## Estimação pontual, estimação intervalar e tamanho de amostras

**Inferência:** por meio das amostras, conhecer informações gerais da população. Problemas de inferência, em geral, se dividem em estimação de parâmetros e testes de hipóteses.

Neste curso consideraremos problemas de inferência envolvendo estimação de parâmetros

**Estimação de parâmetros:** A estimação é o processo que consiste no uso de dados da amostra (dados amostrais) para estimar valores de parâmetros populacionais desconhecidos, tais como média, desvio padrão, proporções, etc.

**Estimador:** é uma função dos elementos da amostra, que será usada no processo de estimação do parâmetro desejado. **O estimador é, como vemos, uma estatística.** Será, portanto, uma variável aleatória caracterizada por uma distribuição de probabilidade e seus respectivos parâmetros próprios.

**Estimativa:** cada valor particular assumido por um estimador.

**ESTIMATIVA PONTUAL:** é uma estimativa de um único valor para um parâmetro populacional.

**ESTIMATIVA INTERVALAR:** é um intervalo de valores usado para estimar um parâmetro populacional.

Exemplo: Considere um amostra aleatória das alturas, em cm, de meninos de 15 anos: 165, 170, 170, 168, 178, 180, 168, 150, 172, 170

Perguntas típicas de estimação de parâmetros em relação a estes dados são:

A partir dos dados amostrais e considerando que a amostra foi retirada de uma população com média $\mu$ desconhecida:

(a) Estime um único valor representando o parâmetro desconhecido $\mu$. –> **estimação pontual**

(b) Estime um intervalo de possíveis valores para $\mu$.–> **estimação intervalar.**

**Estimação Pontual de um parâmetro**

$\theta$ : parâmetro (valor numérico constante e desconhecido da população)

$\hat{\theta}$ : estimador pontual ( estatística visando estimar o parâmetro)

Um estimador $\hat{\theta}$ é uma função dos valores amostrais que é usado para estimar o valor de um parâmetro desconhecido $\theta$. O estimador é uma variável aleatória com uma distribuição de probabilidade. Quando uma amostra aleatória é selecionada de uma população e $\hat{\theta}$ é calculado a partir dos dados, o valor numérico obtido é chamado uma estimativa de da amostra considerada.

Alguns estimadores pontuais:

| **Parâmetro da população ( $\theta$ )** | **Estimador ( $\hat{\theta}$ )** |
|---|---|
| Média ( $\mu$ ) | Média amostral ( $\bar{x}$ ) |
| Proporçao ( p ) | Proporção amostral ( $\bar{p}$ ) |
| Desvio-padrão ( $\sigma$ ) | Desvio-padrão amostral ( S ) |

Como escolher um estimador?

**(1) Deve ser não enviesado (não tendencioso)**

Um estimador $\hat{\theta}$ é não enviesado para o parâmetro $\theta$ se $E(\hat{\theta}) = \theta$, qualquer que seja o verdadeiro valor de $\theta$. Se esta propriedade não é satisfeita, $\hat{\theta}$ é um estimador enviesado.

Lembre-se: E (x) é o valor esperado de x.

## (2) Selecione o estimador que tem menor variância

Selecione o estimador não enviesado de $\theta$ que tem menor variância, qualquer que seja o valor de $\theta$. Se ele existir, ele é chamando de estimador não tendencioso de mínima variância.

O desvio - padrâo (DP) do estimador $\hat{\theta}$ é chamado erro padrão e será denotado : $\varepsilon(\hat{\theta}) = \text{DP}\ (\hat{\theta})$

## Ilustrando a idéia de enviesamento

Considere a variância de uma população de tamanho N:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}, \quad \text{onde } \mu \text{ é é a méa populacional.}$$

Vamos supor que temos uma amostra aleatória de tamanho n e queremos estimar, a partir da amostra, a variância da população.

Consideremos duas situações:

(a) A média populacional µnalonhecida.

$$\text{Vamos definir o estimador } \hat{\theta} = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

Vamos verificar se o estimador definido é ou não enviesado :

$$E(\hat{\theta}) = E\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}\right) = \frac{\sum_{i=1}^{n}E\left((x_i-\mu)^2\right)}{n} = \frac{\sum_{i=1}^{n}\sigma^2}{n} = \frac{n\sigma^2}{n} = \sigma^2$$

por tanto se conhecermos a média populacional

o estimador $\hat{\theta} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$ não enviesado.

(b) Considere que não conhecemos a média populacional.

Podemos estimar a média populacional pela média amostral
e definir, por analogia, o estimador da variância como :

$$\hat{\theta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$

e vamos testar se este estimador é ou não enviesado.

Temos: $E(\hat{\theta}) = E\left(\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}\right)$

e vamos usar a identidade: $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2$

Então: $E(\hat{\theta}) = \dfrac{\sum_{i=1}^{n} E((x_i - \mu)^2) - nE((\bar{x} - \mu)^2)}{n} = \dfrac{\sum_{i=1}^{n}\sigma^2 - n\dfrac{\sigma^2}{n}}{n} = \dfrac{(n-1)\sigma^2}{n} \neq \sigma^2$

e portanto o estimador da variância $\hat{\theta} = \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$ é enviesado.

Dai a definição vista no inicio do curso para a variância amostral:

$\hat{\theta} = S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ que é não enviesado.

