Estimação

Avaliar características da população com base em informações da amostra

$$\Downarrow$$

Estimar os parâmetros

Mais utilizadas:

- média (μ)
- proporção (π)
- variância (σ^2)

Exemplos:

- produção média de determinada cultura;
- proporção média de área foliar atacada por uma praga;
- parâmetros estatísticos genéticos (variância genética, ambiental e fenotípica)...
- proporção de crescimento de um microorganismo;
- proporção de alimento estragado/vencido na prateleira;

Propriedades dos estimadores

Estimadores

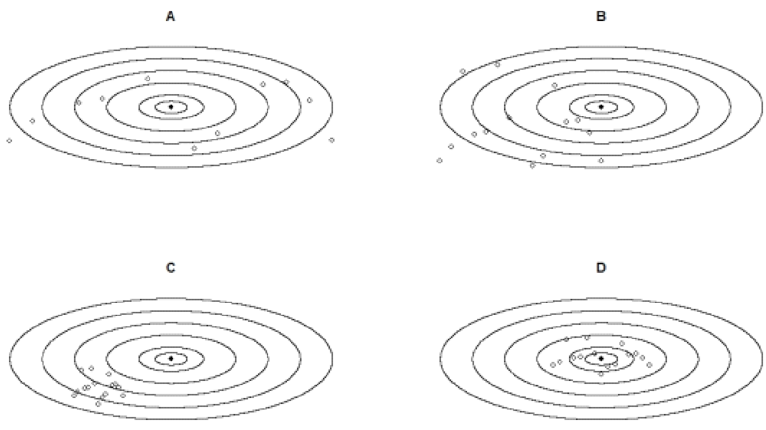
$$\text{Média: } \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\text{Proporção: } P = \frac{\text{número de sucessos}}{n}$$

$$\text{Variância: } S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

Propriedades

- **não viesado**: média da distribuição amostral igual ao parâmetro
- **preciso**: variância amostral pequena
- **acurado**: erro amostral pequeno



A: não viesado, pouca precisão e pouca acurácia
B: viesado, pouca precisão e pouca acurácia
C: viesado, boa precisão e baixa acurácia
D: não viesado, boa precisão e boa acurácia

Estimativas pontuais e intervalares

Modelo probabilístico

$\Downarrow$

Estimar os parâmetros da distribuição

$\Downarrow$

Amostra

Estimadores $\Rightarrow$ **Estatísticas**

Estimativas pontuais

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

$$\text{Proporção: } p = \frac{\text{número de sucessos}}{n}$$

$$\text{Variância: } s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Estimativas intervalares

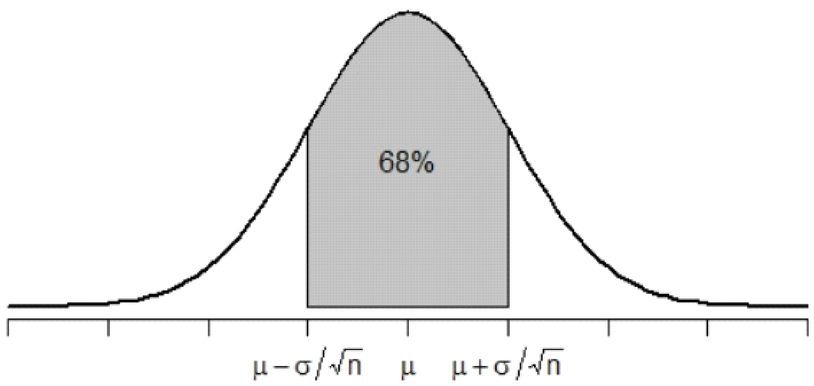
Intervalo de confiança

Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de tamanho n de uma população e θ o parâmetro de interesse. Sejam $\hat{\theta}_1$ e $\hat{\theta}_2$ estatísticas tais que:

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha.$$

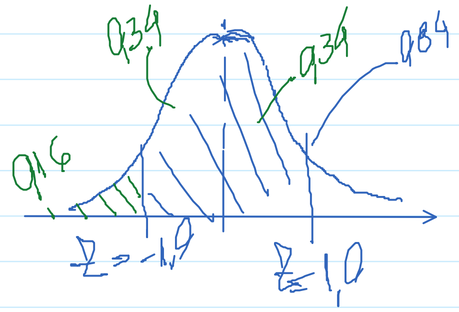
Então o intervalo $(\hat{\theta}_1; \hat{\theta}_2)$ é chamado intervalo de **100(1-α)% de confiança** para o parâmetro θ. Usualmente toma-se $1 - \alpha$=0,95 ou 0,99.

