

Estatística Prática para Docentes e Pós-Graduandos

de Geraldo Maia Campos

Prefácio

Que minhas primeiras palavras sejam para explicar por que este livro foi escrito, e por que o escrevi da maneira como ele se apresenta. Ele não surgiu evidentemente de nenhuma inspiração momentânea. Pelo contrário, é fruto de anos de meditação e a amparar-me tenho a experiência de mais de vinte anos dedicados à orientação e à elaboração da análise estatística de muitas centenas de trabalhos de pesquisa, tanto teses como artigos para publicação, e tanto de minha própria autoria como de outros pesquisadores, muitos destes bastante experientes em pesquisa, mas comumente pouco versados em Estatística. Nesses anos todos, eu era freqüentemente fustigado sempre pela mesma intrigante pergunta: por que teriam as pessoas tanta dificuldade em entender e em aplicar os métodos estatísticos, que a mim me pareciam tão lógicos e tão simples? Cheguei à conclusão de que o problema da Estatística deveria ser o mesmo da Matemática - que a grande maioria dos pesquisadores que a mim recorriam (geralmente da área biológica) ia declarando logo de início detestar cordialmente. Essa aversão generalizada à Matemática estendia-se pois à Estatística porque, por algum motivo, o conceito de Estatística parecia-lhes estar intimamente ligado ao de Matemática. Isso talvez até seja verdadeiro para quem se dedica a criar, desenvolver ou aperfeiçoar métodos e testes estatísticos, ou até mesmo para quem pretende programar esses testes em computador. Mas não o é para aqueles que são apenas usuários dos métodos e testes estatísticos, e não os seus idealizadores ou programadores. Nesse caso, por que a ojeriza generalizada à Estatística? E por que a maioria das pessoas não conseguia entender os seus métodos e a sua lógica, não obstante todas já tivessem feito um ou outro curso de Estatística em sua vida, às vezes até mais de um? Por que seria a Estatística considerada assim tão difícil? Cheguei à conclusão de que, se a Estatística não era na verdade tão difícil, então era fatal concluir que, se os estudantes não conseguiam entendê-la, era porque deveria estar sendo ensinada de uma forma incorreta. Mas, se estava sendo mal ensinada, qual seria a melhor maneira e o modo mais correto de fazê-lo? Muitas horas de meditação levaram-me por fim a desenvolver um método de ensino da Estatística, com base nos procedimentos que vinha adotando ao longo dos muitos anos em que atendi aos que me procuravam para ajudá-los a resolver problemas de Estatística relacionados com a interpretação dos resultados de seus experimentos. Tratava-se de pessoas provindas das mais diversas áreas do conhecimento humano, de modo que me via forçado a fazer muitas perguntas, para tentar entender o que cada uma pretendia com sua pesquisa.

Assim, acabei descobrindo algo muitíssimo importante: para poder ajudar alguém de uma forma eficaz, é forçoso conhecer o seu trabalho tanto quanto ele próprio, ou provavelmente até mais do que ele. Desse modo, com base nas perguntas que fazia, e nas respostas que costumava receber de meus consultentes, acabei aprendendo onde residiam as dúvidas da maioria das pessoas não afeitas aos métodos estatísticos, e o que deveria ensinar-lhes, para que futuramente fossem capazes de resolver elas próprias os seus problemas. Esse método é muito simples: consiste apenas em ensinar alguns poucos conceitos fundamentais e, a partir destes, traçar um roteiro lógico que habilite as pessoas não propriamente a deduzir fórmulas ou a realizar cálculos matemáticos, mas sim a reconhecer o modelo matemático em que se enquadram os seus experimentos e, com base nessas premissas, a decidirem elas mesmas sobre qual o teste estatístico mais adequado ao tratamento estatístico dos seus dados experimentais, a fim de poderem interpretar corretamente os seus resultados, evidenciar o seu verdadeiro valor e a sua real importância e, finalmente, tirar deles conclusões substanciais, pertinentes e relevantes. Esse é pois o objetivo deste livro.

1. A primeira pergunta

A primeira pergunta que um estaticista faz ao seu consulente — e que este deveria estar apto a responder — é esta: Qual (ou o quê) é a sua variável? Parece fácil, mas posso garantir, com base em minha experiência pessoal de muitos anos, que poucos darão uma resposta correta a essa indagação tão simples. Alguns dirão coisas como esta: "Usei três marcas diferentes de gessos, dois métodos diferentes de manipulação e quatro proporções diferentes de água / pó" — ou qualquer coisa semelhante, que varia conforme o campo de atividade do pesquisador. Assim, na opinião do pesquisador do exemplo acima, estaríamos diante de um experimento com três variáveis: a marca do gesso, a técnica de manipulação e a proporção água / pó. Porém, na verdade, nada disso constitui a variável do trabalho em questão. Mas é bastante comum que, quando o estaticista diz isso ao seu consulente, este arregale os olhos de espanto.

A identificação da variável.

No entanto, a identificação da variável deveria ser o primeiro passo na realização de qualquer trabalho de pesquisa — e de fato é, e todo pesquisador sabe perfeitamente qual é essa variável. Todavia, a sua falta de vivência estatística geralmente impede que ele a identifique como tal. Falta-lhe um roteiro lógico, um caminho bem definido, uma orientação clara, que ele possa seguir, sabendo perfeitamente o que faz e por que faz. O planejamento estatístico de uma pesquisa deve delinear-se quando ainda na fase de elaboração do projeto de pesquisa (ou do plano de trabalho), pois já nesse momento o investigador deve pensar sobre qual tratamento estatístico aplicará futuramente aos seus dados experimentais, para conseguir resultados, tirar conclusões, e obter respostas às indagações iniciais que motivam e justificam a realização dos seus experimentos. Voltando à pergunta inicial, é bem provável que o consulente só entenda a pergunta de seu conselheiro estatístico quando este a reformular e indagar: Afinal, o que foi que você mediu (ou contou, ou pesou, ou qualquer coisa do mesmo gênero)? Então, considerando ainda o mesmo caso dos gessos que estamos tomando como exemplo, o pesquisador dirá categoricamente: "Eu medi a dureza dos corpos-de-prova de gesso, construídos com cada marca de gesso, cada técnica de manipulação e cada proporção água / pó!". Eis aí, finalmente, a verdadeira variável do experimento: o grau de dureza dos corpos-de-prova de gesso! A sua variável é portanto precisamente aquele elemento que

permite a comparação entre todas as combinações possíveis das marcas de gesso, técnicas de manipulação e proporções de água/pó utilizadas na pesquisa.

A variável única e o denominador comum.

A variável do experimento, portanto, tem necessariamente de ser única, porque só assim poderá servir como um denominador comum no confronto entre tudo aquilo que se deseja comparar numa pesquisa, seja ela qual for. Mas o que vem a ser um denominador comum? Apesar de ter jurado que não falaria em Matemática neste curso, vou responder a essa pergunta com outra pergunta, de ordem puramente aritmética: qual das duas frações abaixo representa a grandeza maior?

$$\frac{177}{323} \text{ ou } \frac{234}{544}$$

E agora, entre as duas novas frações abaixo, qual seria a de maior grandeza ?

$$\frac{63.646}{175.712} \text{ ou } \frac{75.582}{175.712}$$

Agora sim, tomou-se muito fácil garantir que a segunda fração é maior do que a primeira, mesmo que os números envolvidos no segundo exemplo sejam muito maiores que os do primeiro e isso sem precisar fazer mais do que um simples exame visual das duas frações. Mas... por que seria assim? Na verdade as duas frações do primeiro conjunto são exatamente iguais às duas frações do segundo conjunto. A única diferença é que, neste último, as frações foram reduzidas ao mesmo denominador, calculando-se o seu denominador comum, uma tarefa matemática elementar, que aprendemos no curso primário, ao estudarmos frações ordinárias. A variável de um experimento é pois o denominador comum — ou seja, o termo de comparação — que permite cotejar seja lá o que for que queiramos comparar. Por isso tem de ser única ou então cairíamos no caso do primeiro conjunto de frações ordinárias mostrado em nosso exemplo matemático, em que os denominadores são diferentes.

Fatores de variação.

Muito bem, mas se, na pesquisa sobre gessos que estamos adotando como exemplo no presente capítulo, a variável é a dureza dos corpos-de-prova, o que seriam afinal as marcas de

gesso, as técnicas de manipulação e as proporções água / pó? É evidente que tudo isso é importante, ou não seria considerado na pesquisa. Na verdade, são exatamente esses fatores que fazem com que a variável dureza realmente varie. São portanto fatores de variação. Os fatores de variação, ao contrário da variável, podem ser múltiplos, não havendo teoricamente um limite para o seu número. A experiência, porém, bem como o bom-senso que costuma dela advir, aconselha que esse número não deva ser superior a três, e a razão disso será analisada quando se falar sobre algo muito importante em Estatística, que são as interações entre os diversos fatores de variação.

Os dois primeiros passos.

Resumindo o que foi dito até agora, podemos finalmente indicar os dois primeiros passos a serem dados na preparação da análise estatística dos dados experimentais de uma pesquisa, seja esta qual for, esteja ela ainda na fase inicial de planejamento, ou já no seu final, com todos os experimentos realizados e todos os dados experimentais obtidos e convenientemente anotados nos protocolos elaborados para o registro das observações. Esses dois passos iniciais são:

1º passo - Identificação da variável,

2º passo - Identificação dos fatores de variação.

As repetições (ou réplicas).

Contudo, esses dois elementos — variável e fatores de variação — não são os únicos que devem ser definidos logo no início de um experimento. Há ainda outro, de capital importância, que muitas vezes constitui uma verdadeira dor de cabeça para o pesquisador: o número de repetições (ou réplicas) a ser adotado nos experimentos. Aliás, é preciso dizer que uma das indagações que os estatísticos mais ouvem de seus consultantes, tanto de pesquisadores, como de pós-graduandos, e até mesmo dos orientadores destes, é esta: qual o número ideal de repetições num experimento, para tornar confiáveis os resultados e sua interpretação estatística? Pois bem, vamos responder com a mais absoluta certeza e segurança a essa pergunta: não existe tal número ideal de repetições! Nesse caso, dirão os leitores deste texto, completamente atônitos e decepcionados, o que determinaria qual o número de repetições a ser adotado num experimento? A resposta a esta reformulação da mesma pergunta será dada mais adiante, e não agora, a esta altura deste curso, pois na verdade ainda

há muita coisa importante a ser comentada, antes de chegarmos a esse detalhe. Um detalhe, aliás, importantíssimo, como se verá mais à frente, porque sem repetições é muito provável que a Estatística, tal como a conhecemos hoje, nem sequer existisse.!

2. Tipos de variáveis

Variáveis contínuas e discretas.

Grandezas como comprimento, área, volume, peso, tempo, proporções, porcentagens, ângulos, valores das funções trigonométricas, temperatura, etc., que num determinado intervalo podem tomar quaisquer valores, sejam estes inteiros ou fracionários, são chamadas variáveis contínuas. Mas grandezas outras, tais como a contagem de pessoas, a soma do número de pontos no lançamento simultâneo de três dados, o número de culturas bacterianas positivas, o número de respostas sim, ou de respostas não, o número de gols por rodada de um campeonato de futebol, e assim por diante, não admitem valores fracionários, e por isso são denominadas variáveis discretas, ou seja, variáveis que só podem variar por unidades inteiras. Todavia, no mesmo exemplo dos gols em partidas de futebol (por exemplo), a média de gols por partida de cada equipe é uma variável contínua, porque a média pode tomar valores intermediários, fracionários, ao passo que o número de gols, pura e simplesmente não admite fragmentos de gols, que devem ser contados por unidades inteiras, sendo pois, neste caso, uma variável discreta. Esses tipos de variáveis podem ambos ser utilizados em Estatística, mas o simples fato de a variável de um experimento pertencer a um ou outro tipo já constitui um fator de seleção para a indicação de um ou de outro grupo de testes. É preciso pois definir também em qual desses grupos a variável de um experimento se encaixa.

Variáveis ordenadas.

Mas os tipos de variáveis não se esgotam com esses dois, que na verdade são os mais comuns porém não os únicos. De fato, pesquisas existem em que a única característica dos dados experimentais que poderia ser usada para classificá-los é o fato de eles poderem ser ordenados, de forma crescente ou decrescente. É o caso, por exemplo, de um experimento em que se queira avaliar a intensidade dolorosa, ou a positividade de um teste para sífilis ou qualquer outra doença, usando grupos de sinais (+) para graduar a dor ou a positividade do teste: (++++) maior que (+++) maior que (++) maior que (+), além do caso negativo (-).

Os escores (variáveis subjetivas).

Um recurso muito usado nesse caso para quantificar os dados é substituir os sinais por números: 4, 3, 2, 1 e 0. Esse tipo de graduação (ou de notas, ou de escores) é também bastante usado, quando os dados experimentais traduzem apenas uma impressão subjetiva ou a opinião pessoal — visual, por exemplo — de um grupo de observadores a respeito de um fenômeno qualquer que esteja sendo estudado. Na opinião do autor destas linhas os escores, dada a sua natureza subjetiva, não é uma variável muito forte, mas situações e tipos de pesquisa há em que não há alternativa senão usá-los.

Variáveis nominais.

Além disso, as variáveis podem ser também nominais, como em experimentos que envolvem perguntas que só admitem duas respostas: sim ou não. Também neste caso, os dados nominais acabam se tornando numéricos quando se consideram o número de respostas sim e o de respostas não. Isso significa que, mesmo quando a pesquisa envolve dados de natureza apenas qualitativa, esses dados terão forçosamente de ser transformados em dados quantitativos, para poderem ser analisados estatisticamente.

Dependência ou independência dos dados.

Outra característica importante dos dados experimentais, que deve ser levada em conta, é se eles são independentes ou vinculados (dependentes), que serão comentados mais adiante, com maiores detalhes.

O terceiro passo.

É necessário, pois, saber também, desde o início da pesquisa, quais as características e qual o tipo de variável utilizado, porque essas informações irão sem dúvida condicionar o uso de um ou de outro grupo de testes estatísticos, por ocasião do tratamento dos dados experimentais, com vistas à interpretação correta dos resultados da pesquisa. Este seria portanto o terceiro passo de nosso roteiro para o planejamento estatístico:

3º passo - Identificação do tipo de variável utilizado

3. As repetições

Experimento com uma única observação.

Imaginemos que, num experimento qualquer, se fizesse apenas uma única observação. Esse experimento, é evidente, não teria um valor numérico médio, ou uma média, porque não haveria valores que se pudessem somar, nem um número pelo qual a soma desses valores pudesse ser dividido para calcular essa média. Enfim, em última instância, a soma dos dados (digamos assim) seria igual ao próprio valor único, e o número de dados seria 1, de forma que o valor do dado dividido por 1 seria o seu próprio valor original. Levando esse raciocínio ad absurdum, diríamos então que o valor do dado seria exatamente igual ao valor da média. Isso seria ótimo, poderia pensar alguém menos avisado. Não, não seria. Na verdade, seria péssimo, até mesmo desastroso. Isto porque, se o valor medido estivesse errado, tudo mais estaria também errado, inclusive quaisquer eventuais conclusões que se pudessem tirar desse resultado falso.

Experimento com mais de uma observação.

Todavia, se o número de observações do mesmo fenômeno fosse aumentado para 2, ou 3 ou 10, ou para qualquer outro número maior do que 1, o pesquisador notaria um fato interessante: as medidas apresentariam diferenças entre si, mesmo que ele repetisse sempre os mesmos passos na execução dos experimentos, e mesmo que usasse sempre o mesmo observador para executar as medidas. Enfim, haveria diferenças, mesmo que ele fizesse tudo exatamente igual, desde o começo até o fim de sua pesquisa. Isso seria péssimo, poderia pensar aquele mesmo alguém que já fizera o comentário do parágrafo anterior. Mas na verdade ele estaria novamente enganado, e isso na verdade seria ótimo. Isto porque, mesmo que um, ou alguns, ou mesmo todos os valores medidos estivessem errados, o valor médio desses valores errados estaria sempre mais próximo do valor real daquilo que estava sendo medido, do que muitas vezes qualquer dos dados experimentais considerado isoladamente.

A média dos valores dos dados experimentais.

Do que foi exposto no item anterior, depreende-se que a média tende a aproximar os valores errados do valor real daquilo que se mede. Isto porque a média é uma espécie de

limite central, para o qual tendem naturalmente a convergir os erros de medida, considerando que estes podem ser para maior ou para menor, em relação ao valor real daquilo que é medido. Enfim, se não houvesse erro nenhum de medida, todas as medidas efetuadas seriam iguais à média, pois não haveria diferenças nem para maior, nem para menor, em relação ao valor real. É por esse motivo que os matemáticos chamam a média de esperança matemática, porque a média é o valor que se esperaria obter, caso não houvesse erros e todos os valores medidos fossem iguais.

O número de dados.

De tudo o que se expôs acima, pode-se concluir que, quanto maior o número de repetições, tanto mais o valor médio se aproximará do valor real, o que é absolutamente verdadeiro. Porém é preciso considerar que o número de repetições pode teoricamente estender-se desde 2 até o infinito. Contudo, um número infinito de observações, ou de medidas, é absolutamente impraticável. Aliás, mesmo considerando um número de medidas finito porém demasiadamente grande, ainda que fosse praticável, não seria todavia prático.

Amostragem, probabilidade e significância.

Deve existir portanto um número de repetições que, mesmo sendo finito, e por isso mesmo limitado, seja capaz de permitir que se possam tirar conclusões válidas a respeito de um fenômeno qualquer que se queira estudar. A Estatística procurou resolver esse problema pela associação de duas coisas cujos nomes nos habituamos a ouvir a toda hora, quando lidamos com testes estatísticos: a amostragem e a probabilidade. A significância dos valores calculados pelos diversos testes estatísticos — que é algo também comumente ouvido e discutido em Estatística — em última análise não é mais do que a probabilidade de serem corretas as conclusões tiradas a partir de amostras de dimensões limitadas, reduzidas, retiradas de conjuntos de dados às vezes infinitamente maiores do que a própria amostra analisada pelos testes.

O quarto passo.

A esta altura do planejamento experimental, poder-se-ia acrescentar mais um passo em nosso roteiro estatístico:

4º passo - Estabelecer o número de repetições.

4. As repetições e o experimento-piloto

O número mais adequado de repetições.

O estabelecimento do número de repetições põe novamente em cena a mesma velha pergunta: qual o número mais adequado de repetições? E a resposta seria ainda a mesma já dada anteriormente: não existe tal número. Pelo menos não existe nenhum número mágico — poder-se-ia dizer mesmo cabalístico — que pudesse servir indiferentemente a qualquer experimento. O que há, na verdade, é um número mais adequado de repetições para cada experimento — um número que varia de um experimento para outro, e que precisa portanto ser calculado para cada um deles.

A variabilidade dos dados experimentais.

Mas como fazer esse cálculo? Que leis regem a escolha desse número? A isto, sim, é possível responder: o que rege a escolha do número de repetições mais adequado a um experimento qualquer é a variabilidade dos seus dados experimentais. Mas — poderá objetar alguém — se o experimento ainda não foi realizado, como se pode conhecer o seu grau de variação, ou avaliar a sua variabilidade? É precisamente aí que entram dois novos elementos igualmente muito importantes na execução de qualquer experimento: o experimento piloto e a verificação preliminar da variabilidade dos dados experimentais.

O experimento piloto.

O experimento piloto é aquele que se faz previamente à realização da pesquisa propriamente dita, e visa a testar o método de trabalho e os processos técnicos envolvidos na execução dos experimentos. Em geral o piloto segue o mesmo plano geral de trabalho, que orienta a investigação como um todo. Todavia, difere dele num ponto: no número de repetições — e o faz exatamente porque, a essa altura, o número mais adequado de réplicas ainda não foi fixado de forma definitiva. No experimento piloto, a variável é a mesma já definida, os fatores de variação são os mesmos já estabelecidos para a pesquisa, mas em geral o número de repetições é pequeno, para não tornar o experimento-piloto muito trabalhoso ou muito demorado.

O número inicial de repetições.

É bastante comum a escolha de 3 repetições, como um bom número para começar. Isto porque o número 3 evita a invariabilidade do número 1, foge ao perigo da repetição coincidente representado pelo número 2, e já apresenta alguma variação, a qual no mais das vezes já é suficiente para testar a variabilidade determinada pelos fatores de variação e pelas próprias repetições.

