

ACH3584 - Estatística II

Aula 03: Introdução à Inferência Estatística

Alexandre Ribeiro Leichsenring
alexandre.leichsenring@usp.br



Organização

- 1 Introdução
- 2 População e Amostra
- 3 Problemas de Inferência
- 4 Como selecionar uma amostra

Organização

- 1 Introdução
- 2 População e Amostra
- 3 Problemas de Inferência
- 4 Como selecionar uma amostra

Introdução à Inferência Estatística

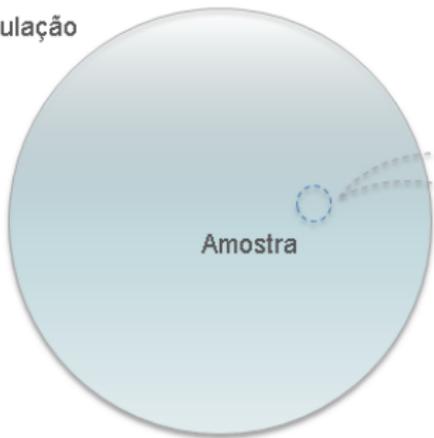
Já vimos:

- 1 Como resumir descritivamente variáveis associadas a conjuntos de dados
- 2 Modelos teóricos (probabilísticos) para algumas variáveis aleatórias

Inferência

- ▶ A inferência está relacionada com o uso de informações de uma amostra para concluir sobre o todo
- ▶ Argumentos estatísticos para fazer afirmações sobre as características de uma população, com base em dados de amostras

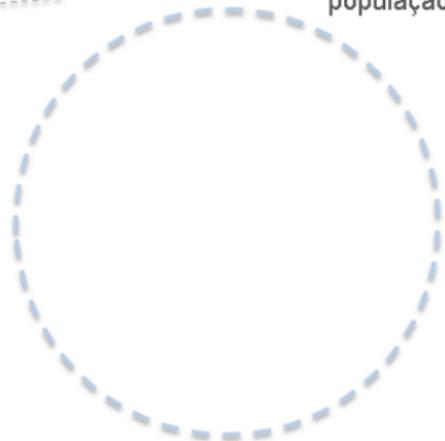
População



Amostra



Inferência sobre a população



Incerteza!



Organização

- 1 Introdução
- 2 População e Amostra**
- 3 Problemas de Inferência
- 4 Como selecionar uma amostra

População e Amostra

Definição

População É o conjunto de todos os elementos ou resultados sob investigação. Também chamado de universo.

Amostra É qualquer subconjunto da população.

Exemplo

Consideremos uma pesquisa para estudar os salários dos 500 funcionários da Companhia MB. Seleciona-se uma amostra de 36 indivíduos, e anotam-se os seus salários.

Variável a ser observada: “salário”;

População: os 500 funcionários da companhia;

Amostra: os 36 indivíduos selecionados.

- ▶ Esperamos que a distribuição dos salários na amostra reflita a distribuição de todos os salários (se a amostra for escolhida com *cuidado*).

Exemplo

Queremos estudar a proporção de indivíduos de uma cidade que são favoráveis a certo projeto governamental. Uma amostra de 200 indivíduos é sorteada, e a opinião de cada uma é registrada (a favor ou contra o projeto).

Variável a ser observada: proporção (ou número) de pessoas favoráveis ao projeto na cidade;

População: todos os moradores da cidade;

Amostra: 200 pessoas selecionadas.

Definindo a seguinte variável:

$$X = \begin{cases} 1, & \text{se a resposta de um morador for favorável} \\ 0, & \text{se a resposta for contrária ao projeto} \end{cases}$$

► Podemos dizer que o objetivo é compreender a distribuição de X .

Exemplo

Em alguns casos, fazemos suposições mais precisas sobre a população (ou sobre a variável definida para os elementos da população).

Digamos que X represente o peso real de pacotes de café, enchidos automaticamente por uma máquina.

Sabe-se que a distribuição de X pode ser representada por uma normal, com parâmetros μ e σ desconhecidos.