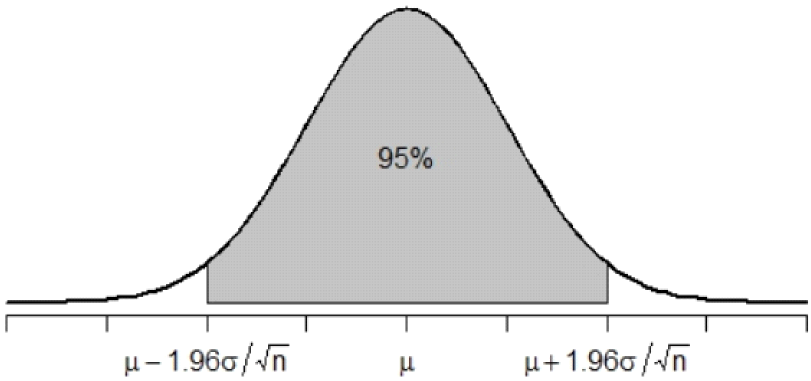
Distribuição normal



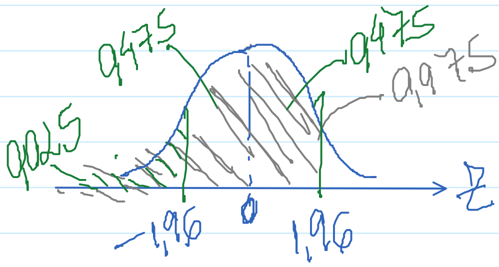
Podemos dizer que 68% dos possíveis valores da média de uma amostra aleatória simples de tamanho n não se afastam mais do que $\sigma/\sqrt{n}$.

0,34 0,34 0,84





Podemos dizer que 95% dos possíveis valores da média de uma amostra aleatória simples de tamanho n não se afastam mais do que $1,96\sigma/\sqrt{n}$.



99%
$Z_t = 2,58$

Intervalos de confiança para média populacional

Casos

- População Normal e Variância da população conhecida;
- População Normal e Variância da população desconhecida;
- População não Normal, grandes amostras (n>30).

Intervalo de confiança para média

- **População normal e variância populacional conhecida**

$$X \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \tfrac{\sigma^2}{n}) \Rightarrow Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

$$P\left(-z_T < \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} < z_T \right) = 1-\alpha$$

$$\cdots$$

$$P\left(\bar{X} - z_T\sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X} + z_T\sqrt{\frac{\sigma^2}{n}} \right) = 1-\alpha$$

$$\boxed{IC(\mu)_{1-\alpha} = \left(\bar{X} - z_T\sqrt{\frac{\sigma^2}{n}} ; \bar{X} + z_T\sqrt{\frac{\sigma^2}{n}} \right)}$$

Exemplo: A distribuição dos pesos de pacotes de determinadas sementes, enchidos automaticamente por uma certa máquina, é normal, com desvio padrão (σ) conhecido e igual a 0,20 kg. Uma amostra de 15 pacotes retirada ao acaso apresentou os seguintes pesos, em kg:

20,05	20,10	20,25	19,78	19,69	19,90	20,20	19,89
19,70	20,30	19,93	20,25	20,18	20,01	20,09	

Construir os intervalos de confiança de 95% e 99% para o peso médio dos pacotes de sementes.

$\bar{x} = 20,02$ $\quad n = 15 \quad$ $\sigma = 0,20$ kg

$IC(\mu)_{95\%}: \ 20,02 \pm 1,96 \cdot \frac{0,20}{\sqrt{15}}$

$$\boxed{IC(\mu)_{95\%}: \ 19,92 < \mu < 20,12}$$

$IC(\mu)_{99\%}: \ 20,02 \pm 2,58 \cdot \frac{0,20}{}$

$IC(\mu)_{99\%}: 20,02 \pm 2,58 \cdot \frac{0,20}{\sqrt{15}}$

$\boxed{IC(\mu)_{99\%}: 19,89 < \mu < 20,15}$

Exercício:

Para uma amostra de 36 observações de uma população normal com variância populacional conhecida $\sigma^2 = 0{,}36kg^2$ *e média* μ *desconhecida, seja* $\bar{x} = 59{,}6\ kg$ *a média amostral. Construir os intervalos de 95 e 99% de confiança para* μ.