**Note que:** não enviesamento não deve ser considerado um critério excessivamente importante para julgar um estimador na medida em que, o valor esperado é apenas uma das possíveis medidas de centro. Existem exemplos de estimadores enviesados que tem concentração de probabilidade próximo do parâmetro maior do que o estimador não enviesado de mínima variância. Entretanto, neste curso, os estimadores mais comumente utilizados são não enviesados.

Estimação pontual da média populacional ( $\mu$ )

parâmetro: $\mu$ (média populacional)

dados: $x_1, x_2, \ldots, x_n$ (amostra aleatória de tamanho n)

estimador: $\bar{x}$ (média amostral)

erro padrão da média: $\varepsilon(\bar{x}) = \dfrac{\sigma}{\sqrt{n}}$

amostragem sem reposição de população finita: $\varepsilon(\bar{x}) = \dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$

se não conhecer $\sigma$, estime o erro usando desvio amostral S

Estimação pontual da proporção populacional ( p ):

parâmetro: p (proporção populacional)

dados: x: número de ocorrências de uma certa característica
numa amostra aleatória de tamanho n

estimador: $\hat{p} = \dfrac{x}{n}$

erro padrão da proporção: $\varepsilon(\hat{p}) = \sqrt{\dfrac{p(1-p)}{n}}$

amostragem sem reposição de população finita: $\varepsilon(\hat{p}) = \sqrt{\dfrac{p(1-p)}{n}}\sqrt{\dfrac{N-n}{N-1}}$

se não conhecer p, estime o erro usando a proporção amostral $\hat{p}$

Note que o erro padrão da **proporção** depende da proporção populacional, ou de alguma estimativa dela como por exemplo a própria proporção amostral.

Entretanto, se não for possível obter alguma estimativa de p, e como para qualquer valor de p temos a desigualdade **p(1-p)≤1/4**, o erro padrão é limitado por

$$\varepsilon(\hat{p}) \leq \frac{1}{2\sqrt{n}}$$

**Estimação intervalar de um parâmetro populacional**

Um estimador intervalar de um parâmetro é um intervalo que é previsto para conter o parâmetro. A confiança que atribuimos ao intervalo é a probabilidade de que ele irá conter o parâmetro.

Considere o parâmetro $\theta$.

Seja $(1-\alpha)$ uma probabilidade especificada e L e U funções dos valores amostrais $X_1,\ldots,X_n$, de modo que

$$P(L<\theta<U)=1-\alpha$$

Então o intervalo $(L,U)$ é chamado intervalo de confiança de $100(1-\alpha)\%$ para o parâmetro $\theta$, e $(1-\alpha)$ chamado de nível de confiança associado ao intervalo.

O nível de confiança (**NC**) é a probabilidade $1-\alpha$, que é a proporção de vezes que o intervalo de confiança realmente contém o parâmetro populacional, supondo que o processo seja repetido um número grande de vezes.

Alguns níveis de confiança geralmente usados.

$$NC = 1-\alpha$$

| | |
|---|---|
| Para NC = 0,90 ou (90%) | $\alpha$ = 0,1 |
| Para NC = 0,95 ou (95%) | $\alpha$ = 0,05 |
| Para NC = 0,99 ou (99%) | $\alpha$ = 0,01 |

**Vamos considerar primeiro a estimativa intervalar para a proporção populacional**

**Exemplo de uma estimativa intervalar**:

A estimativa de **IC** de 95% para a proporção populacional $p$ é

$$0,381 < p < 0,497.$$

Interpretação: Há uma interpretação correta e muitas erradas, diferentes e criativas para o **IC**.

Correta: **Estamos 95% confiantes de que o intervalo de 0,381 a 0,497 realmente contém o verdadeiro valor de *p*.**

**Errada**: Há uma chance de 95% de que o verdadeiro valor de $p$ esteja entre 0,381 e 0,497.

**Para estabelecer um IC são utilizados valores críticos.**

O uso do escore padrão $Z$.

O escore padrão $Z$ pode ser usado para de distinguir entre estatísticas amostrais que têm chance de ocorrer e aquelas que não têm.

Sob certas condições a distribuição amostral das **proporções** amostrais pode ser aproximada por uma distribuição normal.

Há uma probabilidade $1-\alpha$ de que uma proporção amostral caia numa região limitada por dois valores de $Z$, para os quais uma área de $\alpha$ /2 seja definida na cauda esquerda e uma área de $\alpha$ /2 seja definida na cauda direita.

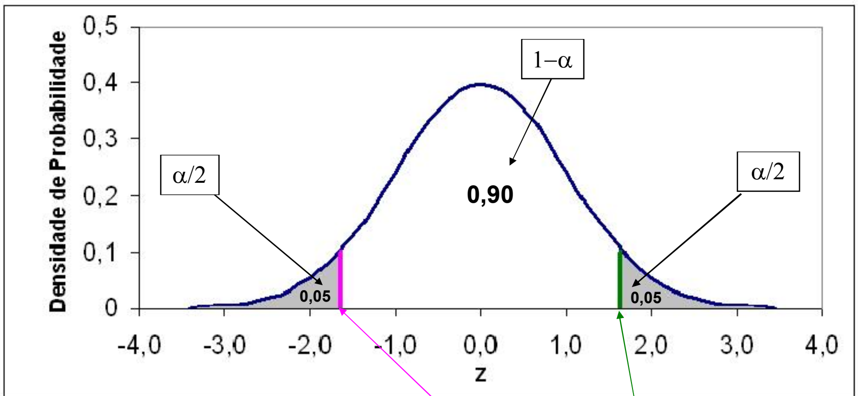
**Valores críticos.**

Um valor crítico é um número na fronteira que separa estatísticas amostrais que têm uma chance de ocorrer daquelas que não têm.

