nos interessa de mais prático é o valor do Pr > t, isto é a probabilidade das médias serem iguais. Neste caso, esta probabilidade é muito pequena, é menor que 0.05. Lembre-se sempre que a probabiliade der menor ou igual a 0,05 dizemos que o evento é pouco provável. Mas como decidir qual dos métodos utilizar? Para tanto devemos fazer o teste de **igualdade das variâncias** (equality of variances). Como neste caso a probabilidade das variâncias serem iguais é de 0.8022 , consideramos esta probabilidade grande e aceitamos que as variâncias "são iguais" isto é não são diferentes.

Bom é assim que se le o output (saída) do SAS. Todos os programas tem saídas mais ou menos iguais. Sendo que poucos informam o teste de igualdade das variâncias. Assim, há necessidade de se pedir o teste de variâncias iguais antes de se fazer o teste t em outros programas estatísticos.

Tamanho de efeito (Effect size) de Coehn, chamado de Coehn´s D. O que significa e como se calcula? È uma medida diferença entre médias em relação ao desvio padrão, e não erro padrão. Imagine se a diferença entre duas médias for muito pequena em relação a variabilidade dos dados, então essa diferença intuitivamente não será muito grande. Admite-se que se essa medida for de 0.2 é uma diferença muito pequena, se for de 0.5 será médio o efeito, e se for de 0.8 para cima seria um efeito bom, ou grande. Imagine se a diferença fo de 1. Neste caso, a diferença seria igual a 1 desvio padrão. O cálculo deste coeficiente de Cohen (Cohen´s D) então compreende diferença de média dividida pelo desvio padrão. O desvio padrão de uma diferença é a soma dos desvios como vimos acima, quando explicamos como calculamos o teste-t. Se os tamanhos das amostras das duas médias forem iguais, o cálculo deste desvio padrão da diferença é mais simples se forem diferentes é mais complicadinho. Na verdade é a fórmula que mostramos anteriormente. Apenas se forem iguais passa a se ser mais simplificada será a soma das variâncias divido por 2.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

$$s^2 = \frac{(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2}}{n_1 + n_2 - 2}$$

O teste t no ***Rstudio*** não fornece todas as informações juntas e este de igualdade de variâncias, assim deve-se pedir cada um separado. Primeiro, depois de verificar os pressupostos de normalidade e independencia dos dados, vamos verificar a igualdade das variâncias com os comandos a seguir.

```
var.test (variaveldesfecho ~ grupo, nomebandodedados, alternative = "two.sided")
var.test (rn1pesonasrc ~ rn1sexo, bancofinal, alternative = "two.sided")
```

A distribuição normal: tipo de variável, gráficos, curtose, simetria e testes estatísticos

O pressuposto de normalidade para o teste de comparação de médias é um tópico muito controverso. Muitas pessoas tem o hábito de sempre avaliar a normalidade de uma distribuição por meio de testes estatísticos como Kolmogorov-Smirnov, ou Shapiro-Wilkins, Anderson-Darling, Cramer-von Mises entre outros. Estes testes comparam a distribuição avaliada com uma distribuição normal, se a probabilidade destas distribuições serem iguais for grande, assume-se normalidade e se não se conclui que a distribuição avaliada não é normal. Acontece que em amostras pequenas a tendência é não encontrar diferenças, e em amostras grandes encontrar diferenças. Assim, devem-se utilizar estes testes com parcimônia. O teste Komogorov-Smirnov é indicado para amostras grandes com n acima de 200, se for menor Shapiro-Wilk é indicado, e em geral o mais utilizado. Anderson-Darling é uma modificação do KS com maior poder de teste.

O mais importante é saber se a variável vem de **uma distribuição conhecidamente normal**, e prestar atenção em suas características. Uma distribuição normal verdadeira vem de uma variável contínua que vai do infinito negativo ao positivo. Por exemplo, altura é uma variável contínua, porém não da para dizer que tenham pessoas com altura negativa, mas se aceita que existam pessoas bem pequenas a pessoas bem grandes sem um limite definido. Além de avaliar o tipo de variável se continua ou discreta, é preciso, especialmente no caso de não se ter certeza se a

amostra vem de uma distribuição conhecidamente normal, avaliar a distribuição da amostra coletada. Assim, observamos o "jeitão" de normalidade por meio de histograma e Box-plot. Além da característica geral de normalidade podemos olhar duas medidas chamadas de curtose e simetria já mencionadas anteriormente e combinar todas as informações que se tem sobre a distribuição. Curtose é uma medida que expressa o quanto as caudas de uma distribuição são "pesadas" em relação à distribuição toda. Mais ou menos se essas caudas são "gordinhas". Uma distribuição normal tem curtose com valor de 3, assim uma distribuição para ser normal deve ter kurtose próxima de 3. Mas isso depende de como é calculada. No SAS kurtose zero é normal porque o valor 3 é descontado do que é apresentado.

Como se calcula a curtose? Basicamente é a relação da variância que é a soma dos desvios ao quadrado dividino por n, e a variabilidade ao poder 4, menos o valor de 3 que deve dar zero. Esta fórmula foi proposta por Snedocor e Cochran (1967), quando não se desconta o valor de 3, assim para os programas que utilizam a fórmula sem descontar o 3, a distribuição normal tem valor referência de 3. Esta forma é encontrada no progama Stata. No SPSS também o valor de referencia é zero. Deve-se verificar em cada programa estatístico como a curtose é calculada. Mas porque a *variabilidade a quarta potência* comparada com a variabilidade a potência 2?. Ao se elevar a quarta potência, os valores grandes crescem mais rapidamente e assim pode-se ter uma ideia de velocidade de peso das caudas de uma distribuição em relação aos valores mais próximos da média.

Se o valor da curtose é menor do que zero ou menor do que 3 isso indica uma curva em que as caudas não contribuem muito para a curva e não pesam na distribuição e é chamado de curva **platicúrtica**. Vamos dizer que temos menos elementos nas caudas e a aparência é uma curva mais "parruda", vamos dizer que seria um Orácio (personagem do Maurício de Souza) de bracinhos curtos e parrudinho. Se o valor é igual a 0 ou 3, então temos uma curva normal com características usuais e é chamado de **mesocúrtica**. Se o valor da curtose é maior do que 0 ou 3, é uma curva chamada de **leptocúrtica**, em que as caudas são longas, por ter mais indivíduos nas caudas bem longas.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

A fórmula da simetria (skewness) é igual a razão de 3 vezes a distância entre média e mediana dividido pelo desvio padrão. A distribuição normal não deve ter diferença entre média e mediana, que deve ser zero. Se a distribuição tem cauda longa para um lado, o resultado é que a mediana irá se distanciar da média. Assim, simetria de valor zero significa que é distribuição normal, e aceita-se até um valor de 1 para simetria ser considerada aceitável dentro de uma distribuição normal. Maior que zero, skewness positivo, e menor skewness negativo.

Mas como vou usar esta informação? Bom, até o valor de 1 ao redor do zero para skewness é razoável, consideramos bom para uma distribuição ser normal e com alguma tolerância até 1, e alguns autores aceitam até mesmo 2. Já para a kurtose, dependende se desconta o zero ou não. Alguns aceitam como compatível com normalidade até 3 pontos ao redor do zero, ou 6 pontos se não descontar o zero. Para algumas técnicas estatísticas como equações estruturais consideram 2 pontos ao redor do zero.

Resumindo, combinamos todas essas informações especialmente visuais e medidas de curtose e simetria para decidir se aceitamos se a distribuição é normal ou não. É fundamental sempre reunir informações sobre o tipo de variável, o que ela representa e mede, para se concluir sobre normalidade. A finalidade é para decidirmos se a média e desvio padrão seriam bons descritores de medidas de tendência central e variabilidade para uma distribuição. Assim, não queremos uma distribuição nem com simetria muito alterada e nem mesmo com caudas muito diferentes. Como em geral nossos dados são amostras o ideal é analisar quantos desvios padrão tem ao redor destas medidas de curtose e simetria e juntar este conhecimento aos aspectos visuais e a finalidade do porque queremos ter certeza de que a distribuição não seja diferente da normal.

Quanto aos testes estatísticos devem servir apenas de guia, pois em amostras grandes em geral sempre rejeitam a normalidade. Existem alguns tipos de testes como Kolmogorov –smirnov que em geral é recomendado para amostras grandes, e o Shapiro-Wilk em geral para amostras menores. Ainda existem o teste de Lilliefors, e o Anderson-Darling.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Antes de correlação é melhor colocar comparação de proporções primeiro aqui (2022).

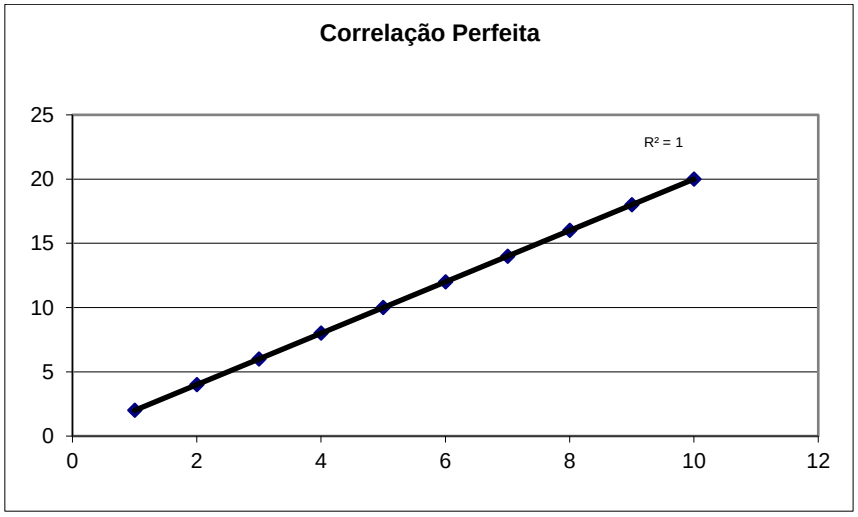
mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Correlação de variáveis

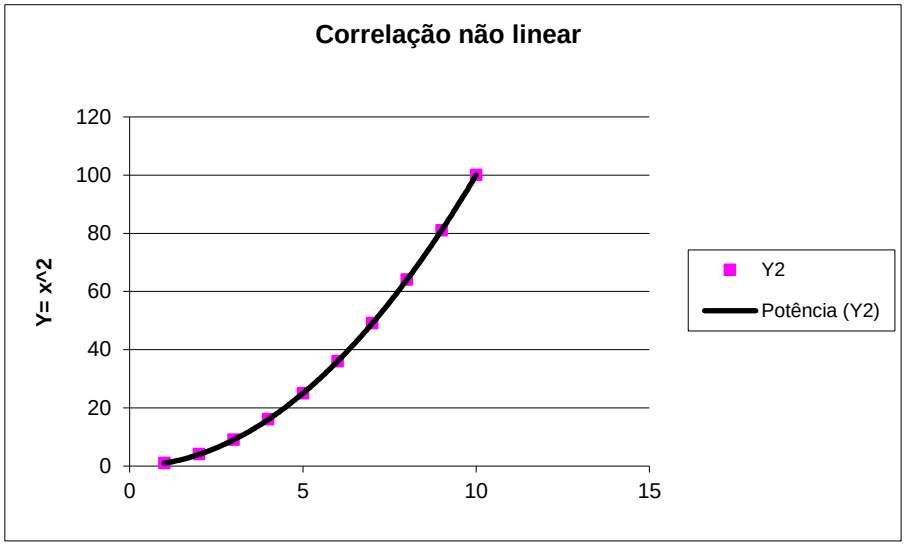
Até o momento comparamos grupos distintos de variáveis categóricas por meio do teste qui-quadrado, ou comparamos médias em grupos distintos testando associações entre variáveis. Agora vamos descrever as análises de correlação de variáveis contínuas como no caso de estudo de correlação entre altura de pais e filhos. Note que aqui utilizamos um termo correlação" ao invés de associação. Devemos deixar o termo correlação apenas para quando formos falar de estudo de variáveis quantitativas concomitantemente, e associação quando estudarmos variáveis qualitativas puras ou qualitativa com quantitativa.

O que nós queremos saber em geral é se uma variável quantitativa varia em relação a outra concomitantemente. As correlações podem ser lineares ou não. Um exemplo de correlação linear pode ser visto na figura abaixo.



Exemplo 1: correlação linear perfeita

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva



Exemplo 2: correlação não linear quadrática.

No exemplo 2 temos uma correlação não linear que no caso é resultante de uma função quadrática. Vamos descrever aqui apenas o estudo das **correlações lineares**. O que queremos ver é se uma variável x está correlacionada *linearmente* com a variável y. É claro que outros tipos de correlações não lineares são de extrema importância, mas isso demanda conhecimento além do nosso curso.

O grau de correlação linear é medido pelo coeficiente de correlação (identificado pela letra *r)* que varia de – 1 a + 1. Esta medida é calculada com base na variação poderada dos valores das variáveis estudadas em relação a média de cada variável. Vamos tentar entender o que isso significa. Se houver correlação perfeita entre duas variáveis quer dizer que quando uma variável aumenta a outra também aumenta proporcionalmente ou diminui proporcionalmente, podendo ser portanto uma correlação positiva ou correlação negativa. Vamos simular um exemplo de duas variáveis ficticias com correlação perfeita. Para cada aumento de 1 unidade da variável X existe o aumento de 2 unidades de Y. Podemos dizer que Y = 2X sempre sem sobrar nenhuma variação não explicada.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Se tabularmos estes dados veremos que a correlação foi de 1, isto é perfeita. Note que na tabela abaixo os desvios padrão de x e y são respectivamente:2,872281 e 5,744563

X	Y	$x_i - \bar{x}$	$y_i - \bar{y}$	$\frac{x_i - \bar{x}}{DP(x)}$	$\frac{yi - \bar{y}}{Dp(y)}$	$Z_1 * Z_2$
1	2	-4,5	-9	-1,5667	-1,5667	2,454546
2	4	-3,5	-7	-1,21854	-1,21854	1,484849
3	6	-2,5	-5	-0,87039	-0,87039	0,757576
4	8	-1,5	-3	-0,52223	-0,52223	0,272727
5	10	-0,5	-1	-0,17408	-0,17408	0,030303
6	12	0,5	1	0,17408	0,17408	0,030303
7	14	1,5	3	0,522233	0,522233	0,272727
8	16	2,5	5	0,87039	0,87039	0,757576
9	18	3,5	7	1,21854	1,21854	1,484849
10	20	4,5	9	1,5667	1,5667	2,454546
						10/10 = 1

O que a tabela acima mostra é que se cada valor de x e y ponderado for sempre o mesmo teremos uma correlação perfeita. Note que a variância de X é menor do que a de Y.

Podemos transformar o que acabamos de fazer na tabela acima em fórmula.

$$corr(xy) = \frac{1}{n}\sum_{n=1}^{n}\left(\frac{x_i - \bar{x}}{dp(x)}\right)\left(\frac{y_i - \bar{y}}{dp(y)}\right)$$

Também, podemos utilizar um cálculo mais simplificado com a seguinte fórmula.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

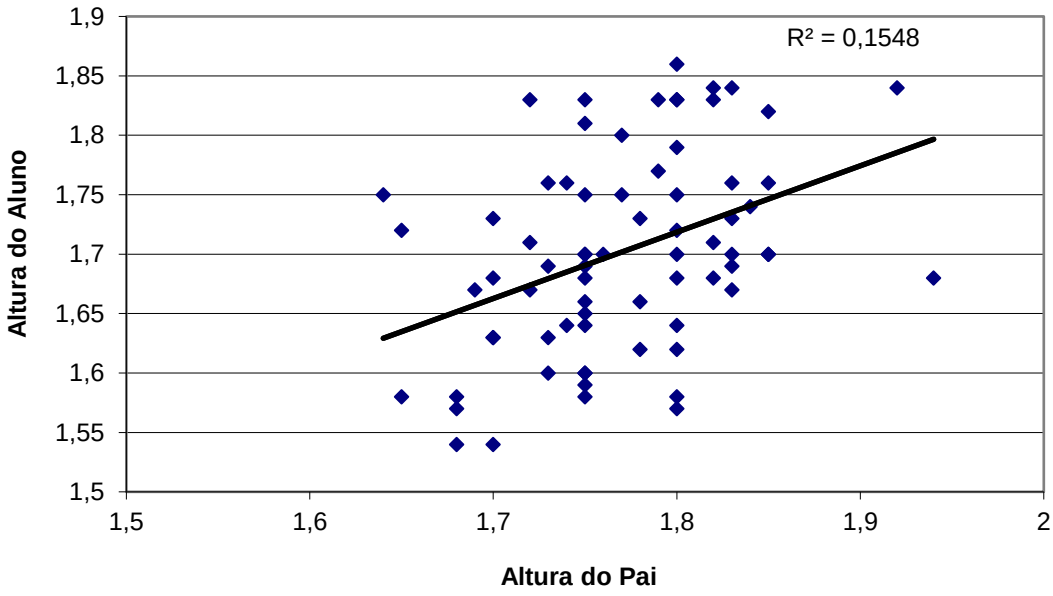
$$r = \frac{\sum_{i=1}^{n}(x_1 - \bar{x})(y_1 - \bar{y})}{\sqrt{\sum (x_1 - \bar{x})^2 \sum (y_1 - \bar{y})^2}}$$

No numerador desta última fórmula, que simplesmente é a simplificação da fórmula anterior, temos o que chamamos de **covariância** de x e y, pois multiplica-se para cada indivíduo o quanto o x está longe da média em relação a quanto o seu y está longe da média. Se considerarmos que no numerador temos a soma de dos desvios da média para x multiplicado pela soma dos desvios para y , podemos fazer um paralelo com as leis da probabilidade vistas anteriormente. É como se no denominador multiplicassemos duas probabilidades independentes resultando na máxima variabilidade que teríamos com o conjunto de dados observados. Já no numerado temos o resultado da covariância entre x e y. Se a covariância representar toda a variabilidade possível dos dados quer dizer que a correlação entre os dados é perfeita e será igual a 1 e se não existir será igual a 0. Lembrando que ela pode ser positiva ou negativa.

Quando analisamos correlações entre variáveis quantitativas, temos que ter muito cuidado. Devemos sempre lembrar de verificar como os pontos estão dispersos, e se realmente a tendência linear é a melhor para descrever a associação, pois o computador sozinho não consegue fazer esta análise.