$n = 36 \qquad \sigma^2 = 0,36 kg^2 \qquad \bar{x} = 59,6 kg$

95%

$z_T = 1,96 \qquad IC(\mu)_{95\%} = \bar{x} \pm z_T \sqrt{\frac{\sigma^2}{n}}$

$= 59,6 \pm 1,96 \cdot \sqrt{\frac{0,36}{36}}$

$\boxed{IC(\mu)_{95\%}: 59,404 < \mu < 59,796}$

99%

$z_T = 2,58 \qquad IC(\mu)_{99\%} = 59,6 \pm 2,58 \sqrt{\frac{0,36}{36}}$

$\boxed{IC(\mu)_{99\%}: 59,342 < \mu < 59,858}$

Exercício:

De um povoamento de eucaliptos, sortearam-se 10 árvores e determinaram-se os diâmetros, em cm, com a finalidade de estimar o diâmetro médio do povoamento, que possui desvio padrão conhecido de 7,8cm. Estes diâmetros foram:

10,1	15,8	18,5	22,3	23,5	17,2	17,8	18,7	16,7	28,0

$\bar{x} = 18,86$ cm

10,1	15,8	18,5	22,3	23,5	17,2	17,8	18,7	16,7	28,0

$\bar{x} = 18,86$ cm

$\sigma = 7,8$ cm

Com base nessa amostra, calcule:

a) intervalo com grau de confiança de 95%;

$Z_T = 1,96$

$IC: \bar{x} \pm Z_T \sqrt{\frac{\sigma^2}{n}}$

$: 18,86 \pm 1,96 \cdot \frac{7,8}{\sqrt{10}}$

$$\boxed{IC(\mu)_{95\%}: 14,03 < \mu < 23,69}$$

b) intervalo com grau de confiança de 99%.

$Z_T = 2,58$

$IC: \bar{x} \pm Z_T \cdot \sqrt{\frac{\sigma^2}{n}}$

$: 18,86 \pm 2,58 \cdot \frac{7,8}{\sqrt{10}}$

$$\boxed{IC(\mu)_{99\%}: 12,50 < \mu < 25,22}$$

***Exercícios**:*

***1)** Um antropólogo mediu as alturas de uma amostra aleatória de 100 homens de determinada população, encontrando uma média amostral de 173 cm. Se o desvio padrão da população for de 9 cm, calcular:*

a) um intervalo de 95% de confiança para a altura média de toda a população; interpretar o intervalo de confiança;

b) um intervalo de 99% de confiança para a altura média de toda a população; interpretar o intervalo de confiança.

2) *Uma máquina enche pacotes de café cum um desvio padrão igual a 10. Ela estava regulada para enchê-los com 500g, em média. Agora ela está desregulada e queremos saber qual a nova média verdadeira (populacional). Uma amostra de 25 pacotes apresentou média igual a 485g.*

a) Construir intervalos de confiança de 95% e de 99% para a média verdadeira; interpretar o intervalo de confiança;

b) Qual o erro máximo associado aos intervalos encontrados em "a"?

c) Qual o tamanho da amostra será necessário para produzir um intervalo de confiança para a verdadeira média populacional, com precisão de 3,5g de café para mais e para menos?

3) *De 1000 lavouras de arroz, foi levantada uma amostra de 25 lavouras e a informação a respeito da produtividade permitiu o cálculo do rendimento médio, por hectare, que foi de 3400kg com desvio padrão populacional conhecido de 150kg.*

a) Determine intervalos com grau de confiança de 95% e de 99% para o verdadeiro rendimento médio;

b) Qual tamanho deve ter a amostra para que seja de 95% o grau de confiança na estimativa intervalar 3400 ± 100?

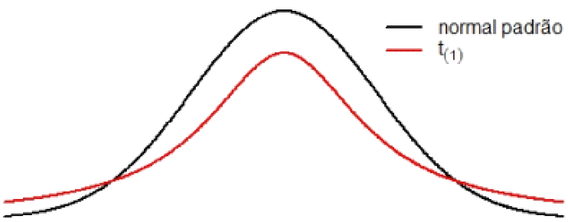
- **População normal e variância populacional desconhecida**

Nova estatística:

$$T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{(n-1)}$$

Distribuição t de Student

- Simétrica em relação ao zero;
- Semelhante à distribuição normal padrão, porém com "caudas mais grossas";
- Para $n \to \infty$ $(n \geq 30)$ a distribuição t tende para a normal padrão



$$IC(\mu)_{1-\alpha} = \left(\bar{X} - t_T\sqrt{\frac{S^2}{n}}; \bar{X} + t_T\sqrt{\frac{S^2}{n}}\right)$$

Utilização da tabela da distribuição t de Student

Exemplos:

(a) número de graus de liberdade $= 5$ e $\alpha = 0,02$. t_T?

(b) número de graus de liberdade $= 15$ e $\alpha = 0,10$. t_T?