A variabilidade dos dados no experimento piloto.

A verificação preliminar da variabilidade dos dados experimentais é feita após a execução do experimento-piloto. Para isso, faz-se uma análise de variância dos dados obtidos nesse piloto, sem qualquer preocupação quanto ao tipo de distribuição dos dados, e seja lá qual for o número de repetições nele fixado. Realizada essa análise de variância preliminar, faz-se então o teste estatístico para determinar, especificamente para esse experimento, qual seria aquele misterioso e tão procurado número mais adequado de repetições.

O número mais adequado de repetições.

Esse teste é do tipo iterativo e requer um programa de computador, pois de outra forma ele seria muito demorado e trabalhoso. O teste é chamado iterativo porque, partindo do número inicial de repetições do experimento-piloto, ele calcula um novo número de repetições, e volta a introduzir no teste esse número calculado, à guisa de novo valor inicial, recalculando tudo e achando outro número de repetições. Assim, sucessivamente, vai recalculando até que o número de entrada do teste seja igual ao número de saída. Quando essa igualdade ocorre, o teste é dado por terminado, e esse último número de repetições constitui o número mais adequado de repetições para aquele experimento. O teste sugere pois que, para aquela variação detectada pelo experimento-piloto, é necessário aquele número mínimo de réplicas, para que se possam perceber diferenças estatisticamente significantes entre os fatores de variação estudados



5. O experimento fatorial

Uma vez definida a variável, estabelecidos os fatores de variação, e fixado o número de repetições necessário para possibilitar a detecção de possíveis diferenças significantes entre os fatores de variação, tem-se perfeitamente delineado todo o plano da pesquisa. A partir dessas informações, torna-se possível conhecer uma série de detalhes importantes, tanto no que respeita ao experimento em si como à análise estatística posterior dos dados.

O número total de dados do experimento.

A essa altura da pesquisa, já é possível calcular o número total de dados (n) a ser obtido após realizados todos os experimentos. Esse número é fornecido pela multiplicação do número de colunas (c), pelo de linhas (l), pelo de blocos (b), e pelo de repetições (r). Matematicamente, poderíamos representá-lo pela seguinte igualdade:

$$n = c \times l \times b \times r$$

Em virtude de essa multiplicação ser um produto, ou seja, uma multiplicação de fatores, esse tipo de experimento é chamado de experimento fatorial.

E se faltar um fator de variação na equação?

A expressão matemática acima transcrita vale sempre, mesmo que falte um ou mais dos seus fatores. Só que, quando falta qualquer dos fatores, este jamais será igual a zero (0) mas igual a um (1), uma vez que, se fosse igual a zero (0), o produto todo se anularia, e o número total de dados seria igual a zero. Assim, quando um fator não existe, ele não existe apenas aparentemente, porque na verdade existe sim, porém possui apenas um elemento, sendo portanto realmente igual a um (1). Desse modo, se um experimento apresenta apenas um fator de variação (colunas), além das repetições (que devem existir sempre), a expressão do número de dados será fornecida pela igualdade:

$$n = c \times 1 \times 1 \times r$$

que equivale a

$$n = c \times r$$

uma vez que a multiplicação por um (1) não altera o produto. O mesmo ocorre, se houver apenas dois fatores (colunas e linhas), caso em que o número de dados será:

$$n = c \times l \times 1 \times r \quad \text{ou} \quad n = c \times l \times r$$

A única alternativa que jamais poderá ocorrer é a existência apenas de blocos como fator de variação, porque, por definição, bloco é um conjunto de colunas e linhas. Se estas forem ambas iguais a um (1), a idéia de bloco se confundiria com a de coluna, porque haveria então apenas um fator de variação. O mesmo pode-se dizer em relação a um experimento que envolva apenas o fator de variação linhas, uma vez que neste caso seria indiferente colocar as repetições em cada linha ou em cada coluna.

A distribuição dos dados numa tabela.

Por convenção, ou apenas por hábito, é comum reunir os dados da seguinte maneira:

a) em colunas, quando há apenas um fator de variação; b) em colunas e linhas, quando há dois fatores de variação; e c) em colunas, linhas e blocos, quando há três fatores de variação.

O protocolo das observações experimentais.

A determinação do número de fatores de variação e do número de repetições possibilita ao pesquisador construir uma tabela de dados, antes mesmo que qualquer desses dados tenha sido obtido. É costume, ao se planejar uma pesquisa, elaborar o chamado protocolo das observações, que em última análise não é mais que a ficha onde são anotadas todas as informações que possam ter interesse na investigação, tais como identificação dos pacientes, dos corpos-de-prova, ou dos animais de laboratório, além de informações complementares relevantes, como idade, peso, sexo, etc., informações essas que variam muito e dependem do tipo de pesquisa realizada. O protocolo das observações é absolutamente necessário, porque é ali que fica registrado praticamente todo o andamento da pesquisa. Todavia, o pesquisador pode elaborar também, paralela e simultaneamente, uma tabela vazia de dados, espécie de grade, onde já está indicado previamente o lugar onde será colocado o valor numérico referente a cada um dos dados experimentais, à medida que estes vão sendo obtidos na fase experimental da pesquisa. Assim, quando o experimento chegar ao fim, o pesquisador terá em mãos a sua tabela geral de dados, já completa e acabada.

6. A tabela geral de dados

O próprio título deste capítulo já sugere claramente que a tabela com os dados experimentais deva ser abrangente, única e completa. Ou, em outras palavras, todos os dados obtidos devem estar contidos numa tabela única, na qual constem todos os elementos que compõem o fator de variação colocado nas colunas, todos os que compõem as linhas, e todos os que integram os blocos, além, é claro, de todas as repetições. A maneira como esses três fatores são distribuídos (como colunas, linhas ou blocos) depende muito do espaço físico disponível, principalmente considerando que modernamente as tabelas são comumente elaboradas em computador, nos quais o espaço é limitado, principalmente no sentido horizontal da tela do monitor, ou seja em relação ao espaço destinado às colunas. Quanto às linhas e blocos, caso seja necessário, podem alongar-se no sentido vertical, podendo passar à página seguinte, e portanto sem qualquer problema de limitação do espaço. O ideal, contudo, seria que a tabela geral de dados ocupasse apenas uma página, pois isso permitiria o exame visual do conjunto de dados experimentais todos de uma só vez. Isso pode ser conseguido, inclusive em computadores, pela redução do tamanho dos caracteres, o que permite escrever um número maior de caracteres por linha, na tela do monitor, e também no papel quando a tabela é impressa. Quando o número de colunas é pequeno, os blocos poderão ser colocados lado a lado (no sentido horizontal, se o espaço permitir, de modo que a tabela terá, verticalmente, a extensão dada pelo número de linhas e de repetições. Se o número de colunas da tabela for muito grande, ocupando uma grande extensão horizontal, inviabilizando a colocação dos blocos lado a lado, estes poderão ser colocados no sentido vertical, um embaixo do outro. Neste caso, a extensão vertical será dada pelo número de linhas multiplicado pelo número de blocos e de repetições. Em suma, confeccionar tabelas é, na verdade, uma questão de prática, uma vez que esta acaba habilitando o pesquisador a decidir rapidamente sobre qual a melhor conformação física para qualquer tabela de dados que tenha eventualmente de construir. O que foi dito acima é apenas uma sugestão de como começar, a fim de vir um dia a adquirir essa prática. A seguir, estão transcritos alguns modelos de tabelas de dados, identificadas estas por letras maiúsculas. Por exemplo, tecnicamente, não se pode dizer que haja diferença entre as tabelas A, B, C e D.

Tabela A. Exemplo de tabela com apenas colunas (4) e repetições (10).

Tratamento 1	Tratamento 2	Tratamento 3	Tratamento 4
123,56	132,77	145,00	234,12
116,72	145,87	175,33	304,21
120,09	177,55	147,55	244,98
123,89	188,73	145,66	234,66
113,87	198,32	165,00	238,54
25,00	145,11	143,54	300,67
177,96	156,59	149,22	341,15
122,46	177,21	147,10	203,72
111,21	145,09	148,22	255,19
122,12	155,01	146,03	245,00

Tabela B. Exemplo de tabela com apenas colunas (4) e repetições (10).

Tratamento 1		Tratamento 2		Tratamento 3		Tratamento 4	
123,56	125,00	132,77	145,11	145,00	143,54	234,12	300,67
116,72	177,96	145,87	156,59	175,33	149,22	304,21	341,15
120,09	122,46	177,55	177,21	147,55	147,10	244,98	203,72
123,89	111,21	188,73	145,09	145,66	148,22	234,66	255,19
113,87	122,12	198,32	155,01	165,00	146,03	238,54	245,00

Tabela C. Exemplo de tabela com apenas linhas (4) e repetições (10).

Tratamento 1	123,56	116,72	120,09	123,89	113,87	125,00	177,96	122,46	111,21	122,12
Tratamento 2	132,77	145,87	177,55	188,73	198,32	145,11	156,59	177,21	145,09	155,01
Tratamento 3	145,00	175,33	147,55	145,66	165,00	143,54	149,22	147,10	148,22	146,03
Tratamento 4	234,12	304,21	244,98	234,66	238,54	300,67	341,15	203,72	255,19	245,00

Tabela D. Exemplo de tabela com apenas linhas (4) e repetições (10).

Tratamento 1	123,56	116,72	120,09	123,89	113,87
	125,00	177,96	122,46	111,21	122,12
Tratamento 2	132,77	145,87	177,55	188,73	198,32
	145,11	156,59	177,21	145,09	155,01
Tratamento 3	145,00	175,33	147,55	145,66	165,00
	143,54	149,22	147,10	148,22	146,03
Tratamento 4	234,12	304,21	244,98	234,66	238,54
	300,67	341,15	203,72	255,19	245,00

A única diferença entre as tabelas A e B reside no fato de as repetições, na tabela B, terem sido divididas em dois grupos de cinco, dentro da mesma coluna, ao invés de um só grupo com todas as dez repetições, como se vê na tabela A. A diferença entre essas duas primeiras tabelas (A e B) em relação às duas outras (C e D) está no fato de as repetições nestas últimas terem sido dispostas em linhas, e não em colunas, como nas duas primeiras. Por seu turno, a

diferença entre as tabelas C e D está também na sua disposição em dois grupos de cinco repetições para cada linha, na tabela D, e em apenas um grupo com as dez repetições, na tabela C. A opção por qualquer desses quatro tipos de tabelas é apenas uma questão de conveniência, tal como a disponibilidade de espaço em função do número de colunas ou de linhas, ou a maior facilidade ou comodidade na introdução dos dados numéricos no computador, ou às vezes até mesmo por simples conveniência estética. Porém, do ponto de vista puramente técnico, todos os quatro tipos de tabelas apresentados são aceitáveis para esse modelo matemático de experimentos, que envolve apenas um fator de variação, esteja este colocado em colunas ou em linhas. Todavia, por uma espécie de convenção, é costume dispor os dados em colunas, e não em linhas, quando há apenas um único fator de variação, tal como se fez nas tabelas A e B. Mas como ficaria uma tabela que envolvesse tanto colunas como linhas? Imagine-se, por exemplo, um modelo experimental que envolvesse quatro Tratamentos aplicados a dois grupos de pacientes (Controle e Tratado), com cinco repetições (pacientes) em cada grupo. Como seria a tabela para esses dados experimentais? Poderia ser assim:

Tabela E. Exemplo de tabela com colunas (4), linhas (2) e repetições (5)

Grupos	Tratamentos			
	Droga A	Droga B	Droga C	Droga D
Controle	123,56	132,77	145,00	234,12
	116,72	145,87	175,33	304,21
	120,09	177,55	147,55	244,98
	123,89	188,73	145,66	234,66
	113,87	198,32	165,00	238,54
Tratado	125,00	145,11	143,54	300,67
	177,96	156,59	149,22	341,15
	122,46	177,21	147,10	203,72
	111,21	145,09	148,22	255,19
	122,12	155,01	146,03	245,00

No caso específico da tabela acima, essa é a configuração mais adequada — com os Tratamentos nas colunas e os grupos experimentais nas linhas. Isto porque, se os Tratamentos estivessem nas linhas e os Grupos experimentais nas colunas, a tabela ficaria muito alongada no sentido vertical, e muito estreita no sentido horizontal, tal como uma lingüiça gráfica a estender-se de cima para baixo — ou seja, antiestética e pouco prática, uma vez que, dependendo do número de repetições, poderia abranger mais de uma página de texto. Mas

nada proíbe que qualquer dos fatores de variação possa ser colocado indiferentemente nas colunas ou nas linhas. É uma simples questão de conveniência gráfica. A única exigência é que as repetições fiquem reunidas na célula da tabela que corresponde ao cruzamento de uma linha com uma coluna. Contudo, há ainda mais um elemento que pode complicar a elaboração de uma tabela de dados: a existência de blocos, ou seja, de um terceiro fator de variação. Quando isso ocorre, cada bloco será uma reedição do modelo para colunas e linhas reproduzido acima, e envolverá tantas novas tabelas (com colunas e linhas) quantos forem os elementos que compõem o fator de variação a que os blocos se referem. Por exemplo: imagine-se que, além dos Tratamentos (A, B, C e D) e dos Grupos experimentais (Controle e Tratado), a pesquisa envolva também três Tempos de observação (1, 3 e 7 dias). Como ficaria a nova tabela de dados, nesse caso? Ainda nesse caso, o critério que vigora é apenas a conveniência gráfica, para decidir qual fator de variação será colocado nas colunas, qual estará nas linhas, e qual ficará nos blocos. É, portanto, pura questão de bom-senso, associado ao bom-gosto, ou ao senso estético de cada um, os quais podem ser comentados e até criticados, mas sem dúvida jamais ensinados. Na página seguinte há uma sugestão para a construção da tabela com os três fatores de variação acima mencionados. Essa tabela ilustrativa foi deixada deliberadamente vazia, sem nenhum dado numérico transcrito, com o propósito único de mostrar que uma tabela vazia de dados pode perfeitamente ser elaborada antes mesmo que qualquer dado experimental tenha sido obtido. À medida que a pesquisa vai se desenvolvendo, os dados irão surgindo e serão anotados na tabela vazia, até preenchê-la toda quando do final da pesquisa. Para elaborar a tabela vazia, basta saber — e isso é sempre possível — quantos são os fatores de variação, quantos elementos integram cada um deles, e qual o número de repetições estabelecido.

Tabela F. Tabela vazia, para 3 fatores de variação, com 3 colunas, 4 linhas, 2 blocos e 5 repetições.

Tratamentos	Grupos Experimentais					
	Controle			Tratado		
	Tempos de Observação			Tempos de Observação		
	1 dia	3 dias	7 dias	1 dia	3 dias	7 dias
Tratamento (A)						
Tratamento (B)						
Tratamento (C)						
Tratamento (D)						

Como observação derradeira no que diz respeito às tabelas, e tomando como exemplo a tabela transcrita acima, é preciso esclarecer que se consideram como pertencentes à mesma linha todas as repetições que se referem mesmo Tratamento, incluindo-se aí as repetições de todas as colunas e de todos os blocos desse Tratamento. Da mesma forma, consideram-se como pertencentes à mesma coluna todas as repetições relativas a cada um dos Tempos, independentemente das linhas e dos blocos a que estejam ligadas. E, por fim, pertencem também ao mesmo bloco todas as repetições de cada um dos grupos (Controle e Tratado, no caso da tabela que serve de exemplo), sem considerar as linhas e as colunas de cada grupo. Em resumo: cada Tratamento (linha) teria, na verdade, 30 repetições; cada Tempo (coluna) teria 40 repetições; e cada Grupo experimental (blocos Controle e Tratado) teria 60 repetições. O produto do número de elementos de cada fator de variação pelo número de repetições correspondente dá sempre o mesmo número, que é o número total de dados experimentais: $4 \times 30 = 120$, $3 \times 40 = 120$; e $2 \times 60 = 120$. Esse conceito de repetições em relação aos fatores de variação é importantíssimo em Estatística, pois é ele que permite avaliar o efeito exercido exclusivamente pelo fator de variação sobre o valor numérico da variável estudada, ou seja, sobre o valor do dado experimental, separando-o dos efeitos determinados pelas chamadas interações, efeitos esses que resultam da associação de cada um dos fatores de

variação com os demais, combinação essa que pode muitas vezes alterar o efeito produzido por qualquer dos fatores de variação, quando considerado individualmente (ou separadamente

7. A fase pós-experimental

A esta altura de nossas considerações, a variável e os fatores de variação já foram identificados, o tipo de variável foi reconhecido, e a tabela de dados construída e preenchida com os dados obtidos na fase experimental. E agora? O que fazer com essa tabela e com os dados numéricos nela contidos?

Pequeno retrospecto das fases iniciais

Ficou dito, em capítulos anteriores, que o tratamento estatístico deve ser cogitado já nos primórdios da pesquisa, quando ainda se está na elaboração do seu projeto inicial, ou mesmo durante a fase de execução do plano-piloto dos experimentos. De fato, há muita coisa que já pode ser pensada e estudada nessas fases iniciais, em termos de tratamento estatístico dos dados. Porém há também outros detalhes que somente podem ser considerados após ter em mãos os dados numéricos colhidos nos ensaios. Entre as coisas que podem ser verificadas ainda na fase inicial da pesquisa está a vinculação, ou independência, dos dados experimentais, mas o conhecimento desse pormenor ainda não é tão relevante a essa altura do desenvolvimento da pesquisa, ou de seu tratamento estatístico. Mas se-lo-á mais adiante, ocasião em que o assunto será abordado novamente, tecendo-se então sobre ele considerações mais elaboradas.

A distribuição dos erros experimentais

Muito mais importante, todavia, seria analisar agora a distribuição de freqüências dos dados experimentais; ou, mais apropriadamente, estudar a maneira como os erros desses dados se distribuem em torno da média. Enfim, é preciso saber se a distribuição dos erros experimentais em torno da média é normal, ou seja, se o seu histograma de freqüências segue a configuração geral da curva matemática conhecida como curva normal.

Mas... poderão perguntar os eventuais leitores deste texto, por que isso seria assim tão importante? A resposta é: porque os primeiros testes estatísticos, talvez os mais importantes de quantos foram desenvolvidos pelos estudiosos, tiveram por premissa que a distribuição dos erros deveria ser normal, ou seja, que ela deveria ter uma distribuição de freqüências semelhante à da curva de Gauss, também chamada curva normal ou curva dos erros.

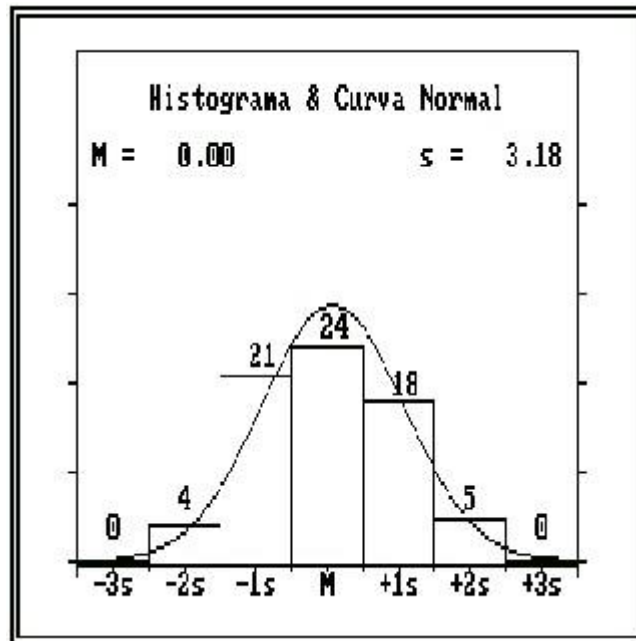


Figura 1. Histograma de freqüências de um conjunto de dados experimentais, sobreposto à curva normal matemática com a mesma média e o mesmo desvio-padrão.

Por que curva "normal"?

A denominação curva de Gauss explica-se porque foi esse notável matemático alemão quem encontrou a sua equação matemática. Da mesma forma, a expressão curva dos erros também se justifica, porque Gauss deduziu a sua equação matemática precisamente a partir de estudos realizados sobre a distribuição dos erros de medida em torno da média, ou seja, a lei matemática que rege a dispersão e o afastamento dos valores de medida em relação ao seu valor médio; ou, mais exatamente, em relação ao valor real da grandeza medida. Sim, tudo isso é compreensível. Mas por que essa curva seria chamada normal?