Além de explorar a correlação geral entre variáveis, em geral estamos na verdade interessados em testar hipóteses de associações. O exemplo da sala de aula, foi sobre a altura dos alunos. Primeiro perguntamos se a altura era diferente entre meninos e meninas. E verificamos pelo teste t que sim. Uma outra pergunta envolveu as alturas dos pais e dos alunos, queríamos saber se a altura dos pais explicava a variação de altura dos alunos. Para tanto, fizemos um gráfico de dispersão. Lembre-se que o que queremos avaliar sempre estará no eixo Y e o que esta “causando” o efeito estudado, deverá estar no eixo x. Neste caso altura do pai estará no eixo X, e altura dos alunos no eixo Y. Em outras palavras a variável dependente é o efeito a ser estudado, e a variável independente será a suposta “causa”.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva



O r^2 neste gráfico corresponde ao quadrado do coeficiente de correlação que é chamado de **coeficiente de determinação**. Este coeficiente tem um signficado diferente do coeficiente de correlação. O r^2 de 0.1548 significa que 15,48% da variação da altura dos alunos foi explicada pela variação da altura do pai. *Note que por vezes você irá encontrar pessoas dizendo que o r^2 significa que o modelo consegue predizer X por cento dos valores observados e não é isso, ele explica a variabilidade apenas, que é o que estamos trabalhando desde o início variabilidade ao redor da média.* Neste gráfico lembre-se que temos meninos e meninas, e pode ser que altura do pai se correlacione de maneira diferente entre meninos e meninas. Veremos como calcular o coeficiente de determinação mais adiante.

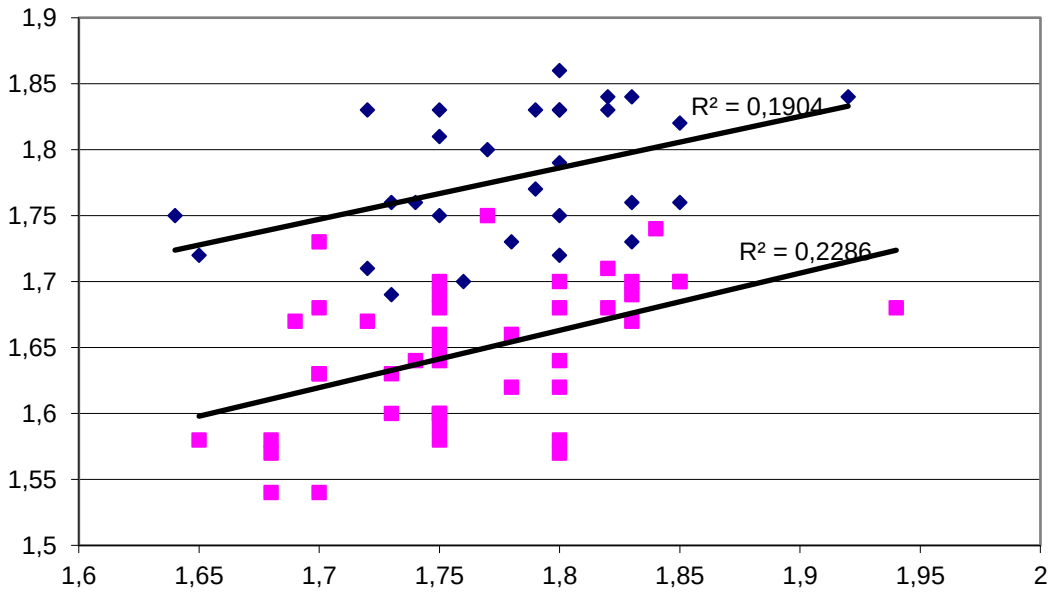
A reta que esta desenhada no gráfico acima é a a reta que melhor descreve a relação de altura de pai e filhos. Seria como se desenhassemos várias retas e escolhessemos aquela em que o quadrado da distância entre cada ponto e a reta fosse o menor. Imagine, que colocassemos uma reta no gráfico e calculassemos as distâncias de cada ponto até a reta, como a reta fica num local no meio dos pontos, vamos ter distâncias positivas e negativas que se somadas dará zero, daí, à semelhança ao cálculo do desvio padrão e há necessidade de se elever as diferenças ao quadrado. Dizemos então que esta é a técnica baseada nos quadrados mínimos (least squares). É óbvio que ninguém fica fazendo inumeras retas e calculando onde ficaria a melhor reta, existem

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

maneiras de se calcular a inclinação da reta e o ponto em que a reta passa pelo intercepto (isto é onde a reta toca o eixo y). Veremos isso nas próximas páginas.

Antes de continuar com a análise de correlação, devemos notar que o gráfico acima tem dois valores que se destacam dos demais. Estes dois valores são chamados de valores extremos ou *outlier* e não devem ser removidos sem uma avaliação criteriosa. Devemos verificar se houve erro de digitação dos valores e concerta-los mas não devemos excluídos de qualquer forma. Veja a discussão anterior sobre *outliers.*

Antes de dar prosseguimento a nossa análise, devemos neste momento de exploração visual dos dados, questionar todas as possibilidades que poderiam invalidar a correlação estudada. Podemos por exemplo questionar a mistura de alturas de meninos e meninas em conjunto. Será que a correlação entre altura de pai e filhos é diferente para meninos e meninas? Se existe esta possibilidade devemos explorar a correlação em gráficos diferentes, ou pelo menos identificar as correlações isoladamente como fizemos no gráfico a seguir.



Neste gráfico os pontos em cor–de-rosa se referem as meninas (reta inferior) e os azuis aos meninos (reta superior). Vemos que o coeficiente de determinação é um pouco maior para as meninas do que para os meninos, mas não sabemos se realmente

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

são estatisticamente diferentes, pois a inclinação da reta não é muito diferente. O que vemos é que a reta para as meninas com angulação semelhante está deslocada para baixo, e isso é porque em média as meninas são mais baixas que os meninos.

Além de simplesmente chegar a conclusão de que uma variável se correlaciona ou não com outra, os estudos de correlação podem ter o objetivo de se calcular a função (fórmula) que descreve esta associação. Isto é qual a função de correlação entre x e y que resulta na reta que melhor descreve a associação observada. Com esta fórmula podemos fazer predições, isto é dado a altura de um pai, qual seria em média a altura esperada do filho? Você deve se lembrar de uma fórmula para descrever uma reta do tipo $y = ax + b$. Pois é, um dos objetivos da analise de correlação pode ser tentar estimar esta fórmula.

A fórmula acima pode ser escrita com qualquer letra, isso não importa muito. No entanto, em geral, na estatística utilizamos y como sendo a variável que chamamos de **dependente,** e x a variável independente. No nosso exemplo em que estudamos influência da altura do pai na altura do filho, altura do pai é a variável independente (x), e a do filho a variável dependente (y).

Além do y e x, vemos que a fórmula tem mais dois elementos, o "a" e o "b". O "b" se refere ao intercepto, isto é onde a reta cruza o eixo Y. E o "a" se refere a inclinação da reta (*slope).* Embora possamos utilizar qualquer letra para descrever esta fórmula, é comum nos livros de estatística utilizar ao invés de "b" o **βo** (beta zero), e ao invés de 'a", o beta $\beta_{1.}$ Tanto β_0 como β_1 são chamados de coeficientes da regressão, e não apenas o βo. Note que este outro beta tem o valor 1 (um) na frente porque se refere a primeira varíavel independente. Podemos sim avaliar num estudo várias variáveis independentes. Nesta nossa análise, por enquanto, estamos avaliando apenas a altura do pai. Neste caso temos apenas β_1 , mas se levarmos em consideração também a altura da mãe, teremos então β_2 , e assim por diante. No caso de estudarmos uma variável apenas chamamos de análise bivariada, e se tivermos mais de uma variável chamamos de análise multivariada.

$Y = \mathbf{\beta o} + \beta_{1,}$ para análise bivariada.

$Y = \mathbf{\beta o} + \beta_1 + \beta_{2.....} + \beta_{k,}$ chamado de análise multivariada

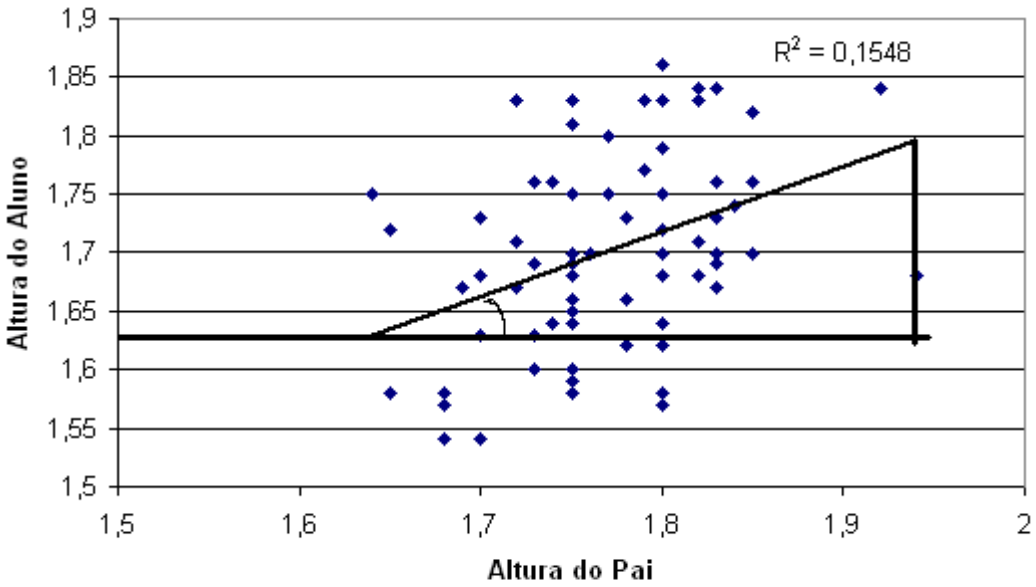
mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Lembre-se que sempre teremos um erro adicionado a estas fórmulas porque, nem todos os pontos de altura estão exatamente em cima da reta que iremos estimar. A reta que traçamos é a melhor reta que minimiza os erros (isto é a distâncias de cada pontinho até a reta).

$$Y = \boldsymbol{\beta o} + \beta_1 + \beta_{2.....} + \beta_{k,} + \varepsilon$$

Ainda para complementar nossa fórmula, lembre-se que estavamos lidando com uma amostra de uma população, logo, estamos falando de estimadores. Seria então mais correto, colocar um acento circunflexo sobre os Betas, pois esses betas não são os verdadeiros e sim estimadores dos verdadeiros Betas. Dai você ja pode concluir uma coisa, se são estimadores, temos erro amostral ao redor de cada um destes estimadores. Sim, podemos calcular intervalos de confiança ao redor de cada beta. Para que? Bom estes intervalos vão nos dizer se o intercepto é igual a zero ou não, e se os betas (inclinações de retas) são iguais a zero ou não.

Voltando ao cálculo da reta, vamos começar calculando a reta baseado no gráfico abaixo já com a reta descrita. O β_1 (ou "a") é o ângulo do triângulo que se forma no gráfico. Vejamos:



mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Este ângulo é o mesmo seja em que ponto estiver do triângulo, ele parecerá mais aberto perto do cateto oposto, mas na verdade é o mesmo ângulo. Como calcular a inclinação desta reta? Para isso precisamos lembrar como se calcula a tangente de um ângulo (justamente esta inclinação da reta) que é o cateto oposto dividido pelo cateto adjacente. Vamos tentar calular esta tângente baseado nos dados da gráfico. A distância do cateto oposto é de 0,17 (de 1,63 até 1,8) , já a distância do inicio da reta deve estar mais ou menos entre 1,64 e 1,94, que é 0,3. logo 0,17 /0.3 = 0,56 metros. Isso quer dizer que devemos dividir a variação das alturas dos alunos pela altura dos pais. Imagine se altura de aluno fosse na verdade quilometros percorridos numa viagem, e altura dos pais minutos. Se dividirmos variação de quilometros por variação de tempo, teríamos quantos quilometros rodados por minuto. Extrapolando isso para altura, temos o quanto a altura do filho varia em relação a variação de um centímetro de altura do pai. Isso nos da uma idéia de "**velocidade**". Para quem é mais interessado em cálculo, a tangente expressa a variação instantância num ponto, que é o mesmo que o cálculo de derivada. Vamos voltar a isso, quando discutirmos taxa na epidemiologia. Matemática, estatística e epidemiologia estão todos relacionados. É bom saber que as coisas que aprendemos no ginásio servem para alguma coisa.

Se você conferir com os dados do SAS verá que a estimação grosseira que fizemos pelo gráfico estava correta. Calculamos 0,56 e o SAS nos deu 0.558 que é aproximadamente 0,56. Como interpretamos esse valor de Beta? Em média a cada unidade de valor do x, que é a altura do pai, neste caso medida em metro, leva a variação de 0,56 metros da altura do filho. Ficou meio estranho, porque os dados estavam em metro e não centímetro que seria melhor. Vamos supor se fosse medido em centimetros, então teríamos tudo multiplicado por 100, mas isso não afetaria a inclinação da reta, apenas a interpretação que não será em metros, mas em centímetros. Desta forma, isso signficaria que a cada centimetro de aumento na altura do pai, haveria em média 0,56 centimetros a mais na altura do filho. No *output* do SAS este valor esta decrito como parametro estimado (parameter estimate) para a variável "altpai". Veja abaixo o *output* do sas para análise de regressão linear (isto é análise que tenta estimar a melhor reta que explique a correlação de uma variável com a outra). Não se assute com o output, vamos entender cada pedacinho dele com o tempo.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Number of Observations Read 72
Number of Observations Used 72

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.07716	0.07716	12.82	0.0006
Error	70	0.42137	0.00602		
Corrected Total	71	0.49853			

Root MSE	0.07759	R-Square	0.1548
Dependent Mean	1.70153	Adj R-Sq	0.1427
Coeff Var	4.55978		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.71384	0.27602	2.59	0.0118
altpai	altpai	1	**0.55815**	0.15589	3.58	0.0006

Desde o início do output o SAS informa, o número de observações lidas (n=72) e a utilizada (n=72) isso significa que existiam informações para todos os alunos na lista. A análise de variância, se refere ao teste estatístico realizado para verificar o quanto da variabilidade da altura do filho é explicada pela altura do pai, e se esta correlação é estatísticamente significante. Veremos a análise variância em detalhes logo a seguir. Por enquanto, vamos nos apreender aos parâmetros estimados da análise de regressão linear. Existem dois estimadores de parâmetros *(Parameter Estimetes),* um deles é o estimador do βo (intercepto), e o outro é o estimador da inclinação da reta (β_1) para altpai. Os dois parâmetros contribuem com 1 grau de liberdade (DF = degrees of freedom), e os parâmetros estimados foram respectivamente 0.71 e 0.56. Cada um destes parâmetros tem um erro padrão (*standard error*) associado, isso porque estamos lidando com uma suposta amostra que certamente tem um erro amostra.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Ao lado de cara erro padrão temos o valor do teste t. Você deve estar indignado, porque falamos anteriormente que teste t servia para calcular diferenças de médias. Bom, de certa forma estamos testando aqui se a média do intercepto e a média da inclinação da reta são iguais a zero (isto é com uma distribuição de média zero). Por isso temos aqui o teste t. O valor de p se refere ao resultado dos testes t. Logo estes valores nos dizem que a probabilidade da inclinação da reta ser igual a zero é de 0,0006, portanto, tão pequena que vamos considerar que β_1 é diferente de zero. Logo a inclinação não é nula.

Mas se lembre uma coisa, nós conseguimos calcular a inclinação da reta (chamada de *slope* em inglês) porque tinhamos a reta já desenhada. Logo, podemos fazer estatística sem contas, apenas fazendo desenho e usando geometria. Isso é verdade, mas conforme vamos fazendo análises com várias variáveis isso se torna bastante complexo, embora o princípio seja o mesmo.

ANÁLISE DE VARIÃNCIA

Vamos entender agora, o que é o termo **análise de variância** que vimos no output do SAS. Esta é a parte da análise que vai nos dizer se a correlação entre a variável dependente e independente é ou não estatísticamente significante. Para fazermos uma análise de variância, necessariamente a variável dependente tem que ser contínua com distribuição normal ou que seja de distribuição conhecidamente normal. Já a variável independente pode tanto ser quantitativa como qualitativa. Logo, você deve ter deduzido que a variável independente não precisa ter distribuição normal, pois é possível que ela seja categórica. No nosso exemplo, temos uma variável quantitativa contínua (altura do pai). Vamos primeiro ver este exemplo, e posteriormente faremos um exemplo com variável categórica.

Para entendermos o que é a análise de variância temos que pensar no objetivo claro de nosso teste. Quando pensamos que queremos saber se existe correlação entre altura do filho e do pai, é porque existe uma considerável variação na altura dos alunos. A variabilidade é o que sempre nos intriga e motiva a fazer pesquisas, isto é, é intrigante tentar desvendar a razão pela qual ou as razões pelas quais as pessoas são diferentes.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Se todos os alunos tivessem exatamente a mesma altura, não teríamos pensado em estudar a variabilidade de altura, porque ela não existiria. Então partiremos de uma referência que será a média de altura dos alunos e tentaremos explicar , "porque todos os alunos não tem altura igual a média"? porque suas alturas variam em torno da média? A média aqui, serve de ponto de referência para todos, porque não sabemos se existe uma altura ideal assim a média passa a ser a esperança. Podemos então começar a esboçar uma fórmula para expressar nossa indignação.

$$Y_i - \bar{Y} = ?$$

A fórmula acima, questiona, o quanto cada aluno se distância da média? Porém, se lembrarmos de situações anteriores as somas das variabilidades em torno da média sempre resulta em zero, porque a média é exatamente o ponto equidistantes de todos os valores. Assim, sempre trabalharemos com a soma destes desvios ao quadrado (*sum of squares*).

$$\sum \left(Y_i - \bar{Y}\right)^2 = ?$$

Assim, estamos tentando explicar o porque desta variabilidade, porque todo mundo não tem altura igual à média? Se nada explica esta variabilidade, então a altura de um indivíduo aconteceria ao acaso, o erro seria apenas aleatório.

$$\sum \left(Y_i - \bar{Y}\right)^2 = \varepsilon$$

No entanto, vamos testar se a altura do pai é capaz de explicar alguma coisa deste erro de forma significante. Será que a altura do pai seria nula na explicação da variabilidade?

$$\sum \left(Y_i - \bar{Y}\right)^2 = \beta_1 + \varepsilon$$

Qual será a parte (ou o quanto) dos desvios ao quadrado que seria explicada pelo modelo (neste caso temos no modelo apenas altura do pai-β1). Porque modelo? Porque estamos tentanto construir um modelo (algo simples e reduzido) para explicar a variabilidade da altura dos alunos. Vamos encontrar esta quantificação da parte explicada pela altura do pai abaixo da palavra "source"(fonte) no output do SAS. Uma

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

das fontes é designada como modelo (*model*), e a outa como *error* (resíduo) que se refere a variabilidade não explicada pelo modelo.

