Inferência

1 – Estimativa pontual de uma média

2 – Estimativa intervalar de uma média

Renata Souza

Centro de Informática UFPE

Aspectos Gerais

- A estatística descritiva tem por objetivo resumir ou descrever características importantes de dados populacionais ou amostrais conhecidos;
- A inferência estatística tem por objetivo fazer generalizações sobre a população em estudo a partir dos dados amostrais

Estimativa pontual da média

- Um estimador é uma estatística amostral (como a média 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.

Estimativa pontual da média da população μ

$\bar{x}$ é uma estimativa de μ.

Seja $(x_1,...,x_n)$ uma amostra de uma variável aleatória X.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Quão boa é esta estimativa?

Estimativa intervalar

- Um intervalo de confiança (estimativa intervalar) é uma amplitude de valores que tem probabilidade de conter o verdadeiro valor da população.
- Um 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- α.
- A construção do intervalo para μ é baseada na distribuição amostral da média amostral e no grau de confiança.
- É necessário que a suposição de normalidade para os dados seja adequada.

Estimativa intervalar: variância conhecida

- Usando o teorema central do limite a média amostral $\overline{X}$ é uma variável aleatória que tem distribuição normal com

Média μ e Desvio padrão $\frac{\sigma}{\sqrt{n}}$

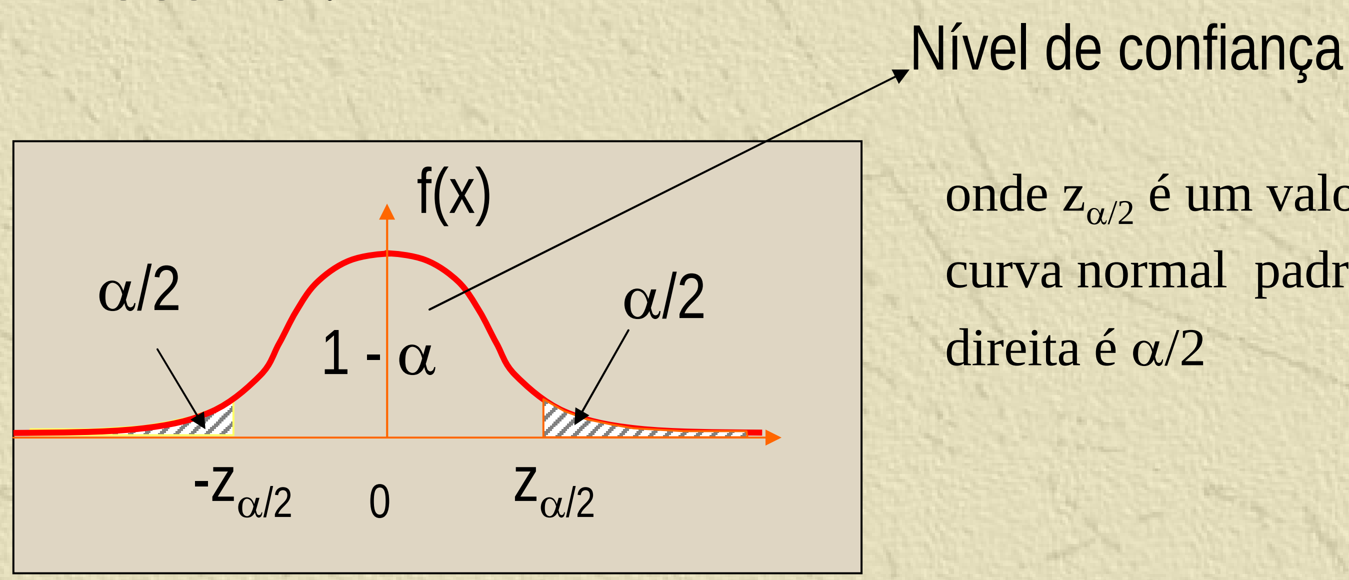
- Transformando $\overline{X}$ em uma variável aleatória (VA) normal padrão temos:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

- As médias amostrais apresentam uma chance relativamente pequena de estar em uma das caudas da distribuição normal.
- α. é a probabilidade da média estar em uma das caudas.

Estimativa intervalar: variância conhecida

- $-z_{\alpha/2}$ e $z_{\alpha/2}$ são valores críticos.
- Um valor crítico é um número na fronteira que os valores das estatísticas amostrais prováveis de ocorrerem do valores que têm pouca chance de ocorrer.



onde $z_{\alpha/2}$ é um valor tal que a área da curva normal padronizada à sua direita é $\alpha/2$

Estimativa intervalar: variância conhecida

- Com ao valores críticos $-z_{\alpha/2}$ e $z_{\alpha/2}$ e $\frac{\sigma}{\sqrt{n}}$ podemos definir os valores do intervalo de confiança para μ.

$$-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

o que equivale a:

$$\bar{X} - z_{\alpha/2}\,\sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_{\alpha/2}\,\sigma/\sqrt{n}$$

Estimativa intervalar: 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 pode ser obtido da seguinte maneira

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Estimativa intervalar: variância desconhecida

- Estima-se a variância populacional σ^2 através da variância amostral.

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

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

$$\bar{x} - t_{\alpha/2} s / \sqrt{n} \le \mu \le \bar{x} + t_{\alpha/2} s / \sqrt{n}$$

Construindo estimativas intervalares das médias usando distribuição t

- Exemplo 2: 16 restaurantes são selecionados ao acaso e mede-se a temperatura do café vendido em cada um. A temperatura média amostral é de 162ºF, com desvio amostral de 10ºF. Obtenha o intervalo de confiança de 95% para a temperatura média.

$$E = t_{\alpha} \frac{s}{\sqrt{n}} = 2,131 \frac{10}{\sqrt{16}} = 5,32$$

$$156,67 \leq \mu \leq 167,32$$

Propriedades da distribuição t

- Se uma variável aleatória X e aproximadamente normal, a distribuição amostral de $\bar{x}$ é uma distribuição t

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

- 1. Tem forma de sino e é simétrica em torno da média.
- 2. É uma família de curvas, cada uma determinada pelo número de graus de liberdade.
  - Graus de liberdade: número de escolhas livres deixados por uma amostra após uma estatística ter sido calculada.

Propriedades da distribuição t

- 3. A forma de sino reflete maior variabilidade que se espera com pequenas amostras.
- Quando o número de graus de liberdade cresce, a distribuição tende para a distribuição normal.
- Tem média zero.

Exercícios

- Registram-se os valores 0,27; 0,26; 0,27; 0,33; e 0,32 em segundos, obtidos em cinco medições do tempo de fabricação de certo produto. Seja $\sigma = 0{,}03$. Determine intervalos de confiança com níveis de 95% e 99%, para o tempo real médio que se requer para fabricar o produto
- Um prefeito de certa cidade turística deseja estimar a média de gastos para os turistas que visitam a cidade. Com este propósito, uma amostra aleatória de 120 turistas foi selecionada para a investigação e encontrou-se que a média foi igual a 800 u.m.(unidades de medidas) com desvio padrão de 200 u.m.
  - Achar o intervalo de confiança, a 99% para a média de gastos dos turistas com a cidade.
  - Qual deve ser o tamanho mínimo da amostra para que, ao nível de confiança de 95%, o erro de estimação seja 20 u.m.