Na verdade, eu não sei nem nunca li qualquer explicação racional para isso. Acredito, porém, que essa denominação tenha algo a ver com os fenômenos naturais, tal como ocorre com outras curvas matemáticas, que traduzem fenômenos normalmente encontrados na natureza. É o caso, por exemplo, da curva chamada catenária (do latim catena = cadeia, corrente), cuja equação expressa matematicamente a curva natural descrita por uma corrente metálica, quando presa pelas extremidades e submetida à ação do próprio peso.

Agora, caros leitores, mentalizem, por exemplo, uma ampulheta e pensem: qual seria a equação matemática que descreve o perfil do montículo de areia que flui dentro dela e se deposita no seu compartimento inferior? Ou qual seria a equação matemática capaz de descrever o perfil do montículo que se forma, quando se despeja sobre o solo um saco de grãos de um cereal qualquer? Eu pessoalmente estou convencido de que, muito provavelmente, seria uma curva dessa família de curvas conhecidas como curvas normais. Talvez derive daí a denominação normal atribuída a esse tipo de curva

9. Os valores de média e do desvio-padrão

O que significa uma mudança no valor da média?

Na representação gráfica da curva normal, a variação do valor da média, em termos práticos, corresponde a um deslocamento da figura ao longo do eixo horizontal das

coordenadas cartesianas: para a esquerda, se a média for negativa; ou para a direita, se o seu valor for positivo. Porém a figura apenas se desloca para um ou outro lado, sem todavia provocar qualquer alteração na sua configuração geral.

E uma mudança no valor do desvio-padrão?

Entretanto, o mesmo não ocorre quando se altera o valor do desvio-padrão, ou seja, dos pontos de inflexão da curva. De fato, mudanças no valor do desvio-padrão tendem a provocar deformações na configuração gráfica da curva normal. A esse fenômeno, dá-se o nome de curtose (palavra de origem grega, que significa curvatura ou convexidade) que, em Estatística, vem a ser uma espécie de medida que avalia o grau de achatamento da curva normal.

Os três tipos de curtose.

Quando a curva normal tem desvio-padrão igual a 1, tal como ocorre na curva matemática teórica, ela é chamada de mesocúrtica (do grego mesos = médio) + cúrtica.

Todavia, quando o desvio-padrão tem valores entre 0 e -1, ou entre 0 e +1, a curva torna-se espigada, alta e estreita, porque os dados tendem a aglomerar-se junto à média, sendo exatamente esse pormenor que determina o pequeno valor do desvio-padrão. Quando ocorre esse alongamento vertical no centro da curva normal, e ao mesmo tempo o seu estreitamento no sentido horizontal, a curva é dita leptocúrtica, do grego (leptós = delgado, fino) + cúrtica.

Por seu turno, se os dados apresentarem valores muito afastados do valor da média, esse detalhe faz aumentar por sua vez o valor do desvio-padrão, provocando ao mesmo tempo o alongamento horizontal da curva normal e o seu achatamento no sentido vertical. A curva normal é então dita platicúrtica, palavra igualmente derivada do grego (platys = largo, amplo) + cúrtica.

Essas deformações tendem a dificultar, ou mesmo a impedir, comparações entre os efeitos dos fatores de variação sobre a variável estudada, de tal forma que muito comumente é preciso realizar transformações dos dados, para tornar factível a sua análise estatística.

A transformação dos dados em “valores de z”.

Uma dessas transformações, chamada “em valores de z”, tem a propriedade de transformar os parâmetros de qualquer distribuição de dados, de tal modo a que a média se torne igual a zero e o desvio-padrão igual a 1 (tal como na normal matemática), o que corresponde, em termos gráficos, a arrastar a curva ao longo do eixo horizontal, de modo a centralizá-la no ponto $x = 0$ das coordenadas cartesianas. Essa transformação em valores de z se faz pela relação: $z_i = (x_i - m) / s$, onde z_i é o novo valor do dado, x_i é o seu valor original, m é a média da amostra, e s o seu desvio-padrão. Essa transformação é particularmente útil quando se realizam testes para verificar se a distribuição dos erros experimentais é normal, uma vez que, após subtraída a média geral de todos os dados amostrais, o que sobra são as diferenças entre os dados originais (x_i) e a média (m) ou, em outras palavras, os erros experimentais, cuja distribuição e normalidade se deseja estudar e verificar.

Como há dados maiores e dados menores que a média, os novos valores serão negativos quando os dados forem menores que a média, e positivos quando maiores que ela. Em consequência, a soma dos dados z_i positivos será igual a soma dos dados z_i negativos, o que produz uma soma de dados igual a zero, soma essa que, dividida pelo número de dados (n), leva a uma média também igual a zero: $0 / n = 0$.

Por sua vez, como todas as diferenças ($x_i - m$) são divididas pelo desvio-padrão (s), quando essa diferença (ou erro experimental) for igual ao próprio desvio-padrão (s), o resultado será

um valor de $z_i = 1$, ou seja, $s / s = 1$, exatamente como na curva normal padrão.

Em suma: a transformação dos dados em valores de z faz com que eles passem a representar diretamente os erros experimentais, o que torna possível o estudo de sua distribuição em torno da média, possibilitando calcular a probabilidade de essa distribuição ser normal. Isso se faz pela avaliação do grau de aderência, ou de ajuste, entre as duas distribuições: a experimental e a normal matemática padrão (ou distribuição teórica de Gauss).

Item especial para quem gosta de matemática.

Em páginas anteriores, foram feitas duas assertivas, a respeito da curva normal, que são muito importantes:

1. que a média é o ponto onde a função atinge o seu valor máximo, ou seja, onde a frequência da distribuição atinge o seu ponto mais elevado; e
2. que o desvio-padrão marca o lugar onde a curva normal muda de côncava para convexa, ou seja, onde sofre uma inflexão.

Essas assertivas não são afirmações vazias, mas podem ser facilmente demonstradas matematicamente.

Embora eu tenha prometido, tanto aos senhores leitores como a mim mesmo, que não falaria em Matemática neste texto, pus-me a pensar que pode haver um ou outro curioso que poderia interessar-se por esses detalhes. Por isso, resolvi incluir no Apêndice que há no final deste texto, a demonstração matemática dessas afirmativas. Para entendê-las é preciso conhecer alguma coisa de cálculo diferencial.

O processo é simples: para comprovar que a média é o valor máximo da função normal, basta derivar a função (derivada primeira), igualar a derivada a zero, e isolar o valor de x . O resultado mostra que isso ocorre quando $x = m$, ou seja, quando x é igual à média.

Por sua vez, para comprovar que o desvio-padrão marca os pontos de inflexão da curva normal, basta por sua vez derivar novamente a própria derivada, e igualar esta segunda derivada também a zero, isolando-se o valor de x . O resultado demonstra que os pontos de inflexão da curva ocorrem quando $x = m \pm s$, ou seja, quando x é igual à média \pm o desvio-padrão. Aos que duvidarem, convido-os a consultar a demonstração matemática, no capítulo aqui referido como Apêndice.

O roteiro, passo a passo, até este ponto.

Retomando, porém, o nosso curso prático de Estatística, vamos recapitular os passos do roteiro que vimos traçando, capítulo após capítulo. Esses passos, até agora, foram:

- 1º Passo - Identificar a variável;*
- 2º Passo - Identificar os fatores de variação;*
- 3º Passo - Identificar o tipo de variável utilizado;*
- 4º Passo - Estabelecer o número de repetições;*
- 5º Passo - Construir uma tabela vazia para os dados (ainda não obtidos);*
- 6º Passo - Preencher essa tabela vazia com os dados já obtidos;*
- 7º Passo - Completar a tabela, testar a normalidade da distribuição.*

Qual seria o antônimo de “curva normal”?

Em geral, os testes utilizados para a finalidade de verificar a normalidade (ou não-normalidade) da distribuição dos dados experimentais apenas esclarecem qual a probabilidade de a distribuição testada ser normal, mas quem decide se essa probabilidade é aceitável ou não é o próprio pesquisador. Comumente, adota-se o limite de 5 por cento para a aceitação da normalidade, mas quanto maior for essa probabilidade, tanto melhor. O que não se pode fazer de modo algum é aceitar probabilidades menores que 5 por cento, pois isso indicaria que a diferença entre a distribuição experimental é significativamente diferente da distribuição normal padrão, ao nível de 5 por cento de probabilidade.

Observem os meus leitores que o oposto de curva normal não é curva anormal, mas sim curva não-normal, e o mesmo se pode dizer em relação à distribuição normal.

Condições complementares à normalidade.

Realizados os testes para julgar da normalidade (ou não-normalidade) da distribuição dos erros amostrais, se essa normalidade for comprovada, ficaria autorizado o uso dos testes chamados paramétricos. Ficaria sim, assim mesmo no condicional, porque há ainda uma condição, talvez até duas ou três, que devem ser preenchidas, antes da decisão final. Essas condições complementares são a homogeneidade das variâncias, a aditividade dos efeitos provocados pelos fatores de variação sobre a variável, e a independência dos erros.

Fica pois aqui mais uma pergunta: o que fazer, se a distribuição não for normal, se não houver homogeneidade das variâncias, ou se os efeitos não forem aditivos? Há duas alternativas: 1) ou tentar uma transformação dos dados originais; ou então 2) utilizar testes que não levam em conta os parâmetros amostrais (média e desvio-padrão), ou seja, usar a estatística por isso mesmo chamada não-paramétrica

10. O erro experimental

Composição do valor numérico do dado experimental.

Embora o dado experimental pareça um número simples, na verdade trata-se de uma entidade bastante complexa, onde há muita coisa embutida, que é preciso decompor e estudar, a fim de entender a sua verdadeira natureza. Por exemplo, num experimento fatorial com três fatores de variação, podemos representar cada dado numérico pela seguinte igualdade:

$$X_i = \mu + \alpha + \beta + \gamma + \alpha\beta + \alpha\gamma + \beta\gamma + \alpha\beta\gamma + \epsilon$$

Nessa expressão, X_i (ou igésimo X) é cada um dos dados numéricos do experimento, μ (μ) é a média geral da amostra, α (alfa), β (beta) e γ (gama) são as variações determinadas pelos três fatores principais de variação, as combinações $\alpha-\beta$ (alfa-beta), $\alpha-\gamma$ (alfa-gama), $\beta-\gamma$ (beta-gama) e $\alpha-\beta-\gamma$ (alfa-beta-gama) representam as variações provocadas pelas interações entre os três fatores de variação, e finalmente epsilon é a variação relativa ao erro experimental casual.

A média amostral e os erros experimentais.

Essa expressão demonstra que a média geral está presente em todos os dados da amostra, na qualidade de grandeza fixa, constante, ao passo que todos os demais símbolos representam grandezas variáveis. Isso quer dizer que, se não houvesse variação alguma, todas estas grandezas variáveis seriam iguais a zero e, em conseqüência, todos os dados seriam iguais à

média. Desse modo, fica evidente que todas essas grandezas variáveis, por representarem diferenças em relação à média, devem ser consideradas também erros experimentais.

Erros controlados.

Só que as variações determinadas pelos fatores de variação e suas interações representam erros introduzidos intencionalmente no experimento, porque é exatamente as diferenças detectadas nessas variações que se deseja estudar. Por esse motivo, tais variações intencionais são chamadas de erros controlados — e são chamados controlados porque é o próprio pesquisador quem determina quais e quantos serão os fatores de variação e quais e quantos serão os elementos componentes de cada um dos fatores.

Erros não-controlados, ou casuais.

Todavia, além dos erros ou variações, propositalmente introduzidos, e portanto controlados, existe também um fator de erro não-controlado, imprevisível, que independe da vontade do pesquisador, e que, na equação acima transcrita, está indicado pela letra grega epsilon.

Esse erro casual, não-controlado, pode decorrer de uma série de circunstâncias, que envolvem as causas mais diversas, que vão desde o fator individual, representado pela própria habilidade pessoal do observador, ou do técnico que realiza as medidas, até erros próprios do equipamento utilizado, ou de condições climáticas e ambientais, entre muitas outras.

Enfim, as causas do erro não-controlado podem ter origens variadas, podendo ser de ordem operacional, de método, de arredondamento dos dados, de aproximação de alguns dos valores envolvidos, além de outras eventuais que, por serem imprevisíveis, às vezes nem sequer são cogitadas, a não ser pelos efeitos que provocam, mas sempre depois que já ocorreram. O erro não-controlado é, antes de tudo, inevitável em qualquer experimento, e ocorre toda vez que se faz uma medida, qualquer que seja a natureza desta.

Na verdade, o que se pode fazer — e mais do que isso, o que se deve fazer — é não praticar erros grosseiros, que estes sim são visíveis e por isso mesmo podem e devem ser controlados e, tanto quanto possível, evitados.

Importância do erro casual, não controlado.

No entanto, apesar de sua inevitabilidade, o erro casual, não-controlável, é tremendamente importante em Estatística, porque é ele que serve como termo de comparação para julgar os demais erros, ditos controlados, que são precisamente aqueles que verdadeiramente interessam ao pesquisador, e que justificam a existência da investigação científica. Esse tema será novamente focalizado mais adiante, quando de nossas considerações sobre significância estatística.

11. Aditividade e homogeneidade

Aditividade dos efeitos dos fatores de variação, e homogeneidade das variâncias.

O termo variância já apareceu diversas vezes neste texto. O que seria variância, afinal de contas? Eu poderia dizer que variância é o quadrado do desvio-padrão. Contudo ressalvaria que essa afirmativa, ainda que matematicamente correta, é ainda estatisticamente incorreta. Isto porque, na verdade, o que se calcula primeiro é a variância da amostra. Só depois é que se extrai a sua raiz quadrada, para conhecer o desvio-padrão, o qual, por isso mesmo, tem duplo sinal: + ou - (\pm s).

Variância e graus de liberdade.

Tecnicamente, a variância vem a ser a soma de todos os desvios dos dados amostrais, em relação à média, elevados ao quadrado, soma essa que depois é dividida por $(n-1)$, ou seja, pelo número de graus de liberdade da amostra. Graus de liberdade, por sua vez, não é mais que o número total de dados da amostra, menos 1. Por que esses desvios são elevados ao quadrado? E por que se divide por $(n-1)$, e não simplesmente por n ? As respostas a essas duas perguntas parecem-me simples:

1. elevam-se os desvios ao quadrado porque, em relação à média, muitos deles são negativos e outros positivos, de modo que se fossem simplesmente somados, o resultado seria zero, tal como ocorre com a média desses mesmos desvios. Elevando-se cada um deles ao quadrado, porém, todos se tornam positivos, inclusive os negativos.
2. os graus de liberdade indicam os espaços entre os dados; e são iguais a $(n-1)$ porque os espaços entre eles estão sempre uma unidade abaixo do número dos próprios dados. Para comprovar essa afirmativa, basta contar os dedos de uma das mãos e depois os espaços existentes entre eles. O mesmo ocorre em qualquer conjunto de dados amostrais.

Isso compreendido, percebe-se que dividir pelo número de graus de liberdade significa dividir pelo número de espaços entre os dados, e não pelo número de dados. A razão de se fazer isso em Estatística é que os estudiosos da Ciência Estatística descobriram que essa operação conduzia a resultados mais coerentes do que a divisão por n , pura e simplesmente.

Variância e desvio-padrão.

Finalmente, torna-se compreensível também a razão da expressão desvio-padrão: é que a extração da raiz quadrada da variância — que, por ser um quadrado, representa uma grandeza em duas dimensões — transforma o quadrado dos desvios em uma grandeza unidimensional, ou seja, em um comprimento, uma espécie de média geométrica dos desvios, a qual pode ser encarada como um desvio realmente padrão. Ou, em outras palavras, um desvio médio em relação à média do conjunto de dados. Quanto a própria variância da amostra, antes da divisão por $(n-1)$ seria uma grandeza representativa da variabilidade total dos dados amostrais em relação a essa mesma média amostral. Após a divisão, seria uma variância média.

Uma vez conhecidos e entendidos esses conceitos básicos, estamos finalmente aptos a entender também o que sejam homogeneidade das variâncias e aditividade dos efeitos causados pelos fatores de variação sobre essas mesmas variâncias. Quanto à independência dos erros, ficará para mais adiante.

Aditividade dos efeitos dos fatores de variação.

Como já foi dito em capítulos anteriores, a aplicação dos testes paramétricos exige, além da normalidade da distribuição dos erros amostrais, que as variâncias sejam homogêneas e que os efeitos dos fatores de variação sejam aditivos; ou, em outras palavras, que sejam passíveis de serem somados uns aos outros, tal como indicam os sinais (+), presentes na expressão matemática transcrita no capítulo anterior. Esses efeitos não devem ser, por exemplo, multiplicativos. Sim, mas quando esses efeitos poderiam ser multiplicativos?

A resposta também nesse caso é simples, como ademais são simples todas as respostas, uma vez que sejam conhecidas, o que nem sempre é possível e nem sempre acontece, um fato

igualmente simples, que aliás constitui a própria razão de existir da pesquisa científica.

Os efeitos de dois ou mais fatores de variação são ditos não-aditivos quando, na associação de um ou mais desses fatores, em vez de se somarem, esses efeitos se multiplicam, de tal forma que o efeito resultante pode ser ampliado (quando o fator multiplicativo é maior que 1), ou reduzido (quando esse fator é menor que 1). É o que comumente ocorre nas chamadas interações entre dois ou mais fatores de variação.

Importância da aditividade.

A aditividade talvez seja a menos rigorosa das restrições que se fazem, quando do emprego da estatística paramétrica, porque se referem às interações entre os fatores de variação, e não aos próprios fatores em si mesmos. Mas não pode ser negligenciada, uma vez que a não-aditividade pode modificar o valor do erro não-controlado, inflando-o ou reduzindo-o, dependendo dessa alternativa de suas dimensões, ou de sua significância. Esse inconveniente deve ter ocorrido muitas vezes no passado, quando o efeito das interações era sistematicamente incorporado ao erro não-controlado do experimento.

A importância de um erro inflado ou reduzido será comentada mais adiante, quando se abordar o tema da significância estatística. Por enquanto, basta saber que os efeitos das interações só podem ser incorporados ao chamado erro residual (ou não-controlado), quando a interação for estatisticamente não-significante. Caso seja significativa, a sua variância deve ser isolada, e tratada como se fosse um fator de variação, pois se torna tão relevante na análise estatística quanto qualquer dos fatores de variação principais.

Homogeneidade das variâncias.

O bom desempenho dos testes paramétricos exige que as variâncias nele envolvidas sejam homogêneas. Isso não implica, porém, que elas devam ser idênticas, porque nada é exatamente igual em Estatística, havendo sempre uma faixa de tolerância em torno de qualquer suposta igualdade. O que os testes exigem é que elas não sejam discrepantes a ponto de ultrapassarem determinados limites de tolerância.

Para entender as razões dessa exigência, basta imaginar o que aconteceria se alguém tentasse comparar a variação do crescimento de melancias com a variação do crescimento de jabuticabas. As variâncias nesse caso seriam heterogêneas, ou seja, tão diferentes que tornariam impossível qualquer comparação direta.

Todavia, mesmo nesse caso, aparentemente absurdo, a comparação estatística não é de todo impossível, bastando para isso que se encontre um denominador comum, capaz de permitir o confronto entre esses dois tipos de crescimento tão diferentes.

Por exemplo, a variável adequada nesse caso poderia perfeitamente ser algo como a taxa de crescimento de cada fruto em relação ao seu próprio peso, ou ao seu próprio volume, taxa essa considerada a intervalos regulares, ao longo do período de tempo estabelecido para as observações.

Em resumo: embora as variações (ou variâncias), nos dois tipos de crescimento considerados em nossa hipótese, pudessem ser heterogêneas e desproporcionais, as variações das taxas de crescimento relativo poderiam ser homogêneas, e portanto compatíveis e passíveis de comparação.

Teste de Cochran para a homogeneidade das variâncias.

Um teste muito simples e de fácil execução, para verificar a homogeneidade das variâncias, é o teste de Cochran — que consiste em calcular todas as variâncias envolvidas no

experimento e dividir a maior delas pela soma de todas. O valor resultante da divisão é então comparado com os valores críticos de uma tabela estatística apropriada, que leva em conta o número de variâncias envolvidas (k) e o número de graus de liberdade (*) utilizado nos cálculos, número esse que evidentemente deve ser o mesmo para todas, pois a tabela é construída dessa forma.