Sorteamos 100 pacotes e medimos seus pesos.

Variável a ser observada: X (peso dos pacotes de café);

População: conjunto de todos os pacotes enchidos ou que virão a ser enchidos;

Amostra: 100 medidas de pacotes selecionado.

- ▶ A amostra pode ser pensada como 100 observações feitas de uma distribuição normal.

Exercício

Para investigar a “honestidade” de uma moeda, a lançamos 50 vezes e contamos o número de caras observadas X .

Variável a ser observada: X (número de caras observadas);

População: Todas as possíveis sequências de caras e coroas;

Amostra: 50 lançamentos da moeda.

Agora considere a variável:

$$X_i = \begin{cases} 1, & \text{se o } i\text{-ésimo lançamento resulta em cara} \\ 0, & \text{se o } i\text{-ésimo lançamento resulta em coroa} \end{cases}$$

Nesse caso, a amostra será uma sequência de 50 *zeros* ou *uns*. O número de caras na amostra, X , será o número de *uns* na sequência. Em outros termos:

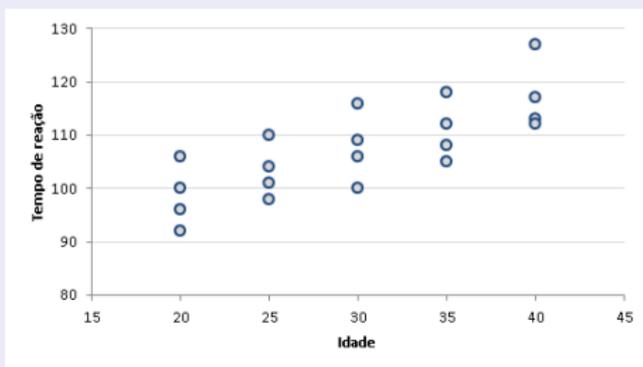
$$X = \sum_{i=1}^{50} X_i$$

Suponha que cada X_i resulta em **1** com probabilidade p . Então:

- Que valor de p traduz o conceito de moeda “honesta”?
- Existe um modelo probabilístico que representa bem X . Qual é?

Exemplo

Há razões para supor que o tempo y de reação a certo estímulo visual dependa da idade do indivíduo. Suponha que haja dependência *linear*. Foram selecionadas 20 pessoas das seguintes faixas de idade: 20, 25, 30, 35 e 40 anos, de ambos os sexos. Cada pessoa foi submetida ao teste e seu tempo de reação y foi medido.



Variável a ser observada: y (tempo de reação ao teste);

População: Todas as pessoas que viessem a ser submetidas ao teste (segundo idade e sexo);

Amostra: as 20 pessoas (ou as 20 medidas correspondentes).

Organização

- 1 Introdução
- 2 População e Amostra
- 3 Problemas de Inferência**
- 4 Como selecionar uma amostra

Problemas de Inferência

- Objetivo é produzir afirmações sobre características da população de interesse a partir de amostras;
- Uma característica pode ser representada por uma variável aleatória;
- Geralmente não temos conhecimento prévio da distribuição da variável aleatória na população. Às vezes temos conhecimento apenas parcial:
 - ▶ No caso da altura dos brasileiros adultos, podemos admitir que siga distribuição normal, mas... desconhecemos parâmetros (**nesse caso, quais são?**);
 - ▶ Às vezes conhecemos os parâmetros mas desconhecemos a forma da curva;
 - ▶ Às vezes não conhecemos nem forma nem parâmetros.
- Em todos os casos, uma amostra nos ajuda a formar opinião sobre o comportamento da variável (população)

Exemplo

Voltemos aos exemplo da moeda.

$$X \sim \text{Bin}(50, p)$$

O modelo é válido, admitindo-se ou não a “honestidade” da moeda ($p = \frac{1}{2}$). Lançada a moeda, vamos supor que tenham ocorrido 36 caras.

- ▶ Esse resultado traz evidência de que a moeda seja “honesta”?