O número $Z_{\alpha/2}$ é um valor crítico, que é um escore $Z$ com a propriedade de separar uma área de $\alpha$ /2 na cauda direita da distribuição normal padronizada.

Vamos considerar um exemplo onde temos um intervalo de confiança de 90%, ou seja, $(1-\alpha)=0{,}90$

# Valores Críticos




-

+

| **Nível de confiança**: NC = $1-\alpha$ |
| --- |
| Por exemplo, NC = 90% |

Valores críticos: $\pm Z_{\alpha/2} = \pm 1{,}6449$

$\alpha = 1 - NC = 0{,}1 \longrightarrow \alpha/2 = 0{,}05 \longrightarrow Z_{\alpha/2} = 1{,}6449$

**Margem de erro** *E*

Quando os dados são usados para estimar um parâmetro populacional $\theta$, a margem de erro (ou erro máximo da estimativa), representada por *E*, pode ser encontrada pela multiplicação do valor crítico pelo erro padrão do estimador (que é o desvio padrão) conforme a expressão:

$$E = Z_{\alpha/2}\, \varepsilon(\hat{\theta})$$

**Margem de erro** *E* **para a proporção**

Quando os dados são usados para estimar uma proporção populacional *p*, a margem de erro (ou erro máximo da estimativa), representada por *E*, pode ser encontrada pela multiplicação do valor crítico pelo desvio padrão das proporções amostrais conforme a expressão:

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

(margem de erro para proporções)

**Note que:** na fórmula original para E, é o valor da proporção populacional p que aparece. Entretanto, não conhecemos p, e usamos uma estimativa conhecida de p, que funciona bem para n grande . E se lembramos que p(1-p)≤1/4, o limite um superior para E é dado por:

$$E \le \frac{Z_{\alpha/2}}{2\sqrt{n}}$$

**Intervalo de Confiança (ou Estimativa Intervalar) para a Proporção Populacional *p***

$$\hat{p} - E < p < \hat{p} + E$$

**ou**

$$\hat{p} \pm E$$

**ou**

$$\left(\hat{p} - E, \hat{p} + E\right)$$

**Procedimento para construção de um IC para a proporção**

(1) Verifique os requisitos (amostra aleatória, tamanho grande,...)

(2) Estabelecido o NC, recorra à Tabela da Normal e encontre $Z_{\alpha/2}$

(3) Calcule a margem de erro:

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

(4) Usando *E* e $\hat{p}$, ache os valores de $\hat{p} - E$ e $\hat{p} + E$.

(5) Substitua estes valores em um formato específico do intervalo de confiança:

$$\hat{p} - E < p < \hat{p} + E$$

ou $\hat{p} \pm E$

ou $\left(\hat{p} - E, \hat{p} + E\right)$

**Determinação do Tamanho Amostral**

Suponha que queremos coletar dados amostrais com o objetivo de estimar alguma proporção populacional.

Quantos itens amostrais ($n$) devem ser obtidos?

Com base na expressão do erro:

-Quando uma estimativa $\hat{p}$ é conhecida:

$$n = \frac{(Z_{\alpha/2})^2 \hat{p}\hat{q}}{E^2}$$

$$\hat{q} = 1 - \hat{p}$$

-Quando não se conhece $\hat{p}$ :

$$n = \frac{(Z_{\alpha/2})^2}{4E^2}$$

**Arredonde o resultado para o próximo inteiro.**

**Amostragem sem reposição de População Finita de tamanho N:**

**Erro:**

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\sqrt{\frac{N-n}{N-1}}$$

**Definição do Tamanho Amostral**

$$n = \frac{N\,\hat{p}\,\hat{q}[z_{\alpha/2}]^2}{\hat{p}\,\hat{q}[z_{\alpha/2}]^2 + (N-1)\,E^2}$$

**Arredonde o resultado para o próximo inteiro.**

Intervalo de confiança para a média populacional ($\mu$)

**( $\sigma$ é conhecido )**

**Requisitos:**

- **O desvio padrão da população é conhecido.**
- A amostra é aleatória simples (todas as amostras de mesmo tamanho tem igual chance de serem selecionadas).
- Uma ou ambas as condições seguintes são satisfeitas: A população é normalmente distribuída ou $n \geq 30$.

**Intervalo de confiança para Média Populacional: σ conhecido**

Vimos que a média amostral $\overline{x}$ é a melhor estimativa pontual da média populacional $\mu$.