Talvez a única dificuldade na execução desse teste seja decidir quais variâncias testar. A experiência de muitos anos acabou me ensinando que as variâncias que melhor se prestam a essa finalidade são as que se referem à interação maior envolvida no plano geral do experimento (binárias ou ternárias, conforme o experimento fatorial tenha dois ou três fatores de variação). Em última análise, essas variâncias da interação maior — que poderíamos chamar de interação de maior grau — referem-se à variação entre as repetições. Assim, se o experimento tiver, digamos, 120 dados numéricos, correspondentes ao produto fatorial de 4 colunas, 3 linhas, 2 blocos e 5 repetições, ($4 \times 3 \times 2 \times 5 = 120$), o teste de Cochran será realizado com 24 variâncias ($k = 24$), cada qual com 4 graus de liberdade ($*=5-1$, $*=4$). O * é a letra n , no alfabeto grego.

Interpretação do resultado do teste de Cochran.

O teste de Cochran é um teste curioso porque nele é de interesse que o valor calculado seja menor do que o valor crítico da tabela, e não maior como ocorre na maioria dos testes estatísticos, pois é exatamente isso que indica que as variâncias são homogêneas. De fato, se o valor calculado fosse maior, o resultado seria significativo, o que negaria a hipótese de igualdade (ou de homogeneidade) das variâncias envolvidas no experimento.

Dessa forma, a essa altura de nossas considerações, já sabemos se a distribuição dos erros de nosso experimento é ou não normal e se as variâncias são ou não homogêneas. Se a distribuição for normal e as variâncias homogêneas, estamos autorizados a usar os testes paramétricos.

Mas... o que fazer, na hipótese de a distribuição não ser normal, ou as variâncias não serem homogêneas...? Voltamos a insistir que só há duas alternativas: ou tentamos uma transformação dos dados, ou usamos testes não-paramétricos.

Antes disso, porém, sobrou ainda um último detalhe, que até agora não foi resolvido: o problema da independência dos erros...

12. Dados independentes ou vinculados

Independência, ou dependência, dos erros.

Diz-se que há independência dos erros, quando os erros controlados de um fator de variação não interferem com os erros controlados de outro. De um modo geral, a dependência dos erros amostrais ocorre quando há algum tipo de vínculo entre os dados que compõem um e outro grupo experimental, ditos então dependentes, ou vinculados.

É o que acontece, por exemplo, em pesquisas em que se estudam os efeitos de algum tipo de tratamento, e nos quais se usam os mesmos corpos-de-prova, os mesmos pacientes, ou os mesmos animais de laboratório, a fim de verificar alguma característica específica, antes e depois do tratamento. Esses grupos experimentais são comumente chamados de grupo-controle e grupo-tratado.

Os grupos Controle e Tratado são apenas um exemplo, porque a vinculação pode tomar as formas mais variadas. Seja como for, a vinculação existe porque os dados, nesses casos, podem sofrer alguma influência individual, seja do paciente humano, seja do processo usado na confecção dos corpos-de-prova, seja nas reações próprias do animal de laboratório.

Tais reações individuais na verdade sempre existem, mesmo nos casos em que os erros são independentes. A vinculação decorre, portanto, não da sua existência pura e simplesmente, mas do fato de essas reações individuais poderem repercutir sensivelmente em ambos os grupos que estão sendo estudados e comparados.

Como reconhecer se há vinculação entre os dados?

Um processo simples de constatar se existe vinculação entre os dados de um experimento é verificar se as repetições dentro de cada grupo podem ser misturadas, ou seja, se a ordem entre elas pode ser alterada, sem que haja prejuízo para o seu inter-relacionamento. Quando existe vinculação entre os dados, isso não pode ser feito, exatamente por causa da correspondência existente entre o dado de um grupo e o dado que ocupa a mesma posição no outro grupo (ou nos outros grupos, se houver mais de um).

Importância da vinculação entre os dados.

A existência de vinculação entre os fatores de variação ou entre os pares de dados não chega a ser propriamente uma restrição capaz de proscrever ou proibir a análise estatística. Apenas é necessário que se saiba de antemão da existência de uma possível vinculação, porque muitos testes estatísticos possuem duas versões, uma para dados independentes e outra para dados vinculados, e é preciso usar a versão correta. Além disso, outros testes há que foram idealizados apenas para um, ou para outro, desses dois tipos de dados experimentais, e é necessário saber se o teste que se pretende utilizar é adequado para o tipo de dados da amostra. Estas últimas observações valem tanto para os testes paramétricos como para os não-paramétricos.

Por fim, após as considerações dos dois últimos capítulos, podemos acrescentar mais alguns passos ao roteiro que estamos paulatinamente construindo, ao longo destas páginas:

8º passo - Testar a homogeneidade das variâncias correspondentes à interação de maior grau;

9º passo - Verificar a existência de vinculação entre dois ou mais dos fatores de variação envolvidos na pesquisa.

Aditividade dos efeitos dos fatores de variação.

Quanto à aditividade, somente após a realização dos testes estatísticos indicados para o modelo matemático do experimento é que se pode saber se ela de fato existe — e sua utilidade consiste apenas em ajudar o pesquisador a decidir se deve ou não isolar a variância de alguma das interações envolvidas no experimento, ou se pode simplesmente juntá-la ao erro residual, com os respectivos graus de liberdade.

13. Transformação dos dados amostrais

Razões para a transformação dos dados.

Quando algum dos requisitos para o emprego da estatística paramétrica — normalidade da distribuição dos erros, homogeneidade das variâncias, e aditividade dos efeitos dos fatores de variação — não puder ser preenchido pelos dados da sua amostra experimental, o pesquisador pode ainda tentar o recurso da transformação dos dados, antes de optar pela aplicação da

estatística não-paramétrica. É um recurso que sempre vale a pena tentar, porque a estatística paramétrica é evidentemente mais poderosa que a não-paramétrica. De fato, esta somente foi desenvolvida como um recurso complementar, destinado a suprir a necessidade de testes estatísticos nos casos em que alguma restrição desaconselhava o uso da estatística paramétrica, ou quando a própria natureza dos dados, muitas vezes não exatamente numéricos, vedava a aplicação desta.

As transformações mais comumente utilizadas.

As transformações diretas dos dados mais comumente utilizadas são: a logarítmica, a logarítmica dos (dados+1), a raiz quadrada dos dados, a raiz quadrada dos (dados + 1, ou mais 1/2), a raiz cúbica dos dados, a transformação angular, a transformação hiperbólica de primeiro grau (ou o inverso dos dados) ou hiperbólica de segundo grau, a transformação percentual, e a transformação em valores de z, já referida quando se comentaram os testes para verificar a normalidade da distribuição dos erros amostrais.

A transformação mais indicada.

Há sempre uma razão objetiva, em geral bem definida matematicamente, para se optar por uma ou outra dessas transformações, tudo dependendo de como ou por que a distribuição amostral está se deformando e fugindo à normalidade. Só a prática, entretanto, acaba ensinando o pesquisador a entrever qual a transformação mais indicada. Todavia, com o advento da informática, essas transformações se tornaram algo tão corriqueiro e tão rápido de realizar, que o estatístico, ou o pesquisador, pode tentar todas elas em seqüência, para ver qual a que produz o melhor resultado, gastando para isso não mais do que alguns poucos minutos de seu precioso tempo.

Na verdade, a transformação mais indicada geralmente coincide com aquela que apresentar a probabilidade mais elevada de a distribuição ser normal, de modo que se torna supérfluo saber a sua justificativa matemática. Se a transformação não for adequada, a probabilidade de normalidade tende a piorar, em vez de melhorar.

Eu próprio elaborei um programa para computador (GMC-software, hoje em sua versão 7.3), que executa todos os testes até aqui mencionados neste texto, bem como os testes mais importantes e mais comumente utilizados em Estatística, e por isso posso assegurar aos meus leitores que não estou argumentando em vão, e que sei perfeitamente do que é que estou afirmando.

Interpretação dos resultados (em dados transformados).

O único cuidado que se deve ter, após transformar os dados experimentais, é passar a raciocinar em termos da natureza dos novos dados, por ocasião da discussão e da interpretação dos resultados. Por exemplo: algumas transformações invertem os valores dos dados, como é o caso da própria transformação inversa (ou hiperbólica de primeiro grau), na qual $X_i = 1/x_i$, e da hiperbólica de segundo grau, em que $X_i = 1 / x_i^2$.

A transformação logarítmica.

Não se deve esquecer portanto que, uma vez transformados os dados em logaritmos, a soma de dados logarítmicos não tem o mesmo valor que a soma de seus antilogaritmos, mas representa o produto destes, de modo que a média dos logaritmos não corresponde ao logaritmo da média de seus antilogaritmos. Na verdade, o antilogaritmo da média dos

logaritmos corresponde à média geométrica dos dados originais, e não à média aritmética destes.

Por isso, no cálculo das médias, após a transformação logarítmica, não se pode esquecer de que os logaritmos passaram a ser tratados como simples dados numéricos, e não mais como logaritmos. Para fazer a conversão para os valores originais, as médias correspondentes às médias dos dados logarítmicos têm de ser calculadas a partir dos dados originais. A única coisa que é mantida nesses casos é a hierarquia dos dados, pois quando um dado original é maior do que outro, os seus logaritmos mantêm essa mesma ordenação hierárquica, ainda que os próprios valores numéricos passem a ser diferentes.

Uma vez normalizada e homogeneizada a distribuição dos dados amostrais, por intermédio da transformação que se comprovar mais conveniente, o pesquisador estará autorizado a utilizar os testes paramétricos. Contudo, se mesmo tendo tentado todos os recursos disponíveis ainda assim a distribuição continua se demonstrando não-normal, ou não-homogênea, ou até mesmo não-aditiva, não há outra alternativa senão utilizar a estatística não-paramétrica.

14. A escolha do teste mais adequado

Testes paramétricos e não-paramétricos.

Os testes estatísticos podem ser divididos em dois grandes grupos, conforme fundamentem ou não os seus cálculos na premissa de que a distribuição de frequências dos erros amostrais é normal, as variâncias são homogêneas, os efeitos dos fatores de variação são aditivos e os erros independentes. Se tudo isso ocorrer, é muito provável que a amostra seja aceitavelmente simétrica, terá com certeza apenas um ponto máximo, centrado no intervalo de classe onde está a média da distribuição, e o seu histograma de frequências terá um contorno que seguirá aproximadamente o desenho em forma de sino da curva normal. O cumprimento desses requisitos condiciona pois a primeira escolha do pesquisador, uma vez que, se forem preenchidos, ele poderá utilizar a estatística paramétrica, cujos testes são em geral mais poderosos do que os da estatística não-paramétrica, e conseqüentemente devem ter a preferência do investigador, quando o seu emprego for permitido.

O que são testes paramétricos?

Os termos paramétrico e não-paramétrico referem-se à média e ao desvio-padrão, que são os parâmetros que definem as populações que apresentam distribuição normal. Essa observação já foi feita e repetida muitas vezes neste texto. Volto a reafirmá-la, todavia, porque tenho visto muitas vezes artigos científicos, além de trabalhos e teses acadêmicas, em que se usaram testes não-paramétricos, mas os resultados eram apresentados em termos de média \pm desvio-padrão da distribuição, ou então em termos de média \pm erro-padrão da média, erro este que é também um valor calculado em função do desvio-padrão da amostra.

Os parâmetros da curva normal.

Ora, de qualquer conjunto de valores numéricos pode-se calcular a média, porém, desvio-padrão, somente as curvas normais o possuem, uma vez que, por definição, "desvio-padrão é o ponto de inflexão da curva normal" — e de mais nenhuma outra. São eles em número de dois e simétricos em relação à média da distribuição. Portanto, curvas assimétricas jamais podem ter desvio-padrão porque, mesmo que tenham pontos de inflexão, como os possuem muitas outras curvas matemáticas, eles dificilmente seriam simétricos em relação à média.

Enfim, mesmo que distribuições experimentais possam apresentar alguma assimetria, esta deve manter-se dentro de certos limites, aceitáveis em termos estatísticos — e aceitáveis porque atribuídos à variação casual determinada pelos erros não-controlados de amostragem, ou seja, à variação do acaso, típica das variáveis e amostras chamadas aleatórias.

Desvio-padrão e testes não-paramétricos.

Quando um pesquisador utiliza testes não-paramétricos, supõe-se que a distribuição de seus dados experimentais não seja normal, ou que ele não tenha elementos suficientes para poder afirmar que seja. Na dúvida quanto a essa informação, nada impede que ele opte pelo uso da estatística não-paramétrica. O que ele não pode fazer, de modo algum, é argumentar em termos de desvios ou erros padrões, embora possa perfeitamente fazê-lo pura e simplesmente em termos de médias.

Qual teste usar, sejam paramétricos ou não-paramétricos?

Qualquer que seja pois a opção do pesquisador, a essa altura de sua investigação científica ele se acha diante de mais um dilema: qual, dentre os muitos testes estatísticos existentes em ambas as categorias acima citadas, seria o mais apropriado, no caso específico de seu trabalho, ou do modelo matemático de seus ensaios? Que elementos desse modelo matemático condicionariam a opção por um ou outro desses testes?

Em geral a resposta está contida no próprio modelo experimental de cada pesquisa. Os detalhes adicionais que devem orientar a escolha do teste são:

- a) a existência ou não de vinculação entre dois ou mais fatores de variação;*
- b) o número de componentes da amostra, que vão ser comparados.*

De fato, seja qual for o tipo de estatística escolhida, paramétrica ou não-paramétrica, há testes especificamente destinados a amostras em que há independência entre os fatores de variação, e outros para amostras em que existe vinculação ou dependência entre eles.

Da mesma forma, o número de comparações a serem realizadas pelo teste é também importante, porque há testes elaborados para comparar apenas duas amostras, e há outros destinados a comparações múltiplas, entendendo-se como múltiplas um número de comparações superior a dois.

Num experimento fatorial, por exemplo, em que há fatores colocados nas colunas, nas linhas e nos blocos, o número de comparações é fornecido pela multiplicação do número de colunas, pelo número de linhas e pelo número de blocos. Enfim, o produto fatorial é semelhante ao usado para calcular o número total de dados da amostra, só não entrando no cálculo o número de repetições.

Assim sendo, no caso do experimento fatorial que, a partir de alguns capítulos atrás, nos vem servindo de exemplo — com 4 colunas, 3 linhas e 2 blocos — o número de comparações possíveis, incluindo-se nele não só os fatores de variação principais mas também todas as interações possíveis entre eles, seria: $4 \times 3 \times 2 = 24$ comparações.

Classificação dos testes estatísticos (GMC versão 7.5): O diagrama abaixo esquematiza as subdivisões dos testes estatísticos, listando os mais comumente utilizados na prática:

Testes Estatísticos			
Paramétricos		Não-Paramétricos	
Independentes	Vinculados	Independentes	Vinculados
2 amostras	2 amostras	2 amostras	2 amostras
		Mann-Whitney	Wilcoxon
		T. da Mediana	T. dos sinais
Teste <i>t</i> (Student)	Teste <i>t</i> (Student)	χ^2 (2 x 2)	Mac Nemar
		Proporções	Binomial
		Exato (Fisher)	
Mais de duas	Mais de duas	Mais de duas	Mais de duas
		Kruskal-Wallis	
		Mediana (m x n)	Cochran
Análise de variância	Análise de variância	χ^2 (m x n)	Friedman
		Nemenyi	

Alguns desses testes usam números como variável, outros usam sinais + e -, outros usam valores fixos, como 1 e 0, e outros ainda utilizam frequências. Esses testes evidentemente estão todos incluídos no grupo dos testes não-paramétricos, simplesmente porque não usam os parâmetros média e desvio-padrão em seus cálculos.

A filosofia de cada teste estatístico.

Após a conclusão destes conceitos iniciais e dos conhecimentos básicos que se deve ter sobre os métodos estatísticos, serão incluídos neste texto alguns breves comentários sobre cada um dos testes listados acima. São resumos sobre o que chamei de Filosofia do Teste, e neles procurei dar uma idéia geral sobre o que tinha em mente o criador de cada um deles, e a quais modelos matemáticos eles se adaptam, bem como em quais circunstâncias cada qual poderia ser utilizado.

Mas são apenas observações condensadas, que evidentemente os interessados poderão ampliar, pela leitura e pelo estudo mais aprofundado em compêndios mais elaborados do que este, sobre a Ciência Estatística, que os há em grande quantidade.

Apresentação dos resultados dos testes.

Uma vez realizados os testes adequados, estes dão o seu parecer, sob a forma de um valor numérico, apresentado (conforme o teste) como valor de *F* (análise de variância), de *t* (teste *t*, de Student), *U* (Mann-Whitney), *Q* (teste de Cochran), χ^2 (letra grega qui, testes diversos, que usam o chamado qui-quadrado), *z* (McNemar e Wilcoxon), *H* (Kruskal-Wallis), ou ρ (letra grega rho, utilizada nos testes de correlação, que serão focalizados mais adiante, neste texto).

Não-significância estatística (H_0).

Seja como for, o valor numérico calculado pelo teste deve ser confrontado com valores críticos, que constam em tabelas apropriadas a cada teste. Essas tabelas geralmente associam dois parâmetros, que permitem localizar o valor crítico tabelado: nível de probabilidades (usualmente 5 % [$\alpha = 0,05$], ou 1 % [$\alpha = 0,01$]), e o número de graus de liberdade das amostras comparadas.

Valores menores que o tabelado indicam que ele não pode ser considerado diferente do que se obteria se as amostras comparadas fossem iguais. Enfim, estaria configurado o que se chama de não-significância estatística, ou de aceitação da hipótese zero, ou de nulidade (H_0).

Significância estatística (H_1).

Porém, se o valor calculado for igual ou maior que o tabelado, aceita-se a chamada hipótese alternativa (H_1), ou seja, a hipótese de que as amostras comparadas não podem ser consideradas iguais, pois o valor calculado supera aquele que se deveria esperar, caso fossem iguais, lembrando sempre que a igualdade, em Estatística, não indica uma identidade. Isso quer dizer que pode eventualmente haver alguma diferença, mas esta não deve ultrapassar determinados limites, dentro dos quais essa diferença decorre apenas da variação natural do acaso, típica da variação entre as repetições do ensaio.

No caso de o valor calculado ser maior do que o valor tabelado, diz-se que há significância estatística, que pode ser ao nível de 5 %, se o valor calculado for maior que o valor tabelado para 5 %, porém menor que o tabelado para 1 %. Ou ao nível de 1 %, caso o valor calculado seja igual ou maior que o valor tabelado para 1 %.

15. Interpretação dos resultados

O que significam valor tabelado e valor calculado?

É possível que alguma dúvida ainda paire no espírito de muitos daqueles que ainda estão se iniciando nos meandros desse mundo misterioso da Estatística: o que significam exatamente os valores calculados pelos diversos testes, e o que quer dizer esse misterioso valor crítico das tabelas estatísticas?

Isso, porém, não é nem tão complicado nem tão difícil de entender, mesmo que para alguns possa parecer assim. E passo a explicar por quê.

Valor calculado, valor tabelado, e significância.

O valor calculado, bem como os valores tabelados, resultam sempre de uma divisão por algum valor, que é tomado como denominador comum, ou termo de comparação entre as grandezas comparadas.

Esse denominador comum, conforme o teste considerado, pode ser tanto o desvio-padrão da amostra, como a variância dos erros não-controlados, ou mesmo um valor teórico esperado. Neste caso, o esperado refere-se ao valor que seria teoricamente encontrado, caso a distribuição amostral seguisse religiosamente uma determinada distribuição matemática teórica previamente conhecida, ou pelo menos prevista por cálculos matemáticos teóricos.

Significância no teste t.

No caso do teste t, por exemplo, o que se divide é a diferença entre as duas médias que se deseja comparar pelo desvio padrão comum às amostras a que elas se referem. Portanto, o valor resultante dessa divisão indica quantas vezes a distância

que vai de uma média à outra contém a distância representada pelo valor do desvio-padrão: $t = (m_1 - m_2) / s$.

Significância na análise de variância.

Na análise de variância, como o próprio nome sugere, a divisão se dá entre variâncias: a variância dos erros controlados (s^2_i) pela variância dos erros não-controlados (s^2_r), esta última conhecida como variância residual, ou simplesmente resíduo: $F = (s^2_i) / (s^2_r)$.

Significância nos testes que usam o χ^2 .

