Estatística

Estimação:

Estimativa Pontual

Estimativa Intervalar

Renata Souza

Centro de Informática UFPE

Estatística Descritiva

- A *Estatística Descritiva* tem por objetivo resumir ou descrever características importantes de dados populacionais ou amostrais conhecidos;

- *Inferência Estatística* é o processo pelo qual tiram-se conclusões ou generalizações acerca de uma população usando informações de uma amostra.

Centro de Informática UFPE

Estimativa

- Um *estimador* é uma estatística amostral utilizada para obter uma aproximação de um parâmetro populacional.

- Uma *estimativa pontual* é um valor (ou ponto) único usado para aproximar um parâmetro populacional.
  - A média amostral é a melhor estimativa pontual para a média populacional.
  - Outra estimativa pontual é a variância amostral para a variância populacional.

Centro de Informática UFPE

1. Estimativa Pontual

- $\bar{x}$ é uma estimativa pontual para μ;

- Isto é, a média amostral é um valor usado para aproximar a média populacional;

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\left(x_1 + \ldots + x_n\right)$$

- Onde $(x_1, \ldots, x_n)$ é uma amostra.

Centro de Informática UFPE

Problema

- A maioria crê que a temperatura média do corpo humano é 98,6ºF. Uma amostra de dados parece sugerir que a média 98,2ºF. Sabemos que as amostras tendem a variar, de forma que talvez a verdadeira temperatura média seja 98,6ºF e a média amostral 98,2ºF seja resultado de uma flutuação aleatória.
  - Para quais valores, em relação à média 98,6ºF, e considerando um determinado erro, a média da amostra seria aceitável? (n = 106)

2. Estimativa Intervalar

- Quão boa é a estimativa pontual da média da população?

Definição

Estimativa intervalar (ou intervalo de confiança) é o intervalo de valores que contém a média da população com uma determinada probabilidade de acerto.

O intervalo de confiança está associado a um grau de confiança que é uma medida de nossa certeza de que o intervalo contém o parâmetro populacional. É a probabilidade *1-a*.

2. Estimativa Intervalar

- A construção do intervalo para μ é baseada na distribuição amostral da média amostral e no grau de confiança.

- Não é necessário que a suposição de normalidade para os dados seja adequada.

- A variância pode ou não ser conhecida. Para cada caso, usa-se diferentes distribuições.
  - 2.1 Estimativa Intervalar para variância conhecida;
  - 2.2 Estimativa Intervalar para variância desconhecida.

2.1 Variância conhecida

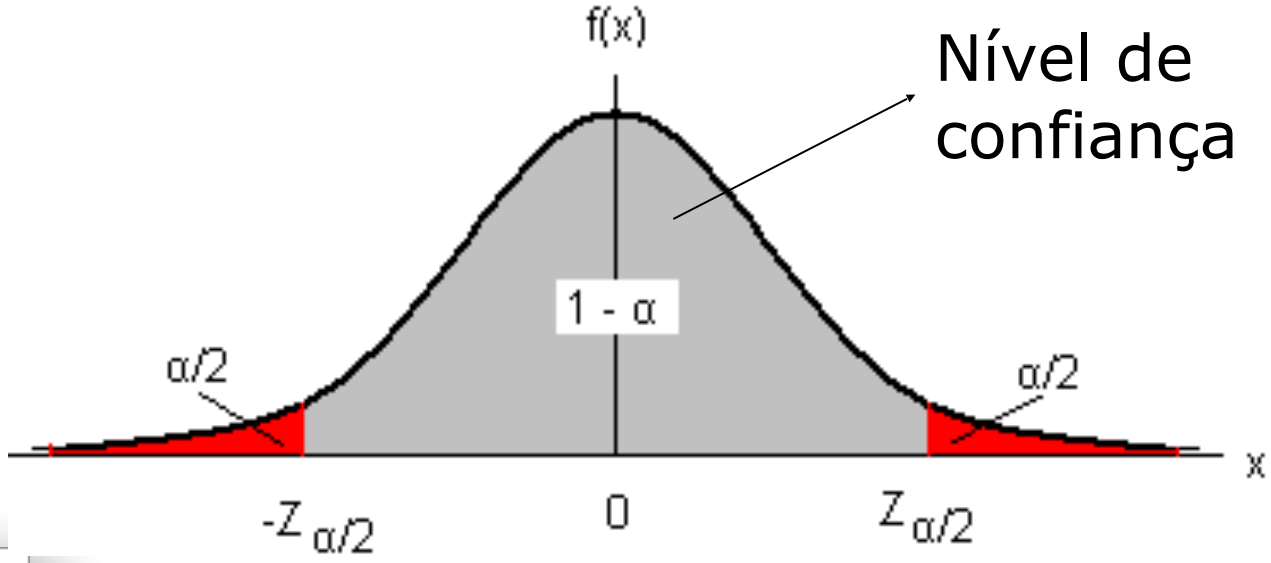
- Usando o *teorema central do limite*, a média amostral $\bar{x}$ é uma variável aleatória que tem distribuição normal com:
  - Média μ;
  - Desvio padrão $\frac{\sigma}{\sqrt{n}}$;