**Intervalos de confiança:**

**Estimativa de intervalo de confiança para a média populacional $\mu$ (com $\sigma$ conhecido):**

$$\overline{x} - E < \mu < \overline{x} + E \quad \text{ou} \quad \left(\overline{x} - E, \overline{x} + E\right) \quad \text{ou} \quad \overline{x} \pm E$$

**Margem de erro:**

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

**Tamanho amostral:**

$$n = \left[ Z_{\alpha/2} \frac{\sigma}{E} \right]^2$$

Arredondar para o inteiro maior mais próximo

**Amostragem sem reposição dePopulação Finita de tamanho N:**

**Margem de Erro:** $E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

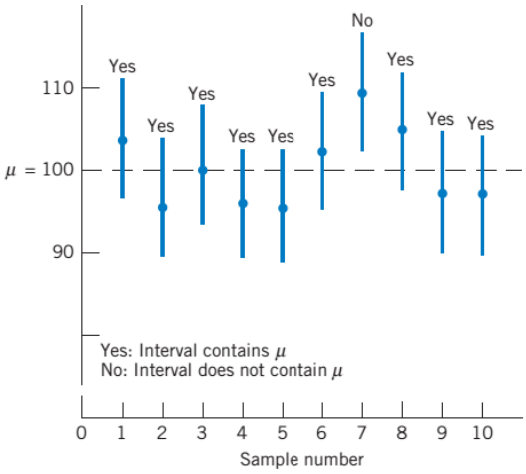
**Definição do Tamanho Amostral**

$$n = \frac{N\sigma^2 (z_{\alpha/2})^2}{(N-1)E^2 + \sigma^2 (z_{\alpha/2})^2}$$

**Arredonde o resultado para o inteiro maior mais próximo.**

Intervalos de de 95% de confiança para 10 amostras de tamanho n = 7 retiradas de uma população com $\mu = 100$ e $\sigma = 10$.

neste caso : $\left(\bar{x} - 1.96 \times \frac{10}{\sqrt{7}}, \bar{x} - 1.96 \times \frac{10}{\sqrt{7}}\right)$ para cada amostra



**Intervalo de confiança da Média Populacional: σ Desconhecido**

Se conhecemos o desvio - padrão populacional $\sigma$, vimos que

a variável padronizada $Z = \frac{\bar{x} - \mu}{\left(\sigma / \sqrt{n}\right)}$, do estimador pontual

da média $\bar{x}$ tem distribuição normal padronizada.

Em geral, $\sigma$ não é conhecido, e é natural usar como estimativa para $\sigma$,
o desvio - padrão amostral S. Neste caso introduzimos variável

$$t = \frac{\bar{x} - \mu}{\left(\frac{S}{\sqrt{n}}\right)}$$

William Sealy Gosset (Conhecido como : Student) mostrou que :

Se a amostra é obtida de uma população com distribuição normal, então
a variável t é distribuida segundo a conhecida "distribuição de Student".

# Distribuição t-Student

**William Sealy Gosset**

Conhecido como: **Student**

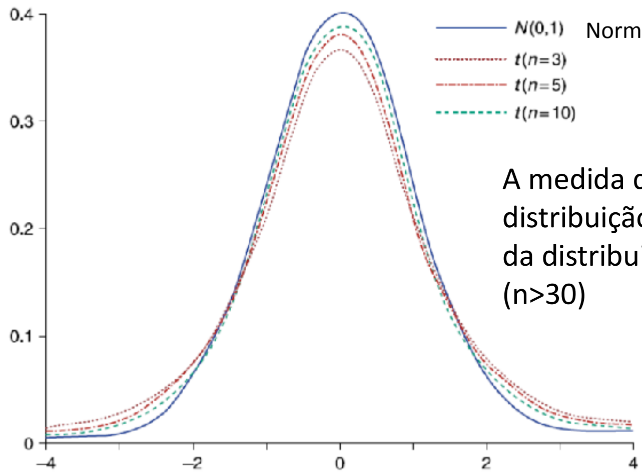
Quimico e Matemático (New College, Oxford). Foi trabalhar para a destilaria de Arthur Guinness & Son.

Gosset iria aplicar os seus conhecimentos de estatística na cervejaria - para a seleção das melhores amostras de cevada. Gosset adquiriu o seu conhecimento no Laboratório Biométrico Karl Pearson. Pearson dava pouca importância aos resultados obtidos por Gosset, pois eram baseados em pequenas amostras na cervejeira, ele (Pearson) por norma tinha centenas de observações e não via urgência em desenvolver um método que tratasse com pequenas amostras

Para prevenir fugas de informação e futuras revelações dos "segredos" da marca, a Guinness proibiu que os seus empregados pudessem publicar quaisquer trabalhos independente do conteúdo. Gosset não tinha como publicar os trabalhos com o seu nome. Então, **usou o pseudonimo Student** para as suas publicações evitando ser detectado pela entidade empregadora. Seu feito mais conhecido, é hoje conhecido com a **Distribuição t-Student**, que noutras circunstâncias seria conhecida como a Distribuição t-Gosset.

**Distribuição-t de Student com gl = (n-1) graus de liberdade**

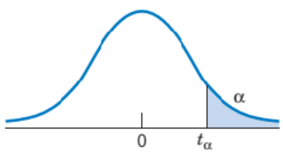
Distribuição simétrica em torno de zero
Para cada valor de gl existe uma distribuição-t