Como fazemos todas estas contas? Primeiro devemos calcular a soma dos quadrados dos desvios da variável dependente o que não tem segredo. O mais difícil é calcular a soma dos desvios atribuídos ao modelo (no caso somente altura do pai), e verificar se esta contribuição será significativamente grande. Mas vamos primeiro expressar estas somas em termos:

SST se refere a soma de quadrados total (Sum of Squares Total), SSM é a soma de quadrados devido ao modelo, e SSE soma de quadrados devido ao resíduo.

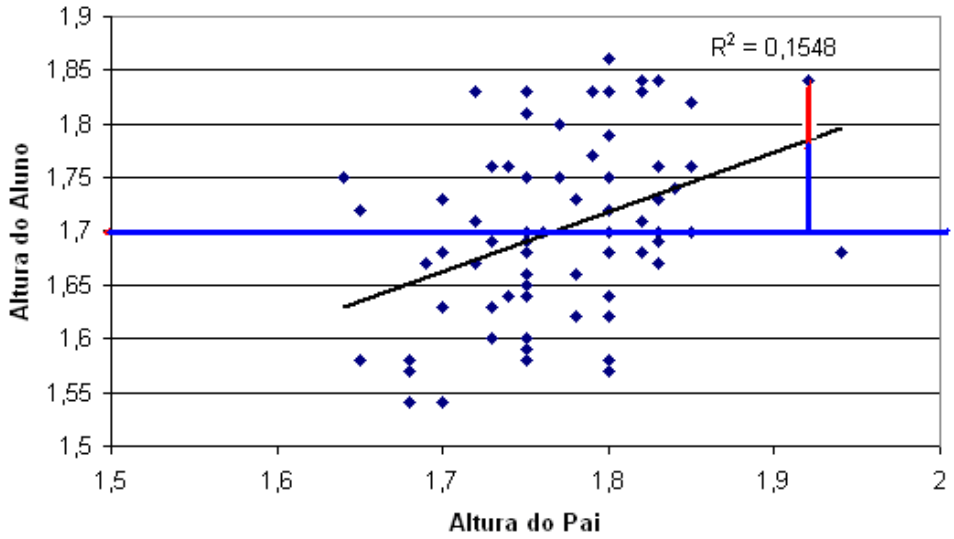
SST = SSM + SSE

Sabemos como calcular o SST, o grande misterio é calcular o SSM. Podemos dizer que SSM = SST – SSE. Mas como calcular o SSE?

Bom, uma vez estimada a reta que melhor explica a relação entre altura do pai e do filho, podemos calcular o que sobra sem explicação. Voltemos ao gráfico. Lembre-se que o modelo está tentando explicar porque as pessoas não tem valor médio de altura. Assim, a distância da reta até a média, é o quanto o modelo conseguiu explicar a variabilidade de altura, o que sobra é o erro.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva



Se sabemos a inclinação da reta, seremos capazes de calcular para cada valor de Y (isto é para cada aluno) um valor esperado dado a inclinação da reta que foi estimada. O que sobra é o erro (resíduo), e ai podemos somar todos os erros. Lembre-se que você já viu algo semelhante (mais ou menos) ao calcular o qui-quadrado. Nós calculavamos os valores esperados e comparavamos o quão distantes estes estavam dos valores observados. A situação aqui é semelhante. Vamos chamar de valor esperado, ou com predição pela reta, o valor $\widehat{Y}$. Assim se subtrairmos cada Yi de $\widehat{Y}$ teremos a somatória dos erros. Já que estamos trabalhando no "mundo" dos quadrados, elevaremos ao quadrado. Logo:

$$SSE = \sum_{i=1}^{n=k} (Y_i - \widehat{Y}_i)^2$$

Portanto, precisamos estimar a inclinação da reta e calcular os valores esperados, para estimar os resíduos. Uma vez feito isso, o que resta de vai ser a soma dos quadrados atribuídos ao modelo. Ai teremos que testar se esta soma de quadrados atribuido ao modelo, será ou não significante. Isto é, será que adicionar altura do pai explicará alguma coisa da variação de altura dos alunos?

Teremos que construir um teste de hipóteses para β1.

Ho : β1 = 0

Ha : β1 $\neq 0$

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Algo interessante, mas óbvio, é que se o modelo não tiver variável para predizer, teremos apenas o βo que será igual a média de y. Tente rodar este modelo no SAS com o seguinte comando:

```
proc reg;
model altaluno = ; run;
```

Você verá que o valor de Bo será igual a média de altura de 1.701.

```
Model: MODEL1
                          Dependent Variable: altaluno altaluno

                         Number of Observations Read          72
                         Number of Observations Used          72


                                  Analysis of Variance

                                            Sum of           Mean
        Source                   DF        Squares         Square    F Value    Pr > F

        Model                     0              0              .        .       .
        Error                    71        0.49853        0.00702
        Corrected Total          71        0.49853


                     Root MSE              0.08379    R-Square     0.0000
                     Dependent Mean        1.70153    Adj R-Sq     0.0000
                     Coeff Var             4.92468


                                  Parameter Estimates

                                     Parameter       Standard
      Variable     Label        DF     Estimate          Error    t Value    Pr > |t|

      Intercept    Intercept     1      1.70153        0.00988     172.30      <.0001
```

Este modelo que acabamos de rodar, sem nenhuma variável independente é chamado de **modelo reduzido**, que terá a SST total. Em relação a este modelo compararemos o nosso modelo a ser testados que tem a variável altura do pai.

È claro que existem fórmulas para se calcular as somas dos quadrados o que veremos em breve. Na verdade como falamos anteriormente que o SSM é calculado indiretamente depois que se calcula o SSE, então o teste estatístico será realizado comparando o SSE do modelo completo (com altura do pai), que chamaremos de SSE(f) (***f*** *de "full" mode*l) em relação ao SSE do modelo reduzido – SSE(r). Se esta diferença for pequena aceitamos Ho, se for grande rejeitamos. Este teste estatístico é feito utilizando uma distribuição que é diferente da distribuição normal que você conhece até agora, e

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

diferente da distribuição t ou qui-quadrado. Vamos utilizar uma distribuição chamada **distribuição F** (de Fisher). Esta é uma distribuição resultante da razão de duas variâncias . No caso, estamos lidando com soma de quadrados de desvios, logo é uma variância. A fórmula é a seguinte.

$$F^* = \frac{SSE(R) - SSE(F)}{df_r - df_f} \div \frac{SSE(F)}{df_d}$$

O que esta fórmula nos dá na verdade é a diferença entre o reduzido e o completo que seria igual a SS do modelo, dividido pelo SSE do modelo completo. Mas note que tudo é dividido pelos graus de liberdade (df degrees of freedom). Podemos simplificar esta fórmula assim:

$$F^* = \frac{SSM}{1} \div \frac{SSE(F)}{n-2}$$

O SSM e SSE divididos pelos graus de liberadade são chamados respectivamente de médias de soma dos quadrados do modelo e de resíduo. No output do SAS esta sob a expressão "*mean square".*

Note que esta fórmula é apenas para quando tivermos uma só variável no modelo que seja contínua ou categórica dicotômica. Se o modelo for representado por duas variáveis contínuas o SSM será dividido por 2 graus de liberade. Porque dois? Porque cada variável contínua corresponde a perda de um grau de liberdade.

O valor de F resultante dos cálculos acima, deverá ser procurado numa tabela específica para distribuição F. Nesta tabela devemos procurar especificamente o valor de *p* específica para 1 e 70 graus de liberdade. Com isso encontraremos qual a probabilidade de β1 ser igual a zero. Se pequena a probabiliade, então rejeitamos H0 e aceitamos Ha.

Se você quiser verificar como isso funciona, utilize o Excel fazendo conta por conta, você verá que somando os desvios em relação a média e elevando ao quadrado terá o valor de 0,49. Com o valor de beta que calculamos no gráfico você pode estimar o valor predito de Y, e calcular a soma dos desvios dos erros e assim por diante. Pelo menos para a construção da tabela de análise de variância de uma variável dependente você pode fazer isso com cálculos simples.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Uma medida que nos ajuda a interpretar nossa estatísticas é o **Coeficiente de Determinação.** Este coeficiente, nos diz o quanto da variação total é explicada pela variação da variável indepente (ou das variáveis indepentes) no modelo. Parece ser fácil se a variação total é o SST, então é só dividir o SSM pelo SST. Sim, é simples.

$$r^2 = \frac{SSM}{SST} \quad \textbf{ou ainda} \quad r^2 = 1 - \frac{SSE}{SST}$$

Note que a segunda fórmula de 1 menos a razão entre o SSE e SST, é simplesmente o valor 1 menos o complemento que não é explicado.

Lembre-se que ja falamos do coeficiente de correlação antes, que era denominado pela letra *r* . Calculando-se o coeficiente de determinação é só tirar a raiz quadrada que vamos ter o coeficiente de correlação, mas devemos entender a correlação para atribuir-lhe sinal de negativo ou positivo, pois o coeficiente de determinação estaria elevado ao quadrado.

Embora sejam úteis os coeficientes de correlação e determinação são passíveis de interpretações erradas. Dentre elas :

1. Um alto coeficiente de correlação leva a uma alta predição, nem sempre depende de cada caso. Um r de 0.91 equivale a um modelo que explica 0.82 % da variabilidade e vai depender da precisão que necessitamos para que o modelo seja ou não útil.
2. Um alto coeficiente de correlação não significa que o modelo de regressão é o melhor para os dados. Podemos ter modelos que não são lineares, mas que fornecem altos coeficientes de correlação. Lembre-se que precisamos ver os gráficos de dispersão.
3. Da mesma forma um coeficiente de correlação baixo não significa necessáriamente que x não explique o y mas apenas que o modelo linear utilizado não é o adequado.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Algo que você pode ter imaginado até agora, é como seria a fórmula para calcular SSE, betas etc de forma mais direta. Se para calcular o SSE tivemos que utilizar o valor de predição baseado em β1 (inclinação da reta) então todos estes etimadores estão interligados. Sim tudo é interligado.

$$r = \left[\frac{\sum (X_i - \bar{X}) \sum (Y_i - \bar{Y})}{\left[\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2 \right]^{1/2}} \right]$$

Ou ainda veja a semelhança entr r e b1.

$$b1 = \left[\frac{\sum (X_i - \bar{X}) \sum (Y_i - \bar{Y})}{\left[\sum (X_i - \bar{X})^2 \right]} \right]$$

Esta fórmula de b1 nos diz qual a razão da covariâncias de x e y com relação a variância apenas de x , que é a variável independente. Na fórmula do r , era a razão da mesma covariância em relação a variabilidade total de x e de y.

$$b_1 = \left[\frac{\sum (Y_i - \bar{Y})^2}{\sum (X_i - \bar{X})^2} \right]^{1/2} * r$$

logo,

$$b_1 = \left(\frac{Sy}{Sx} \right) * r$$

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Lembra-se como calcular a inclinação da reta? Seria a distância no gráfico da variação no triângulo. Isso se repete mais ou menos nesta fórmula, poém com os desvios em relação a média de cada variável. No numerador temos a variabilidade de Y e no numerador a de X. Essa some é toda ao quadrado, mas você nota que o b1 se origina da raiz quadrada (elevado a ½, é o mesmo que tirar raiz quadrada).

Podemos também dizer que no numerado temos a variância em relação a Y, e no numerador a variância em relação a X. A única coisa que você deve estar estranhando é que não dividimos pelo número de indivíduos. Porém como o número será o mesmo, não é necessário. Assim, β1 será igual ao desvio padrão de y divido pelo desvio de x vezes o coeficiente de correlação.

Na análise de regressão linear que acabamos de ver, vimos que o output do SAS teve dois tipos de função. Primeiro fez um teste para ver o quanto a altura do pai explicava a variabilidade de altura dos alunos. Nesta parte podemos dizer que fizemos uma análise de variância. Depois o SAS cálculou a inclinação da reta para que possamos calcular valores preditivos. Neste momento testamos se β1 seria ou não igual a zero, e qual a magnitude desta inclinação. Embora algumas pessoas considerem processos distintos a análise de regressão com estimação dos coeficientes e a análise de variância, estas análises tem a mesma origem. Análise de regressão ou análise de variância são da familia dos modelos lineares generalizados (generalized linear models).

No nosso exemplo usamos uma variável contínua como variável independente, o que é o que se utiliza classicamente na análise de regressão linear. No entanto, a variável independente poderia ser uma variável categórica, e os princípios seriam os mesmos. No entanto, quando a variavel independente é categorica, costuma-se denominar de análise de variância, principalmente porque na maioria das vezes o pesquisador não está interessado em estimar a inclinação da reta, apenas de saber se a variável independente é importante para explicar a variabilidade encontrada na variável dependente.

Embora pareça tudo muito fácil, análise de regressão não é tão simples, e este curso não pretende chegar a muitos detalhes. Sempre que for fazer uma analise de verdade procure um estatístico. Você pode até brincar com os dados, para entender

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

melhor como as variáveis se relacionam, mas não se esqueça de procurar um estatístico para melhor orientá-lo.

Assim como no teste t, temos pressupostos para ser fazer uma análise de regressão linear. Entre os pressupostos temos

Pressupostos da análise de regressão linear

1. Indepêndencia dos dados
2. A variável dependente precisa ter uma distribuição normal, ou pelo menos vir de uma população com conhecida distribuição normal.
3. Homoscedasticidade – que significa que a variância ao redor de cada valor de x deve ser semelhante.
4. Os resíduos devem se distribuir aleatóriamente ao redor do zero.

Você deve estar percebendo algumas semelhanças com o teste t de comparação de médias. A independencia é comum a quase todos os testes. A questão da distribuição normal também é semelhante. Ainda no teste t tinhamos um pressuposto de que a variância seria igual para os dois grupos a serem comparados. Na análise de regressão não temos grupos e sim valores contínuos, mas existe um pressuposto semelhante a este , pois para cada ponto da variável x, teremos uma distribuição de Y, que deverá ter variâncias iguais, cujo errro deve ser igual a zero, e ainda os erros de cada indivíduo não deve ser correlacionado aos erros dos demais (isto implica na questao da independência dos dados). Esta propriedade aqui recebe o nome de **homoscedasticidade**.

Em alguns livros ressalta-se que ter distribuição normal da variável dependente não seria um pressuposto, e isso é verdadeiro em parte, o principal é sair de uma distribuição supostamente normal em que a média seja um ponto de referência desejável. Se tivermos uma distribuição exponetial cuja média não traz muito significado, nao faz sentido ter a média como ponto de estudo da variabilidade. E de forma geral, os resíduos não terão também distribuição normal o que é um pressuposto. Se uma distribuição é de Poisson, embora média seja também um dos parâmetros da distribuição a regressão linear não deverá ser aplicada mesmo porque os pressupostos

em relação aos resíduos tendem a não serem satisfeitos. A regressão deverá ser log-linear que acompanhará a distribuição de Poisson.

Percebendo estas semelhanças você deve estar imaginando então que teste t tem a ver com outros tipos de estatísticas, pois é isso mesmo, os princípio são os mesmos.

Uma vez que temos pressupostos, o teste estatístico somente será valido, se estes pressupostos existirem para os dados que estamos analizando, caso contrário teremos uma estatística que não tem validade alguma. Teremos então que verificar a existência destes pressupostos. Alguns pressupostos vamos verificar antes de fazer a análise e outros como distribuição normal de resíduos somente podemos fazer depois de efetuar a estatística. Mas e se você descobre que os pressupostos não existem? Bem, existem inúmeras técnicas para concertar e existem inúmeras outras técnicas estatísticas. È assim que funciona a estatística, um trabalho arduo, que precisa de bastante conhecimento.

Como podemos analisar estes pressupostos? A independência você avalia sabendo como os dados foram coletados e a que se referem. Por exemplo num estudo de comparação de tempo de sobrevida de restaurações de resina, onde cada indivíduo contribui com 4 dentes, os dentes não podem ser considerados unidades independentes. Pois os dentes estão agrupados em cada indivíduo. Um outro exemplo seria comparar média de perda de inserção de tecido periodontal antes e depois de um estudo num mesmo indivíduo. As medidas antes e depois estão dentro de um mesmo indivíduo. Existem técnicas especiais para se resolver estes problemas, mas não na análise de variância comum nem na análise de regressão.

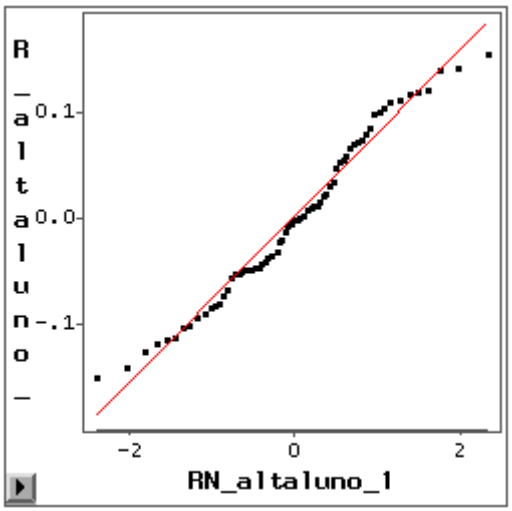
A distribuição normal de Y pode ser checada antes de começar a fazer o estudo. Já a variância comum a todos os valores de x deve ser checada no que chamamos de análise de resíduos ou análises de diagnóstico. Quando “supostamente” acabamos uma análise apertando os botões do computador, na verdade a análise está somente começando, pois se algum pressuposto for violado, o βeta estimado, e os resultados não terão validade alguma.