- Transformando $\bar{x}$ em uma variável aleatória normal padrão, temos: $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Centro de Informática UFPE

2.1 Variância conhecida

- $-Z_{\frac{\alpha}{2}}$ e $Z_{\frac{\alpha}{2}}$ são valores críticos;
- Um *valor crítico* é um número na fronteira que separa estatísticas amostrais que têm chance de ocorrer daquelas que não têm.



Centro de Informática UFPE

2.1 Variância conhecida

Com o desvio padrão $\frac{\sigma}{\sqrt{n}}$ e valores críticos

$-Z_{\frac{\alpha}{2}}$ e $Z_{\frac{\alpha}{2}}$, podemos definir os valores do intervalo de confiança para a média populacional μ:

$$-Z_{\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{\frac{\alpha}{2}}$$

Centro de Informática UFPE

2.1 Variância conhecida

A *margem de erro E* é a diferença máxima provável (com probabilidade *1-α*) entre a média observada (a média amostral) e a verdadeira média (média populacional);

O erro máximo é dado por:

$$E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Logo,

$$\bar{x} - \mathrm{E} \leq \mu \leq \bar{x} + E$$

Centro de Informática UFPE

Exemplo

As medidas dos pesos de uma amostra aleatória de 100 caminhões que foram medidos pela Polícia Rodoviária tem média de 3,2 toneladas e desvio padrão de 0,8 toneladas. Qual o intervalo de confiança para o grau de confiança de 95%?

$$E = 1{,}96 \cdot \frac{0{,}8}{10} = 0{,}157$$

$$3{,}2 - 0{,}157 \leq \mu \leq 3{,}2 + 0{,}157$$

Intervalo: $3{,}043 \leq \mu \leq 3{,}357$

Centro de Informática UFPE

2.2 Variância desconhecida

1. Estima-se a variância populacional através da variância amostral;

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

2. Usa-se *s* para calcular o intervalo de confiança para a média populacional e o valor $t_{\frac{\alpha}{2}}$ da tabela *t-Student* com *n-1 graus de liberdade*.

$$-t_{\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \leq t_{\frac{\alpha}{2}}$$

Centro de Informática UFPE

Grau de Liberdade

O *número de graus de liberdade* para uma coleção de dados amostrais é o número de valores amostrais que podem variar depois que certas restrições tiverem sido impostas aos dados amostrais.

Exemplo

Se 10 estudantes têm escores de testes com uma média de 80, podemos livremente atribuir valores aos nove primeiros escores, mas o 10º escore está, então, determinado. A soma dos 10 escores deve ser 800, de modo que o 10º escore deve ser 800 menos a soma dos 9 primeiros escores.

Centro de Informática UFPE

(...continuação)

Como esses 9 primeiros escores podemos ter valores escolhidos livremente, dizemos que há 9 graus de liberdade disponíveis. Portanto, o número de grau de liberdade é simplesmente o tamanho amostral menos 1.

Graus de liberdade = n - 1

Solução do Problema

$$E = 1,98 \cdot \frac{s}{\sqrt{n}} = 1,98 \cdot \frac{0,62}{\sqrt{106}} = 0,12$$

$$\bar{x} - \mathrm{E} \leq \mu \leq \bar{x} + E$$

$$98,20 - 0,12 \leq \mu \leq 98,20 + 0,12$$

$$98,08 \leq \mu \leq 98,32$$

Como o intervalo acima não contém 98,6ºF, parece muito pouco provável que o valor correto de μ seja 98,6ºF (poderemos dar uma resposta mais segura com o uso de Testes de Hipótese).