Para tomarmos uma decisão, podemos adotar o seguinte procedimento:

- Partimos do princípio de que $p = \frac{1}{2}$ (moeda honesta). A partir dessa suposição podemos encontrar a probabilidade de se obterem 36 caras ou mais (usando o modelo binomial). Avaliamos o resultado.
- Suponha que a decisão tenha sido rejeitar a “honestidade” da moeda. Qual seria então a melhor estimativa para p considerando os resultados obtidos?

Acabamos de descrever os problemas básico da Inferência Estatística:

- é chamado de *Teste de hipóteses*
- é chamado de *Estimação*

Exemplo (pacotes de café)

Suponha que a máquina esteja regulada para encher os pacotes segundo uma distribuição normal com média 500 e desvio padrão 10 gramas.

$$X \sim N(500, 10^2)$$

Para verificar se a produção continua sob controle, vamos colher uma amostra de 100 pacotes e pesá-los.

Como a amostra pode nos ajudar a tomar uma decisão?

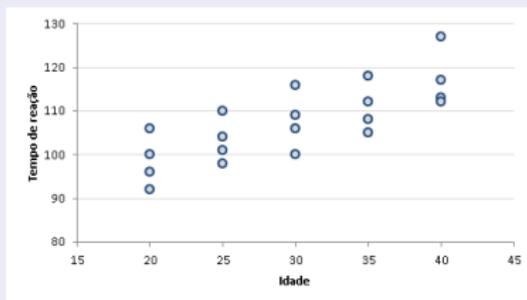
- ▶ Podemos usar a média amostral \bar{x} !
- ▶ Se \bar{x} se afastar muito de 500 gramas, suspeitaremos da regulagem da máquina.

Problema: decidir o que é próximo ou distante de 500 gramas...

- ▶ Importância de se entender as propriedades da distribuição de \bar{X} .

Exemplo (Acuidade visual)

Y : tempo de reação ao estímulo



Podemos supor que esse tempo, para cada idade x , seja uma v.a. com distribuição normal, com média dependendo da idade x .

$$Y \sim N(\mu(x), \sigma^2)$$

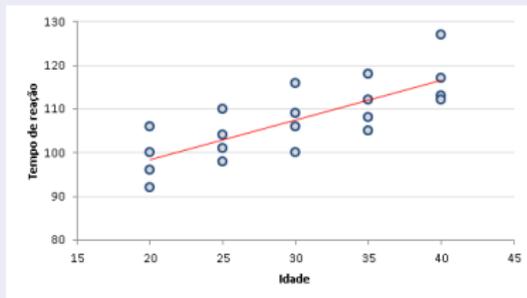
A *linearidade* expressa no problema pode ser incluída na média: $\mu(x) = \alpha + \beta x$

Questões de interesse:

- Estimar os parâmetros α e β , baseados na amostra de 20 dados;
- Investigar se a idade realmente afeta o tempo de reação ($\beta = 0$?);
- Considerando um grupo de pessoas com 40 anos, como deverá se comportar o tempo de reação? (previsão).

Exemplo (Acuidade visual)

Y : tempo de reação ao estímulo



Podemos supor que esse tempo, para cada idade x , seja uma v.a. com distribuição normal, com média dependendo da idade x .

$$Y \sim N(\mu(x), \sigma^2)$$

A *linearidade* expressa no problema pode ser incluída na média: $\mu(x) = \alpha + \beta x$

Questões de interesse:

- Estimar os parâmetros α e β , baseados na amostra de 20 dados;
- Investigar se a idade realmente afeta o tempo de reação ($\beta = 0$?);
- Considerando um grupo de pessoas com 40 anos, como deverá se comportar o tempo de reação? (previsão).

- No caso dos pacotes de café, se pudéssemos repetir o procedimento de colher a amostra de 100 pacotes um número grande vezes, sob a condição da máquina estar regulada, teríamos ideia do comportamento da v.a. \bar{x} . (e saberíamos dizer se um determinado valor observado é ou não raro de ocorrer);
- O mesmo vale para dados de acuidade visual;
- Repetir um experimento muitas vezes, sob as mesmas condições, nem sempre é possível...
- Mas em determinadas condições é possível determinar teoricamente o comportamento de algumas medidas feitas a partir da amostra (como, por exemplo, a média);
- Mas isso depende em grande parte do procedimento adotado para selecionar a amostra...