Começamos fazendo um gráfico dos resíduos. O SAS , na função de análise interativa fornece o gráfico automaticamente. No gráfico abaixo, vemos os resíduos cuja

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

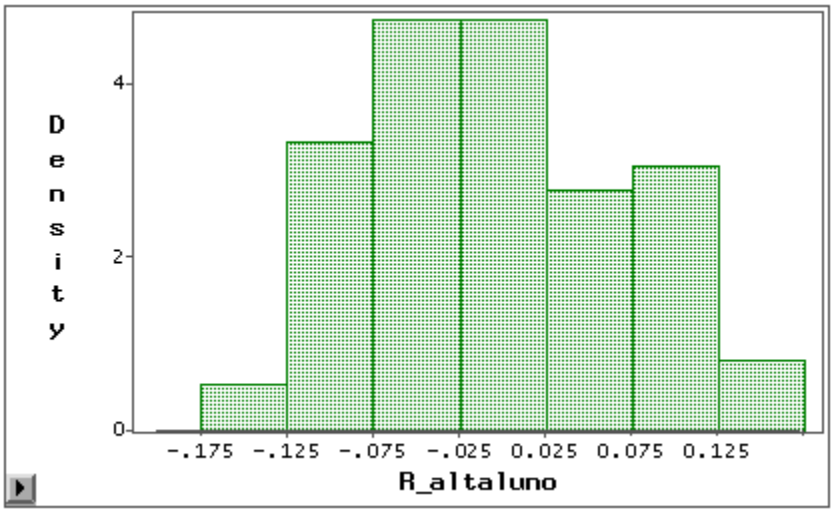
média é zero, distribuidos de acordo com o valor preditivo de Y. Se você fo curioso e abrir o arquivo de dados do SAS verá que ao pedir uma análise de regressão o SAS gera variáveis chamadas P_altaluno e R_altaluno que são o valor predito com a fórmula gerada e o valor do resíduo para cada indivíduo.

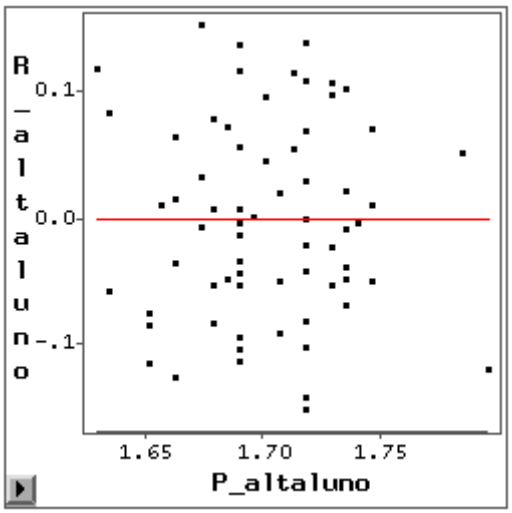
Também, os resíduos devem ter distribuição normal com média zero. Podemos pedir ao SAS que faça um gráfico para verificar se os resíduos tem distribuição normal. Este gráfico é chamado de Q-Q plot (plot de normalidade).



Este é um QQ plot. Tanto mais será normal, quanto a linha de pontinhos se aproximar da linha reta em 45 graus. Vemos que os resíduos mais ou menos são normais. Se quisermos podemos pedir ao SAS que faca um histograma e alguns testes de normalidade.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

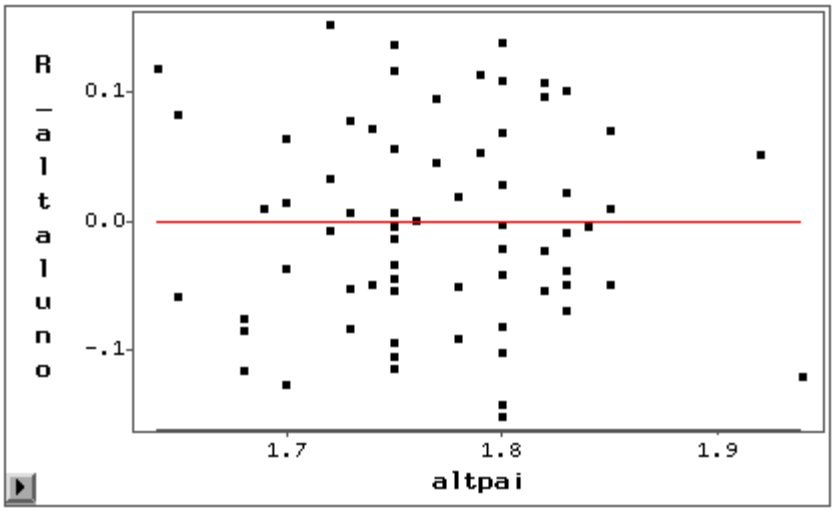




Notamos neste gráfico que mais ou menos os resíduos estão distribuídos aleatóriamente ao redor do zero. Temos então a impressao, talvez pelos dois “outliers”, de que a variância diminui conforme a altura vai se tornando maior. Não temos muitos pontos por isso fica difícil ver realmente uma tendência forte de alteração de variância de acordo com o valor de Y. Mas é possível que isso aconteça. Os alunos mais altos, devem vir de famílias mais altas, mas como temos um limite de crescimento na população a variância diminui. Mas o mais interessante seria fazer um gráfico dos resíduos com o valor de X. Isso o SAS não nos da automaticamente, temos que pedir

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

para ele fazer. Se você estiver no SAS interactive, é só pedir para fazer um gráfico de dispersão (scatter plot) e checar.



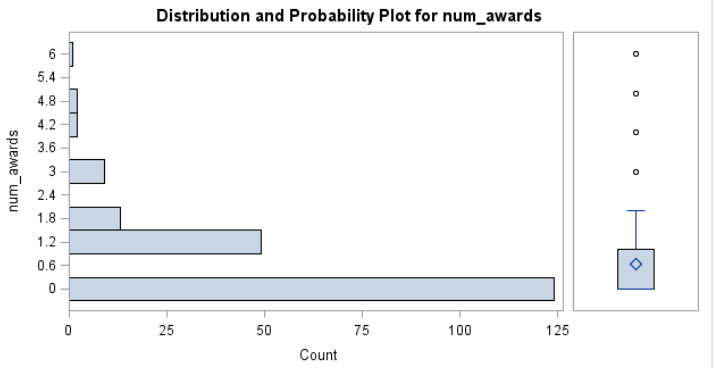
Temos ainda a mesma impressão talvez a variância seja menor conforme tem pessoas mais alta. No entanto, está razoavelmente bom para ser aceito. Talvez se tivessemos uma população maior isso seria mais nítido. Se ficasse claro esta variância não constante, deveríamos adotar outro tipo de estatística, pois não poderiamos acreditar nso resultados dados por esta estatística.

Vamos ficar por aqui em termos de diagnóstico, lembrando que sempre formos fazer uma análise de regressão, devemos no mínimo verificar como os resíduos se distribuem ao redor da média, e procurar um estatístico experiente.

Retornando a necessidade de termos a variável desfecho com distribuição normal ou aproximadamente normal ou ainda pelo menos que venha de uma distribuição conhecidamente normal. E não se levar pela questão dos pressupostos não mencionarem a distribuição de Y e achar que tanto faz qual a distribuição que tudo bem. Na verdade ela irá refletir na distribuição dos residuos ou aleatoriedade dos resíduos. Vejam um exemplo abaixo. O banco de dados tirei um exemplo do site : **stats.idre.ucla.edu/stat/data/poisson_sim.csv,** apenas para servir de exemplo. Note a distribuição de eventos raros, como muitos zeros , evidenciado no histograma e vários supostamente outliers no boxplot. São valores extremos porque a distribuição é de eventos raros e não uma distribuição normal. No quadro seguinte de resumo de resíduos e diagnosticos

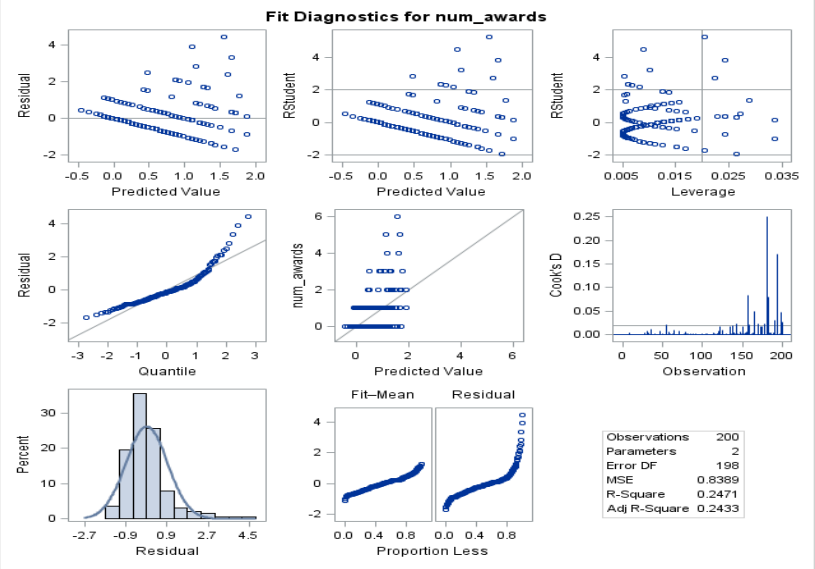
mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

da regressão linear note que embora a distribuição do resíduo observado pelo histograma pareça não muito não normal, notamos que a distribuição aleatória do mesmo não se mantem. O que quer dizer isso, desde o início a distribuição não era normal, e portanto a regressão linear que assume distribuição normal não se aplica. Eu nem começaria a fazer a analise com regressão linear usual. Uma alternativa usar o logarítmo, mas em geral não funciona bem especialmente porque log de zero não é definido e esses valores são perdidos. Portanto, ainda considero que devemos avaliar a distribuição do desfecho e então decidir qual estatística ou regressão utilizar. Existem modelos que levam em consideração distribuições específicas como por exemplo regressão de Poisson (o tema é complexo) que utiliza modelos lineares generalizados que não se basea em ordinario quadrados mínimos, mas utiliza maxima verossimilhança, e nestes modelos onde uma distribuição específica é definida, regressão binomial negativa para dados com superdispersão (muitos zeros onde a variância é maior do que a média). Portanto, quando fugimos da distribuição normal do desfecho (Y) temos opções mas precisam ser estudados para serem utilizados adequadamente.



mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva



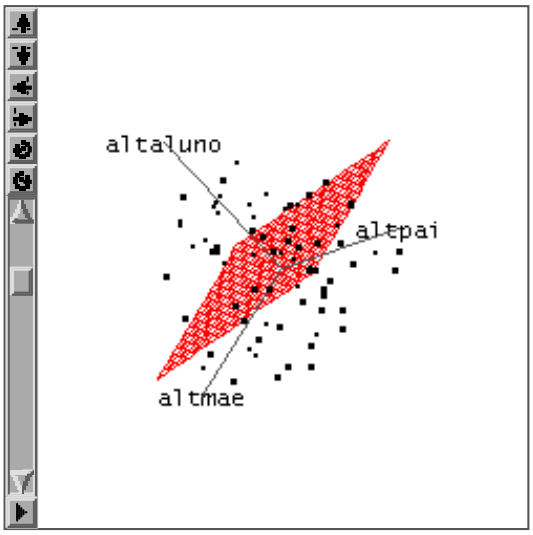
Análise de mais de uma variável (análise multivariável)

Vamos agora adicionar a variável idade da mãe e verificar como fica nossa análise. Apenas como ilustração, antes tinhamos dois eixos (altura do aluno e altura do pai) agora teremos um terceiro eixo que se encaixa tridimensionalmente na análise, e os pontos ficam flutuando no espaço. Mesmo assim teremos que calcular a melhor reta, mas agora na verdade falamos de retas, e não uma só. Não é intensão ensinar analise multivariada, pois existem muitas particularidades de construção de modelos e diagnósticos, mas vamos adicionar mais variáveis explicativas a altura do aluno apenas para que tenham idea do utilidade de uma análise multivariada, ou como alguns dizem multivariável.

Veja o gráfico do SAS, numa tentativa de nos mostrar uma análise tridimencional. Esta função não existe mais no SAS infelizmente, mas era bem didático, pois podiamos brincar e girar o gráfico observando os valores no espaço.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva



Agora vamos ao output do SAS.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.12111242	0.06055621	11.07	<.0001
Error	69	0.37741953	0.00546985		
Corrected Total	71	0.49853194			

R-Square	Coeff Var	Root MSE	altaluno Mean
0.242938	4.346589	0.073958	1.701528

Source	DF	Type I SS	Mean Square	F Value	Pr > F
altpai	1	0.07716138	0.07716138	14.11	0.0004
altmae	1	0.04395103	0.04395103	8.04	0.0060

Source	DF	Type III SS	Mean Square	F Value	Pr > F
altpai	1	0.06814977	0.06814977	12.46	0.0007
altmae	1	0.04395103	0.04395103	8.04	0.0060

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.0581183306	0.35034390	0.17	0.8687
altpai	0.5260612026	0.14903619	3.53	0.0007
altmae	0.4345618159	0.15330441	2.83	0.0060

Ao adicionar a altura da mãe no modelo, vemos que o SST continua o mesmo, é claro, ele não deverá mudar. O que mudou foi o valor do SSM que era de e passou a ser 0.07716 e passou para 0.121. Porém o valor de F não subiu muito, passou de 12.82 para 14,11. Isso quer dizer que altura da mãe não é tão importante, ou então que altura do

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

pai e mãe são muito correlacionadas, e uma tira valor explanatório da outra. Uma falha de nossa análise é que deveriamos ter estudado também a associação bivariada da altura da mãe com altura do aluno e não fizemos isso. Eu fiz isso de propósito para lembrar que temos que ser cautelosos e entender bem os dados antes de começar uma análise multivariada, isso é com várias variáveis independentes.

Mesmo assim, vamos continuar com a descrição do output. O r^2 subiu de 15,48 para 24,29. Ja o beta de altura do pai no modelo anterior era de 0.56, e passou para 0,52. Não alterou muito, mas o modelo não melhorou de mais, melhorou, mas poderia ter melhorado mais. Vamos ver como se comporta o modelo somente com a altura da mãe.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.12111242	0.06055621	11.07	<.0001
Error	69	0.37741953	0.00546985		
Corrected Total	71	0.49853194			

The GLM Procedure

Dependent Variable: altaluno altaluno

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.05296265	0.05296265	8.32	0.0052
Error	70	0.44556929	0.00636528		
Corrected Total	71	0.49853194			

R-Square	Coeff Var	Root MSE	altaluno Mean
0.106237	4.688885	0.079783	1.701528

Source	DF	Type I SS	Mean Square	F Value	Pr > F
altmae	1	0.05296265	0.05296265	8.32	0.0052

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Source	DF	Type III SS	Mean Square	F Value	Pr > F
altmae	1	0.05296265	0.05296265	8.32	0.0052

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.9216451588	0.27053001	3.41	0.0011
altmae	0.4756590306	0.16489956	2.88	0.0052

Realmente a altura do pai, parece ser mais importante para determinar a altura do filho do que a altura da mãe. Vemos aqui que o modelo do pai sozinho explicava 15,48 % da variação de altura do filho, e a altura da mãe sozinha estima cerca de 10,62%. Porém os dois juntos elevam a predição não de forma somatória mais para um meio termo ao redor de 24,29%. Algo importante notar é que a altura da mãe não havia provocado grandes mudanças do coeficiente de β do pai, e agora vemos que isso não aconteceu para com o coeficiente da mãe que passou de 0.47 para 0.43. Parece, então que no todo podemos ficar no modelo com altura da mãe e do pai, pois os dois ajudam a predizer a altura do filho.

Falta ainda pensar sobre a questão do sexo, pois temos meninas e meninos juntos. Podemos então adicionar no model a variável sexo, e ver o que acontece. Mas a variável sexo é categórica tem algum problema? Não, sem problemas apenas temos que interpretar de maneira adequada.

The GLM Procedure

Dependent Variable: altaluno altaluno

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.37851580	0.12617193	71.49	<.0001
Error	68	0.12001615	0.00176494		
Corrected Total	71	0.49853194			

R-Square	Coeff Var	Root MSE	altaluno Mean

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

0.759261 2.469030 0.042011 1.701528

Source	DF	Type I SS	Mean Square	F Value	Pr > F
altpai	1	0.07716138	0.07716138	43.72	<.0001
altmae	1	0.04395103	0.04395103	24.90	<.0001
Sexo	1	0.25740338	0.25740338	145.84	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
altpai	1	0.03670530	0.03670530	20.80	<.0001
altmae	1	0.03562815	0.03562815	20.19	<.0001
Sexo	1	0.25740338	0.25740338	145.84	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.5642177121	0.20337346	2.77	0.0071
altpai	0.3895009436	0.08541015	4.56	<.0001
altmae	0.3915849788	0.08715549	4.49	<.0001
Sexo	-.1225140108	0.01014481	-12.08	<.0001

Vejam que o modelo preditivo aumenta consideravelmente. Isso porque antes tinham uma salada de meninos e meninas, embora a salada esteja ainda aqui, a variável sexo, é fortemente associada a altura. Talvez estejamos fazendo besteira e deveríamos analisar meninos e meninas separadamente. Mas depende, se nosso objetivo é verificar se altura de pai e mãe afetam a altura dos filhos, a não ser que estas alturas de pai e mãe afetem de maneira diferente meninos e meninas não temos razão para estudar meninos e meninas separadamente. Vamos estudar separadamente apenas e as inclinações das retas forem diferentes para meninos e meninas. Vamos testar isso daqui a pouco. Por enquanto vamos voltar ao nosso output. Com a adição de sexo, o modelo passa a explicar 75,93% da variabilidade das alturas. Mas o mais importante é verificar o que aconteceu com os coeficientes (β1 e β2). Agora , o coeficiente do pai que era de 0.49 para 0.38 e da mãe de 0.42 para 0.39. A diferença para o pai foi maior mas nem tanto.

Para saber se as retas são diferentes para meninos e meninas, precisamos construir um modelo com interação desta maneira no SAS.

Pode-se tanto usar proc glm ou proc reg.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva


```
proc glm data = alt.altura2;
        model altaluno =  altpai altmae sexo sexo*altmae sexo*altpai; run;
```

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.5395280775	0.67242220	0.80	0.4252
altpai	0.0292003774	0.30312834	0.10	0.9236
altmae	0.7962817888	0.34716541	2.29	0.0250
Sexo	-.1258314104	0.40707548	-0.31	0.7582
altmae*Sexo	-.2275625639	0.19556387	-1.16	0.2488
altpai*Sexo	0.2124509134	0.17893937	1.19	0.2394

Note que a interação altmae* sexo ou altpai*sexo não foram significantes, portanto, não deve ter interação. Para tirar a prova dos nove, podemos construir modelos separados para meninos e meninas e veremos que nada deve mudar muito quanto a inclinação das retas.

O modelo a seguir foi construído apenas para os meninos. Veja que antes o modelo explicava aproximadamente 15 % da variabilidade de Y, quando adicionado a mãe subiu para cerca de 20%. Agora que fizemos um modelo apenas com β1 e β2 para meninos o modelo explica 45%. Isso porque o que atrapalhava era a mistura com as meninas. O SST eram calculados com base na média de altura das meninas que eram bem mais baixas. Mas o mais importante agora é verificar os coeficientes para os Betas.