Por sua vez, nos testes que utilizam a distribuição conhecida como distribuição do Qui-quadrado (χ^2), o que se divide é a diferença entre dois valores — o obtido (o_i) e o esperado (e_i), que comumente, porém não sempre, são frequências — pelo valor teoricamente esperado para a variação casual, e portanto não-significante: $\chi^2 = (o_i - e_i)^2 / e_i$.

Significância: observação final.

Seja qual for o teste, portanto, o resultado será sempre o quociente de uma divisão, e o quociente de qualquer divisão sempre traduz o quantas vezes o numerador é maior (ou menor, se for inferior a 1) do que o denominador, ou seja, quantas vezes este está contido naquele.

Assim, o resultado do teste, em última análise, apenas indica a proporção entre os erros controlado e não-controlado, embora estes erros possam receber outros nomes, dependendo do tipo de teste estatístico considerado. Mas, no fundo, são apenas variações do mesmo conceito.

A hipótese nula (H_0), símbolo da igualdade estatística.

Do exposto, é fácil concluir que, em Estatística, a igualdade não está representada propriamente pelo valor 0 (zero), mas sim pelo valor 1 (um). Esse valor 1 acontece quando o dividendo (numerador da fração que representa a razão proporcional entre os dois erros comparados) e o divisor (representado pelo denominador dessa mesma razão proporcional) são exatamente iguais. De fato, a divisão de qualquer número por si mesmo é sempre igual a 1. Ora, se o erro relativo aos fatores de variação (controlado) é igual ao erro detectado entre as repetições (não-controlado), é evidente que não pode ser considerado diferente deste último, cuja natureza é puramente casual.

Contudo, poderá alguém argumentar, e com razão: se é assim, se a igualdade estatística ocorre quando o quociente da divisão é igual a 1, o que significa a chamada Hipótese nula (H_0), que é o próprio símbolo da igualdade estatística?

É que o zero da hipótese nula é uma reminiscência que nos ficou dos testes que avaliavam a significância das diferenças entre as médias de duas amostras comparadas — mais especificamente, do teste t de Student. Nesse caso, o zero referia-se à diferença que existiria quando as duas médias comparadas fossem iguais, circunstância em que a diferença entre elas seria 0 (zero). De fato, $m_1 - m_2 = 0$ somente quando $m_1 = m_2$. Nesse caso, o "nula" referia-se evidentemente à hipótese de essa diferença entre as médias ser igual a zero.

Em tempo: o símbolo H_0 lê-se Hzero e não Hó, como já tenho ouvido algumas vezes. A hipótese de nulidade, portanto, refere-se à hipótese de que essas grandezas, ou essas médias, sejam estatisticamente iguais, naturalmente a um certo grau de probabilidade de que essa igualdade seja real. Por sua vez, a expressão H_1 indica apenas a hipótese alternativa para H_0 ,

ou seja, a hipótese de que não haja igualdade estatística entre as grandezas confrontadas $\frac{3}{4}$ e isso, evidentemente, também a um determinado grau de probabilidade de que sejam de fato diferentes (ou não-iguais, como os estatísticos preferem dizer).

Nível de significância estatística: probabilidade.

Mas disso tudo talvez ainda reste uma dúvida: o que seria, afinal, nível de significância? O que significaria exatamente significância estatística ao nível de 5 % de probabilidade? A expressão indica apenas que o valor calculado pelo teste (qualquer que seja este) só poderia ser encontrado, por simples variação natural do acaso, no máximo 5 vezes em 100 amostras aleatórias semelhantes. No caso da significância ao nível de 1 %, o valor encontrado pelo teste seria ainda mais difícil de obter por mero acaso, pois seria da ordem de 1 caso em 100 amostras do mesmo tipo.

Interpretação da significância e dos resultados dos testes.

Como já foi observado páginas atrás, quando se falou do teste de Cochran para a homogeneidade das variâncias, nem sempre o mais interessante para uma determinada pesquisa é que os testes estatísticos dêem resultados significantes.

No caso específico da homogeneidade, somente um valor não-significante seria vantajoso, pois só assim indicaria não haver diferenças estatisticamente relevantes entre as variâncias, sendo estas, pois, homogêneas, ou seja, não-discrepantes.

Mas esse não seria um caso isolado, pois há muitos trabalhos de pesquisa em que uma não-diferença estatística seria desejável, e eu próprio já me vi diante de inúmeros casos assim. Por exemplo, imaginem dois métodos, um caríssimo e outro muito mais barato, para realizar um ensaio qualquer. Nesse caso, o mais conveniente para o pesquisador seria que o teste comparativo entre os resultados fornecidos por ambos fossem não-significantes, pois nesse caso o pesquisador estaria autorizado a usar indiferentemente um ou outro, e por certo daria preferência ao mais barato, uma vez que os resultados seriam equivalentes, a um preço menor, o que em países como o nosso, e em muitos outros, que carecem de recursos para a investigação científica, é importantíssimo.

Outra observação importante é a que se refere ao distanciamento que muitas vezes existe entre a significância puramente matemática dos resultados estatísticos e a relevância desses mesmo resultados em termos de aplicação prática, seja em clínica, seja na vida prática em qualquer campo da atividade científica, ou simplesmente no dia-a-dia da atividade humana

16. Variância e covariância

Foi dito, em algum lugar deste texto, que, nos testes estatísticos, a variável é sempre única, ao passo que os fatores de variação podem ser múltiplos.

Todavia, há uma circunstância em que tal unicidade da variável pode não ocorrer: é quando a finalidade de um experimento é precisamente confrontar duas ou mais variáveis, a fim de verificar se existe algum tipo de variação proporcional entre elas, seja esta direta ou inversa. Dá-se o nome de covariação a esse tipo de variação simultânea entre duas ou mais variáveis, e de covariância à grandeza estatística que serve para medi-la.

Por sua vez, os testes utilizados para detectar essa covariância entre variáveis independentes envolvem duas operações importantes:

- 1) uma delas é quase gráfica, embora utilize cálculos matemáticos para realizá-la. É a operação chamada regressão, que pode ser linear (ou reta), ou curvilínea;*

2) e a outra, calculada a partir da primeira, é a correlação, que tem como unidade convencional de medida uma grandeza chamada coeficiente de correlação, em geral indicada pela letra grega ρ , que se lê "rô" (ou rho).

O coeficiente de variação é uma grandeza que varia de -1 a $+1$, valores estes que traduzem a correlação perfeita entre a variação de uma variável em relação à variação da outra. Por seu turno, a ausência completa de correlação entre as variáveis confrontadas é indicada pelo valor zero do coeficiente de correlação ($\rho = 0$). Os valores positivos do coeficiente de correlação ($0 < \rho \leq +1$), indicam a existência de uma relação diretamente proporcional entre as variáveis, enquanto que os valores negativos ($-1 \leq \rho < 0$) traduzem uma relação inversamente proporcional entre as variáveis em estudo. Por sua vez, o valor numérico de ρ traduz o grau de correlação entre elas, sendo tanto mais significativo quanto mais próximo de $+1$ (correlação direta), ou de -1 (correlação inversa). Hoje em dia, por uma questão de comodidade, costuma-se usar a letra r (erre minúsculo), em lugar de ρ , para o coeficiente de correlação.

Diferença matemática entre variância e covariância.

Basicamente não existe esse tipo de diferença, mas isso só pode ser percebido pela comparação das equações matemáticas que definem essas duas grandezas. Por isso, embora não seja objetivo deste texto falar nos fundamentos matemáticos da Estatística, essas duas equações serão transcritas a seguir, apenas para frisar suas diferenças e semelhanças. Mas, naturalmente, apenas aqueles que tenham alguma noção de álgebra e de somatórios poderão entendê-las.

A equação usada para calcular a variância de uma amostra é esta:

$$s^2_x = \sum x^2 - (\sum x)^2 / n \text{ (Equação 1)}$$

Essa expressão pode ser transformada em outra equivalente, substituindo-se o seu x^2 pelo produto $x \cdot x$, assim como o seu $(\sum x)^2$ por $\sum x \cdot \sum x$, escrevendo-se então:

$$s^2_x = \sum x \cdot x - \sum x \cdot \sum x / n \text{ (Equação 2)}$$

No caso da covariância, a única diferença é que as variáveis são duas (x e y), e não apenas uma (x), como no caso acima. Assim, quando se introduz a segunda variável (y), basta substituir um dos dois x por y para se ter a equação da covariância:

$$s^2_{xy} = \sum x \cdot y - \sum x \cdot \sum y / n \text{ (Equação 3)}$$

Por sua vez, a variância de y seria dada pela relação:

$$s^2_y = \sum y^2 - (\sum y)^2 / n \text{ (Equação 4)}$$

A regressão linear consiste em determinar qual a linha reta que passa, ao mesmo tempo, o mais perto possível de todos os pontos determinados no sistema cartesiano pelos pares x - y dispo

17. Os testes de Regressão e Correlação

Teste de regressão: as duas retas de regressão.

Consideremos a equação matemática da linha reta: $y = a + bx$. Para traçar o gráfico dessa reta, colocam-se os valores de x no eixo das abscissas e y no das ordenadas do sistema de eixos cartesianos. Todavia, é possível traçar outra reta com esses mesmos parâmetros a e b , agora em função de y , e não de x . Para isso, basta isolar o valor de x na equação acima transcrita, que ficará assim: $x = (y - a) / b$. O novo gráfico mostrará uma reta que, no caso da regressão ora focalizada, poderá apresentar um ângulo de inclinação diferente do da primeira reta, conforme se explicará mais adiante.

Cálculo dos parâmetros a e b da reta de regressão.

Para se calcular o valor de b , basta dividir o valor da covariância (Equação 3) pelo da variância da variável que estiver no eixo das abscissas (Equações 1 ou 4). Calculam-se, pois dois valores para b , que podem ser identificados como b_x e b_y .

Por sua vez, os valores de a_x e a_y são calculados pelas relações: $a_x = m_y - m_x \cdot b_x$, e $a_y = m_x - m_y \cdot b_y$, nas quais m_x e m_y são respectivamente as médias dos valores de x e de y .

Correlação: cálculo do valor de r (ou ρ).

O valor de r (ou de ρ) é basicamente a média geométrica dos dois valores de b calculados (b_x e b_y), sendo portanto fornecido pela expressão:

$$r = \pm \sqrt{b_x \cdot b_y}$$

Todavia, a raiz quadrada acima indicada, embora forneça o valor numérico de r , não indica se esse valor é positivo ou negativo. A definição do sinal depende da expressão da covariância (Equação 3): se, nessa expressão, $\sum x \cdot y$ for maior que $\sum x \cdot \sum y / n$, o valor de r será positivo; e, se for menor, o r será negativo.

O que indica o valor de r (ou de ρ)?

O valor de r (ou de ρ) igual a $+1$ ou -1 somente ocorre quando a reta de regressão calculada passa exatamente sobre todos os pontos disponíveis. Graficamente, isso quer dizer que as duas retas de regressão (de x em y e de y em x) se sobrepõem plenamente, de modo que aparecem no gráfico como uma reta única. Conforme o valor de r se afaste de $+1$ ou -1 , aproximando-se de 0 (zero), as duas retas já não mais se sobrepõem, aparecendo no gráfico como duas retas que se cruzam, num ângulo que se abre cada vez mais, até que, quando o valor de r é igual a 0 (que indica a falta total de correlação entre as variáveis), elas se cruzam perpendicularmente uma à outra.

Exemplo de não-correlação entre variáveis.

Para se ter uma idéia do que isso significa, imaginemos duas equações: $y = a + bx^0$ e $x = (y^0 + a) / b$. Como qualquer número elevado a zero é igual a 1 , as mesmas equações se reduziram a $y = a + b$ e $x = a / b$. Fazendo $a = 10$ e $b = 2$, elas ficariam assim: $y = 12$ e $x = 5$. Isto quer dizer que, na primeira equação, y seria igual a 12 , qualquer que fosse o valor de x . Logo, o valor de y não depende do valor de x , uma vez que x^0 será sempre igual a 1 , não influenciando no valor de y . O mesmo vale para a outra equação.

Se essas duas equações fossem representadas graficamente no sistema de coordenadas cartesianas, a primeira seria uma linha reta horizontal, paralela ao eixo das abscissas, passando pelo ponto $y = 12$; e a segunda seria uma reta vertical, paralela ao eixo das ordenadas, passando pelo ponto $x = 5$. Essas duas retas seriam perpendiculares entre si, cruzando-se no ponto $x = 5, y = 12$. Mas não haveria qualquer correlação entre elas, uma vez que os valores de x e y de uma não teria nada a ver com os valores x e y da outra, e vice-versa. O único ponto comum a ambas seria o ponto de cruzamento das duas linhas.

A tangente do ângulo de inclinação da reta horizontal teria um valor igual a 0 (zero), tangente essa que corresponde ao ângulo de 0° ; e a do ângulo de inclinação da reta vertical teria um valor igual ao ∞ (infinito), que corresponde ao ângulo de 90° . Isso indica que as retas se cruzam em ângulo reto, sendo portanto perpendiculares. É por isso que as retas de regressão perpendiculares entre si representam a ausência completa de correlação entre as variáveis x e y , tal como ocorre no exemplo acima.

Comparações entre coeficientes de correlação.

Quando se têm mais de uma reta de regressão, é possível comparar os seus coeficientes de regressão, para verificar estatisticamente se a relação entre as duas variáveis reunidas em pares para o traçado das linhas de regressão, bem como para o cálculo dos coeficientes de correlação correspondentes, é a mesma nas duas ou mais retas em estudo. O coeficiente de correlação avalia o grau de relacionamento entre causa e efeito de um fenômeno qualquer. Assim, a comparação entre dois coeficientes de correlação define se dois fenômenos mostram a mesma resposta de uma das variáveis (y), quando de faz variar a outra (x), ou se elas respondem de maneira diversa, mostrando diferentes tendências de variação, de um fenômeno para outro.

Outro detalhe importante a respeito dos testes de regressão e correlação é que os três parâmetros calculados por eles — isto é, os parâmetros a e b da reta de regressão, e o coeficiente de correlação (r) — podem eventualmente ser usados como variáveis, quando o espaço amostral é representado por um conjunto de retas, cada qual com a , b e r diferentes de uma para outra reta.

O emprego de parâmetros muitas vezes se torna absolutamente necessário, nos casos em que, sem esse recurso, a análise estatística seria totalmente impossível.

Sei bem disso, porque eu próprio já tive necessidade de lançar mão desse artifício técnico para tornar possível análises estatísticas aparentemente inviáveis. Acabei imaginando uma porção delas, por absoluta necessidade prática. Ao processo que visa à criação desse tipo de variável, pelo qual se altera a própria natureza íntima dessas variáveis, a fim de adequá-las matematicamente ao tratamento estatístico e torná-lo viável, batizei-o de mudança de variável, que será o tema do capítulo que vem a seguir.

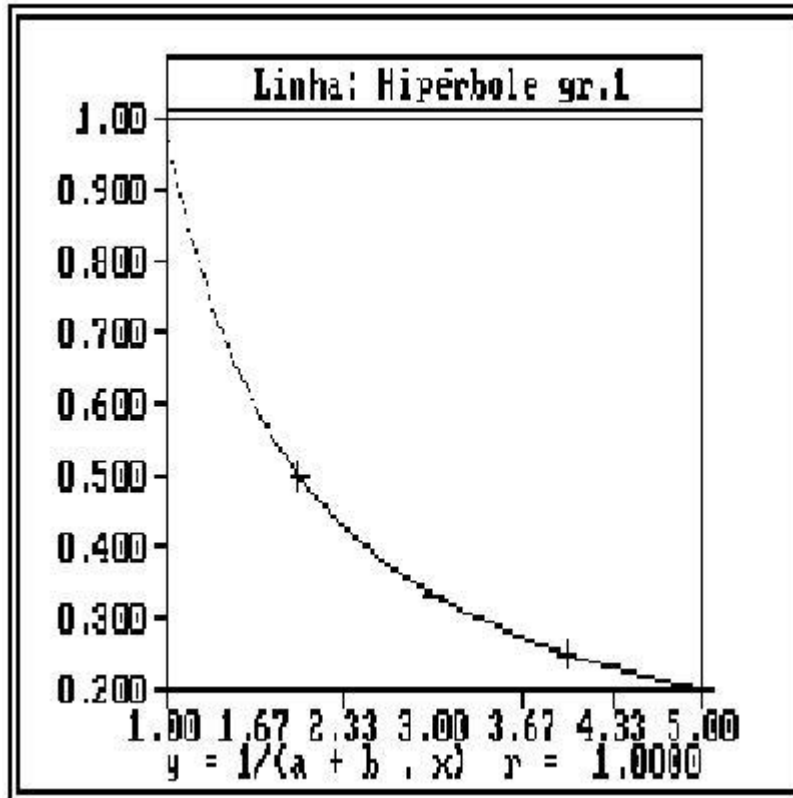


Figura 2. Linha de regressão hiperbólica, na qual $y' = 1/y$. O valor de r é igual a 1 porque os pares foram deliberadamente escolhidos para fornecer uma correlação direta perfeita

18. Mudança de variável (exemplos reais)

Neste texto, já foram comentados, em capítulos anteriores, as transformações simples a que se podem submeter os dados experimentais, visando a normalizar a distribuição dos erros amostrais e a homogeneizar as variâncias, com a finalidade de tornar possível a aplicação da estatística paramétrica.

Todavia, o que, no presente capítulo, é chamado de mudança de variável, são transformações mais profundas e mais complexas, que não têm quaisquer regras, fórmulas ou modelos fixos de transformação, como seria o caso, se a transformação fosse logarítmica, raiz quadrada, angular, ou qualquer das já comentadas anteriormente neste texto.

Na verdade, a mudança de variável é um recurso que se aplica a cada caso, individualmente, variando conforme a natureza de cada experimento. São artifícios técnicos da mesma natureza dos célebres artifícios de cálculo usados em Matemática para resolver certos problemas, os quais só valem para aquele determinado problema em pauta, ou, quando muito, para problemas semelhantes.

Os exemplos que seguem não são hipóteses, mas são todos recursos já empregados de fato pelo autor destas páginas, para resolver problemas específicos e reais, de pesquisadores diversos, que o procuraram em busca de auxílio.

Exemplo no.1: retas diferentes, como variável.

Um desses artifícios já foi citado no capítulo anterior, quando se comentou o emprego dos parâmetros da linha reta como variáveis. Já utilizei esse tipo de mudança de variável, quando fiz o tratamento estatístico de uma tese em que os dados experimentais eram medidas de

densidade óptica feitas em radiografias tomadas de um penetrômetro de alumínio apoiado sobre o filme radiográfico.

Esse dispositivo (penetrômetro) tem forma de escada, na qual os degraus tem espessuras crescentes, aumentando dois milímetros em cada degrau ascendente. A imagem radiográfica dessa escada de alumínio é uma série de faixas com radiopacidade proporcional à espessura de cada degrau, cuja densidade óptica é então medida em aparelho adequado a essa finalidade.

A dificuldade, nesse tipo de trabalho, é que uma radiografia não mostrava apenas um valor numérico, mas vários, cada qual correspondente à densidade óptica de um degrau do dispositivo. Parecia impossível tratar estatisticamente os dados numéricos obtidos.

O artifício que tornou possível a análise estatística envolveu algumas etapas, que passarei a comentar, apenas para ilustrar a maneira como funcionou o raciocínio do estatístico num caso como esse.

O primeiro passo foi realizar um teste de regressão para múltiplas curvas, a fim de determinar qual a curva matemática capaz de descrever a variação da densidade óptica nos oito ou nove degraus do penetrômetro utilizado nos experimentos.

Ficou esclarecido assim que, naquele caso específico, a curva era uma hipérbole de primeiro ou segundo grau (já não me lembro), traduzida pela equação matemática $y = 1 / (a + bx)$ (hipérbole de primeiro grau), ou então $y = 1 / (a + bx)^2$ (hipérbole de segundo grau). O segundo passo foi realizar uma transformação hiperbólica dos dados, que consistia em utilizar o inverso do valor dos dados experimentais ($1/y$, no caso da hipérbole de primeiro grau, ou então $1/y^2$, no caso da hipérbole de segundo grau), e não o valor original (y).

Após essa transformação, um novo teste de regressão mostrou que a relação entre x e y era agora uma linha reta crescente da esquerda para a direita, o que já era matematicamente de se esperar.

De fato, considerem a relação que traduz a transformação hiperbólica de primeiro grau: $y = 1/(a+bx)$.

