



Teoria da Estimação

Prof. Dr. Rogério de Melo Costa Pinto

Curso de Pós-Graduação
Especialização em Estatística Empresarial

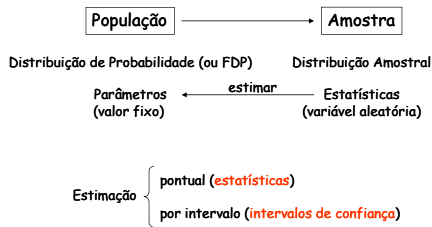


✓ **Estimação:** É o processo que consiste no uso de dados amostrais para estimar valores de parâmetros populacionais desconhecidos, tais como média, desvio padrão, proporções, etc.

Estimação por ponto
vs
Estimação por intervalo



Estimação de Parâmetros



OBS: **estatística**: é a v.a. que estima (pontualmente) um parâmetro (populacional) as vezes é chamada simplesmente de **estimador**
estimativa: é o valor do estimador obtido para uma amostra específica

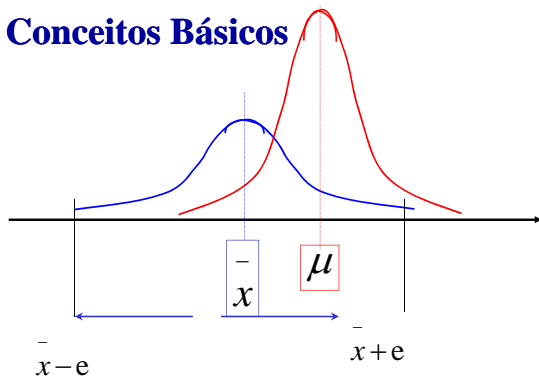
Estimação por ponto: neste caso obtém-se um único valor amostral que serve como uma aproximação do parâmetro estimado.

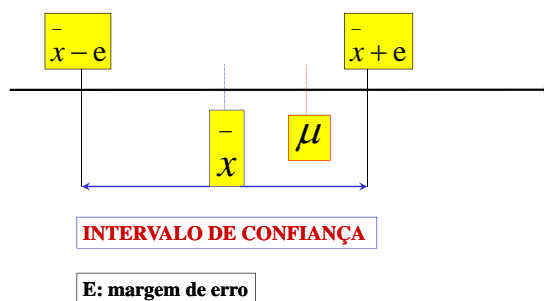
Ex: o resultado da média amostral \bar{x} é uma estimativa por ponto da média populacional μ

Estimação por intervalo: fazemos uma estimativa de um intervalo de possíveis valores, no qual se admite que o parâmetro populacional, com uma determinada confiança.

Ex: $\bar{x} = 50$ *estimar* $(40 < \mu < 60) = 0,95$ ou seja a verdadeira média populacional está dentro do intervalo de 40 a 60 com uma certeza de 95%.

Inferência Conceitos Básicos






Universidade Federal de Uberlândia
 Faculdade de Matemática
 Curso de Especialização em Estatística Empresarial

Questões Comuns.....

- ✓ Como estimar os parâmetros de uma grandeza com base numa simples amostra?
- ✓ Qual a confiança neste resultado?
- ✓ Como amostrar? se

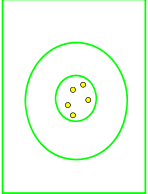


Universidade Federal de Uberlândia
 Faculdade de Matemática
 Curso de Especialização em Estatística Empresarial

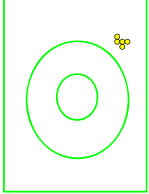
Propriedades dos estimadores

Acurácia vs precisão.

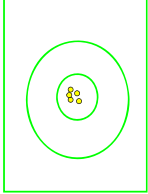
Acurácia



Precisão



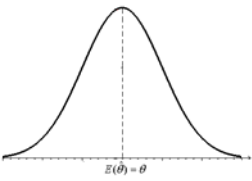
Acurácia e Precisão



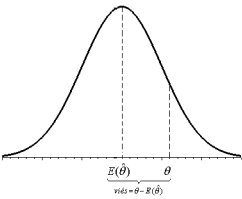
Universidade Federal de Uberlândia
 Faculdade de Matemática
 Curso de Especialização em Estatística Empresarial

Não Tendenciosidade

$$E(\hat{\theta}) = \theta$$



$E(\hat{\theta}) = \theta$



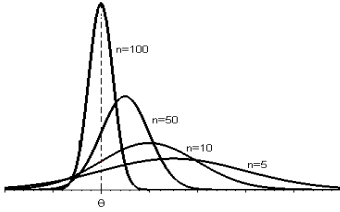
$E(\hat{\theta}) < \theta$
 $\text{viés} = \theta - E(\hat{\theta})$