The GLM Procedure

Dependent Variable: altaluno altaluno

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.03546441	0.01773221	11.40	0.0003
Error	27	0.04201559	0.00155613		
Corrected Total	29	0.07748000			

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

R-Square	Coeff Var	Root MSE	altaluno Mean
0.457723	2.218664	0.039448	1.778000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
altpai	1	0.01475086	0.01475086	9.48	0.0047
altmae	1	0.02071355	0.02071355	13.31	0.0011

Source	DF	Type III SS	Mean Square	F Value	Pr > F
altpai	1	0.00513964	0.00513964	3.30	0.0803
altmae	1	0.02071355	0.02071355	13.31	0.0011

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.4136966671	0.29084943	1.42	0.1664
altpai	0.2416512908	0.13296766	1.82	0.0803
altmae	0.5687192249	0.15588127	3.65	0.0011

Para predizer altura dos meninos, o beta do pai passou de 0.46 para 0.24 e da mãe subiu para 0.56. Antes de tirar conclusões vamos ver o que aconteceu com as meninas.

The GLM Procedure

Dependent Variable: altaluno altaluno

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.04639054	0.02319527	12.24	<.0001
Error	39	0.07390708	0.00189505		
Corrected Total	41	0.12029762			

R-Square	Coeff Var	Root MSE	altaluno Mean
0.385631	2.643274	0.043532	1.646905

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Source	DF	Type I SS	Mean Square	F Value	Pr > F
altpai	1	0.02749566	0.02749566	14.51	0.0005
altmae	1	0.01889487	0.01889487	9.97	0.0031

Source	DF	Type III SS	Mean Square	F Value	Pr > F
altpai	1	0.03002337	0.03002337	15.84	0.0003
altmae	1	0.01889487	0.01889487	9.97	0.0031

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.2878652567	0.27527071	1.05	0.3021
altpai	0.4541022043	0.11408659	3.98	0.0003
altmae	0.3411566610	0.10804199	3.16	0.0031

Para as meninas, o coeficiente de correlação ficou um pouco menor do que para os meninos,38,56%. Mas quando olhamos os **β ,** o coeficiente para altura do pai é mais ou menos a mesma e da mãe é um pouco menor. Bom, chegamos a um impasse, será que realmente deveremos reportar os valores separados? Vamos colocar em foco todos os dados que temos até agora, e também fazer modelos com altura de pai e mãe separados para meninos e meninas. Depois decidiremos.

	r^2	β1	EP	*p*	β2	EP	*p*
Meninos e Meninas							
Pai	0,15	0,56	0,15	0,0006			
Mãe	0,10				0,47	0,16	0,0052
Pai e Mãe	0,24	0,53	0,15	0,0007	0,43	0,15	0,0060
Pai e Mãe + Sexo	0,75	0,39	0,08	<0.0001	0,39	0,09	<0.0001
Meninos							
Pai	0,19	0,39	0,15	0,0159			
Mãe	0,39				0,65	0,15	0,0002
Pai e Mãe	0,46	0,24	0,3	0,0803	0,56	0,15	0,0011
Meninas							
Pai	0,22	0,43	0,12	0,0014			
Mãe	0,13				0,31	0,12	0,0162
Pai e Mãe	0,38	0,45	0,11	0,0003	0,34	0,10	0,0031

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

r^2 = coeficiente de determinação, β1 inclinação da reta para o pai, β2 inclinação da reta para altura da mãe, EP = erro padrão.

Ao compararmos todas as estatisticas, realmente vimos que existe uma certa diferença de se avaliar a altura de meninos e meninas separadamente. Porem, temos que comparar os betas dos modelos separados com aqueles obtidos com o modelo em conjunto quando a variável sexo estava no modelo. Para este modelo, mãe e pai tiveram retas bem semelhantes de inclinação de 0.39. Ao separar os modelos, temos que entender também que o número de indivíduos cai. Portanto as variações podem ser mais evidentes. De qualquer forma, parece que altura de pai e mãe consegue predizer melhor a altura do menino do que das meninas.Embora os valores de p para os modelos dos meninos seja menor, o mais importante é olhar os valores dos betas e do coeficiente de correlação. Embora altura do pai tenha valor de p não significante de 0.0803 no modelos dos meninos, a inclinação da reta é de 0.46 maior que 0.38 das meninas. Portanto os modelos devem ser mais ou menos equivalentes.

Uma vez que os betas do modelo com meninos e meninas não são muito diferentes dos modelos separados para meninos e meninas, podemos “acreditar” que o teste de interação era apropriado, pois ele não conseguiu evidenciar interação entre o sexo e altura de pai ou mãe. Isto é ser menino ou menina não levou a retas com inclinações diferentes. Lembre-se de um gráfico acima onde traçamos duas retas, uma para meninos e outra para meninas, e as reta eram paralelas, não tinham inclinação diferentes. È isso que chamamos de teste de interação:verificar se as inclinações são estatisticamente diferentes.

Temos um grande problema neste “estudo”, que é a falta de uma população definida. Os alunos da sala de odontologia de 2008 não constituem uma população definida, e ainda nem todos os alunos colocaram seus dados no banco. Estas diferenças podem estar sendo influenciadas pela amostra inadequada. Por isso devemos sempre coletar os dados de maneira adequada, planejar a amostragem, o tipo de estudo (que veremos na epidemiologia), para que possamos acreditar nos resultados que estamos obtendo. Será que a diferença observada entre influencia de altura para meninos e meninas é verdadeira, ou somente resultado de amostra mal feita sem represetnar uma população definida? Não sabemos!

Ficou faltando discutir como fica a interpretação do β3 que se referia ao sexo, no modelo

$$Y = \beta o + \beta 1 + \beta 2 + \beta 3.$$

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Para a variável sexo que tem dois valores 1 para meninos e 2 para meninas a leitura é basicamente a mesma que feita para uma variável contínua. Neste exemplo o valor de beta para sexo foi de - 0.12. Isso significa que não ser menino (logo ser menina) corresponde a um decrescimo de 0.12 cm na altura depois de ajustado pela altura do pai e da mãe.

Ainda tem mais um pedado da análise de variância que não exclarecida. No output do Sas temos os Tipo I SS e Tipo III SS. Isto é Sum of Squares Tipo I e o Tipo III. Estes SS mostram a contribuição independente de cada variável no modelo. No caso do Tipo I a primeira variável tem sua contribuição sozinha como se o modelo tivesse apenas a altura do pai. Porém, a altura da mãe que vem a seguir já é levando-se em consideração a existência da altura do pai no modelo. Se você verificar o output do SAS com as 3 variáveis (altura do pai, altura da mãe e sexo), verá que a altura do pai é a original, a altura da mãe é aquela onde ja se encontrava a altura do pai, e o do sexo é a contribuição de sexo uma vez tendo no modelo altura do pai e da mãe.

Para o tipo III, temos a contribuição de cada variável dada a outra variável no modelo, por isso a contribuição de sexo é a mesma para SS Tipo I e Tipo III.

ANOVA – Teste de Análise de Variância

O que nós vimos até agora não deixa de ser análise de variância, vocês viram que nos outputs do SAS estava escrito análise de variância. Classicamente os livros separam a análise de regressão linear do que chamam da ANOVA (ou seja análise de variância), isso porque para não misturar um objetivo com o outros passou-se a chamar de análise de regressão linear quando tanto a variàvel dependente como a independente de interesse são contínuas, e quando a variável de dependente é contínua e a variável independente é categórica. No entanto, regressão ou ANOVA são da mmesma família, a grande diferença é que quando se tem as duas variáveis (dependente e independente) contínuas normalmente estamos interessadas na estimação da reta, dos valores de

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

betas. Já na análise de variância a reta não nos interessa. Nos interessa apenas se as médias são difernetes entre os grupos.

É muito comum utilizar análise de variância em estudos de laboratório de ratinhos e em estudos de corpos de prova em Materiais Dentários. Isso porque são comuns estudos do tipo comparação de resistência entre dois ou mais materiais. Efeito de uma droga específica na pressão arterial de ratinhos etc. Nestes exemplos o que temos são comparações de médias entre dois ou mais grupos. Para estes tipos de estudos não nos interessa a inclinação da reta que uniria este grupos, nos interessa apenas saber se os grupos têm médias semelhantes ou não.

Vimos que o teste t serve exatamente para compara dois grupos mas não mais do que isso. Vimos também que o test t equivale a análise de variância de dois grupos. Mas a análise de variância serve também para compara mais de dois grupos. Podemos fazer testes de comparação de mais de dois grupos por meio de simulações para calcular os intervalos de confiança para cada grupo, ou podemos utilizar a estatística mais comum que seria a chamada Análise de Variância (ANOVA). Na verdade, já vimos o que é ANOVA, vamos apenas agora ver a mesma estatística em contextos diferentes que se assemelham mais ao que vocês terão que fazer na disciplina de Materiais Dentários.

A grande diferença entre análise de regressão e análise de variância é conceitual. Na análise de variância não estamos interessados em saber a inclinação da reta. O foco maior é no teste de hipoteses de associação ou não. A maneira como se encara a análise de variância também é um pouco diferente. Classicamente anova é realizado em experimentos em que se decide previamente quem serão os grupos a serem comparados. Por exemplo, num estudo de laboratório separam-se dois grupos de animais que iram receber uma droga e o outro grupo irá receber um placebo. A maneira de se pensar nos erros e desvios é a seguinte. Imagine que existe sempre uma variabilidade de resposta entre os indivídios independente da droga que foi utilizada. Se a droga for eficaz, além da variabilidade indivídual vamos ter uma variabilidade recorrente da droga. Assim, aquela fórmula que montamos para análise de regressao de $Y - \overline{Y}$, passa a ser diferente. Conceitualmente teremos o seguinte. A variabilidade entre

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

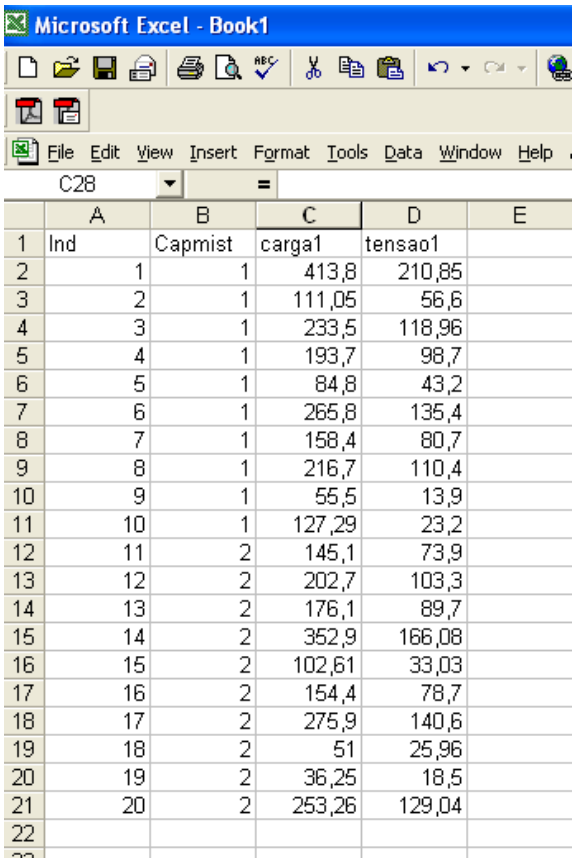
cada indivíduo de cada grupo até a média geral. No final a conta vai ser a mesma, mas, conceitualmente isso é diferente.

Condições ideais para a ANOVA. Os pressupostos para a ANOVA basicamente são os mesmo para a regressão linear, se o estudo é um experimento de materiais dentários ou de ratinhos ou qualquer outro falamos a mesma coisa, só que aplicada para o experiemento. Por exemplo, aleatoriedade pe um dos pressupostos isto é

1. O tratamento deve ser aleatoriamente distribuido entre os animais ou corpos de prova.
2. A ordem de processamento das unidades do experiemento tem que ser de forma aleatória. Pois a ordem em que se faz por exemplo uma cirurgia em ratinhos pode envisar os resultados, assim como a ordem de leitura de material também. Por exemplo, se estamos medindo a tensao de amalgama manipulado com amalgamador e manual, temos que determinar aleatóriamente a ordemd e leitura dos corpos de prova.
3. Independencia dos resíduos em cada um dos grupos.
    a. A falta de independencia pode acontecer
    b. Os residuos podem ser tornar dependentes se a sequencia de processamento e leitra forem não aleatorios
    c. Quado duas ou mais leituras sao realizadas num mesmo corpo de prova ou unidade de experiemtnacao.
4. Homogeneidade das variâncias residuais entre grupos
    a. Testes de levine ou Barlett e outros sao apropriados
5. Residuos normais entre os grupos

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Nas aulas de Materiais Dentários, o que os alunos mais enfrentam dificuldade é em montar os dados no excel. Não é tão difícil quanto se imagina, apenas você precisa pensar um pouquinho. Para alguns a entrada de dados no Excel é obvia, mas para outros alunos não. Primeiro existem regras gerais , sempre coloque em cada linha um corpo de prova diferente. Se foram analizados 20 corpos de prova divididos em dois grupos de tipo de amalgama para os quais se mediu tração teremos então 3 colunas. Uma para identificar o corpo de prova, outra coluna para o tipo de almagama, e outra onde se colocará o resultado da força. Ex:



	A	B	C	D	E
1	Ind	Capmist	carga1	tensao1	
2	1	1	413,8	210,85	
3	2	1	111,05	56,6	
4	3	1	233,5	118,96	
5	4	1	193,7	98,7	
6	5	1	84,8	43,2	
7	6	1	265,8	135,4	
8	7	1	158,4	80,7	
9	8	1	216,7	110,4	
10	9	1	55,5	13,9	
11	10	1	127,29	23,2	
12	11	2	145,1	73,9	
13	12	2	202,7	103,3	
14	13	2	176,1	89,7	
15	14	2	352,9	166,08	
16	15	2	102,61	33,03	
17	16	2	154,4	78,7	
18	17	2	275,9	140,6	
19	18	2	51	25,96	
20	19	2	36,25	18,5	
21	20	2	253,26	129,04	
22					

Vejam como é simples. A primeira linha será reservada para colocar o nome da variável. Veja que os nomes que colocamos são simples e curtos. Tente não ultrapassar 8 letras. Não coloque acentos, hifens ou espaços nos nomes. A variável Ind aqui, significa indivíduo, mas na verdade temos corpos de prova. De qualquer forma Ind signifca para nós corpo de prova. São 20 corpos de prova e numeramos os mesmos de 1 a 20. Em geral quando trabalhamos com pacientes ou indivíduos de uma população é possível que sejam identificados por números grandes e fora de ordem. Tanto faz se os números

estejam em ordem ou não, o importante é servir como identidade e que não se repitam, cada número deve ser único para cada indivíduo.

Capmist foi o nome dado a variável que identifica se o amalgama foi manipulado numa capsula ou se foi dosado no amalgamador diretamente. Atribuiu-se o número 1 para capsula e 2 para sem capsula. As outras duas variáveis são carga e tensão, que são as variáveis dependentes. Note capmist (tipo de capsula) é a variável independente.

Nunca monte a tabela da seguinte maneira, que é o erro mais comum entre os alunos que me procuram.

Capmist	carga1	Capmist	carga1
1	413,8	2	145,1
1	111,05	2	202,7
1	233,5	2	176,1
1	193,7	2	352,9
1	84,8	2	102,61
1	265,8	2	154,4
1	158,4	2	275,9
1	216,7	2	51
1	55,5	2	36,25
1	127,29	2	253,26

Com este tipo de entrada temos mais de um corpo de prova por linha. Lembre-se que teremos que escrever uma fórmula daquilo que investigamos. Y = Capmist + erro. O computador reconhece apenas uma variável por coluna. Embora a ANOVA não tenha intensão de estimar os betas etc, a fórmula que comanda nosso estudo cotinua a ser o mesmo. Não precisamos estimar o beta para capmist, apenas temos que saber se capmist é importante para predizer a variabilidade encontrada em Y.

Uma vez organizados os seus dados. Você deve importar os dados para o SAS e la proceder suas estatísticas. No nosso exemplo, a pergunta será que a tensão varia se usarmos amalgama em cápsula ou sem cápsula? Deverá ser traduzida como hipóteses:

Ho = não existe diferença entre tensão de amalgama em cápsula e sem capsula.

Ha = as tensões resultantes são diferentes entre amalgama em cápsula e sem cápsula.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Com este teste de hipótese em mente agente já sabe que devemos construir um intervalo de confiança para um grupo e para o outro e compará-los, ou ainda, devemos determinar a variância ao redor do zero que é a falta de diferença. Ainda podemos testar esta hipótese com base na análise de variância, a variância encontrada nos 20 corpos de prova, se deve em parte ou totalmente ao tipo de cápsula.

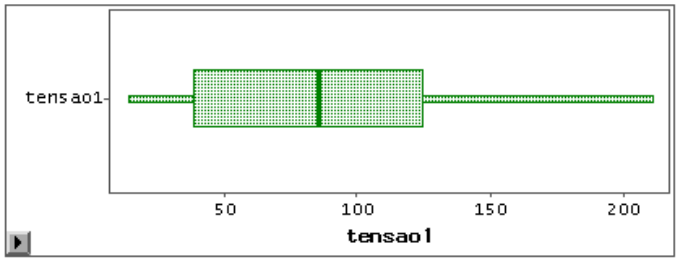
Usando a análise de variância devemos calcular o quanto a tensão de cada corpo de prova esta da média total encontrada e somar todos estes desvios, obtendo o SST. Apartir dai, temos que calcular o quanto deste SST pode ser atribuído as cápsulas e o quanto vai ser resíduo.

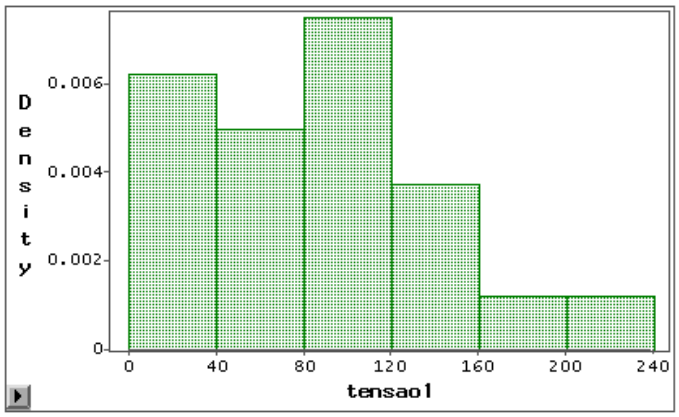
Mas antes de fazer isso precisamos verificar alguns pressupostos da análise de variância (1) os corpos de prova são independentes (2) a distribuição de tensão é normal, ou que pelo menos se possa assumir que vem de uma distribuição normal.