Se invertermos a posição de y e $(a+bx)$, o que resulta é sem dúvida uma linha reta: $a+bx$ (linha reta) = $1/y$.

No caso da hipérbole de segundo grau, ocorre o mesmo: $y = 1/(a+bx)^2$.

Invertendo-se as posições de y e $(a+bx)^2$, tem-se: $(a+bx)^2$ (parábola) = $1/y$

Finalmente, extraindo-se a raiz quadrada de ambos os membros da equação, tem-se:

$$a + bx(\text{linha reta}) = \sqrt{\frac{1}{y}} = \frac{1}{\sqrt{y}}$$

Essas operações algébricas mostram claramente que, se for utilizado o inverso do valor do dado, em lugar do dado original, a linha de regressão será indubitavelmente uma reta, e não mais uma hipérbole de primeiro grau. Da mesma forma, o uso do inverso da raiz quadrada do dado original transforma uma hipérbole de segundo grau numa linha reta.

Mas qual seria a importância disso no caso das radiografias? A importância está em que se pode mudar a variável original (densidade óptica) e utilizar os dois parâmetros (a e b) que definem a reta de regressão de cada radiografia como duas novas variáveis.

A primeira delas, o parâmetro a da equação da reta, traduz a densidade óptica de fundo da radiografia, ou seja a densidade óptica do filme na região não interceptada pela presença do penetrômetro. Graficamente, seria o ponto onde a reta corta o eixo das ordenadas das coordenadas cartesianas, onde x (espessura do degrau da escada de alumínio) é igual a zero. E o parâmetro b nada mais é do que a tangente do ângulo de inclinação da reta, ângulo esse que traduz radiograficamente o grau de contraste do filme exposto.

De fato, se fizermos $a = 0$, o gráfico da reta passará pela origem das coordenadas

cartesianas, onde x e y são iguais a 0 (zero). Essa reta, inclinada, formará com o eixo horizontal um ângulo θ , cuja tangente será: $\text{tang } \theta = y / x$. Chamando a $\text{tang } \theta$ de b , ter-se-ia: $b = y / x$. Ou seja, a tangente do ângulo de inclinação da reta é realmente o b da equação da reta. Isolando-se o y , a equação ficaria assim: $y = bx$, sem o a , porque estamos considerando que a reta passa por $y = 0$. Se $y > 0$, então a equação terá de incluir o a , ficando assim: $y = a + bx$.

Essas duas novas variáveis, a e b permitiram, portanto, estudar os filmes sob dois aspectos importantes em qualquer radiografia: a densidade óptica geral e o contraste radiográfico dos filmes (vencidos, não-vencidos, conservados ou não em geladeira, armazenados ou não em estufas a 37/38 graus, para simular condições ambientais favoráveis ou adversas à sua conservação.

Sem o artifício da mudança de variável, de densidade óptica para os parâmetros a e b das diversas retas de regressão, correspondentes a cada filme exposto, a análise estatística dos resultados da pesquisa teria sido impraticável.

Exemplo no.2: associação de variáveis.

Contudo, a mudança de variável muitas vezes pode ser utilizada também para diminuir o número de variáveis de um trabalho de pesquisa, o que se consegue quando duas ou mais dessas variáveis podem ser combinadas para dar origem a uma outra variável, única, resultante dessa associação entre duas ou mais delas.

Por exemplo, imagine-se um experimento em que se desejasse saber qual, dentre uma série de soluções solventes, seria a mais eficaz para dissolver uma determinada massa de uma substância qualquer. O pesquisador poderia determinar a massa (m) de cada corpo-de-prova (variável 1), medir o tempo (t) gasto para a dissolução completa da massa correspondente a cada um deles (variável 2), e calcular a velocidade de dissolução (v), fornecida pela quociente massa dividida pelo seu correspondente tempo de dissolução — $v = m / t$ (variável 3).

No entanto, essas três variáveis, que exigiriam testes isolados para cada uma, poderiam ser associadas, resultando numa variável única a ser analisada, que combinaria os efeitos de todas as três.

Realmente, há na Física uma grandeza que associa essas três variáveis: é a chamada Força de Impulsão, definida pela expressão: $F = m \cdot v / t$.

Mas o raciocínio matemático e físico pode ir mais além.

De fato, partindo de três equações da Física: uma que define a Força ($F = m \cdot a$), outra que define a velocidade de um móvel ($v = a \cdot t$), e finalmente a que define o Trabalho ($T = F \cdot e$), nas quais F = força, a = aceleração, m = massa, t = tempo, v = velocidade, T = trabalho (ou energia despendida) e e = espaço percorrido.

Podem fazer-se diversas transformações algébricas simples: Se $v = a \cdot t$, então $a = v / t$; e se $a = v / t$ e $F = m \cdot a$, então $F = m \cdot v / t$.

Contudo, $m / t = v$, e a equação ficaria assim $F = v \cdot v$, ou $F = v^2$.

Considerando, porém a equação do trabalho ($T = F \cdot e$), e tendo em mente que, no caso da dissolução do tecido, o espaço percorrido (e) corresponde à massa dissolvida (m), pode-se fazer a substituição do espaço pela massa na equação do trabalho, uma vez que, nesse caso, $e = m$. Conseqüentemente, $T = F \cdot m$.

Substituindo agora, na equação $T = Fm$, o valor de F , tem-se: $T = m \cdot v^2$.

Como Trabalho e Energia são grandezas da mesma natureza, uma vez que são avaliadas pela mesma unidade física (Joule, erg), pode-se dizer indiferentemente: $T = m \cdot v^2$, ou $E = m \cdot v^2$.

Finalmente, se a velocidade v fosse a velocidade da luz (c), cairíamos na velha equação da

liberação da energia, descoberta por Einstein: $E = m \cdot c^2$!

Portanto, a nova variável de trabalho, calculada a partir de variáveis medidas nos experimentos, seria agora a energia (E) consumida na dissolução do tecido da polpa bovina, energia essa que difere de uma para outra das soluções utilizadas nos experimentos.

Exemplo: um massa de uma substância qualquer com peso = 2,33g, dissolvida pela solução A em 53 segundos teria uma velocidade de dissolução $2,33 / 53 = 0,044$ g / seg. A energia despendida, ou o trabalho realizado, nessa dissolução seria $0,044^2 \times 2,33 = 0,0045$ ergs, ou, pelo SI (MKF) 45×10^{-7} Joules.

Em termos de Força, teríamos $F = m \cdot v^2$, ou $F = 0,044^2$, igual a 0,0019 dinas no sistema CGS, que corresponde a 19×10^{-5} Newtons, aproximadamente, no sistema SI (MKF). Ter-se-ia de multiplicar o resultado da operação por 9,80665, mas a diferença é irrelevante para a análise estatística, porque todos os dados seriam então multiplicados pelo mesmo valor escalar.

Uma observação importante: após a mudança das variáveis, a discussão dos resultados da análise estatística terá forçosamente de ser feita em termos da nova variável.

Exemplo no.3: a variável área (produto de 2 variáveis).

Imaginemos um trabalho de pesquisa em que se estuda a velocidade de resfriamento de corpos-de-prova, deixados expostos ao meio ambiente, após terem sido previamente aquecidos a temperaturas diferentes. Para avaliar esse resfriamento, sua temperatura seria medida de minuto em minuto, variando de um para outro corpo-de-prova.

Este tipo de pesquisa é um exemplo típico de como se pode usar áreas como variável, em vez das duas realmente utilizadas no decorrer do trabalho experimental, variáveis essas que seriam o tempo gasto no resfriamento até a volta à temperatura ambiente, e as medidas de temperatura do corpo-de-prova minuto após minuto.

Lançadas em gráfico essas duas variáveis, associadas como pares de tempo/temperatura, nos quais o tempo seria marcado no eixo das abscissas (eixo de x), e as temperaturas nas ordenadas (eixo de y), o resultado seria uma área fechada, limitada por três linhas: duas retas (os eixos de x e y) e uma curva (curva de decrescimento da temperatura ao longo do tempo).

As áreas determinadas por essas linhas podem ser usadas como a variável do experimento, com a vantagem de associar as duas variáveis utilizadas simultaneamente, e não isoladamente. Essa nova variável traduziria numericamente a quantidade total de calor perdido durante todo o tempo gasto no resfriamento dos corpo-de-prova.

Não se trata de uma sugestão puramente teórica. Já usei pessoalmente esse recurso em um trabalho de tese, para o qual os meus préstimos foram solicitados. No caso real, registravam-se as temperaturas no interior de canais radiculares, após a aplicação de irrigações com soda clorada, e media-se, por meio de um par termoeletrico, a queda de temperatura da solução, dentro do conduto, minuto a minuto, durante o tempo decorrido até que a temperatura voltasse àquela que o canal apresentava no início do experimento.

Às vezes, quando os erros experimentais relativos às áreas, não apresentam distribuição normal, torna-se necessária a transformação dos dados pela raiz quadrada dos valores numéricos dos dados realmente obtidos.

Nessa transformação, o que se faz de fato é encarar todas áreas calculadas como se fossem quadrados equivalentes — ou seja, com a mesma área da figura de contorno irregular projetada em gráfico, figura essa já comentada em parágrafo anterior — de tal forma que a raiz quadrada desses quadrados transformaria uma grandeza bidimensional (área dos quadrados) em uma grandeza unidimensional, que seria o comprimento dos lados desses quadrados. Esse tipo de transformação costuma tornar normal uma distribuição e erros antes não-normal, porque tende a reduzir a amplitude da variação dos dados amostrais originais.

Exemplo no.4: variável área (métodos estereológicos).

Um recurso muito prático para calcular áreas, principalmente de figuras fechadas e de contorno irregular, é o uso da Estereologia, que basicamente consiste em utilizar uma grade de pontos com dimensões conhecidas, para calcular a superfície contida no interior de uma linha de contorno qualquer, a partir do número de pontos que incidem sobre a superfície fechada que está sendo avaliada.

É evidente que, quanto maior for a área da figura, tanto maior será a probabilidade de um número maior de pontos da grade caírem dentro dela. Na verdade, há uma proporcionalidade matemática entre o número de pontos que recaem no interior da figura e a sua área real. Assim, esta pode ser calculada por comparação com a superfície total da grade de pontos utilizada, que é uma área conhecida e representa, em termos de dimensão real, 100 % da área da grade de pontos, traduzidos pelo número total de pontos nela contidos.

A grade de pontos pode ser adaptada à ocular de um microscópio, ou opcionalmente traçada em papel, projetando-se sobre este a imagem microscópica da área que se quer medir, por meio de uma câmara clara. Este segundo método tem a vantagem de possibilitar a contagem de pontos posteriormente, abreviando o tempo em que o pesquisador fica preso ao microscópio, um processo muitas vezes cansativo para os olhos.

Um terceiro método consiste em obter slides das áreas a serem medidas (ou cópias transparentes dessas áreas), as quais poderão ser aumentadas, pela projeção das transparências sobre um anteparo, o que permite o emprego de grades com maior número de pontos, dando maior precisão à avaliação das áreas, e maior comodidade visual na contagem de pontos.

Os métodos estereológicos podem ser empregados também na contagem diferencial de elementos componentes de uma estrutura qualquer, sejam eles elementos tissulares simples, tais como células num processo inflamatório ou neoplásico, sejam estruturas mais complexas, como vasos sanguíneos, trabéculas ósseas ou fibras colágenas, num processo de cicatrização de uma ferida qualquer.

Com auxílio dos métodos estereológicos, caso seja levada em consideração a espessura dos cortes histológicos, por exemplo, é possível avaliar também o volume das estruturas estudadas, aparentemente a partir de imagens tomadas em duas dimensões. São portanto métodos extremamente úteis, porque permitem ao pesquisador transformar em valores numéricos algo que é basicamente de natureza qualitativa, e não quantitativa, como é o caso dos cortes histológicos. Esse mesmo recurso estereológico foi utilizado no artifício técnico descrito a seguir.

Exemplo no.5: transformação de áreas em vetores.

Além dos quadros histológicos, as radiografias são também exemplos de quadros cuja natureza é basicamente qualitativa. O autor destas linhas já teve em mãos um caso em que os resultados do trabalho de pesquisa de um pós-graduando consistiam numa série de radiografias da articulação temporomandibular (ATM). Quando me procurou, travamos o seguinte diálogo:

— E agora, professor, o que faço com os resultados da minha pesquisa?!, perguntou-me, completamente desarvorado.

Examinei uma das radiografias contra a luz, e perguntei, por meu turno:

— Você se lembra de uma coisa que aprendeu no colegial (ou mesmo no cursinho para o vestibular) chamada números complexos?

— Não me lembro, confessou-me ele, sem saber aonde eu queria chegar.

— Mas eu me lembro, repliquei, e é exatamente o fato de me lembrar que vai resolver o seu

problema...

— O que vem a ser número complexo, professor?

— Talvez você o conheça pelo nome de número imaginário, representado por aquele i usado em equações de segundo grau quando, após a aplicação da fórmula de Bhaskara, resultam raízes quadradas de números negativos. Por exemplo, a raiz de -4 , que é transcrita como $\pm 2i$. O número é dito imaginário porque não há raiz quadrada de números negativos, uma vez que qualquer número positivo ou negativo, quando elevado ao quadrado produz apenas números positivos.

— E como posso reconhecer um número complexo?

— A forma geral de um número complexo é $n = a + bi$, onde n é um número complexo qualquer, a e b são números reais, e i é a raiz quadrada de -1 , ou seja:

$$i = \sqrt{-1}, \text{ e portanto } n = a + b\sqrt{-1}$$

— Mas o que tem isso a ver com as minhas radiografias?!

— Tem tudo a ver. Se você aplicar o teorema de Pitágoras, usando os valores de a e b , terá a amplitude do deslocamento do côndilo de sua posição central; e se dividir b por a , terá a tangente do ângulo em que esse deslocamento se deu. Para saber que ângulo é esse, basta consultar uma tabela da função tangente.

— E como consigo esses valores de a e b ?

— Precisamos criar um método para obter esses valores, e nada melhor do que criar dois vetores para representá-los.

Eu julgava estar esclarecendo o assunto, mas ele parecia cada vez mais confuso. Na verdade, eu ainda não imaginara o método, mas já estava pensando nele exatamente naquele momento — e ele me surgiu por inteiro, de um instante para o outro: bastaria criar dois índices estereológicos, um horizontal e um vertical, e usá-los à guisa de vetores, como se fossem um sistema vetorial com dois deslocamentos ortogonais, do qual se determinaria a resultante, calculando-lhe o módulo e o ângulo de inclinação correspondentes.

A maneira como isso foi feito é bastante simples:

Primeiramente, selecionaram-se dois pontos de referência anatômicos cuja posição fosse relativamente estável nas radiografias da ATM, ou seja, que variasse pouco, em função de pequenas variações decorrentes do ângulo de incidência dos raios-x. Os pontos de referência escolhidos foram a imagem do meato auditivo externo e a crista anterior da cavidade articular da ATM.

Em seguida, por meio de um projetor comum de slides, projetava-se a imagem da radiografia sobre uma folha de papel presa a um anteparo vertical plano, colocado sempre à mesma distância do projetor, para que a ampliação fosse sempre a mesma em todas as radiografias, e traçava-se a lápis o contorno do côndilo, da cavidade articular e do conduto auditivo.

Uma vez obtido o desenho ampliado da ATM (com os pontos de referência citados nos itens anteriores), traçavam-se seis linhas retas sobre esse desenho, sendo três verticais e três horizontais.

A linha básica horizontal era uma reta que tangenciava ao mesmo tempo a crista da parede anterior da cavidade articular da ATM e a borda inferior do meato auditivo. As outras duas retas horizontais eram paralelas a essa linha básica, e tangenciavam respectivamente o contorno superior da cabeça do côndilo e o contorno da cavidade articular em seu ponto mais elevado.

A linha básica vertical era uma perpendicular à linha básica horizontal, e passava sobre o ponto em que esta tangenciava o contorno do côndilo. As outras retas verticais eram paralelas

a essa vertical básica e passavam sobre os ponto de interseção da linha básica horizontal com o contorno da cavidade articular, sendo portanto um anterior e outro posterior a essa linha vertical básica.

Desse modo, as seis linhas assim traçadas delimitavam uma área retangular subdividida em quadrantes, sendo dois destes superiores e dois inferiores, e ao mesmo tempo dois anteriores e dois posteriores, conforme considerados no sentido vertical ou horizontal do desenho.

Sobre esse esboço da ATM era colocada uma grade de pontos, e contados os pontos que incidiam em cada um dos quadrantes. A soma do número de pontos contidos nos dois quadrantes superiores, dividida pela soma do número de pontos incidentes sobre os dois quadrantes inferiores, fornecia o valor do vetor vertical do sistema vetorial buscado.

Da mesma forma, a soma dos pontos contidos nos dois quadrantes anteriores, dividida pela soma dos pontos referentes aos dois quadrantes posteriores, fornecia o valor do vetor vertical desse sistema vetorial. Esses dois valores numéricos eram, em suma, o a e o b procurados para definir o número complexo que caracterizava cada uma das radiografias da ATM, que eram assim transformadas em valores numéricos, o que as tornava passíveis de uma análise estatística coerente, que antes parecia uma tarefa tecnicamente irrealizável. E assim foi feito...

Uma observação interessante sobre o método acima descrito é que, quando o número de pontos contidos nos quatro quadrantes é exatamente o mesmo em todos eles, isso resulta em dois vetores iguais a 1, que teoricamente deveria representar a posição centrada do côndilo no interior da cavidade articular. Entretanto, o cálculo do módulo do vetor resultante revela que essa posição, dada pela raiz quadrada de $1^2 + 1^2$ ($\sqrt{2}$), é igual a 1,4241356, e não 1 ou 0, como se poderia pensar. Da mesma forma, o ângulo cuja tangente é igual a 1 é o de 45° , e não 0° ... Assim, todos os deslocamentos do côndilo deverão ser estudados em relação a esses valores referenciais teóricos, a fim de se avaliarem corretamente os valores reais desses desvios de posição (extensão e angulação).

Exemplo no.6: a probabilidade binomial como variável.

Outro caso curioso envolvia o emprego do teste bacteriológico conhecido como BANA. A pós-graduanda, autora do trabalho, dividia cada arcada dentária em três regiões, sendo duas posteriores e uma anterior, o que resultava na divisão das duas arcadas em seis sextantes. De um dente pertencente a cada um desses sextantes, colhia-se uma amostra do conteúdo de bolsas periodontais ali existentes, e com esse material realizavam-se os testes bacteriológicos, que poderiam dar resultados exclusivamente positivos (+) ou negativos (-).

O projeto inicial de trabalho previa a contagem e a comparação do número de resultados positivos nos dois grupos estudados, que reuniam pacientes diabéticos do tipo I (insulino-dependentes) e do tipo II (não-insulino-dependentes). Tudo estaria bem, não fossem dois detalhes, dos quais a autora do trabalho aparentemente não se dera conta ao planejar sua pesquisa.

a) O primeiro desses detalhes dizia respeito ao fato de nem sempre os seis sextantes estarem presentes, uma vez que muitos pacientes eram parcialmente desdentados, o que fazia variar o número de sextantes, e conseqüentemente o número de testes, por paciente. Essa variabilidade do número total de testes por paciente desaconselhava a contagem pura e simples do número de resultados positivos do teste BANA, uma vez que dois casos positivos em três testes realizados, por exemplo, não significam a mesma coisa que quatro, cinco ou seis casos positivos obtidos em seis testes realizados.

Para resolver o problema, sugeri um artifício estatístico que não me consta ter sido usado jamais por alguém anteriormente: adotar a probabilidade binomial de, em n testes realizados,

serem obtidas m respostas positivas (+); ou afirmativas, caso a variável inicial consistisse em respostas afirmativas (sim), ou negativas (não).

Com essa mudança de variável, os dados numéricos deixavam de ser valores discretos, que podiam ser apenas contados, produzindo freqüências que variavam de 0 a 6, para se transformarem em grandezas contínuas, que variavam de 0 a 1, que é a variação da probabilidade, ou de 0 a 100, se essas probabilidades fossem transformadas em probabilidades percentuais, uma escolha que, em termos estatísticos, é totalmente indiferente.

E assim foi feito, com o mais absoluto sucesso.