A medida que n aumenta a distribuição de t se aproxima da distribuição de Z (normal) (n>30)

| d.f \ α | .25 | .10 | .05 | .025 | .01 | .00833 | .00625 | .005 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.000 | 3.078 | 6.314 | 12.706 | 31.821 | 38.204 | 50.923 | 63.657 |
| 2 | .816 | 1.886 | 2.920 | 4.303 | 6.965 | 7.649 | 8.860 | 9.925 |
| 3 | .765 | 1.638 | 2.353 | 3.182 | 4.541 | 4.857 | 5.392 | 5.841 |
| 4 | .741 | 1.533 | 2.132 | 2.776 | 3.747 | 3.961 | 4.315 | 4.604 |
| 5 | .727 | 1.476 | 2.015 | 2.571 | 3.365 | 3.534 | 3.810 | 4.032 |
| 6 | .718 | 1.440 | 1.943 | 2.447 | 3.143 | 3.287 | 3.521 | 3.707 |
| 7 | .711 | 1.415 | 1.895 | 2.365 | 2.998 | 3.128 | 3.335 | 3.499 |
| 8 | .706 | 1.397 | 1.860 | 2.306 | 2.896 | 3.016 | 3.206 | 3.355 |
| 9 | .703 | 1.383 | 1.833 | 2.262 | 2.821 | 2.933 | 3.111 | 3.250 |
| 10 | .700 | 1.372 | 1.812 | 2.228 | 2.764 | 2.870 | 3.038 | 3.169 |
| 11 | .697 | 1.363 | 1.796 | 2.201 | 2.718 | 2.820 | 2.981 | 3.106 |
| 12 | .695 | 1.356 | 1.782 | 2.179 | 2.681 | 2.779 | 2.934 | 3.055 |
| 13 | .694 | 1.350 | 1.771 | 2.160 | 2.650 | 2.746 | 2.896 | 3.012 |
| 14 | .692 | 1.345 | 1.761 | 2.145 | 2.624 | 2.718 | 2.864 | 2.977 |
| 15 | .691 | 1.341 | 1.753 | 2.131 | 2.602 | 2.694 | 2.837 | 2.947 |
| 16 | .690 | 1.337 | 1.746 | 2.120 | 2.583 | 2.673 | 2.813 | 2.921 |
| 17 | .689 | 1.333 | 1.740 | 2.110 | 2.567 | 2.655 | 2.793 | 2.898 |
| 18 | .688 | 1.330 | 1.734 | 2.101 | 2.552 | 2.639 | 2.775 | 2.878 |
| 19 | .688 | 1.328 | 1.729 | 2.093 | 2.539 | 2.625 | 2.759 | 2.861 |
| 20 | .687 | 1.325 | 1.725 | 2.086 | 2.528 | 2.613 | 2.744 | 2.845 |
| 21 | .686 | 1.323 | 1.721 | 2.080 | 2.518 | 2.601 | 2.732 | 2.831 |
| 22 | .686 | 1.321 | 1.717 | 2.074 | 2.508 | 2.591 | 2.720 | 2.819 |
| 23 | .685 | 1.319 | 1.714 | 2.069 | 2.500 | 2.582 | 2.710 | 2.807 |
| 24 | .685 | 1.318 | 1.711 | 2.064 | 2.492 | 2.574 | 2.700 | 2.797 |
| 25 | .684 | 1.316 | 1.708 | 2.060 | 2.485 | 2.566 | 2.692 | 2.787 |
| 26 | .684 | 1.315 | 1.706 | 2.056 | 2.479 | 2.559 | 2.684 | 2.779 |
| 27 | .684 | 1.314 | 1.703 | 2.052 | 2.473 | 2.552 | 2.676 | 2.771 |
| 28 | .683 | 1.313 | 1.701 | 2.048 | 2.467 | 2.546 | 2.669 | 2.763 |
| 29 | .683 | 1.311 | 1.699 | 2.045 | 2.462 | 2.541 | 2.663 | 2.756 |
| 30 | .683 | 1.310 | 1.697 | 2.042 | 2.457 | 2.536 | 2.657 | 2.750 |
| 40 | .681 | 1.303 | 1.684 | 2.021 | 2.423 | 2.499 | 2.616 | 2.704 |
| 60 | .679 | 1.296 | 1.671 | 2.000 | 2.390 | 2.463 | 2.575 | 2.660 |
| 120 | .677 | 1.289 | 1.658 | 1.980 | 2.358 | 2.428 | 2.536 | 2.617 |
| $\infty$ | .674 | 1.282 | 1.645 | 1.960 | 2.326 | 2.394 | 2.498 | 2.576 |

$\alpha$ é a probabilidade (área) na cauda direita



**Estimativa intervalar da Média Populacional: σ Desconhecido**

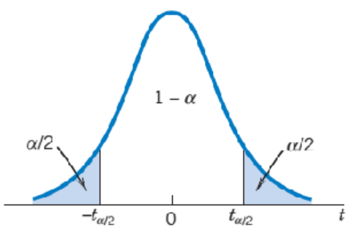
**Para população com distribuição normal e com σ desconhecido, um intervalo de confiança com (1-α).100% para a média populacional μ é**

$$\bar{x} - E < \mu < \bar{x} + E \quad \text{ou} \quad (\bar{x} - E, \bar{x} + E) \quad \text{ou} \quad \bar{x} \pm E$$

**a margem de erro:**

$$E = t_{\alpha/2} \frac{S}{\sqrt{n}}$$

**Valor de t: obtido da distribuição-*t* de student**




$$1 - \alpha = P\left[-t_{\alpha/2} < \frac{X - \mu}{S/\sqrt{n}} < t_{\alpha/2}\right] = P\left[-t_{\alpha/2}\frac{S}{\sqrt{n}} < X - \mu < t_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