Antes de começar a fazer o teste ANOVA é importante, verificar a frequência geral, para ver se não existe nenhum dado digitado errado. Depois vá ao SAS interactive e peça a distribuição da variável tensão, o que resultará num box-plot e um histograma seguido de resultados descritivos a respeito da variável. Ai você poderá ver se a distribuição é normal. Poderá também pedir ao SAS testes para verificar se a distribuição é normal. Veja detalhes no manual do SAS que está na home-page.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva





Moments

N	20.0000	Sum Wgts	20.0000
Mean	87.5360	Sum	1750.7200
Std Dev	53.5165	Variance	2864.0110
Skewness	0.4686	Kurtosis	-0.1570
USS	207667.235	CSS	54416.2087
CV	61.1365	Std Mean	11.9666

A distribuição não parece ser muito normal. Pode-se então pedir ajuda do SAS para realizar os testes de normalidade. É so ir ao comando *Table >> tests for normality*. Irá aparecer abaixo do gráfico um quadro com vários testes. Prefira o teste Shapiro-wilk, embora cada teste sirva para uma situação específica.

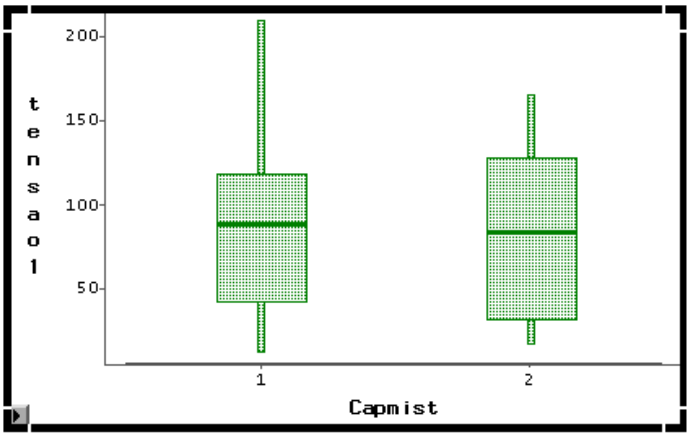
Tests for Normality

Test Statistic	Value	p-value
Shapiro-Wilk	0.958350	0.5114
Kolmogorov-Smirnov	0.096294	>.1500
Cramer-von Mises	0.026161	>.2500
Anderson-Darling	0.239440	>.2500

Pelo resultado do teste Shapiro-Wilk, a probabilidade de ter distribuição normal é de 0.5114, portanto, podemos assumir normalidade, e podemos prosseguir com o nosso teste ANOVA. Na verdade a melhor maneira de avaliar a normalidade é uma somatória de tipo de variável (contínua) formato do histograma, com formato do box-plot e do chamado Gráfico de normalidade (Normal QQ Plot). Estas evidências são

melhores do que os testes estatísticos acima. Embora normalidade seja muito importante, pequenos devios quando a amostra é razoavelmente grande são “aceitos” na análise de variância ou regressão, por causa do Teorema do Limite Central.

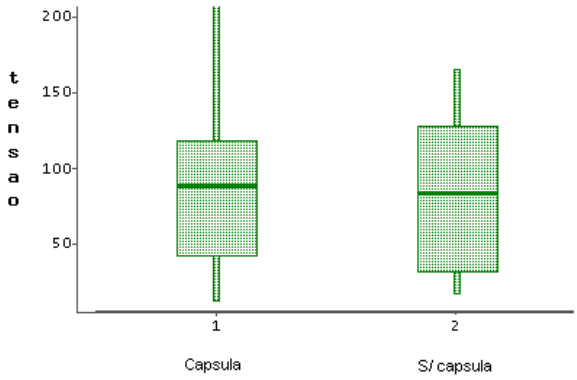
Uma vez confirmado a distribuição normal, independência dos dados e ausência de erros, você pode pedir ao SAS para fazer a análise univariada para cada tipo de amalgama (capsula e sem capsula). Você também pode pedir ao SAS para fazer dois boxplots juntos, este tipo de gráfico ajuda bastante a comparar os dados. Você consegue estes dois box-plots em “Solutions” >> BoxPlot/Mosaic Plot. Ai coloque no eixo Y sua variável dependente, e no eixo x sua variável independente, respectivamente tensao e capmist. Uma dica para copiar o gráfico do SAS para seu documento word é, selecionar o gráfico com o mouse, e pedir para copiar. Preste atenção que infelizmente, não sei porque control C, não funciona no SAS. Então, selecione, vá ao Edit >> copy. Uma vez feito isso cole no programinha Paint Brush e depois selecione e cole no seu word. É o que eu fiz para colar aqui os gráficos do SAS.



No Paint Brush você pode apagar estas linhas feias do SAS e até mesmo reescreer o nome das variáveis. Veja como melhorou, mas se você quiser trabalhar mais a aparência no Paint Brush sinta-se à vontade.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva



Pelo box-plot acima parece que as distribuições não são muito diferentes. As medianas são próximas, sendo que o grupo 1 teve maior variabilidade de tensão. Se perdirmos ao sas na parte do editor as tabelas de frquencia de tensão para cada grupo podemos ver que existem alguns valores bem altos para o tipo 1 em comparação com o tipo 2. Principalmente o valor de 210.85 é bastante diferente dos demais.

Tensão para capsula tipo 1.

tensao1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
13.9	1	10.00	1	10.00
23.2	1	10.00	2	20.00
43.2	1	10.00	3	30.00
56.6	1	10.00	4	40.00
80.7	1	10.00	5	50.00
98.7	1	10.00	6	60.00
110.4	1	10.00	7	70.00
118.96	1	10.00	8	80.00
135.4	1	10.00	9	90.00
210.85	1	10.00	10	100.00

Tensão para capsula tipo 2.

tensao1	Frequency	Percent	Cumulative Frequency	Cumulative Percent

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

18.5	1	10.00	1	10.00
25.96	1	10.00	2	20.00
33.03	1	10.00	3	30.00
73.9	1	10.00	4	40.00
78.7	1	10.00	5	50.00
89.7	1	10.00	6	60.00
103.3	1	10.00	7	70.00
129.04	1	10.00	8	80.00
140.6	1	10.00	9	90.00
166.08	1	10.00	10	100.00

Os comandos para obter os valores acima são:

```
proc freq;
tables tensao1;
where capmist = 1;run;

proc freq;
tables tensao1;
where capmist = 2;run;
```

Preste atenção no significado destes comandos. Proc freq é o comanda que pede para fazer uma tabela de frequência. No caso a variável é tensao1 que é especificado em *tables* . Como queremos ver esta variável apenas para os corpos de prova que foram feitos com amalgama em cápsula, então usamos o comando *where (onde).* Assim nós dissemos ao SAS " SAS, faça uma tabela de frequencia de tensão, mas especificamente apenas para aqueles onde capmist é igual a 1". O mesmo foi feito para capmist = 2.

Agora podemos proceguir com nosso teste de anova. No SAS existem várias maneira usado proc anova, prog glm, ou proc reg.

A maneira mais clássica de se pedir uma anova no SAS, é utilizando o proc anova. Para este comando precisamos avisar ao SAS qual é nossa variável categorica (a variável independente).

```
*****************************************;
* Proc Anova
```

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva


```
*****************************************;

proc anova;
class capmist;
model tensao1 = capmist;
run;
```

The ANOVA Procedure

Dependent Variable: tensao1 tensao1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	54.78050	54.78050	0.02	0.8944
Error	18	54361.42818	3020.07934		
Corrected Total	19	54416.20868			

R-Square	Coeff Var	Root MSE	tensao1 Mean
0.001007	62.78017	54.95525	87.53600

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Capmist	1	54.78050000	54.78050000	0.02	0.8944

Veja o resultado acima do proc anova. Temos 54416,20 de SST, e o modelo explica 54,78. Quando verificado se esta explicação era significante em relação ao que deixou de ser explicado no resíduos verifica-se que não. A probabilidade de capmist não explicar alguma coisa da variância total foi de 0,8944 isto é muito grande por isso, concluímos que não foi possível encontrar diferenças entre a tensão entre os grupos estudados.

Se você utilizar o proc glm, você terá mais detalhes da tabela de variância.

The GLM Procedure

Dependent Variable: tensao1 tensao1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	54.78050	54.78050	0.02	0.8944

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Error	18	54361.42818	3020.07934
Corrected Total	19	54416.20868	

R-Square	Coeff Var	Root MSE	tensao1 Mean
0.001007	62.78017	54.95525	87.53600

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Capmist	1	54.78050000	54.78050000	0.02	0.8944

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Capmist	1	54.78050000	54.78050000	0.02	0.8944

A diferença agora é que o proc glm da informações sobre os tipos de sum of squares tipo I e II, que neste caso é a mesma coisa, pois temos apenas uma variável.

Para o uso do proc reg, não há necessidade de se especificar que a variável dependente é categorica. A diferença é que no proc reg que é específico para calcular a fórmula de predição vamos encontrar os estimadores dos parametros intercepto e inclinações de reta. Mas note que o resultado é o mesmo.

```
*******************************************;
* Proc reg
*******************************************;

proc reg;
model tensao1 = capmist;
run;
```

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Number of Observations Read 20
Number of Observations Used 20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	54.78050	54.78050	0.02	0.8944
Error	18	54361	3020.07934		
Corrected Total	19	54416			

Root MSE 54.95525 R-Square 0.0010
Dependent Mean 87.53600 Adj R-Sq -0.0545
Coeff Var 62.78017

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	92.50100	38.85923	2.38	0.0286
Capmist	Capmist	1	-3.31000	24.57673	-0.13	0.8944

Da mesma forma que temos aqui a comparação de dois grupos podemos ter de 3 ou mais grupos.

Tolerance and Variance Inflation

Estas duas medidas aparecem no output do SAS para analise de regressão. A tolerancia é igual 1 – R2. Logo se R2 é o quanto é explicado pela variável então tolerancia eh o que falta a ser explicado.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Comentamos de forma geral como comparar proporções utilizando intervalos de confiança, mas existem outras maneiras formais de se comparar proporções. Uma delas é comparação de proporções (adicionar aqui pxq e etc) utilizando uma aproximação normal e outra é por meio do tese do qui-quadrado.

Uma outra forma de olhar para um teste de proporções é tabulando os dados e trabalhando com tabelas de contingência. Tabelas de contingência é o nome dado a tabelas que cruzam duas variáveis categóricas e que podem ter tamanhos variados de r vs c [r = row (linha) e c column (colunas)]. Vamos supor que numa comunidade de 10 mil crianças coletamos dados de uma amostra aleatória de mil crianças para testar a hipótese de associação entre asma e cárie dentária, isto é se crianças com asma tem maior prevalência de cárie do que crianças normais. Os resultados se encontram na tabela abaixo.

		Cárie +	Cárie -	
Asma	+	132	88	220
	-	390	390	780
		522	478	

Vemos que 132 (60%) dos asmáticos e 390 (50 %) de não asmáticos tem cárie, mas será que esta diferença é realmente significante? Podemos realizar este teste considerando algumas regras. O teste de comparação de proporções pode ser realizado por meio de simulação, o que estimaria o intervalo de confiança ao redor de cada uma destas estimativas pontuais ou por meio de cálculos. Para se realizar o teste por meio de cálculos, temos que considerar alguns pressupostos. Primeiro, assumimos que as prevalências de cárie e asma encontradas na amostra refletem as verdadeiras prevalências na população. Assim, com estas condições nos perguntamos, como a distribuição de cárie aconteceria se não houvesse associação com asma? **Esperamos** que se não houver associação, a porcentagem de cárie entre asmáticos e não asmáticos seria a mesma. Assim, considerando-se as porcentagens gerais de asmáticos e crianças com

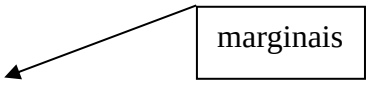
cárie na amostra acima, podemos calcular estas **porcentagens esperadas** que serão comparadas com as porcentagens observadas.

Uma vez qua a prevalência geral de cárie é de 55,2% (552/1000) logo, espera-se que 55,2% de asmáticos e, também, de não asmáticos tenham cárie. Uma forma de se calcular os valores esperados dado a não associação, é utilizando alguns princípios de probabilidade que você aprendeu durante o segundo grau. Lembre-se que os valores encontrados (valores observados) no interior da tabela (132, 88, 390, 390) refletem a ocorrência conjunta dos eventos, isto é, respectivamente asmáticos com cárie, asmáticos sem cárie, não asmáticos com cárie e não asmáticos sem cárie. Lembramos que quando temos eventos independentes a probabilidade conjunta de dois eventos é obtida multiplicando-se a probabilidade de um evento pelo outro evento. Isso você aprendeu la no ginásio ou segundo grau, isso é a probabilidade conjunta independente de dois eventos pode ser dada pela probabilidade de um evento vezes a probabilidade de outro evento. Vamos supor que o Brasil tenha 30% de negros, e 50% de indivíduos pobres, qual seria a probabilidade de você encontrar um Negro Pobre, seria 0.30 x 0,50 que é igual a 0.15, isso é a probabilidade de você sortear alguem no Brasil e esta pessoa ser negro e pobre é de 15%. Se os negros são muito mais pobres do que os brancos, a probabilidade passa a ser condicional e não independente como no exemplo que foi mencionado aqui. Não iremos fazer estas contas aqui, pois nos interessa mais no momento, a probabilidade não condicional. No exemplo da asma, multiplicando-se a probabilidade de ter asma pela probabilidade de ter cárie.

Sendo **Pa** probabilidade de ter asma e **Pc** probabilidade de ter cárie temos

Pa x Px = 0.52 x 0.22 = 0,1144 ou seja 11,44 % dos indivíduos da amostra todal tem cárie e são asmáticos. Dado que a amostra tem mil indivíduos, 114,4 indivíduos teriam asma e cárie ao mesmo tempo. Substituindo este valor para compor a tabela de números esperados conseguimos calcular os demais valores, pois as marginais são fixas.

Cárie



mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

		+	-	
Asma	+	I	II	220
	-	III	IV	780
		522	478	1000

Cárie

		+	-	
Asma	+	114,8		220
	-			780
		522	478	1000

Cárie

		+	-	
Asma	+	114,8	105,2	220
	-	407,2	372,8	780
		522	478	1000

Com os valores esperados numa condição de não associação entre asma e cárie, podemos comparar os números observados com aqueles esperados. Se os números observados estiverem próximos dos esperados numa não associação podemos concluir que não deve existir associação. Quanto mais longe estiverem os números maior será a associação. Assim, vamos verificar a distância dos números observados em relação aos números esperados tendo como base os próprios números esperados.

Observado	Esperado	O - E	$(O – E)^2$	$(O – E)^2 / E$
132	114,8	17,2	295,84	295,84/114,8 = 2,577003
88	105,2	-17,2	295,84	295,84/105,2 = 2,812167
390	407,2	-17,2	295,84	295,84/407,2 = 0,726523
390	372,8	-17,2	295,84	295,84/372,8 = 0,793562

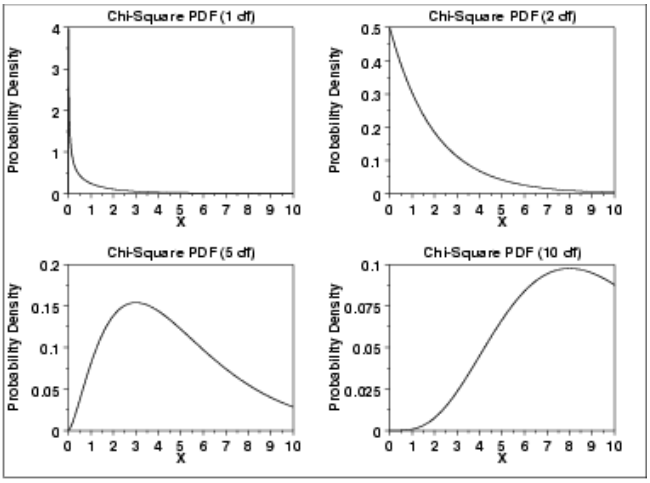
mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Soma		0	1181,6	6,909256

Este valor de 6,909256 que pode ser aproximado para 6,01 é uma estimativa de distância entre números observados e esperados ponderados pelos números esperados. Este é o valor do teste qui-quadrado cuja fórmula é :

$$\text{Teste estatístico} : X^2 = \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$$

Este valor faz parte de uma distribuição chamada qui-quadrado, que tem um formato um pouco diferente da curva normal. A figura abaixo mostra distribuições tipo chi-quadrado.



Podemos dizer “grosseiramente” que esta distribuição é como se fosse a distribuição normal elevada ao quadrado. Assim, os números negativos se transformariam em positivos, e a distribuição começa no valor zero. Além do mais, para cada grau de liberdade (veja explicação de graus de liberado mais a frente) existe uma distribuição de qui-quadrado diferente. Esta distribuição é utilizada da seguinte forma, escolhe-se a curva adequada para os graus de liberdade apropriados e verifica-se a probabilidade referente ao valor. Neste caso, procuramos onde ficaria o valor 6,91 para o grau de liberdade de 1 . Na tabela abaixo vemos que com 1 grau de liberdade o valor 6,91 está entre 0.01 e 0.001. Portanto, a probabilidade dos números observados serem iguais aos esperados está entre 0.01 e 0.001. Dizemos que a probabilidade de associação

entre asma e cárie é menor que 0,01 logo é uma probabilidade muito pequena dos números observados serem iguais aos esperados dado a não associação, portanto, concluímos que existe associação entre asma e cárie.

Valores críticos superiors da distribuição do qui-quadrado com ν graus de liberdade.