Para aqueles que possam algum dia ter diante de si o mesmo problema, transcrevemos abaixo a equação utilizada para efetuar a transformação das freqüências de respostas + e - (ou sim e não) em probabilidades de ocorrência dessas freqüências em n número de casos:

$$\text{prob} = \frac{(p + q)!}{p! \cdot q!} \cdot 2^{-(p+q)}$$

onde q = número de respostas negativas (-), e p = número de respostas positivas (+).

b) O segundo detalhe acima mencionado, que entrevi logo de início no plano de pesquisa ora comentado, envolvia um problema talvez bastante comum entre os pesquisadores: a escolha do grupo controle, principalmente quando, como no caso focalizado, duas condições patológicas estão simultaneamente presentes no mesmo paciente, e se deseja estudar uma delas exatamente em função da presença concomitante da outra.

Nesse caso, convém que o grupo controle não seja formado por indivíduos sadios, mas sim por pessoas portadoras de apenas uma das condições patológicas estudadas, para que se possam avaliar convenientemente os efeitos da outra sobre esta, que os pacientes controles também apresentam. No caso da associação diabetes/doença periodontal, é evidente que o interesse maior concentra-se nesta última, e que aquilo que se quer verificar é de que maneira os dois tipos diferentes de diabetes poderiam influir no desenvolvimento, ou no agravamento, da condição periodontal.

Assim, o ponto de referência (grupo controle) seria representado por pacientes não-diabéticos, porém igualmente portadores de doença periodontal, mesmo porque já está perfeitamente estabelecido que os pacientes diabéticos tendem a desenvolver doença periodontal, mais cedo ou mais tarde, de modo que é sempre mais fácil encontrar pacientes não-diabéticos com doença periodontal, do que achar pacientes diabéticos sem doença periodontal.

Exemplo no.7: escores, uma variável que se deve evitar (sempre que possível).

Tenho tanta fé nos escores, como variável capaz de avaliar um fenômeno qualquer, como tenho nas notas de avaliação como meio eficaz para julgar o desempenho de um aluno na escola. Tanto aqueles como estas implicam um grau de subjetividade que é sempre grande demais para o gosto de um estaticista. Os estaticistas, de um modo geral, preferem tratar com variáveis que sejam mais objetivas do que uma simples opinião pessoal, a qual nunca possui a imparcialidade fria de um instrumento de medida.

De fato, a opinião humana, por melhor que seja o avaliador, é sempre mais sujeita a falhas de interpretação do que um instrumento de medida, seja este qual for e seja qual for o seu grau de precisão. Em termos puramente estatísticos, isso quer dizer que a variabilidade da opinião humana tende a aumentar o valor do erro experimental, o que conduz fatalmente a uma

redução na capacidade de julgamento de pequenas diferenças entre as grandezas comparadas.

Realmente, é preciso ter sempre em mente que a significância estatística é a consequência direta de uma divisão de variâncias; ou seja, uma fração ordinária na qual o numerador (ou o dividendo) é a variância observada entre as grandezas comparadas, e o denominador (ou o divisor) é a variância entre as repetições (ou seja, a variância do erro experimental). Ora, se o denominador da fração for demasiadamente grande, o quociente da divisão será pequeno demais; e se for pequeno demais, o quociente será demasiadamente grande. Em qualquer das alternativas, o resultado estará prejudicado, produzindo falsas não-significâncias no primeiro caso, e falsas significâncias no segundo. Por isso, o erro tem de ser razoável, nem exageradamente pequeno, nem desmesuradamente grande. Os escores tendem a produzir erros experimentais grandes demais, no caso de avaliadores determinados aleatoriamente; e pequenos demais, no caso dos avaliadores ditos calibrados.

Do ponto de vista estatístico, um resultado significativo, no caso de erros experimentais grandes demais, seria altamente confiável, uma vez que revelou significância mesmo com o tamanho do erro experimental trabalhando contra. Todavia, o mesmo não se poderia dizer com relação aos resultados não-significantes, que poderiam caracterizar aquilo que se convencionou chamar de falsos negativos. Neste caso, a diferença entre os grupos comparados estaria sendo mascarada pelas diferenças muito grandes encontradas entre as próprias repetições realizadas dentro de cada grupo. Tecnicamente, em jargão estatístico, se diria que a variação entre grupos seria mais ou menos igual à variação intra-grupo $\frac{3}{4}$ e é exatamente essa quase-igualdade que caracteriza a não-significância estatística.

O raciocínio expresso no parágrafo anterior vale também para o caso de resultados não-significantes, em caso de erros experimentais demasiadamente pequenos. Neste caso, os resultados seriam válidos para a não-significância, mas poderiam acarretar erros nos casos de significância (falsos positivos).

A razão é basicamente a mesma já exposta no parágrafo referido: a significância aparente correria por conta apenas da divisão de uma variância relativamente grande entre grupos por uma variância intra-grupo inadequada exatamente por ser pequena demais. Seria como querer avaliar a variação do tamanho de melancias tomando como base a variação do tamanho de jaboticabas. Haveria, nesse caso, incompatibilidade entre o objeto medido e a unidade de medida utilizada. É por esse motivo, exatamente, que se mede tecido em metros, estrada em quilômetros, e célula em micrometros.

Contudo, por uma questão de coerência, devo observar que aquilo que foi dito acima traduz também apenas uma opinião pessoal do autor destas linhas, com tudo que uma opinião pessoal possa implicar, de acordo com o próprio texto em que essa opinião foi exposta. Aliás, o próprio Cristo já prevenia seus apóstolos contra o perigo do julgamento humano, quando sabiamente ensinou: "Não julgueis, para não serdes julgados, pois com o julgamento com que julgais sereis julgados, e com a medida com que medis sereis medidos" (Mateus, 7:1-2).

Apesar do risco, todavia, não posso deixar de expressar minha opinião, e estou disposto a agüentar o tranco que disso advier. Mesmo porque o ensinamento do Mestre apenas confirma essa opinião: aquele que usa escores deve estar também preparado para enfrentar as consequências dos erros de julgamento de seus avaliadores.

O diabo (como provavelmente diria o próprio Criador) é que às vezes não há como evitar usá-los, porque a natureza do experimento pode tornar incontornável o seu emprego na avaliação experimental... Mas, pelo amor de Deus!, se o uso de escores for inevitável, jamais calibre os seus avaliadores, porque isso tornaria a coisa ainda pior!

Uma observação final sobre os escores: evitem o escore 0 (zero). Procurem começar com o escore 1 para indicar a ausência seja lá do que for. Como a gradação é uma classificação meramente convencional, isso pode ser feito sem nenhum problema. O grau 0 não esclarece coisa alguma, mas pode acarretar alguma dificuldade, em caso de divisão por 0, ou se houver

necessidade de transformação logarítmica dos dados.

Aproveitando o ensejo, deve-se, tanto quanto possível, evitar dados com valores negativos, porque eles poderiam complicar as coisas em caso de ser necessário extrair a raiz quadrada desses valores negativos. Os zeros e os valores negativos não são dificuldades incontornáveis, quando presentes, mas a sua inexistência pode poupar tempo ao investigador, quando do tratamento estatístico dos dados obtidos em sua pesquisa.

Exemplo no 8: uso e abuso.

Da variável porcentagem.

O principal, e provavelmente o mais comum dos abusos e das liberdades que se tomam com a variável porcentagem, talvez consista em usá-la para números de dados inferiores a 100. De fato, esse procedimento raia pelos domínios da profecia, ou da adivinhação, uma vez que, a partir de um número reduzido de dados, pretende-se extrapolar frequências e achados, observados em amostras reduzidas, para amostras de tamanho igual ou maior que 100, amostras estas não existentes, e que podem não vir a apresentar as mesmas características dos dados que já foram obtidos até aquele momento, podendo na verdade fugir completamente a essas características, uma vez que porcentagens atuais não garantem porcentagens iguais no futuro.

Porcentagens são portanto dados que falam de fatos passados, e não de fatos que ainda não aconteceram. Quando se diz tantos por cento, o que se quer dizer é que para cada grupo de cem dos dados (que já se tem em mãos) uma certa parte tem uma determinada característica, dentre as que se estão estudando. Nada garante que o dobro do número desses dados virá a apresentar o dobro dessa frequência.

Em caso de amostras pequenas, é preferível falar em proporção, e não em porcentagem. Por exemplo: 6 casos em 36 estudados ($6 / 36 = 0,167$). A porcentagem seria essa mesma proporção multiplicada por 100 (16,67%), mas só teria sentido se se tratasse de 60 em 360 dados, ou, na pior das hipóteses, de 17 em 100 para a mesma porcentagem.

(Falta escrever) A transformação angular.

Exemplo no10: as variáveis multidimensionais.

Raciocínio idêntico ao apresentado no Exemplo no4 pode ser feito em relação a variáveis tridimensionais, das quais o volume é o exemplo representativo mais simples. A raiz cúbica dos dados transforma essa grandeza tridimensional e uma grandeza unidimensional, que seria o comprimento da aresta de um cubo, mesmo que o volume inicial não seja exatamente um cubo, mas uma esfera, um elipsóide, ou um sólido sem forma definida, mas cujo volume pode ser medido. Qualquer que seja o sólido, o valor do seu volume pode ser transformado numericamente em um valor equivalente ao de um cubo com uma aresta de comprimento x , aresta essa que, elevada ao cubo, reproduz o volume tanto do próprio cubo como do volume do sólido inicialmente considerado. Nesse caso, o tratamento estatístico pode ser feito considerando as arestas dos diversos cubos de volumes equivalentes aos dos sólidos originais, independentemente da forma real destes.

(Abaixo estão os itens que faltam ainda para escrever:)

Resumos dos testes que constam no software GMC

Exemplo no. 9: o denominador comum nas comparações.

Número suficiente de dados da amostra. Como calcular?

19. Filosofia de alguns testes estatísticos

A. Testes paramétricos, para duas amostras (independentes ou vinculadas).

1. Filosofia do teste *t* de Student.

Uma população é definida por dois parâmetros: a média e o desvio-padrão, que são únicos para essa população como um todo. Entretanto, a Estatística não lida com populações inteiras, mas utiliza subconjuntos dessas populações, aos quais chama amostras.

Como as amostras envolvem um número reduzido de elementos representativos da população da qual fazem parte, é natural que a sua média e o seu desvio-padrão difiram alguma coisa em relação aos parâmetros da população considerada em seu todo.

Os métodos estatísticos visam a possibilitar que se possam tirar conclusões sobre os parâmetros populacionais, partindo de informações obtidas a partir de amostras dela retiradas.

Como a média e o desvio-padrão das amostras, mesmo pertencendo estas à mesma população, sempre divergem alguma coisa em relação aos parâmetros reais da população, é compreensível que, se forem traçados os gráficos das distribuições amostrais e da população original, por certo haverá alguma discrepância entre todos eles.

Como as amostras pertencem todas à mesma população, e mesmo assim há diferenças, é natural que a variação decorrente da própria variabilidade casual da amostragem deva ser considerada, ao se avaliar a igualdade entre os parâmetros da amostra e os da população original, ou mesmo entre os de duas amostras entre si.

Em resumo: é preciso respeitar uma certa faixa de variação, dentro da qual as amostras são consideradas como provindas de uma mesma população, ou como iguais entre si.

Os testes estatísticos em geral, e entre eles o teste *t*, visam a estabelecer precisamente os limites além dos quais duas amostras já não devam ser consideradas como retiradas de uma mesma população, e sim como pertencentes a populações diferentes.

Quando as amostras comparadas são independentes, o teste *t* destina-se a verificar se mesmo assim pertencem à mesma população, apenas com variações casuais de amostragem. Quando são vinculadas, visam a verificar se algum tratamento realizado teve o dom de modificar os parâmetros amostrais, fazendo nascer assim uma nova população, com parâmetros típicos diferentes da inicial.

B. Testes paramétricos, comparações múltiplas.

2. Filosofia da análise de variância (fatores de variação independentes).

Para entender a análise de variância, é preciso distinguir dois conceitos fundamentais: a) variável e b) fator de variação.

Variável: é a medida pela qual alguma coisa é avaliada, tal como o peso, a altura, a área, o volume, o teor de alguma substância, etc.

Fator de variação: é tudo aquilo que faz a variável realmente variar. Por exemplo, um tratamento que faça variar o peso, a altura, o volume, o teor de glicose no sangue, etc.

A variável é sempre uma só, mas o fator de variação pode ser múltiplo. De fato, fatores diversos podem atuar ao mesmo tempo, influenciando todos sobre uma característica qualquer da amostra.

A filosofia do teste admite que o efeito final dos múltiplos fatores de variação que atuam ao mesmo tempo sobre uma variável pode ser decomposto e analisado por partes (daí o termo análise aplicada ao teste).

Esses efeitos parciais referem-se a três tipos de variação:

- a) a variação causada pelos fatores intencionalmente introduzidos no experimento, até certo ponto controlada pelo pesquisador;
- b) a variação determinada pelas possíveis interações entre alguns ou entre todos esses fatores experimentais controlados; e
- c) a variação ocasional, não-controlada, decorrente de causas estranhas, muitas vezes desconhecidas, que em conjunto constituem o erro experimental, presente em qualquer ensaio.

O erro experimental — chamado de variação residual, ou simplesmente resíduo — é importantíssimo, porque é por ele que se mede a significância estatística de um experimento. A significância estatística não é mais do que uma comparação entre a variação experimental controlada e a variação não-controlada (erro).

A filosofia geral do teste é muito simples: ela admite que, se um fator de variação é realmente ativo num experimento, os seus efeitos aparecem e podem ser isolados, mesmo quando diversos fatores atuam simultaneamente. A variância final do experimento seria, assim, uma adição de três variâncias separáveis: a dos fatores principais, a de suas interações, e a do erro.

Na análise de variância com múltiplos fatores, chamados estes de critérios de variação, ou de classificação, cada fator (ou interação) é analisado separadamente, ignorando-se os demais, considerados estes, para todos os efeitos, como simples repetições.

Por exemplo: se forem estudadas duas drogas (A e B), administradas em duas circunstâncias diferentes (C e D), ao se analisarem os efeitos das drogas A e B, o dados de C e D são classificados apenas em relação a A ou a B. O efeito conjugado droga/circunstância seria analisado também, é claro, mas como um fator secundário denominado interação.

O número de fatores de variação estudados deve limitar-se a 3 no máximo, porque o número de interações possíveis aumenta consideravelmente com números maiores, uma vez que a quantidade de variâncias a serem analisadas é dada pela equação: $N = 2^n - 1$, onde N é o número de variâncias, e n é o número de fatores de variação. Um excesso de variâncias (principalmente de interações) leva a um emaranhado de interligações, quase sempre de difícil interpretação. Num ensaio com 5 variáveis, por exemplo, o número de variâncias a serem estudadas seria: $2^5 - 1$, ou seja, $32 - 1 = 31$ variâncias!

3. Filosofia da análise de variância (fatores de variação vinculados).

Imagine-se uma pesquisa feita para comparar a dureza da dentina em cortes transversais de raízes dentais, nos terços cervical, médio e apical, nas regiões junto ao canal radicular, perto do cimento e a meia distância entre elas, após tratamento das secções com diversas soluções auxiliares da instrumentação dos canais radiculares, aplicadas durante tempos diferentes.

Esse é um exemplo de um experimento em que os fatores de variação estão todos vinculados, com exceção das repetições (que seriam as diversas raízes usadas para repetir o ensaio).

A experiência mostra que, quando existe vinculação, há também uma certa hierarquia na dependência entre os fatores vinculados. Por exemplo: no caso citado, os três terços referem-se à mesma raiz dental, as três regiões da dentina ao mesmo terço da raiz, e as soluções irrigantes atuam durante tempos diferentes, porém sobre as mesmas regiões de cada corte dental. Neste caso, começando com o mais dependente, a hierarquia da vinculação seria:

tempos de ação, regiões da dentina e terços da raiz.

O modelo matemático-estatístico e a forma de programação no computador (GMC Software) exigem que os dados sejam introduzidos obedecendo a essa hierarquia. Assim, o fator mais dependente deve ser sempre colocado nos blocos da tabela de dados, seguindo-se as linhas e as colunas (a organização da tabela pode ser vista no programa estatístico GMC, onde se explica como os dados devem ser introduzidos no computador, para que a programação funcione corretamente).

O programa estatístico GMC abrange modelos estatísticos com dois ou com três fatores de variação, podendo o primeiro ter um ou dois fatores vinculados, e o segundo um, dois ou três fatores mutuamente vinculados. Quando todos os fatores são interdependentes, o único fator que sempre permanece independente são as repetições, cuja variação pode ser isolada e o seu efeito avaliado.

Organize corretamente a sua tabela de dados, de modo a introduzir os valores numéricos na ordem adequada, caso contrário o programa fornecerá resultados incorretos, uma vez que os dados estarão misturados.

4. Filosofia da análise de variância (1 fator de variação com repetições).

A análise de variância geralmente envolve uma amostra populacional equilibrada, na qual os grupos estudados têm um número igual de repetições, principalmente quando há diversos fatores de variação (ou critérios de classificação dos dados) envolvidos.

Todavia, quando há apenas um fator de variação, ou seja, quando o conjunto de dados consiste de vários grupos que devem ser comparados entre si, é possível realizar uma análise de variância desse conjunto de dados, mesmo que cada um dos grupos tenha um número diferente de repetições, o que ocorre frequentemente.

Esses grupos poderiam ser comparados dois a dois, pelo teste t de Student, por exemplo, mas isso às vezes envolve a realização de um grande número de testes, dependendo do número de grupos a serem comparados. A análise de variância tem a vantagem de comparar todos os grupos com um único teste.

Quando uma pesquisa envolve mais de um fator de variação, ainda que a análise de variância não seja de todo impossível, ela seria muito complicada. Mais prático será então usar amostras equilibradas, onde todos os grupos tenham o mesmo número de repetições.

Testes não-paramétricos, amostras independentes (uma só variável, duas amostras comparadas).

5. Filosofia do teste de Mann-Whitney.

Se duas amostras forem retiradas ao acaso de uma mesma população, a ordenação crescente e conjunta dos dados das duas amostras tende a misturá-los uniformemente. Isso faz com que os dados se encaixem de maneira equitativa, tal como se intercalam os números pares e ímpares na seqüência natural dos números reais.

À medida em que os valores ordenados das duas amostras se separam e se afastam, a probabilidade de elas pertencerem à mesma população vai se tornando cada vez mais remota. Os valores de U calculados pelo teste avaliam o grau de entrelaçamento dos dois conjuntos de valores numéricos confrontados.

O caso extremo ocorre quando as duas amostras já não se intercalam — isto é, são disjuntas — o que indica tratar-se de amostras provindas de populações diferentes. A disjunção dos dados traduz a significância estatística máxima do teste, e ocorre quando o U menor é igual a 0 (zero).

6. Filosofia do teste da mediana (para 2 amostras).

O teste da mediana visa a verificar se duas amostras diferem em relação às suas tendências centrais, uma vez que a mediana e o valor que marca o centro da distribuição amostral.

Assim, o teste exige que as amostras possam ser pelo menos passíveis de uma ordenação por valores ascendentes dos dados, para que se possa calcular o valor que divide o conjunto de dados das amostras reunidas exatamente ao meio, ou seja, com 50% dos dados acima e 50% abaixo desse valor. Esse valor é a mediana.

A filosofia do teste admite que, se duas amostras provêm de uma mesma população (isto é, se são estatisticamente iguais), a mediana do conjunto de dados reunidos não difere significativamente da mediana de cada uma delas considerada isoladamente.

O teste é, no final, um teste de χ^2 (qui-quadrado) em que as frequências comparadas se referem ao número de dados — em cada uma das amostras comparadas — que se encontram acima ou abaixo da mediana comum, calculada para o conjunto das amostras reunidas.

7. Filosofia do teste do χ^2 (qui-quadrado), 2 x 2.

O teste do χ^2 (qui-quadrado) é um teste que compara frequências obtidas experimentalmente com frequências teóricas, calculadas matematicamente para o mesmo número de dados da amostra.

Os dados devem portanto ser grandezas discretas, isto é, alguma coisa que possa ser contada e reduzida a uma tabela de frequências, tabela essa denominada tabela de contingência.

A tabela de contingência é formada de duas linhas e duas colunas. O grau de liberdade é dado pelo produto de $(2-1) \times (2-1) = 1 \times 1 = 1$.

O teste calcula a relação: quadrado da diferença entre as frequências obtida e esperada em cada uma das quatro células da tabela de contingência, dividido pela frequência esperada, e soma esses quadrados.

O teste é considerado significativo quando essa soma ultrapassa determinados valores, relacionados em tabelas apropriadas, valores esses que dependem do grau de liberdade da amostra.