(c) Para n^o de graus de liberdade $= 10$, determinar t_T tal que $P(-t_T < T < t_T) = 0,95$

(d) Para n^o de graus de liberdade $= 4$, determinar t_T tal que $P(-t_T < T < t_T) = 0,80$

(e) Para n^o de graus de liberdade $= 10$, determinar t_T tal que $P(T > t_T) = 0,05$

(f) Para n^o de graus de liberdade $= 4$, determinar t_T tal que $P(T < -t_T) = 0,20$

(g) Para n^o de graus de liberdade $= 24$, determinar t_T tal que $P(T < -t_T) = 0,01$

(h) Para n^o de graus de liberdade = 13, determinar t_T tal que $P(T > t_T) = 0,005$

(i) Para n^o de graus de liberdade = 11, determinar t_T tal que $P(T < t_T) = 0,80$

(j) Para n^o de graus de liberdade = 12, determinar t_T tal que $P(T > -t_T) = 0,90$

(k) Para 10 graus de liberdade, achar P(-3,169<T<3,169), P(T<3,169), P(T<-3,169)

(l) Para 5 graus de liberdade, achar P(-1,476<T<1,476), P(T<1,476), P(T<-1,476)

- **População normal e variância populacional desconhecida**

Exemplo:

Os resíduos industriais jogados nos rios, muitas vezes, absorvem o oxigênio necessário à respiração dos peixes e de outras formas de vida aquática. Uma lei estadual exige um valor médio não inferior a 5ppm de oxigênio dissolvido, cujo conteúdo seja suficiente para manter a vida aquática. Seis amostras de água retiradas de um rio revelaram os índices:

4,9	5,1	4,9	5,0	5,0	4,7

Construir o intervalo com 95% de confiança para a verdadeira média do oxigênio dissolvido, em ppm, e interpretar.

- **População normal e variância populacional desconhecida**

Exemplo:
Para avaliar o peso médio ao nascer de bezerros da raça Ibagé foi examinada uma amostra de 20 partos, obtendo os dados a seguir:

24,58	26,64	28,01	23,76	26,98	23,47	26,92	27,53	26,69	23,34
24,38	28,31	26,21	29,92	28,93	26,34	28,14	28,91	25,35	28,23

Supondo que a distribuição dos dados de peso ao nascer é aproximadamente normal,

(a) Determinar estimativas por ponto para a média e para a variância dos pesos para essa amostra;

(b) Construir um intervalo de 95% de confiança para μ;

(c) Calcule o tamanho de n da amostra necessária para que se obtenha um intervalo de confiança de 95% com precisão de 2% da média.

- **População não normal, grandes amostras ($n > 30$)**

Pelo Teorema Central do Limite, se n for razoavelmente grande ($n > 30$), então

$$\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim N(0, 1)$$

e o intervalo de $100(1 - \alpha)\%$ de confiança para a média μ da população é dada por:

$$IC(\mu)_{1-\alpha} = \left(\bar{X} - z_T\sqrt{\frac{s^2}{n}}; \bar{X} + z_T\sqrt{\frac{s^2}{n}}\right)$$

Exemplo: Para se avaliar a intensidade da infestação de uma área por uma espécie de lagarta, foram observadas 32 parcelas quanto ao número de lagartas, obtendo-se uma média de 3,3 lagartas por parcela e variância 3,2 (lagartas por parcela)^2. Construir os intervalos de 95% e 99% de confiança para o número médio de lagartas na área total.

(a) Calcular o tamanho n da amostra necessária para que se obtenha um intervalo de 95% de confiança com precisão $d = 0,4$ lagartas por parcela.

(b) Calcular o tamanho n da amostra necessária para que se obtenha um intervalo de 99% de confiança com precisão $d = 0,4$ lagartas por parcela.

Intervalo de confiança para proporção

$$IC(\pi)_{1-\alpha} = \left(\hat{\pi} - z_T\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}; \hat{\pi} + z_T\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right)$$

Exemplo: Coletou-se uma amostra de 35 peixes da espécie *Xenomelaniris brasiliensis*, na localidade da praia da Barra da Lagoa, Florianópolis, SC, a qual apresentou 45,7% de peixes com comprimento total acima de 50 mm. Encontre um intervalo com 95% de confiança, dentro do qual deve estar a verdadeira proporção de peixes dessa espécie com comprimento acima de 50 mm.
Qual o tamanho da amostra necessário para que tenhamos 95% de confiança de que o erro de nossa estimativa não seja superior a cinco pontos percentuais (0,05)?

Exemplo: Em um experimento, 320 de 400 sementes germinaram. Determine o intervalo de confiança de 99% para a verdadeira proporção de sementes que germinaram. Para realizar o teste de germinação, quantas sementes serão necessárias utilizar, se desejarmos um intervalo de confiança de 99%, com precisão de quatro pontos percentuais?