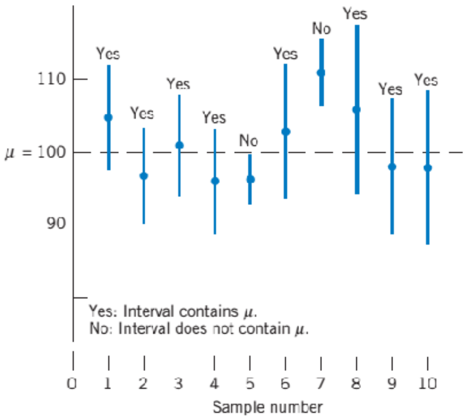
$$P\left[\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

interpretação : (1 - α) é a probabilidade de que o intervalo $(\bar{x} - E, \bar{x} + E)$ contenha a média

Intervalos de de 95% de confiança para 10 amostras de tamanho $n = 7$ retiradas de uma população com distribuição normal com $\mu = 100$ e $\sigma = 10$. No caso, $\alpha = 0.05$ e $t_{0.025} = 2.447$ (dist. t com gl $= n - 1 = 6$).

Neste caso: $\left(\bar{x} - 2.447 \times \frac{s}{\sqrt{7}}, \bar{x} + 2.447 \times \frac{s}{\sqrt{7}}\right)$ para cada amostra,

onde s é o desvio padrão de cada amostra

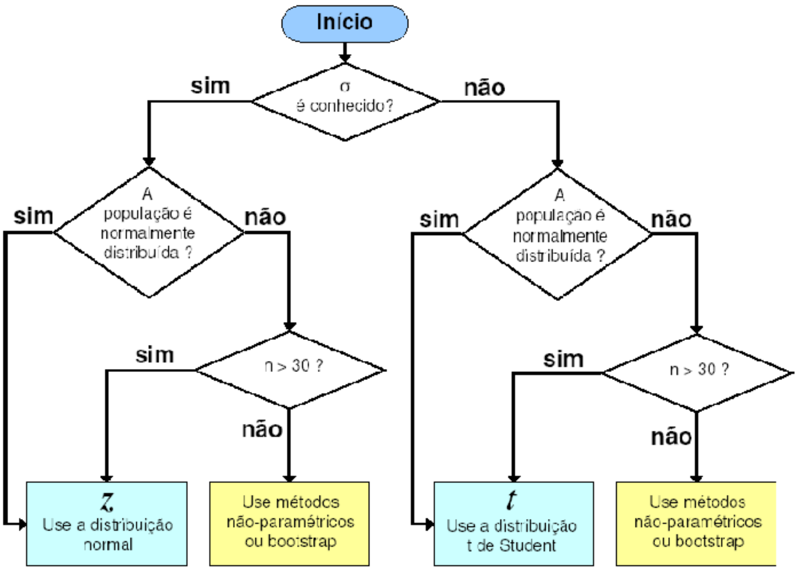



**A Estimativa intervalar é a mais utilizada.**

Note que:

Numa estimativa intervalar, o centro do intervalo de confiança não é visto como o melhor estimador pontual, o intervalo é a estimativa. Ele é obtido de modo que temos uma % de confiança dele conter o parâmetro e nenhuma preferência é atribuída a algum de seus pontos. Devido a simetria da distribuição aproximada de $\bar{x}$, ocorre que o centro do intervalo de confiança para $\mu$ é $\bar{x}$. Entretanto, se a distribuição amostral não é simétrica, o ponto médio do intervalo não necessariamente coincide com a melhor estimativa pontual do parâmetro.

## Estimação da Média Populacional: Resumo

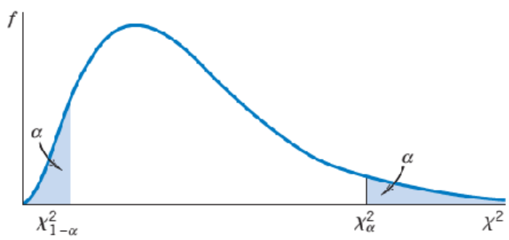



Distribuição $\chi^2$

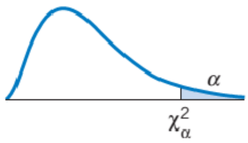
Seja $x_1, x_2, \ldots, x_n$ uma amostra aleatória de uma população normal $N(\mu(\sigma))$. A distribuição de

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

é chamada distribuição $\chi^2$ com (n-1) graus de liberdade.