	Probability of exceeding the critical value				
ν	0.10	0.05	0.025	0.01	0.001
1	2.706	3.841	5.024	**6.635**	**10.828**
2	4.605	5.991	7.378	9.210	13.816
3	76.251	7.815	9.348	11.345	16.266
4	.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588
11	17.275	19.675	21.920	24.725	31.264
12	18.549	21.026	23.337	26.217	32.910
13	19.812	22.362	24.736	27.688	34.528
14	21.064	23.685	26.119	29.141	36.123
15	22.307	24.996	27.488	30.578	37.697
16	23.542	26.296	28.845	32.000	39.252
17	24.769	27.587	30.191	33.409	40.790
18	25.989	28.869	31.526	34.805	42.312
19	27.204	30.144	32.852	36.191	43.820
20	28.412	31.410	34.170	37.566	45.315
21	29.615	32.671	35.479	38.932	46.797

Como sabemos que temos apenas 1 grau de liberdade? O grau de liberadade em tabelas de contigência é dado pela multiplicação do número de colunas menos 1 vezes o número de linhas menos 1. Logo (2 linhas menos 1) vs (2 colunas menos 1) ou seja :

GL = (r-1) (c-1)

Para uma tabela 2 vs 2 temos um grau de liberdade, para uma tabela 2 vs 3 temos dois graus de liberdade, 3 vs 3 temos 4 graus de liberdade. `Mas o que significa este grau de liberdade com esta fórmula? Numa tabela 2 vs 2, se calcularmos um número

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

esperado as demais caselas serão automáticamente deduzidas, logo temos (ou na verdade perdemos) apenas um grau de liberdade. Se você montar uma tabela 2 vs 3 , verá que precisará calcular dois valores de caselas esperados para deduzir os demais, logo 2 graus de liberade. Quando maior as possibilidades de intersecção entre as duas variáveis maior será a quantidade de caselas minimas calculadas para que as demais sejam deduzidas.

Notas importantes:

1. É importante ressaltar que o teste de qui-quadrado deve somente ser utilizado quando houver uma amostras suficientemente grande. Como sabemos se a amostra é suficientemente grande? Isto é revelado, calculando-se os números esperados, se algum número esperado for menor que 5 então deve-se utilizar um outro teste mais apropriado para amostras pequenas. O teste de escolha é o chamado teste de Fisher que se baseia numa distribuição diferente do qui-quadrado que é a distribuição chamada de hipergeométrica. Esta distribuição é baseada em sorteios com reposição de dados (veremos isso posteriormente).
2. Existe uma correção que é utilizada chamada correção de Yates para quando a amostra é relativamente pequena. No entanto, seu uso é controverso pois acaba substimando as associações. Cuidado porque alguns programas como o R tem como padrão liberar os resultados do teste com correção, se você estiver interessado no qui-quadrado original deverá inativar a correção.
3. Teste qui-quadrado é utilizado em diversas situações neste exemplo que demos é chamado de teste de qui-quadrado de independecia dos dados. Por vezes podemos utilizar a mesma logica com distribuição do qui-quadrado para verificar se uma variável é compatível com uma certa distribuição, e ai é chamado de qui-quadrado de aderência. Porque aderência? Aderência a distribuição que se imagina que pode conter. Em inglês seria o chi-square goodness of fit (o quando se adequa bem a uma distribuição). Alguns outros testes baseados nesta lógica de comparar observados e esperados no mundo do quadrado têmm sido comentados em sites de estatística como o teste de

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Cochran Q. No entanto, este teste é utilizado para dados dependentes e portanto não é o foco desta apostila. Mas já que comentamos , o teste chamado Cochram Q é um teste de concordância de eventos ao longo do tempo.

4. Conforme você for aprendendo algumas estatísticas básicas vai comecar a ler mais e compreender que existem inúmeros teste com nome semelhantes para situações particulares de dados e amostra.

5.

6. Até o momento comentamos apenas sobre estes que não levam em consideração a dependência de dados. O Cochran Q teste concordância de proporções.

7. Ao relatar os resultados devemos nos expressar da seguinte maneira " o estudo demonstra que existe associação estatisticamente significativa entre asma e cárie dentária". Caso, não encontremos associação estatísticamente significante, devemos dizer " neste estudo não foi possível demonstrar associação estatísticametne significante entre asma e cárie dentária". Note que não podemos concluir que não existe associação, pois um teste estatístico depende do tamanho da amostra. Pode ser que com uma amostra maior encontrássemos associação estatística. Lembre-se que com amostras infinitamente grandes o intervalo de confiança diminui e pode tornar-se "artificialmente" pequeno de mais e tornar as associações estatísticamente significantes. Esta é uma limitação da estatística e temos que lidar com ela com bom senso.

Assim, podemos dizer que para aplicar o teste do qui-quadrado os pressupostos são: que os dados sejam independentes, isso é cada indivíduo é um indivíduo (não temos conglomerados de dados), e que nenhuma casela de número esperado possa ser menor do que 5. Se a primeira condição acontecer, precisamos usar programas estatisticos especiais para levar em consideração a dependência dos indivíduos, então não usaremos o qui-quadrado como ensinado aqui. Se a segunda opção acontecer, não usaremos o teste do qui-quadrado, e teremos que usar um teste chamado teste exato de Fisher, que não

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

ensinaremos neste curso. Mas saibam , que se a amostra é pequena e a distribuição binomial teremos que utilizar um teste exato de Fisher.

Explicar as distribuições

Uma outra forma de testar diferenças de proporções

Uma outra forma de testar a hipótese de associação entre asma e cárie testando se as proporção de cárie entre asmáticos é a mesma que a observada entre não asmáticos, chamdo de teste de comparação de proporções de proporções. Mas para que se utilize este tipo de teste aqui apresentado, é necessário assumir que a amostra é razoavelmente grande. Existem duas possibilidades para se testar hipótese de igualdade de duas proporções. Podemos ter uma situação em que as proporções pertencem a uma mesma população porém em grupos diferentes desta população e temos o caso em que as prorporções podem vir de diferentes populações.

Os cálculos para esta estatística somente fazem sentido se tivermos uma noção de probabilidade, vamos aos poucos introduzir alguns princípios de probabilidade. Em particular temos que entender o que é uma distribuição binomial, jogo de Bernouli, e o que quer dizer Teorema Central . Na verdade, já explicamos o Teorema Central quando explicamos como funciona a simulação para se obter a distribuição de esperanças para proporções. Voltaremos a este assunto posteriormente.

A distribuição binomial resulta de inúmeros “jogos” independentes, onde existe apenas dois tipos de resultados sucesso ou falha. Um exemplo, seria uma pessoa jogar inúmeras vezes uma moeda para cima, e verificar qual o resultado. Se fizermos isso vamos ter que a probabilidade de cara ou coroa é de 0.5 (isto é cinquenta por cento), se jogarmos 2 vezes qual seriam as possibilidades? Poderimaos ter uma cara-coroa, coroa-coroa, ou cara-cara. Se a probabilidade de sucesso é identificada pela letra p então podemos dizer que a possibilidade de falha é igual a 1- p que identificamos como q. Logo, q é complemento de p.

Se eventos são independentes, podemos utilizar este conhecimento para saber qual a probabilidade de encontrar um determinado evento. Imagine que numa

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

contagem de células do sangue , sabemos que a possibilidade da célula a ser um neutrófilo é de 0.6, isto é de 60% logo 1- p = 0.4. Para nosso exemplo, vamos chamar o neutrofilo de “a” e “b” outra célula qualquer. Se formos contar 5 células, podemos responder por exemplo a pergunta de qual a probabilidade deste conjunto ser constituído por babba. Sabendo o valor das probabilidades podemos calcular qpqqp = $p^2 q^3 = (0.6)^2 (0.4)^3$. Isso dá uma probabilidade de 0,02304, isto é de aproximadamente 2,3%. Mas este resultado é apenas a probabilidade na verdade de se ter 2 neutrofilos e 3 outras células. Na verdade, podemos com estas ocorrências ter várias combinações. Lembre-se de análise combinatória que você aprendeu no segundo grau. Aquilo servia para alguma coisa, caso você não tenha entendido na época.

Voltando ao exemplo, existem 10 possíveis combinações de 2 neutrofilos e 3 outras células. Podemos fazer esta conta escrevendo cada uma das combinações ou podemos calcular este número por meio de análise combinatória. $_5C_2$. Se você não se lembra, esta anotação significa análise combinatória de 5 elementos organizados dois a dois.

$$_5C_2 = \frac{n!}{k!(n-k)!}$$

Portanto, n! = 5 x 4 x 3 x 2 x 1 , e no denominador temos k! = 2 x 1 , e (n-k)! = 3 x 2 x 1. Portanto, 10 combinações. Isso é, n! Deu igual a 120, e no demominador as contas resultam em 12. Portanto, existem 10 possibilidades destas combinações e logo temos a probabilidade 10 x 0.02304 = 0.2304.

Podemos reescrever as duas fórmulas até agora como

$$\binom{n}{k} = p^k (1-p)^{n-k} = \binom{n}{k} p^k q^{n-k}$$

Assim podemos calcular qual a probabilidade de cada combinação possível em um jogo. Lembre-se que quando falamos da construção de uma curva de esperanças para proporções para se estimar o intervalo de confiança, cada amostra é uma combinação de x elementos. A distribuição resultante de diversos “jogos” , no caso seleção ao acaso de indivíduos numa população resulta numa distribuição de combinações que chamamos de distribuição de Binomial.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Preste atenção que a variável que trabalhamos numa distribuição binomial, é uma variável pura do tipo sim/não , isto é qualitativa. Porém a distribuição não é feita com a própria variável sim e não, mas com inúmeros "jogos" com estas variáveis. Assim, a distribuição binomial assume a forma de probabilidades que são, vamos dizer, "exatas" e consideradas como variável discreta, que na verdade é uma variável quantitativa. Não confunda o tipo de variável inicial que era do tipo sim/não, com a variável que compoe a distribuição de probabilidades de vindas de jogos.

Já que estamos falando de jogos cujas possibilidades são sucessos ou falhas, a esperança da distribuição binomial será igual a probabilidade de sucesso.

A esperança da distribuição binomial será dada por :

$$E(X) = \sum_{i=1}^{K} x_i \Pr(X = x_i)$$

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

se resolvermos a equação veremos que a esperança se reduz a expressão sim de E (x) = np.

A variância que seria a diferença ao quadrado de cada resultado de cada jogo, seria dado pela fórmula.

$$E(X) = \sum_{k=0}^{n} (k - np)^2 \binom{n}{k} p^k q^{n-k}$$

O que é simplificado para variância igual a npq.

A variância é máxima quando p = 0.5 e diminu quando se aproxima dos extremos das probabilidades, isto é zero ou um.

A estatística parte do princípio que vamos comparar duas proporções, portanto, se subtrairmos uma proporção da outra e encontrarmos o valor zero quer dizer que não existirá associação. Para realizar o teste, a idéia seria construir uma curva de proporções

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

esperadas cuja esperança seja zero. Já sabemos então que a "média" ou melhor esperança de nossa curva será zero, mas precisamos saber qual a largura desta curva, isto é o erro padrão. Para tanto, teremos que construir uma curva com esperança zero, e com uma certa variância, ou melhor erro padrão. Vamos ter que estimar o erro padrão apartir dos dados que temos, isto é, proporções de 50 e 60% e com os respectivas quantidades de indivíduos sem e com asma 780 e 220. Temos duas possibilidades ao comparar proporções: (1) elas podem vir de duas populações distintas (2) ou de dois grupo de uma única população. No caso de ser uma única população como por exemplo, num estudo de uma cidade identificamos crianças com e sem asma e queremos ver a associação com cárie dentária, consideramso uma população. Neste caso, a variâncias será calculada de forma geral. No caso de amostrarmos crianças com asma e depois criancas sem asma ai poderemos considerar duas populações a variância será a combinação das duas populações.

$$\frac{pq}{n1}+\frac{pq}{n2}=pq\left(\frac{1}{n1}+\frac{1}{n2}\right)$$

Na fórmula acima o p é a proporção de eventos que estamos estimando, e "q" é o complemento de p. Isto é se a prevalência é de 0.4 o complemento é de 0.6. Já sabemos que variância é dado pela multiplicação de p*q.

Se nós tirarmos a raiz quadrada da variância termos o erro padrão logo o erro padrão. Podemos dizer também que este é o "desvio padrão" da distribuição binomial que é resultado de "inúmeros jogos". Lembre-se que anteriormente falamos de desvio padrão era apenas calculado para os dados reais com distribuição normal, e que o erro padrão era calculado para a distribuição de médias, e que podiamos dizer que o erro padrão seria o desvio padrão da distribuição de médias. Note que estamos falando agora de uma distribuição de esperanças, portanto este "desvio padrão" da distribuição de esperanças representa o erro padrão. Sendo assim, temos agora o cálculo do erro padrão para a diferença entre proporções.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

$$EP = \sqrt{pq\left(\frac{1}{n1}+\frac{1}{n2}\right)}$$

Vamos retornar ao exemplo da asma e cárie. A prevalência de cárie entre asmáticos era de 0.60 e entre não asmáticos de 0.5. Utilizandose a fórmula acima temos

$$\frac{0.6(0.4)}{220}+\frac{0.5(0.5)}{780}=0{,}00109+0{,}000321=0{,}001411$$

Portanto, a variância é de 0,00141 e tirando-se a raiz quadrada temos 0,0376. Com o valor do erro padrão, podemos construir uma curva ao redor do zero que seria a diferença esperada caso não existisse diferenças entre as duas proporções de cárie entre asmáticos e não asmáticos. Assim temos 0,6 – 0,5 / 0,0376 = 13,309. Este valor é o que chamamos de z score numa aproximação da distribuição normal. Agora temos que verificar onde fica este valor numa tabela de distribuição binomial evamos verificar que este valor é significante com p > 0.0001

Se você quiser calcular o intervalo de confiança para a diferenca de proporção dado acima, você terá a diferença de 0.10 + 1.96 (erro padrão) que seria 0,0376. Isto é: aproximando 0.074

Limite inferior : 0.10 – 0,074 = 0.026

Limite superior : 0.10 + 0,074 = 0.174

Portanto, este intervalo de confiança não inclui o valor zero (0), que seria a não diferença completa.

Devemos lembrar que dependendo do que estamos estudando, não estamos interessados em diferenças de zero, mas sim de alguma porcentagem. Por exemplo, um pesquisador pode testar um novo produto que somente compensa substituir um produto ja existente se este for pelo menos 20% melhor do que o já existente. Para isso podemos estabelecer que nosso teste leve isso em consideração.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Outro banco de dados de alunos

Diferencas entre proc glm e pro reg

Proc reg

The REG Procedure

Model: MODEL1

Dependent Variable: aluno aluno

Number of Observations Read 57

Number of Observations Used 57

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.02716	0.02716	3.09	0.0844
Error	55	0.48367	0.00879		
Corrected Total	56	0.51083			

Root MSE	0.09378	R-Square	0.0532
Dependent Mean	1.68684	Adj R-Sq	0.0360
Coeff Var	5.55930		

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.01956	0.37991	2.68	0.0096
pai	pai	1	0.38142	0.21704	1.76	0.0844

**********;

Mais um exemplo inventado

**********;

Os dados abaixo representam perda de inserção entre indivíduos fumantes e não fumantes. Queremos saber se existe diferença estatisticamente significante de pi entre fumantes e não fumantes. Copie os dados para o editor do SAS e acompanhe.

```
data perio;
imput ind fumo pi;

cards;

1 1 2.7
2 1 1.5
3 1 0.1
4 1 5.3
5 1 4.2
6 1 3.7
7 1 2.1
8 1 5.9
9 1 1.4
10 1 5.9
11 2 0.5
12 2 1.7
13 2 1.8
14 2 1.1
15 2 0.5
```

```
16      2       0.4
17      2       0.2
18      2       0.7
19      2       2
20      2       1.5
;
run;
```


Para se testar se as médias são iguais precisamos primeiro decidir se média é uma medida de tendência central adequada para ser avaliada. Isso vai acontecer se a distribuição for normal. Então primeiro precisamos fazer um box-plot e um histograma para ver como é esta distribuição.

```
**********************************************************************;
Proc univariate;
Var pi;
Run;
```

Teste de não inferioridade, que o novo tratamento não seja consideravelmente pior do que o existente. Rejeita a hipótese de que o novo teste é apreciavelmente pior do que o novo teste. Deve-se evitar erro tipo 1, portanto diminui-se o erro para 0.01

Mantel Haenzel

Exemplos do Kleimbaum et al 1982

		Catecolamina +	Catecolamina -	
DCV	+	27	44	71
	-	95	443	538
		95	487	609

RR = (27/122)/(44/487) = 2,45

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

		< 55 anos				>= 55 anos	
		Catecolamina				Catecolamina	
		+	-			+	-
DCV	+	4	24			23	20
	-	21	309			74	134
RR = 2.22						RR = 1.83	

Cuidados, quando dois estratos estão em direção oposta, chi-quadrado de Mantel Haenzel não deve ser utilizado.

Em construção .....

Distribuições

Distribuição Binomial

Distribuição de sucesso e fracasso (0 ou 1)

Variavel categórica (sim não), quantos os indivíduos seria uma variável discreta, modelo binomial porque pode ser 1 ou 2 ou 3. O experimento é repetido n vezes, eventos independentes (um individuo não influencia no outro), a cada repetição apenas 2 eventos, probabilidade de sucesso p, e de fracasso em cada repeticao se mantem constante. Número de sucesso em n repetições.

Probabilidade VS Maxima verossimilhança

Probabilidade é a área sob uma curva de uma distribuição conhecida, oq eu seria a probabilidade dos dados coletados dado uma distribuição.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Likelihood é medido no eixo Y, com fixed data points with distributions that can be move. L (distribution/data).

Likelihood, assume que você já fez um trabalho e tem um rato com x , e a media dos ratos deu 32 , ai falamos da likelihood da distribuição ser de 32 com o desvio padrão que foi observado

Distribuição Exponencial

Modela tempo between events.

Tipos de seleções numa regressão

Embora existam vários tipos de seleção, quando se trata de dados em biologia e ciência sociais, esta seleção é mais bem realizada dentro de um modelo teórico. Para algumas áreas pode até fazer sentido, mas achamos que deve ser com parcimônia.

Alguns models de seleção que são utilizados no SAS. Os programas estatísticos tem diferenças. Escrever com mais detalhes, e verifica no SAS os modelos de seleção que vamo desde rsquare selecion, Mallows CP selection etc.

Valliante pesos amostrais

Pesos amostrais são fundamentais para produzir estimadores populacionais. Um estimador tem a forma de t= ∑s wiyi onde yi pe a resposta dada pelo ith individuo na

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

amostra e wi é o peso amostral daquele individuo. Sem pesos amostrais estimadores podem refletir apenas nuances de uma amostra em particular e não sua função primaria de representar a população e pode conter significante níveis de vieses.

São vários passos que são comuns a surveys : calculo de pesos básicos que são chamados de pesos de desenho (design weights) 2)ajustes para eligibilidade não conhecida, ajustes de não resposta, uso auxiliar de dados para reduzir variâncias ou corrigir deficiências de estrutura amostral.

Em amostras probabilísticas pesos de base são inversos das probabilidades de seleção. Estes pesos podem ser calculados para amostras de populações completas e se todas as pessoas responderem. Em algumas vezes a complete frame units are avaibalbe for sampling and grame problems are not a concern. Em outras the frame pode conter algumas unidades que nao são elegíveis ou pode omitir unidades que são raras. Tendo unidades inelegíveis numa frame é um tipo de overcoverage, e existe uma forma de ajustar para inelegíveis e também outra para overcoverage.