Consistência

$$i) \lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$$

$$ii) \lim_{n \rightarrow \infty} V(\hat{\theta}) = 0$$





\bar{X} é consistente.

$$E(\bar{X}) = \mu$$

$$V(\bar{X}) = \frac{\sigma^2}{n} \lim_{n \rightarrow \infty} V(\bar{X}) = 0$$



Eficiência relativa

$\hat{\theta}_1$ é mais eficiente que $\hat{\theta}_2$ se:

$$V(\hat{\theta}_1) < V(\hat{\theta}_2)$$

$$Ef_{\hat{\theta}_1/\hat{\theta}_2} = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

Intervalo de Confiança para μ

$X \sim N(\mu, \sigma^2)$ μ desconhecido, mas σ^2 conhecido

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

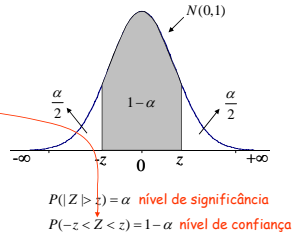
$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \text{ (Normal Padrão)}$$

$$P\left(-z < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z\right) = 1 - \alpha$$

$$P\left(-z \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(\bar{X} - z \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

IC para μ



Intervalo de Confiança para μ

1º Caso - Amostras grandes ($n > 30$)

(i) o IC $IC(\mu)_{1-\alpha} : \bar{x} \pm e$

(ii) erro da estimativa $e = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} ; e = z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$

z depende da confiança que se queira do intervalo

Valores típicos de z

GRAU DE CONFIANÇA	$Z_{\alpha/2}$
90%	1,64
95%	1,96
99%	2,57



Exemplo

Uma pesquisa nos bancos de dados de um *call-center* mostrou que, em 121 chamadas amostradas, a venda média foi de R\$ 700, com desvio padrão de R\$ 100. Como estimar a venda média deste negócio?

Resp: intervalo de confiança da média



Solução do exemplo

$$\text{venda} : 700 \pm 1,96 \frac{100}{\sqrt{121}}$$

$$\text{venda} : R\$700 \pm R\$17,8$$

(confiança : 95%)

Repetir para confiança de 90% e 99%



Tamanho da Amostra (n) para Estimativa da Média

Volta ao exemplo do call-center (média R\$700, desvio-padrão R\$100, 95% confiança)

e: R\$ 10 tamanho do intervalo desejado (semi-amplitude);

$$(690) \text{---} (700) \text{---} (710)$$

$$e = 1,96 \frac{100}{\sqrt{n}}$$



$$n = \left(1,96 \frac{100}{10}\right)^2$$

$$n = 385$$

Intervalo de Confiança para μ

2º Caso - Amostras pequenas ($n \leq 30$)

(i) O IC $IC(\mu)_{1-\alpha} : \bar{x} \pm e$

(ii) erro da estimativa $e = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$

GL= n-1

Intervalo de Confiança para μ

$X \sim N(\mu, \sigma^2)$ μ e σ^2 desconhecidos

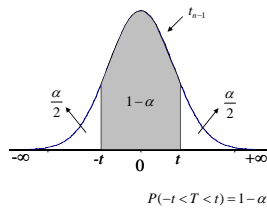
$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$

$P(-t < \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} < t) = 1 - \alpha$

$P(-t \frac{s}{\sqrt{n}} < \bar{X} - \mu < t \frac{s}{\sqrt{n}}) = 1 - \alpha$

$P(\bar{X} - t \frac{s}{\sqrt{n}} < \mu < \bar{X} + t \frac{s}{\sqrt{n}}) = 1 - \alpha$

IC para μ



Intervalo de Confiança para μ

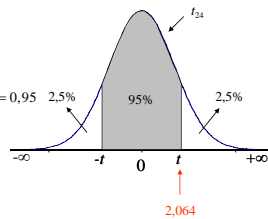
Exemplo: uma v.a. qualquer tem uma distribuição desconhecida com média μ e variância σ^2 também desconhecidas. Retira-se uma amostra de 25 valores e calcula-se a média amostral e a variância amostral. Construa um IC de 95% para μ supondo que $\bar{X} = 12,7$ e $s^2 = 16$.