| d.f. \ α | .99 | .975 | .95 | .90 | .50 | .10 | .05 | .025 | .01 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | .0002 | .001 | .004 | .02 | .45 | 2.71 | 3.84 | 5.02 | 6.63 |
| 2 | .02 | .05 | .10 | .21 | 1.39 | 4.61 | 5.99 | 7.38 | 9.21 |
| 3 | .11 | .22 | .35 | .58 | 2.37 | 6.25 | 7.81 | 9.35 | 11.34 |
| 4 | .30 | .48 | .71 | 1.06 | 3.36 | 7.78 | 9.49 | 11.14 | 13.28 |
| 5 | .55 | .83 | 1.15 | 1.61 | 4.35 | 9.24 | 11.07 | 12.83 | 15.09 |
| 6 | .87 | 1.24 | 1.64 | 2.20 | 5.35 | 10.64 | 12.59 | 14.45 | 16.81 |
| 7 | 1.24 | 1.69 | 2.17 | 2.83 | 6.35 | 12.02 | 14.07 | 16.01 | 18.48 |
| 8 | 1.65 | 2.18 | 2.73 | 3.49 | 7.34 | 13.36 | 15.51 | 17.53 | 20.09 |
| 9 | 2.09 | 2.70 | 3.33 | 4.17 | 8.34 | 14.68 | 16.92 | 19.02 | 21.67 |
| 10 | 2.56 | 3.24 | 3.94 | 4.87 | 9.34 | 15.99 | 18.31 | 20.48 | 23.21 |
| 11 | 3.05 | 3.81 | 4.57 | 5.58 | 10.34 | 17.28 | 19.68 | 21.92 | 24.72 |
| 12 | 3.57 | 4.40 | 5.23 | 6.30 | 11.34 | 18.55 | 21.03 | 23.34 | 26.22 |
| 13 | 4.11 | 5.01 | 5.89 | 7.04 | 12.34 | 19.81 | 22.36 | 24.74 | 27.69 |
| 14 | 4.66 | 5.62 | 6.57 | 7.79 | 13.34 | 21.06 | 23.68 | 26.12 | 29.14 |
| 15 | 5.23 | 6.26 | 7.26 | 8.55 | 14.34 | 22.31 | 25.00 | 27.49 | 30.58 |
| 16 | 5.81 | 6.90 | 7.96 | 9.31 | 15.34 | 23.54 | 26.30 | 28.85 | 32.00 |
| 17 | 6.41 | 7.56 | 8.67 | 10.09 | 16.34 | 24.77 | 27.59 | 30.19 | 33.41 |
| 18 | 7.01 | 8.23 | 9.39 | 10.86 | 17.34 | 25.99 | 28.87 | 31.53 | 34.81 |
| 19 | 7.63 | 8.90 | 10.12 | 11.65 | 18.34 | 27.20 | 30.14 | 32.85 | 36.19 |
| 20 | 8.26 | 9.59 | 10.85 | 12.44 | 19.34 | 28.41 | 31.41 | 34.17 | 37.57 |
| 21 | 8.90 | 10.28 | 11.59 | 13.24 | 20.34 | 29.62 | 32.67 | 35.48 | 38.93 |
| 22 | 9.54 | 10.98 | 12.34 | 14.04 | 21.34 | 30.81 | 33.92 | 36.78 | 40.29 |
| 23 | 10.20 | 11.69 | 13.09 | 14.85 | 22.34 | 32.01 | 35.17 | 38.08 | 41.64 |
| 24 | 10.86 | 12.40 | 13.85 | 15.66 | 23.34 | 33.20 | 36.42 | 39.36 | 42.98 |
| 25 | 11.52 | 13.11 | 14.61 | 16.47 | 24.34 | 34.38 | 37.65 | 40.65 | 44.31 |
| 26 | 12.20 | 13.84 | 15.38 | 17.29 | 25.34 | 35.56 | 38.89 | 41.92 | 45.64 |
| 27 | 12.88 | 14.57 | 16.15 | 18.11 | 26.34 | 36.74 | 40.11 | 43.19 | 46.96 |
| 28 | 13.56 | 15.30 | 16.93 | 18.94 | 27.34 | 37.92 | 41.34 | 44.46 | 48.28 |
| 29 | 14.26 | 16.04 | 17.71 | 19.77 | 28.34 | 39.09 | 42.56 | 45.72 | 49.59 |
| 30 | 14.95 | 16.78 | 18.49 | 20.60 | 29.34 | 40.26 | 43.77 | 46.98 | 50.89 |
| 40 | 22.16 | 24.42 | 26.51 | 29.05 | 39.34 | 51.81 | 55.76 | 59.34 | 63.69 |
| 50 | 29.71 | 32.35 | 34.76 | 37.69 | 49.33 | 63.17 | 67.50 | 71.42 | 76.15 |
| 60 | 37.48 | 40.47 | 43.19 | 46.46 | 59.33 | 74.40 | 79.08 | 83.30 | 88.38 |
| 70 | 45.44 | 48.75 | 51.74 | 55.33 | 69.33 | 85.53 | 90.53 | 95.02 | 100.43 |
| 80 | 53.54 | 57.15 | 60.39 | 64.28 | 79.33 | 96.58 | 101.88 | 106.63 | 112.33 |
| 90 | 61.75 | 65.64 | 69.13 | 73.29 | 89.33 | 107.57 | 113.15 | 118.14 | 124.12 |
| 100 | 70.06 | 74.22 | 77.93 | 82.36 | 99.33 | 118.50 | 124.34 | 129.56 | 135.81 |




Intervalo de (1-α)% de confiança para σ

$$\left( s\sqrt{\frac{(n-1)}{\chi^2_{\alpha/2}}},\ s\sqrt{\frac{(n-1)}{\chi^2_{(1-\alpha/2)}}} \right)$$

Estimativas de σ usando a distribuição $\chi^2$ são muito afetadas (não confiáveis) para amostras de populações não normais.

- A Stewart Aviation Products Company vem fabricando altímetros de aviões com erros distribuídos normalmente com média 0 (obtida por calibragem) e desvio padrão de 43,7 pés. Após a instalação da nova produção de equipamentos selecionaram-se aleatoriamente 30 altímetros da nova linha de produção. Esta amostra acusou erros com desvio padrão de 54,7 pés. Baseado na amostra, vamos fazer uma estimativa intervalar com 95% de confiança para o desvio-padrao dos erros dos altímetros

n=30 amostras retiradas de população normal
com desvio- padrão s= 54,7 pés.