A falha de resposta é algo para se preocupar e sem ajuste para não resposta , os estimadores porem ter vieses significantes. Existem vários métodos de ajuste (13.5).

O objetivo geral de se pesar é encontrar uma conjunto de pesos wi, que possa ser usado em vitualmente todas as analises para produzir estimadores para a população alvo do estudo. Por exemplo a media y(médio estimador) = soma wiyi/ soma wi for um conjunto de unidades na amostra. Outras estatísticas que podem ser escritas como combinações de total estimado usaria os mesmos sets de pesos. Analise de regressão por exemplo começa com um tipo de estimated total que é utilizado para derivar paramestros estimados. Estimadores de medianas e quantis usam os mesmos pesos.

Se propriamente construídos um set de pesos podem provide estimadores consistentes e não enviesados de muitas diferentes quantidades populacionais.

No diagrama apresentado por Valliante et al, existem dois passos w1 e w2, sendo que w1 admite não elegíveis, se amostra é toda conhecida então W1 não precisa ser calculado. O próximo passo W2 é para fazer ajustes de não resposta. Uma forma é colocar respondentes e não respondentes em classes e e fazer ajustes comuns dentro de cada classe. Classes podem ser formadas baseadas em estimated response

propensities ou classification algorithms. Em alguns surveys os pesos finais são os nonresponse adjusted weights. Em outros cabibration to population values (step W3). Design based approach propriedades do estimadores como viés e variância são avaliados com respeito a repeated sampling. Esta abordagem requer probability sampling de verdade. Se amostra probabilística é utilizada vieses concientes e não concientes são eliminados ao se selecionar a amostra. E assim amostragem probabilística da fundamentos matemáticos para se calcular as propriedades dos estimadores.Porem por causa de não resposta, a amostra que começou probabilística não termina probabilística. Assim, strictly design-based inferência usualmente não é possível (feaseable). Assim modelos para não resposta undercoverage and nonsampling erros são necessários. Entretando, calculo de baseweights (inverse probabilities) é o primeiro passo.

Ter propriedades boas desgn-based is confortante. No entanto a relação entre response variables e preditores não é formalmente considerada na inferência design-based inferece. Pensar em modelos que descrevem a variável na população fornece alguma estrutura que pode ser usada como guia. By contrast, a strictly model-based approach ignora o desanho da amostra e considera apenas a estrutura populacional (isso é o modelo) para decidir o estimador e os pesos correspondentes. Esse método pode ser aplicado tanto a amostras probabilísticas como não probabilísticas. Por exemplo cursos de matemática estatística usa estimadores com a pressuposição de que unidades são amostradas de populações infinitas. Os estimadores resultantes são não-enviesados sob o modelo usado para contruir os estimadores mas podem ser enviesados se o modelos for misspecified ou se o modelos que fits the sample é diferente do que descreve a população como um todo. Em alguns casos model base no entanto, é o único de escolha. Em um survey pela internet com participantes voluntários, não existe nenhum probability sampling design, e os estimadores devem ser construídos usando modelos. Se os voluntários são muito diferentes da população os estimadores serão problemáticos.

No entanto, modelos inevitavelmente precisam ser considerados quando desenvolvendo os peso, mesmo com probability samples. Qualquer amostra com

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

algum grau de não resposta necessita pressupostos a respeito da natureza das variáveis para não resposta e sobre o mecanismo.

Existe bons, embora muito técnicos, argumentos para explicar porque a distribuição aleatória por si soh não deve ser a bbase para inferência, mesmo na ausência de não-resposta. A linha geral é que averaging over a randomization distribution envolve averaging over samples that can be much different from ehte one selected (melhor estudar para frente essa questao).

A abordagem hibrida use tanto model based e o design based thinking e e chamado de modelo assistido (assisted model). Isso é amostra probabilística é selecionada, pesos são calculados e modelos guiam a escolha do estimador. Inferencias são feitas utilizando distribuições geradas por um plano de amostragem probabilística não um modelo. Pesquisas sugerem que pesos fornecem um certo nível de proteção contra misespecification do modelo.

13.3 (livro)

Pesos básicos ou de desenho são calculados quando a amostra vem de uma população finita. Amostra probabilística é aquela realizada sob 4 condicoes.

1. Um set de todas as amostras possíveis S { = s1....sM) que pode ser selecionada de uma população finita U pode ser definida dado um procedimento especifico de amostragem.
2. A probabilidade conhecida de seleção p(s) é associada com cada possível amostra s em S.
3. Cada elemento da população alvo tem uma probabilidade não zero de seleção com um especifico procedimento de amostragem aleatório
4. Uma amostra s é selecionada por mecanismo aleatório sobe o qual s esta em S recebe a proabilidade p(s).

   A função p(s) define a distribuição de probabilidade em S, que é o set de todas as possíveis amostras. O valor para p(s) é asscoado com cada amostra s, e difere da seleção de probabilidade do unidade individual dentro da amostra. Para calcular base weighjts não é necessário ser capas de computar o p(s) apenas precisamos saber a probabilidade de seleção dos elementos individuais.

   π i = probabilidade de seleção ou inclusão do elemento i.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Os pesos base d0i = 1/ π i são probabilidade de seleção inversas. As probabilidades de seleção podem ser calculadas como produto de probabilidades condicionais em diferentes estágios da seleção.

Pesos bases devem ser criados assim que a amostra e selecionada. Isso facilita analise preliminares como por exemplo a taxa de performance.

Probabilidades de seleção são todas entre 0 e 1, Base weights devem somar o total de número de elemtbos na população ou a um estimador do tamanho da população. Checagem também em que ser feita para subgrupos maiores (raca, sexo etc).

Base weights uma exceção. Sempre devem ser calculados primeiro, exceto quando podem ser selecionais mais de uma vez. Esses métodos são utilizados as vezes no primeiro estagio de amostras multi-estagios. Ex. distritos são selecionados primeiro com probabilidade proporcional ao número de alunos em cada distrito. Grandes distritos podem ser selecionados mais de uma vez, no qual neste caso uma grande amostra de escolas neste distrito pode ser selecionado. Quando algumas unidades são permitidas serem selecionadas mais de uma vez o número experado de seleccoes devem ser tracked, e ai pode ser maior do que 1. Ai o base weight deve ser o inverso do número esperado de seleções. (não entendi isso muito bem porque poderia ser selecionado mais de uma vez).

SRS sem reposição (seswor). Quando n (fixesd) unidades são selecionadas de uma população N, a seleção de probabilidade de cada unidade ipe a mesma = n/N; Nessa situação d0i – π-1 = N/n. Srswor é chamado de self weihing ou epsem (equal probabilidade sampling and estimation method).

Exemplo de estratified simple random sampling withour replacement (stsr-swor). A populacao é dificida em h = 1 …. H mutualmente exclusive estratos que cobrem a população inteira. A srswor em cada estrato de tamanho n_h e selecionado de uma população Nh. A probabilidade de seleção da unidade i no estrato h vai ser π_{hi} = nh/Nh e o base weights = Nh/nh. Isso dentro de 1 estrato, entre estratos os sampling weights serão diferentes.

Dois estatios levando a epsem. Imagine se amostra de estudantes é selecionada em 2 estagios, escolas no primeiro e estudandos no segundo. Neste caso PSU são as escolas, assuma que mPSUs são selecionados com probabilidade proporcional ao tamanho (pps) do corpo de estudantes e que probabilidade e que equal probability seja utilizada em cada PSU. Assim inclusion probability será:

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Π= mNi?N para escola i

Ni = número de estudantes em PSU i

Ver contas, mas se dentro das escolas a probabilidade de selecionar os estudantes for a mesma dentro de cada escola, então de novo tem se selfweighing. Note escola é ao acaso, e dentro da escola outro ao acaso.

Amostras de dois estágios for domains . Uma amostra de m PSU é selecionada e ai n secundarias sampling units (SSU) são selecionadas dentro do PSU i com probabilidade proporcional ao número de pessoas em cada SSU. Exemplo, PSU seria pequenos segmentos geográficos, e SSU residências dentro destes seguimentos . Cada casa tem uma ou mais pessoas com idades diferentes com 4 age groups. Cada pessoas dentro de cada SSU e grupo de idade é selecionado com a mesma taxa.

Ajustes para elegibilidade desconhecida, porque por mais que se tente limpar o sample frame vai ter unidades que não são elegíveis e passam. Em household survey de imunização de crianças, aquelas que não tem crianças são inelegíveis. E as vezes mesmo depois de se fazer o survey não vamos saber se era ou não inelegível. Exemplo seria ring/no answe in telefone survey, carta que retorna e não se sabe do survey, e aqueles que nunca estão em casa num survey de households. Ao final temos respondentes que se conhece elegibilidade que incluem elegíveis, elegíveis não respondentes, inelegíveis e e eligibilidade desconhecida. Uma forma simples e distribuir a ineligibilidade entre todos os participantes. Portanto a teria um frame maior e a probabilidade do individuo pertencer a amostra sera um pouco menor do que pode-se verificar efetivamente. No exemplo, de 110 pessoas em que 10 é desconhecido a elegibilidade então, soh tem elegibilidade os 100 dos 110 e isso significa que a probabilide de ser elegível eh de 90,0 porcento e logo o inverso disso é 1,1. Ou seja 110/100, dilui-se entre tods e ai o ajuste será de 1.1. Assim 50 respondentes passa de 45,5 para 55 como peso com ajuste.

13.5 Ajuste para não resposta:

O ajuste para não resposta pode ser simples ou elaborado dependendo do quanto se sabe dos respondentes.

Respostas podem ser tanto determinística como estocástica.

Deterministica: cada unidade pode ser tanto respondente ou não respondente. A escolha não é ao acaso so that que as unidades poderiam até serem sorteadas previamente entre respondentes e não respondentes.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Estochastica : cada unidade tem uma probabilidade não zero de responder. Quando perguntado para participar a unidade faz uma escolha aleatória de participar ou não.

Se for determinístico o viés seria a diferenta entre o estimador dos respondentes menos dos não respondentes. A idea de weighing class adjustment é que os pesos iriam igualar as medias dos respondentes e não respondentes.

A probabilidade de estar na amostra é Pr (Ii = 1) = πi enquanto a probabilidade de responder dado que a unidade i esta na amostra é Pr (Ri = 1/Ij= 1) = Øi chamado por Rosembaum e Rubin de propensity score para a unidade i. Se o propensite for 0 para algumas unidades, isso e probabilidade de nunca responder então pode causar bias. Se todas as unidades tem probabilidade não zero de responder ai sim podemos produzir estimadores não enviesados.

Suponha que doi = 1/ πi é o peso de base que foi desiginado a unidade i e considere que este simples estimador de uma media . Sob o setup de quase-randomization onde amostragem e resposta are both considerados serem at random, o viés será: formula do livro.

Em palavras o viés depende da covariance da variável de resposta e sua propensity score. Se yi e o propensity scores não são relacionadados, não tem bias e a não resposta não precisa ser corrigida pelo menos quando apenas estimando a media. Geralmente precisamos fazer alguma coisa para reduzir viés.

Missing completamente at random – se a probabilidade de resposta não depende de Yi ou de xi, então MCAR. Se a não tiver associado a nada então qualquer não respondente é missimg completely at random

Missing at random – se nao depende de Y mas depende de alguma variável coletada então missing data is MAR. As vezes é chamado de ignorable non-response, significando que se modelado corretamente e ajustes forem feitos , então a inferência é possível.

Nonginorable nonresponse – se observamos que a não resposta além das variáveis que foram coletadas para todos, esta também associado a algo que foi coletado dos respondentes, ai fica difícil dectar e corrigir bias.

Weighing class adjustments :

Se você puder criar grupos ou classes que tanto totas as unidades tem a mesma probabilidade de resposta ou about the same y-values, ai o viés de não resposta ira ser

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

aproximadamente eliminado. ASssi, a o set ideal de classes iera estar relacionadao tanto ao y e probabilidade de resposta. Porem a dificuldade pratica eh que não temos o y para quem não responde. E ainda mais, uma set de classes não será igualmente efetivo para todos os y´s. Consequentemente um set de classes is usually identified based em probabilidades. Se os covariantes foram também preditores de Y variables isso é um bônus.

Mecanismos de de utilizar classes para ajustar não resposta. Existem diferentes formas de formar classes. We index as classes by c = 1 .....C. O objetivo em formar classes is top ut units together que teria a mesma response propensity. Como ressaltado, é também desejável ter associação entre as medias de variáveis analisadas e a forma em que as classes são formadas. Se todas as unidades numa classse tiver o mesmo valor de covariante , xc, e propensão de respsta e uma função de cv, então oi = o (xc) para todas as unidades em c.

mdsaraiv@umich.edu

Prof. Dr. Maria da Conceição P. Saraiva

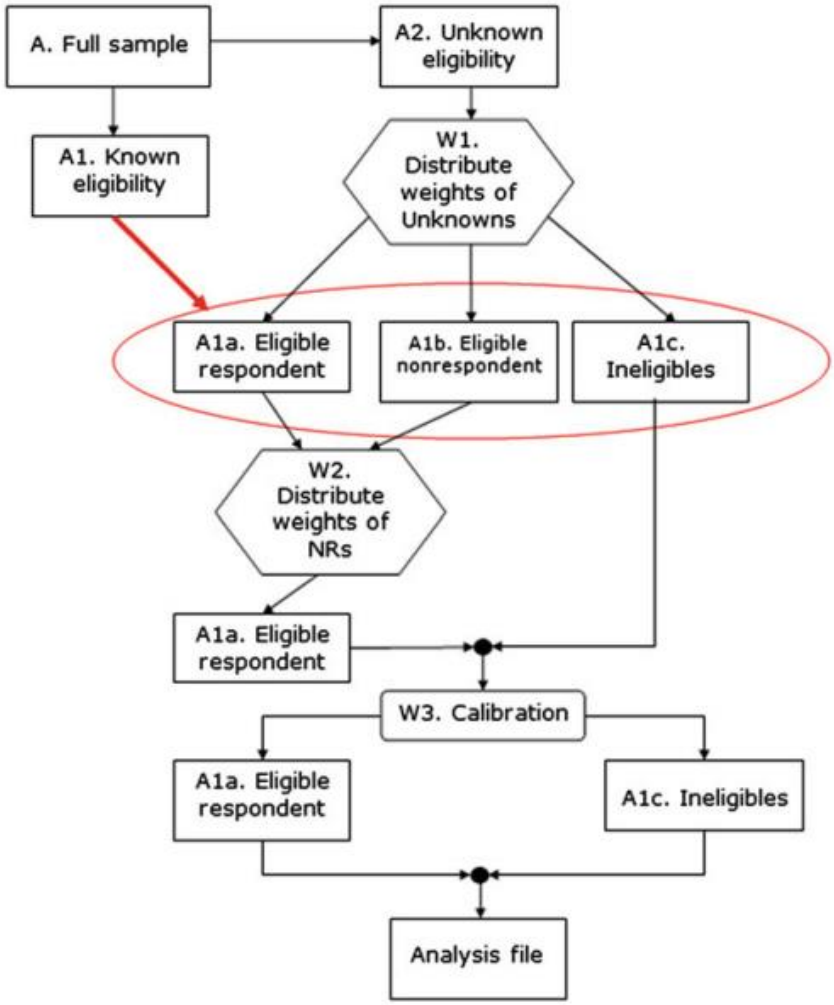
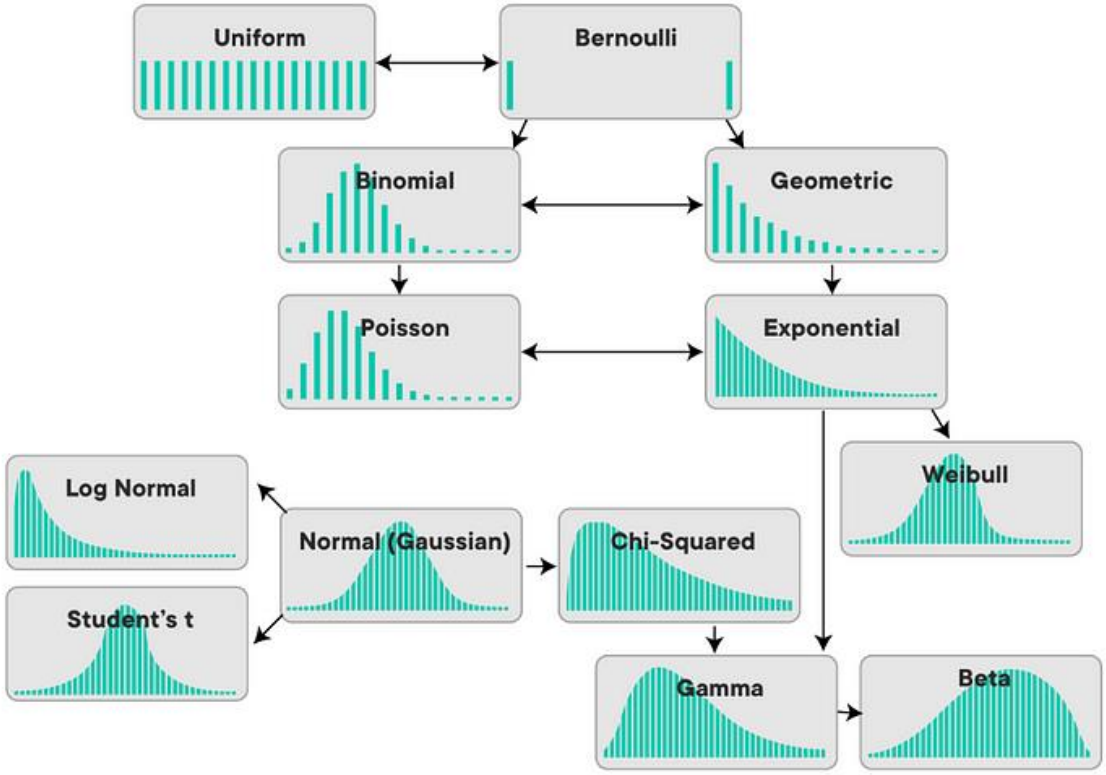


Fig. 13.1: General steps used in weighting.

mdsaraiv@umich.edu
Prof. Dr. Maria da Conceição P. Saraiva

Distribuições



De amatula

https://medium.com/@amanatulla1606/understanding-probability-distributions-in-statistics-and-its-application-in-machine-learning-e62751e83c39