$P(\bar{X} - t \frac{s}{\sqrt{n}} < \mu < \bar{X} + t \frac{s}{\sqrt{n}}) = 0,95$

$P(12,7 - 2,064 \frac{4}{\sqrt{25}} < \mu < 12,7 + 2,064 \frac{4}{\sqrt{25}}) = 0,95$

$P(12,7 - 1,6512 < \mu < 12,7 + 1,6512) = 0,95$

$P(11,0488 < \mu < 14,3512) = 0,95$



Dimensionamento da Amostra

$$n_o = \frac{t_{\alpha/2}^2 \cdot S^2}{e^2}$$

$$n = \frac{n_o}{1 + \frac{n_o - 1}{N}}$$

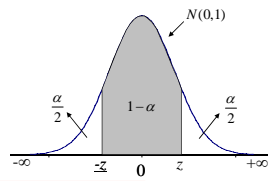
Intervalo de Confiança para $\mu_1 - \mu_2$

$X_1 \sim N(\mu_1, \sigma_1^2)$ $X_2 \sim N(\mu_2, \sigma_2^2)$ μ_i desconhecidas, mas σ_i^2 conhecidas

$\bar{X}_1 \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$ $\bar{X}_2 \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$ $\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$P(-z < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z) = 1 - \alpha$$



$$P(\bar{X}_1 - \bar{X}_2 - z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}) = 1 - \alpha$$

IC para $\mu_1 - \mu_2$

$$P(-z < Z < z) = 1 - \alpha$$

Intervalo de Confiança para $\mu_1 - \mu_2$

1º Caso - Amostras Grandes ($n > 30$)

(i) O IC $IC(\mu_1 - \mu_2)_{1-\alpha} : (\bar{x}_1 - \bar{x}_2) \pm e$

(ii) erro da estimativa

$$e = z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} ; e = z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$



Exemplo

Estão sendo estudados dois processos (A e B) para conservar alimentos, cuja principal variável de interesse é o tempo de duração dos mesmos. Sorteiam-se duas amostras independentes. Admita que as populações apresentam variâncias iguais.

Para uma amostra $n_A=n_B=100$ latas que apresentou tempo médio de duração de 50 e 60 dias e variância de 100 e 75 dias² para A e B, respectivamente. Construa um IC de 98% para a diferença entre médias. O que se conclui com este IC?



Intervalo de Confiança para $\mu_1 - \mu_2$

2º Caso - Amostras pequenas ($n \leq 30$), independentes e variâncias populacionais estatisticamente iguais

(i) O IC

$$IC(\mu_1 - \mu_2)_{1-\alpha} : (\bar{x}_1 - \bar{x}_2) \pm e$$

(ii) erro da estimativa

$$e = t_{\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \therefore GL = n_1 + n_2 - 2$$

$$s_p = \sqrt{\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}}$$



Intervalo de Confiança para $\mu_1 - \mu_2$

3º Caso - Amostras pequenas ($n \leq 30$), independentes e variâncias populacionais estatisticamente desiguais

(i) O IC $IC(\mu_1 - \mu_2)_{1-\alpha} : (\bar{x}_1 - \bar{x}_2) \pm e$

(ii) erro da estimativa

$$e = t_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \therefore g = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$



Exemplo

Os dados a seguir referem-se aos pesos médios de latas de alumínio (gr) produzidas por 2 empresas.

A:	34	30	31	33	34	34	33	32
B:	25	20	34	28	32	30	27	31
	20	29						

a) Encontre o IC para a diferença entre médias das 2 empresas com uma confiança de 95% (considere que as populações apresentam variâncias diferentes);

b) Encontre o IC para a diferença entre médias das 2 empresas com uma confiança de 99% (considere que as populações apresentam variâncias iguais);

interprete os resultados dos itens (a) e (b).



Intervalo de Confiança para amostras relacionadas

4º Caso - Amostras dependentes, dados aparecem em pares (Ex.: antes e depois)

(i) O IC $IC(\mu_D)_{1-\alpha} : \bar{D} \pm e \quad \therefore \bar{D} = \frac{\sum d_i}{n} \quad \therefore d_i = X_{1i} - X_{2i}$

(ii) erro da estimativa

$$e = t_{\alpha/2} \cdot \frac{s_D}{\sqrt{n}} \quad \therefore s_D = \sqrt{\frac{\sum d_i^2 - (\sum d_i)^2}{n-1}}$$



Exemplo

A seguir são mostradas as perdas semanais médias de horas-homem devido a acidentes em 10 indústrias, antes e depois da adoção de um programa de segurança.

Antes:	45	73	46	124	33	57	83	34
	26	17						
Depois:	36	60	44	119	35	51	77	29
	24	11						

a) Encontre um IC de 90% para o efeito da intervenção;

b) Quantas indústrias você recomendaria usar numa próxima pesquisa para estimar a eficiência do programa com 90% de confiança e um erro de 0,5 horas-homem.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Intervalo de Confiança para proporção p

$\hat{p} \sim N\left(p, \frac{pq}{n}\right)$

$\frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$

$P(-z < \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} < z) = 1 - \alpha$

$P(\hat{p} - z \sqrt{\frac{pq}{n}} < p < \hat{p} + z \sqrt{\frac{pq}{n}}) = 1 - \alpha$

$P(\hat{p} - z \sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z \sqrt{\frac{\hat{p}\hat{q}}{n}}) = 1 - \alpha$

IC para p

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Intervalo de Confiança para proporção p

IC(p)_{1-α} : $\hat{p} \pm e$

$p = \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Dimensionamento da Amostra

$n = \frac{z_{\alpha/2}^2 \hat{p} \hat{q}}{e^2}$

$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$



Exemplo

Uma pesquisa de mercado com 90 consumidores mostrou que o market-share de sua empresa é de 25%. Como estimar o market-share do mercado, com 95% de confiança?