95% de confiãnça significa $\alpha/2=0.025$

Temos que encontrar na distribuição $\chi^2$ com gl=n-1=29
Os valores correspondentes a

$$\chi^2_{\alpha/2} = \chi^2_{0.025} = 45.72$$

$$\chi^2_{(1-\alpha/2)} = \chi^2_{0.975} = 16.4$$

| d.f. \ α | .99 | .975 | .95 | .90 | .50 | .10 | .05 | .025 | .01 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | .0002 | .001 | .004 | .02 | .45 | 2.71 | 3.84 | 5.02 | 6.63 |
| 2 | .02 | .05 | .10 | .21 | 1.39 | 4.61 | 5.99 | 7.38 | 9.21 |
| 3 | .11 | .22 | .35 | .58 | 2.37 | 6.25 | 7.81 | 9.35 | 11.34 |
| 4 | .30 | .48 | .71 | 1.06 | 3.36 | 7.78 | 9.49 | 11.14 | 13.28 |
| 5 | .55 | .83 | 1.15 | 1.61 | 4.35 | 9.24 | 11.07 | 12.83 | 15.09 |
| 6 | .87 | 1.24 | 1.64 | 2.20 | 5.35 | 10.64 | 12.59 | 14.45 | 16.81 |
| 7 | 1.24 | 1.69 | 2.17 | 2.83 | 6.35 | 12.02 | 14.07 | 16.01 | 18.48 |
| 8 | 1.65 | 2.18 | 2.73 | 3.49 | 7.34 | 13.36 | 15.51 | 17.53 | 20.09 |
| 9 | 2.09 | 2.70 | 3.33 | 4.17 | 8.34 | 14.68 | 16.92 | 19.02 | 21.67 |
| 10 | 2.56 | 3.24 | 3.94 | 4.87 | 9.34 | 15.99 | 18.31 | 20.48 | 23.21 |
| 11 | 3.05 | 3.81 | 4.57 | 5.58 | 10.34 | 17.28 | 19.68 | 21.92 | 24.72 |
| 12 | 3.57 | 4.40 | 5.23 | 6.30 | 11.34 | 18.55 | 21.03 | 23.34 | 26.22 |
| 13 | 4.11 | 5.01 | 5.89 | 7.04 | 12.34 | 19.81 | 22.36 | 24.74 | 27.69 |
| 14 | 4.66 | 5.62 | 6.57 | 7.79 | 13.34 | 21.06 | 23.68 | 26.12 | 29.14 |
| 15 | 5.23 | 6.26 | 7.26 | 8.55 | 14.34 | 22.31 | 25.00 | 27.49 | 30.58 |
| 16 | 5.81 | 6.90 | 7.96 | 9.31 | 15.34 | 23.54 | 26.30 | 28.85 | 32.00 |
| 17 | 6.41 | 7.56 | 8.67 | 10.09 | 16.34 | 24.77 | 27.59 | 30.19 | 33.41 |
| 18 | 7.01 | 8.23 | 9.39 | 10.86 | 17.34 | 25.99 | 28.87 | 31.53 | 34.81 |
| 19 | 7.63 | 8.90 | 10.12 | 11.65 | 18.34 | 27.20 | 30.14 | 32.85 | 36.19 |
| 20 | 8.26 | 9.59 | 10.85 | 12.44 | 19.34 | 28.41 | 31.41 | 34.17 | 37.57 |
| 21 | 8.90 | 10.28 | 11.59 | 13.24 | 20.34 | 29.62 | 32.67 | 35.48 | 38.93 |
| 22 | 9.54 | 10.98 | 12.34 | 14.04 | 21.34 | 30.81 | 33.92 | 36.78 | 40.29 |
| 23 | 10.20 | 11.69 | 13.09 | 14.85 | 22.34 | 32.01 | 35.17 | 38.08 | 41.64 |
| 24 | 10.86 | 12.40 | 13.85 | 15.66 | 23.34 | 33.20 | 36.42 | 39.36 | 42.98 |
| 25 | 11.52 | 13.11 | 14.61 | 16.47 | 24.34 | 34.38 | 37.65 | 40.65 | 44.31 |
| 26 | 12.20 | 13.84 | 15.38 | 17.29 | 25.34 | 35.56 | 38.89 | 41.92 | 45.64 |
| 27 | 12.88 | 14.57 | 16.15 | 18.11 | 26.34 | 36.74 | 40.11 | 43.19 | 46.96 |
| 28 | 13.56 | 15.30 | 16.93 | 18.94 | 27.34 | 37.92 | 41.34 | 44.46 | 48.28 |
| 29 | 14.26 | 16.04 | 17.71 | 19.77 | 28.34 | 39.09 | 42.56 | 45.72 | 49.59 |

O intervalo de 95% de confiança para o desvio

padrão será : $\left( s\sqrt{\dfrac{(n-1)}{\chi^2_{\alpha/2}}},\ s\sqrt{\dfrac{(n-1)}{\chi^2_{(1-\alpha/2)}}} \right) = \left( 54.7\sqrt{\dfrac{29}{45.72}},\ 54.7\sqrt{\dfrac{29}{16.4}} \right)$

ou seja : $(\ 43.6,\ 72.7)$