Os testes não são exatamente iguais para tabelas com apenas 1 ou mais de 1 grau de liberdade, e por isso essas duas possibilidades são focalizadas separadamente neste programa estatístico (para 2 x 2 ou para m x n frequências).

Além disso, há uma série de restrições:

A. Para 1 grau de liberdade:

- a) pode ser aplicado para n maior que 40 (n = número total de dados);
- b) para n entre 20 e 40, o teste só pode ser aplicado se todas as frequências esperadas forem maiores ou iguais a 5;
- c) se a menor frequência for menor que 5, ou se n for menor que 20, será preferível usar o teste exato de Fisher.

B. Para mais de 1 grau de liberdade:

- a) nenhuma casela pode ter valor menor que 1;
- b) o número de caselas com valores esperados menores do que 5 não pode ultrapassar 20 % do número total de caselas; e
- c) se isso ocorrer, reformule a tabela (somando caselas vizinhas).

8. Filosofia do teste de igualdade entre proporções.

Esse teste é praticamente igual ao teste do χ^2 (qui-quadrado), com a diferença de que as frequências são transformadas em proporções, dividindo-se as frequências obtidas em cada uma de duas amostras pelo respectivo número total de dados dessa amostra.

A finalidade do teste é verificar se duas proporções podem ser consideradas iguais, quando resultantes de amostragens com números diferentes de dados. Por exemplo: será que as proporções de 34 dados numa amostra com 147 dados, e de 167 dados em outra com 985 dados, seriam iguais? As duas proporções são respectivamente 0,23129 e 0,16954. Seriam elas estatisticamente equivalentes?

Nesse caso, a resposta poderia ser dada tanto por este teste como por um simples teste de χ^2 numa tabela 2 x 2.

Pode ocorrer, entretanto, que numa ou outra circunstância não se possa usar o teste do χ^2 . Nesse caso, o teste da diferença entre duas proporções poderia ser utilizado, como uma alternativa para o teste do χ^2 .

Este teste, todavia, tem também as suas restrições, tal como as tem o próprio teste do χ^2 .

Assim, como ocorre no χ^2 , convém que as amostras sejam grandes, de tal modo que as frequências obtidas (F_n), ou os seus complementos ($N - F_n$), sejam todas maiores que 5.

O teste do χ^2 também faz o mesmo tipo de exigência. A única diferença é que o teste entre proporções não tem a correção de Yates.

Diante disso, toda vez em que for possível, é preferível usar o teste do χ^2 ou, se as frequências forem muito pequenas, o teste exato de Fisher.

Outra opção é usar a distribuição de Poisson (para eventos raros). Um evento é considerado raro quando sua probabilidade de ocorrência está próxima de 0 (zero). Praticamente, considera-se raro o evento cuja ocorrência é de 5 vezes (ou menos) em 50 (ou mais) tentativas ($p \leq 0,1$). Isto é, quando a probabilidade de 1 evento x o número de tentativas (n) é igual a 5, ou menor que 5 ($p.n \leq 5$).

9. Filosofia do teste de Fisher.

O teste exato de Fisher testa diferenças entre dois grupos independentes (G_1 e G_2), em relação a uma variável qualquer que só admita duas alternativas como resposta: Sim/Não, Positivo/Negativo, ou +/- . Isso leva à construção de uma tabela de contingência 2 x 2.

O teste é basicamente um χ^2 (qui-quadrado), porém o teste de Fisher é particularmente adequado para pequenas amostras (com 20 dados ou menos), caso em que o teste do χ^2 estaria contra-indicado.

Em compensação quando o número de dados da amostra é grande, o teste de Fisher é que não deve ser usado, porque envolve o cálculo de fatoriais, o que pode conduzir a números excessivamente elevados. Nesses casos, a opção deve ser pelo teste do χ^2 .

Testes não-paramétricos, amostras vinculadas (uma só variável, duas amostras comparadas).

10. Filosofia do teste de Wilcoxon.

Uma amostra A_1 , submetida a um tratamento T_1 , e o seu efeito medido. Posteriormente, essa mesma amostra, chamada agora de A_2 , é submetida a um segundo tratamento T_2 , medindo-se o seu efeito pela mesma variável usada no primeiro tratamento.

Comparando-se o efeito dos dois tratamentos em cada elemento da amostra, podem ocorrer 3 alternativas:

- a) O efeito aumentou (+);
- b) O efeito diminuiu (-); e
- c) O efeito permaneceu o mesmo (=).

Até este ponto, o teste seria idêntico ao chamado teste dos sinais. A diferença porém é que, no teste de Wilcoxon, leva-se em conta a magnitude do aumento ou da diminuição, e não apenas a direção da variação para mais ou para menos.

Assim, para cada par vinculado A1/A2, calcula-se a diferença numérica $T1 - T2$. Essa diferença poderá ser positiva, negativa, ou igual a zero (quando não houver variação, sendo $T1 = T2$).

Uma vez calculadas todas as diferenças entre os valores obtidos para cada par de dados, essas diferenças são ordenadas pelo seu valor absoluto (sem considerar o sinal), substituindo-se então os valores originais pelo posto que ocupam na escala ordenada.

Feito isso, atribui-se a cada um desses novos valores dos dados o mesmo sinal que eles tinham antes da transformação em postos.

A filosofia do teste presume que, se os tratamentos forem idênticos, a soma dos postos com sinais positivos será equivalente à soma dos postos com sinais negativos.

O teste de Wilcoxon calcula um valor z , ao qual está associada um valor de probabilidade. Essa probabilidade traduz o grau de possibilidade de ocorrência desse valor de z por mero acaso, e não por efeito dos tratamentos efetuados ($T1 = T2$). No caso do GMC software, o programa já faz automaticamente o cálculo da probabilidade do z obtido pelo teste, não havendo necessidade de consultar qualquer tabela.

11. Filosofia do teste dos sinais.

Uma amostra A1, submetida a um tratamento T1, e o seu efeito medido. Posteriormente, essa mesma amostra, chamada agora de A2, é submetida a um segundo tratamento T2, medindo-se o seu efeito pela mesma variável usada no primeiro tratamento.

Comparando-se o efeito dos dois tratamentos em cada elemento da amostra, podem ocorrer 3 alternativas:

- a) O efeito aumentou (+);
- b) O efeito diminuiu (-); e
- c) O efeito permaneceu o mesmo (=).

Os dados serão codificados apenas como 1 ou 0, para os valores maior e menor de cada par. O valor real do dado não afeta o teste.

Calculando-se a frequência em cada uma das duas primeiras alternativas e desprezando-se a terceira, em que não houve alteração, pode-se estimar se as frequências dos sinais + e - devem ser consideradas estatisticamente diferentes ou não.

A decisão estatística envolve o cálculo binomial da probabilidade de os sinais + e - terem aquelas frequências por mero acaso.

Quando os pares vinculados puderem ser medidos quantitativamente, de forma que seja possível estabelecer não só a hierarquia, mas também o quanto um membro do par é maior ou menor do que o outro, o teste mais preciso seria o de Wilcoxon, e não este.

12. Filosofia do teste de McNemar.

Uma amostra A1, submetida a um tratamento T1, e o seu efeito medido. Posteriormente, essa mesma amostra, chamada agora de A2, é submetida a um segundo tratamento T2,

medindo-se o seu efeito pela mesma variável usada no primeiro tratamento.

Comparando-se o efeito dos dois tratamentos em cada elemento da amostra, podem ocorrer 4 alternativas:

- a) Foi positivo em A1 e A2 : T1+ e T2+ ;
- b) Foi negativo em A1 e A2 : T1- e T2- ; e
- c) Foi negativo em A1 e positivo em A2 : T1- e T2+ .
- d) Foi positivo em A2 e negativo em A1 : T1+ e T2-

Calculando-se a frequência em cada uma das 4 alternativas, constrói-se uma tabela de contingência 2 x 2.

A decisão estatística é dada por um teste de χ^2 (qui-quadrado), cujo resultado dirá se a distribuição de frequências encontrada pode ser considerada puramente casual, ou se as diferenças de frequência devem ser atribuídas realmente ao tratamento realizado.

13. Filosofia do teste binomial.

O teste binomial é particularmente útil em experimentos que apenas admitem duas alternativas como resposta, tais como certo ou errado, sim ou não, verdadeiro ou falso, masculino ou feminino, positivo ou negativo, e assim por diante.

O teste utiliza o desenvolvimento matemático binomial de duas frequências relativas complementares p e q (sendo $p + q = 1$) para avaliar a probabilidade de elas poderem ser consideradas estatisticamente não-diferentes, ainda que desiguais em termos puramente numéricos.

Assim, os dados experimentais utilizados pelo teste são as frequências relativas p e q, referentes às duas alternativas possíveis naquele determinado experimento. A frequência esperada para p e q, em caso de igualdade perfeita, seria $\frac{1}{2}$ para ambos.

Como, num experimento, dificilmente p é igual a q, o teste avalia, em última análise, até que ponto os valores de p e q podem diferir, sem deixarem de ser estatisticamente iguais.

Testes não-paramétricos, amostras independentes (uma só variável, comparações múltiplas).

14. Filosofia do teste de Kruskal-Wallis.

O teste de Kruskal-Wallis é uma espécie de análise de variância a um critério de variação, para dados amostrais independentes.

Por exemplo: a superfície de n corpos-de-prova construídos com k marcas comerciais de gesso para modelos seria igualmente lisa?

A variável testada, nesse caso, é o grau de lisura da superfície dos corpos-de-prova, e o único fator que faz essa variável alterar os seus valores é a marca comercial dos gessos.

O erro experimental é dado pela variação casual determinada por diferenças eventuais ocorridas durante a confecção dos diversos corpos-de-prova (repetições) que constituem a amostra referente a cada um dos materiais envolvidos.

A filosofia do teste considera que, se os materiais forem todos igualmente lisos, a única variação será aquela decorrente dessa variabilidade natural, que sempre existe, mesmo entre elementos de uma mesma população.

O teste não utiliza os valores numéricos diretamente, mas sim os postos que eles ocupam numa série de dados ordenados por valores crescentes, série essa que reúne num só conjunto os dados de todas as amostras que vão ser comparadas. Os dados são introduzidos amostra

após amostra.

Ainda segundo a filosofia do teste, se as k amostras comparadas provierem da mesma população (amostras iguais), a média dos postos correspondentes a cada amostra será aproximadamente igual.

Se isso não ocorrer, as amostras pertencerão provavelmente a populações diferentes * ou seja, serão diferentes entre si.

Embora o teste tenha sido idealizado para testar um único fator de variação, parece viável utilizá-lo também em casos de mais de um critério de variação, desde que se faça a análise de um deles de cada vez, reunindo em grupos todos os dados que tenham em comum esse fator, considerando os demais como simples repetições.

15. Filosofia do teste da mediana (para k amostras).

O teste da mediana visa a verificar se duas ou mais (k) amostras diferem em relação às suas tendências centrais, uma vez que a mediana e o valor que marca o centro da distribuição amostral.

Assim, o teste exige que as amostras possam ser pelo menos passíveis de uma ordenação por valores ascendentes dos dados, para que se possa calcular o valor que divide o conjunto de dados das amostras reunidas exatamente ao meio, ou seja, com 50 % dos dados acima e 50 % abaixo desse valor. Esse valor é a mediana.

A filosofia do teste admite que, se duas ou mais amostras provêm de uma mesma população (isto é, se são estatisticamente iguais), a mediana do conjunto de dados reunidos não difere significativamente da mediana de cada uma delas considerada isoladamente.

O teste é, no final, um teste de χ^2 (qui-quadrado) em que as frequências comparadas se referem ao número de dados — em cada uma das amostras comparadas — que se encontram acima ou abaixo da mediana comum, calculada para o conjunto das amostras reunidas.

16. Filosofia do teste do χ^2 (qui-quadrado), $m \times n$.

O teste do χ^2 (qui-quadrado) é um teste que compara frequências obtidas experimentalmente com frequências teóricas, calculadas matematicamente para o mesmo número de dados da amostra.

Os dados devem portanto ser grandezas discretas, isto é, alguma coisa que possa ser contada e reduzida a uma tabela de frequências, tabela essa denominada tabela de contingência.

A tabela de contingência é formada de (m) linhas e (n) colunas, sendo que a menor tabela que se pode formar seria uma tabela com 1 linha x 2 colunas, ou 2 linhas x 1 coluna.

O grau de liberdade é dado pelo produto de $(m-1) \times (n-1)$, quando m e n são iguais ou maiores do que 2; e por $(m-1)$ ou $(n-1)$, caso um deles (n ou m) for igual a 1.

O teste calcula a relação: quadrado da diferença entre as frequências obtida e esperada em cada casa da tabela de contingência, dividido pela frequência esperada, e soma esses quadrados.

O teste é considerado significativo quando essa soma ultrapassa determinados valores, relacionados em tabelas apropriadas, valores esses que dependem do grau de liberdade da amostra.

Os testes não são exatamente iguais para tabelas com apenas 1 ou mais de 1 grau de liberdade, e por isso essas duas possibilidades são focalizadas separadamente neste programa estatístico.

Além disso, há uma série de restrições:

A. Para 1 grau de liberdade:

- a) pode ser aplicado para n maior que 40 (n = número total de dados);
- b) para n entre 20 e 40, o teste só pode ser aplicado se todas as frequências esperadas forem maiores ou iguais a 5;
- c) se a menor frequência for menor que 5, ou se n for menor que 20, será preferível usar o teste exato de Fisher.

B. Para mais de 1 grau de liberdade:

- a) nenhuma casela pode ter valor menor que 1;
- b) o número de caselas com valores esperados menores do que 5 não pode ultrapassar 20 % do número total de caselas;
- c) se isso ocorrer, reformule a tabela (somando caselas vizinhas).

17. Filosofia do teste de Nemenyi.

O teste de Nemenyi é uma espécie de análise de variância não-paramétrica, para um fator único de variação, que faz comparações entre várias amostras independentes.

O fator de variação estudado é colocado nas colunas, com as repetições dispostas verticalmente, ao longo das colunas. Os dados são introduzidos no computador seguindo o sentido vertical da tabela, repetição após repetição, e não no sentido horizontal.

Os dados de todas as amostras são ordenados por valores crescentes, sendo os valores originais substituídos pelo número de ordem ocupado por eles na série do conjunto ordenado. Em caso de empates, faz-se a média dos postos correspondentes, e se atribui esse mesmo valor a todos os dados empatados.

Se as amostras pertencerem à mesma população — isto é, se forem iguais — as médias dos seus postos serão mais ou menos iguais. A avaliação estatística é feita pela comparação dessas médias.

Testes não-paramétricos, amostras vinculadas (uma só variável, comparações múltiplas).

18. Filosofia do teste de Cochran.

Os (n) elementos de uma mesma amostra (A) são julgados segundo (k) padrões ou métodos diferentes de avaliação ($P_1, P_2, P_3, \dots, P_k$).

Os dados experimentais devem apresentar-se como respostas do tipo (+/-), (Sim/Não), ou (Positivo/Negativo).

A aplicação do teste, porém, exige que essas respostas, seja como for que se apresentem, sejam convertidas em valores numéricos 1 (um) para os Sim, Positivo ou (+), e em 0 (zero) para os Não, Negativo, ou (-).

O teste procura responder a perguntas do tipo: Os resultados dos diversos métodos de julgamento testados seriam equivalentes?

A filosofia do teste considera que, se os diversos métodos produzem efeitos semelhantes sobre os elementos que compõem a amostra, a distribuição dos 1 e 0 nos vários métodos comparados será aproximadamente igual (a não ser, é claro, pelas variações casuais, presentes em qualquer experimento).

Os dados amostrais (reduzidos a 0 e 1) devem ser reunidos em uma tabela com (n) linhas e (k) colunas.

Torna-se possível, assim, definir se a proporção (ou frequência) de respostas é a mesma em cada uma das (k) colunas comparadas, ou se, pelo contrário, houve influência sobre ela dos métodos ou dos padrões de julgamento utilizados para avaliá-las.

O teste aplica-se a uma grande variedade de situações, bastando para isso que os dados possam ser reduzidos a valores 0 e 1, e possam ser reunidos em tabelas desse tipo ($n \times k$).

O teste é, no fundo, um teste de χ^2 (qui-quadrado), para $(k-1)$ graus de liberdade.

O programa já calcula automaticamente qual a probabilidade de haver igualdade entre as amostras comparadas. Indica também o nível de significância estatística, quando forem detectadas diferenças entre as amostras.

19. Filosofia do teste de Friedman.

O teste de Friedman é uma espécie de análise de variância a dois critérios de variação, para dados amostrais vinculados.

Por exemplo: a superfície de corpos-de-prova construídos com diversos tipos de materiais poderia ser avaliada sucessivamente por dois ou mais métodos diferentes.

Nesse caso, os dois critérios de variação seriam: 1) os métodos de avaliação; e 2) os materiais utilizados. As amostras são vinculadas porque as avaliações se fazem na mesma superfície de cada corpo-de-prova.

O teste responde a este tipo de pergunta: seria idêntica a avaliação da superfície pelos vários métodos, em relação aos diversos materiais? Ou então: responderiam os materiais igualmente aos diversos métodos de avaliação? Ou ainda: haveria concordância entre os diversos métodos em relação à avaliação da superfície dos corpos-de-prova?

A resposta do teste depende de qual dos fatores esteja colocado nas colunas de uma tabela de dados com k colunas e n linhas.

Desse modo, a organização da tabela de dados é muito importante, uma vez que depende dela a interpretação do resultado do teste. O fator comparado principal deve ser colocado nas colunas, e os dados serão introduzidos no sentido das linhas da tabela.

O teste de Friedman não utiliza os dados numéricos diretamente, mas sim os postos ocupados por eles, após a ordenação por valores ascendentes desses dados. A ordenação numérica é feita separadamente em cada uma das amostras, e não em conjunto.

A filosofia do teste considera que, se as diversas amostras provêm de uma mesma população, isto é, se elas são estatisticamente iguais (hipótese de nulidade, ou de H_0), a distribuição dos postos nas diversas colunas será mais ou menos equivalente, de modo que a soma dos postos em cada coluna será aproximadamente igual.

A hipótese alternativa (H_1) seria de que as amostras não pertenceriam à mesma população — isto é, seriam diferentes — e nesse caso haveria diferenças entre as somas das diversas colunas.

Teste para mais de uma variável (regressão e correlação).

20. Filosofia dos testes de regressão e correlação.

O teste de regressão linear — e seu complemento natural, que é o teste de correlação — são testes estatísticos extremamente úteis porque permitem estudar o comportamento de duas (ou mais) variáveis ao mesmo tempo, buscando detectar uma possível relação proporcional coerente entre a variação de uma em função da variação da outra (ou das outras, quando mais de duas).

As variáveis podem ser as mais heterogêneas, ao contrário de outros testes que, para que duas ou mais amostras possam ser comparadas, exigem que a variável seja única. Assim, esse teste de regressão e correlação pode reunir variáveis tão heterogêneas quanto o tamanho das melancias de uma plantação e o teor de cálcio ou de potássio do adubo utilizado para fertilizar a terra onde elas crescem.

O teste é também particularmente útil quando se deseja avaliar ou comparar tendências, tais como o comportamento da inflação ao longo do ano, ou a tendência da queda ou do aumento da inflação num determinado ano, em relação ao de outro ano qualquer.

A regressão linear refere-se sempre à linha reta. Contudo, nem sempre a equação matemática que traduz um fenômeno científico se traduz por uma linha reta. Porém muitas delas podem ser reduzidas a uma reta, por meio de transformações algébricas adequadas.

Por exemplo: um determinado fenômeno natural pode ser representado por uma hipérbole, cuja equação matemática é $y = 1 / a + bx$. Caso se faça a inversão dos termos (y) e ($a + bx$), obter-se-á uma nova expressão algébrica para a mesma igualdade: $1 / y = a + bx$. Chamando y' ao termo $1 / y$, tem-se: $y' = a + bx$, que é a expressão algébrica da linha reta. Assim, a transformação $y' = 1 / y$ tende a retificar uma linha originalmente curva, como é o caso da hipérbole.

Essas transformações, que tornam possível a regressão linear de algumas curvas comumente encontradas em pesquisa científica, é o objeto do presente teste. A correlação entre duas variáveis é expressa por r , cujo valor varia de +1 (correlação direta) a -1 (correlação inversa). O valor $r = 0$ indica ausência de correlação.