Organização

- 1 Introdução
- 2 População e Amostra
- 3 Problemas de Inferência
- 4 Como selecionar uma amostra**

Como selecionar uma amostra

- Quanto mais conhecemos a população, mais informativa é a amostra;
- Nem sempre a escolha de uma amostra adequada é imediata;
- No exemplo sobre a proporção de pessoas favoráveis a um determinado projeto governamental:
 - ▶ Escolhendo na amostra 200 pessoas moradoras do bairro que será beneficiado pelo projeto, o resultado conterà o que chamamos de *viés de seleção*;
 - ▶ Nesta amostra, a proporção de pessoas favoráveis ao projeto deve ser maior do que no todo.
- É necessário que a elaboração da amostra seja guiada por procedimentos estatísticos.

- A maneira de se obter a amostra é tão importante que esses procedimentos constituem especialidades dentro da Estatística:
 - ▶ Amostragem;
 - ▶ Planejamento de Experimentos;
 - ▶ Etc...
- Podemos dividir os procedimentos científicos em 3 grandes grupos:
 - a) Levantamentos Amostrais;
 - b) Planejamento de Experimentos;
 - c) Levantamentos Observacionais.

Levantamentos Amostrais

- Amostra é obtida de uma população bem definida por meio de protocolos bem definidos e controlados

Probabilísticos

- ▶ Usam mecanismos aleatórios de seleção dos elementos da amostra;
- ▶ Cada elemento tem uma probabilidade definida a *priori* de pertencer à amostra

Não probabilísticos

- ▶ Amostras intencionais: elementos são selecionados com ajuda de especialistas;
- ▶ Amostras de voluntários: ocorre em alguns testes de medicamentos e vacinas;
- ▶ Amostras probabilísticas possibilitam medição da precisão da amostra obtida; medidas são bem mais difíceis para os procedimentos não probabilísticos.

Planejamento de Experimentos

- Principal objetivo é analisar o efeito de uma variável sobre outra;
- Requer interferências do pesquisador sobre o ambiente em estudo e sobre fatores externos;
- Exemplos:
 - ▶ Bastante comuns em experimentação agrônômica;
 - ▶ Bastante comuns em ensaios clínicos em medicina, e.g., testar eficácia de um novo medicamento sobre uma doença;
 - ▶ Efeito da altura da gôndola sobre vendas de um produto no supermercado.

Levantamentos Observacionais

- Dados são coletados sem que pesquisador tenha controle sobre as informações obtidas;
- São exemplos típicos:
 - ▶ Pesquisas utilizando levantamentos de dados dos órgãos de pesquisa oficiais (IBGE, SEADE, etc);
 - ▶ Estudos com base em dados administrativos (INEP, Cadastro Único, RAIS, etc);
 - ▶ Séries temporais.
- Bastante suscetíveis a viés de seleção;
- Nesses casos, especificação de modelos estatísticos têm papel crucial.

Exemplo

Dê sua opinião sobre os tipos de problemas que surgiriam nos seguintes planos amostrais:

- a) Para investigar a proporção dos operários de uma fábrica favoráveis à mudança do início das atividades das 7h para as 7h30, decidiu-se entrevistar os 30 primeiros operários que chegassem à fábrica na quarta-feira.
- b) Mesmo procedimento, só que o objetivo é estimar a altura média dos operários.
- c) Para estimar a porcentagem média da receita municipal investida em lazer, enviaram-se questionários a todas as prefeituras, e a amostra foi formada pelas prefeituras que enviaram as respostas.
- d) Para verificar o fato de oferecer brindes nas vendas de sabão em pó, tomaram-se quatro supermercados na zona sul e quatro na zona norte de uma cidade. Nas quatro lojas da zona sul, o produto era vendido com brinde, enquanto nas outras quatro era vendido sem brinde. No fim do mês, compararam-se as vendas da zona sul com as da zona norte.