Centro de Informática UFPE

Intervalo de Confiança

Devemos ser cuidadosos para interpretar corretamente os intervalos de confiança. Considere o intervalo de confiança $0{,}476 \leq \mu \leq 0{,}544$.

Correta: "Estamos 95% confiantes de que o intervalo de 0,476 a 0,544 realmente contém o verdadeiro valor de μ."

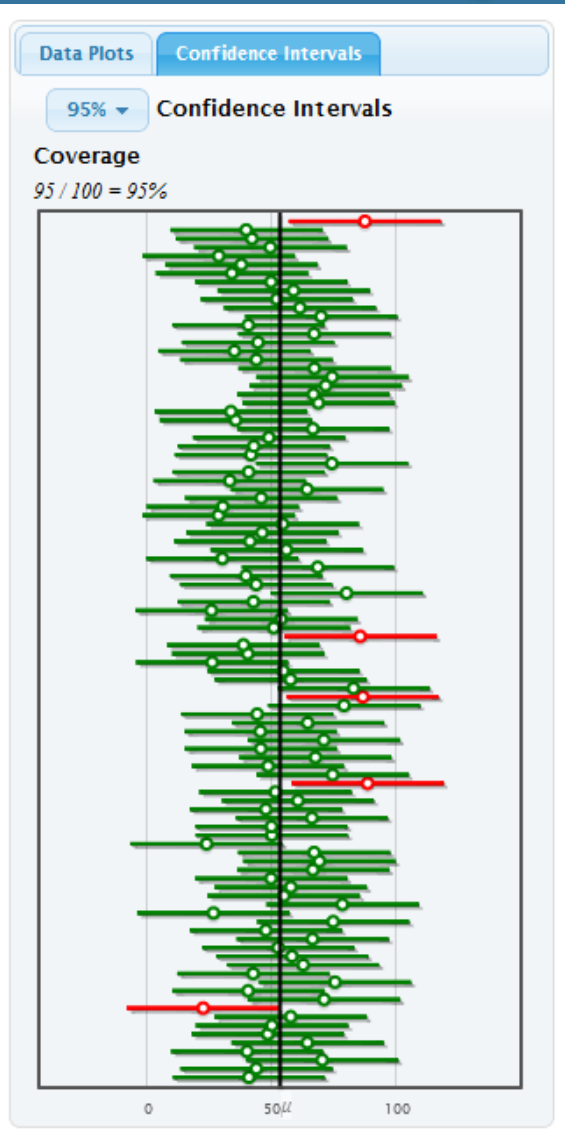
Errada: "Há uma chance de 95% de que o verdadeiro valor de μ estará entre 0,476 e 0,544." ou "95% de todos os valores amostrais estão entre 0,476 e 0,544".

Justificativa: μ é uma constante fixa (embora desconhecida), não uma variável aleatória. E o intervalo de confiança não descreve o comportamento de médias amostrais individuais.

Centro de Informática UFPE

Interpretando Intervalo de Confiança

- O parâmetro é fixo
- A estatística é aleatória (depende da amostra).
- O intervalo é aleatório (depende da estatística).
- 95% dos intervalos conterão o parâmetro.



Centro de Informática UFPE

Exercício

1. As medidas dos diâmetros de uma amostra aleatória de 200 rolamentos esféricos produzidos por certa máquina, durante uma semana, apresentam a média de 0,824 polegada e o desvio padrão de 0,042 polegada. Determine os limites de confiança de (a) 95%, (b) 99%, para o diâmetro médio de todos os rolamentos esféricos.

Centro de Informática UFPE

Exercício

2. Um armazém desconfiado com o seu fornecedor de tijolos (alegando má qualidade) fez medidas do seu peso de uma amostra, encontrando média de 0,8kg para cada um e desvio padrão de 0,05kg. Quantas medidas de peso de tijolos foram feitas pelo armazém, sabendo que o normal é que o tijolo pese 0,89kg? (considere o grau de confiança de 95%)