Solução do exemplo

$$p = 0,25 \pm 1,96 \sqrt{\frac{0,25(1-0,25)}{90}}$$

$$p = 0,25 \pm 0,09$$

Repetir p/outros níveis de confiança



Tamanho da Amostra (n) para Estimativa da Proporção

Volta ao exemplo do market-share ($\hat{p} = 0,25$, 95% confiança)

e: 2% tamanho do intervalo desejado (semi-amplitude);

(23%) _____ (25%) _____ (27%)

$$n = \frac{1,96^2 \cdot 0,25 \cdot 0,75}{0,02^2}$$

$$n = 1801$$





Intervalo de Confiança para $p_1 - p_2$

$$IC(p_1 - p_2)_{1-\alpha} : (p_1 - p_2) \pm e$$

$$e = z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$



Exemplo

Em uma pesquisa de proprietários de carros em uma universidade, entre alunos e alunas, foram obtidos: 24 de 100 alunos possuem automóveis e 19 de 100 alunas possuem automóveis. Encontre um IC de 98% e 90% para a diferença entre proporções.

Interprete os resultados.

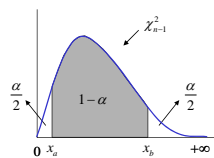


Intervalo de Confiança para σ^2

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

$$P(x_a < \frac{(n-1)s^2}{\sigma^2} < x_b) = 1 - \alpha$$

$$P\left(\frac{1}{x_b} < \frac{\sigma^2}{(n-1)s^2} < \frac{1}{x_a}\right) = 1 - \alpha$$



$$P(x_a < \chi_{n-1}^2 < x_b) = 1 - \alpha$$

$$P\left(\frac{(n-1)s^2}{x_b} < \sigma^2 < \frac{(n-1)s^2}{x_a}\right) = 1 - \alpha$$

IC para σ^2



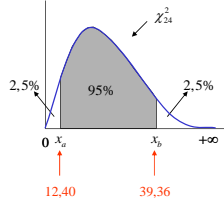
Intervalo de Confiança para σ^2

Exemplo: uma v.a. qualquer tem uma distribuição desconhecida com média μ e variância σ^2 desconhecidas. Retira-se uma amostra de 25 valores e calcula-se a variância amostral. Construa um IC de 95% para σ^2 supondo que $s^2 = 2,34$.

$$P\left(\frac{(n-1)s^2}{x_b} < \sigma^2 < \frac{(n-1)s^2}{x_a}\right) = 0,95$$

$$P\left(\frac{24 \cdot 2,34}{39,36} < \sigma^2 < \frac{24 \cdot 2,34}{12,40}\right) = 0,95$$

$$P(1,43 < \sigma^2 < 4,53) = 0,95$$





CASE

- 1) Se a proporção de pessoas que pretendem votar no Partido Verde deve ser estimada com um nível de confiança de 95%, que tamanho de amostra deverá ser definido:
 - a) Se o erro for no máximo 0,01?
 - b) Se o erro for no máximo 0,03?
 - c) Se o erro for no máximo 0,06?
 - d) Repita os itens acima com um nível de confiança de 90%.



- 2) A gerência de um restaurante local pretende determinar o gasto mensal médio das famílias em restaurantes de luxo. Algumas famílias nada gastam, enquanto outras gastam até R\$300,00 por mês, sendo o desvio-padrão de R\$40,00. A gerência pretende ter 95% de confiança nos resultados e não quer que o erro exceda mais ou menos R\$5,00.
 - a) Qual o tamanho amostral que deve ser usado para determinar a despesa mensal média por família?
 - b) Após feita a pesquisa, obteve-se uma despesa média de R\$90,30, com desvio padrão de R\$45,00. Construa um intervalo de 95% de confiança. Que se pode dizer quanto ao nível de precisão?



3) Para determinar a efetividade da campanha de propaganda para um novo vídeo cassete, a gerência está interessada em saber que porcentagem das residências conhece realmente a nova marca. A agência de propaganda acha que essa cifra chega aos 70%. A gerência deseja um intervalo de confiança de 95% e uma margem de erro não superior a mais ou menos 2%.

a) Qual deve ser o tamanho da amostra para esse estudo?

b) Suponha que a gerência deseja 99% de confiança, mas esteja disposta a tolerar um erro de mais ou menos 3%. Como se modificaria o tamanho da amostra?
