

Intervalos *de Confiança*

Prof. Adriano Mendonça Souza, Dr.

Departamento de Estatística

- PPGEMQ / PPGEF - UFSM -

Estimação de Parâmetros

O objetivo da Estatística é a realização de inferências acerca de uma população, baseadas nas informações amostrais. Como as populações são caracterizadas por medidas numéricas descritivas, denominadas parâmetros, a inferência estatística diz respeito à realização de inferências sobre esses parâmetros populacionais.

Os métodos de realizar inferências a respeito dos parâmetros pertencem a duas categorias.

- ◆ Pode-se tomar decisões relativas ao valor do parâmetro, através de um teste de hipótese;
- ◆ Pode-se estimar ou prever o valor do parâmetro.

A estimação é o processo que consiste em utilizar dados amostrais para estimar os valores de parâmetros populacionais desconhecidos.

Qualquer característica de uma população pode ser estimada a partir de uma amostra aleatória.

Entre as mais comuns, estão a média e o desvio padrão de uma população e a proporção populacional.

Estimação Pontual

As estatísticas amostrais são utilizadas como estimadores de parâmetros populacionais.

Assim uma média amostral é usada como estimativa de uma média populacional.

Tais estimativas chamam-se estimativas pontuais, porque originam uma única estimativa do parâmetro.

Média e variância de uma população normal $N(\mu; \sigma)$

O melhor estimador da média populacional μ é a média amostral ($\bar{X}$) :

$$\bar{X} = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$

sendo X_i variáveis aleatórias independentes tem-se:

$$E(X_i) = \mu, \quad i = 1, 2, \dots, n.$$

$$Var(X_i) = \sigma^2$$

Assim,

$$\begin{aligned} E(\bar{X}) &= \frac{1}{n} E(X_1 + X_2 + \dots + X_n) = \frac{1}{n} [E(X_1) + E(X_2) + \dots + E(X_n)] \\ &= \frac{1}{n} [\mu + \mu + \dots + \mu] = \frac{n\mu}{n} = \mu \end{aligned}$$

$$\begin{aligned} Var(\bar{X}) &= \frac{1}{n^2} Var(X_1 + X_2 + \dots + X_n) = \frac{1}{n^2} [Var(X_1) + Var(X_2) + \dots + Var(X_n)] \\ &= \frac{1}{n^2} [\sigma^2 + \sigma^2 + \dots + \sigma^2] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

$E(\bar{X}) = \mu$ revela que $\bar{X}$ é um estimador não tendencioso de μ . $Var(\bar{X}) = \sigma^2/n$ revela que quanto maior o valor de n , menor a sua variabilidade. Assim, neste caso, sendo $\theta = \mu$, adotaremos $\hat{\theta} = \bar{X}$

A variância populacional (σ^2) o estimador S^2 , com $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
 $\hat{\theta} = S^2$, é não tendencioso, conforme já foi exemplificado.

Adotaremos assim, para $\theta = \sigma^2$.

Exemplo:

- ① Se de uma população normal extraímos amostras cujos valores são: 1.1; 0.9; 0.3; -0.2; -3.1; 1.5; -2.7; 0.5; -1.5; 2.1, obtenha estimativas pontuais da μ ; σ^2 e $P(X > 2.5)$.

Estimativa de μ :

$$\bar{X} = \frac{1}{10} [1.1 + 0.9 + 0.3 + \dots + 2.1] = -0.11$$

Estimativa de σ^2 :

$$S^2 = \frac{1}{9} [(1.1 + 0.11)^2 + (0.9 + 0.11)^2 + \dots + (2.1 + 0.11)^2] = 3.17$$

Estimativa de $P(X > 2.5)$:

Sendo $\sigma^2 \simeq 3.17$ $\sigma = 1.78$

Assim para $X = 2.5$ $z = \frac{X - \mu}{\sigma} = \frac{2.5 - (-0.11)}{1.78} = 1.466$

e

$$P(X > 2.5) = P(Z > 1.466) = 1 - P(Z < 1.466) = 1 - 0.9286 = 7.14\%$$

Média e variância de uma proporção (p)

Consideremos agora o caso em que $\theta = p$, de uma população que apresenta certa característica. Extraí-se da população uma amostra de tamanho n . X será o número de elementos da amostra que apresentam a característica em estudo.

É intuitivo que um estimador da proporção p seja a proporção amostral $\hat{p}$:

$$\hat{p} = \frac{X}{n}$$

As observações dos n elementos podem ser considerados como n provas de Bernoulli com probabilidade de sucesso p , ou seja, X tem distribuição binomial com média np e variância npq , temos:

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}(np) = p$$

$$Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}npq = \frac{pq}{n}$$

Assim $\hat{p}$ é não tendencioso.

O desvio padrão de $\hat{p}$ é também denominado de *Erro-Padrão* de $\hat{p}$,
representado por $EP(\hat{p}) = \sqrt{\frac{pq}{n}}$ sendo $p \cong \hat{p}$ e $q \cong \hat{q} = 1 - p$.

❗ Para se avaliar a taxa de desemprego em determinado Estado, escolhe-se uma amostra aleatória de 1000 habitantes em idade de trabalho e contam-se os desempregados: 87. Estimar a proporção de desempregados em todo o Estado (população). Avaliar o erro padrão de estimativa.

$$\hat{p} = \frac{X}{n} = \frac{87}{1000} = 0.087 = 8.7\%$$

$$p \cong \hat{p} = 8.7\%$$

$$\hat{q} = 1 - \hat{p} = 0.913$$

$$EP(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(0.087)(0.913)}{1000}} = 0.009$$

$$\sigma \simeq 0.009$$

Estimação Intervalar

Sabemos que a estimação por ponto é em geral insuficiente, pois a probabilidade, de que a estimativa adotada venha a coincidir com o verdadeiro valor do parâmetro é praticamente nula.

Isso decorre dos estimadores serem muitas vezes **VA** contínuas, logo as estimativas serão diferentes do valor do parâmetro, então temos um erro de estimação.

Em virtude da variabilidade amostral, é usual incluir uma estimativa intervalar, com certo **nível de confiança $(1-\alpha)$** ou de **significância α** , para acompanhar a estimativa pontual.

Essa nova estimativa proporciona um intervalo, de possíveis valores do parâmetro populacional.

Constroe-se um **intervalo em torno da estimativa por ponto**, de modo que este intervalo contenha o verdadeiro parâmetro populacional.

Seja então $X_1 + X_2 + \dots + X_n$ uma amostra aleatória de uma população e θ o parâmetro de interesse. Sejam $\hat{\theta}_0$ e $\hat{\theta}_1$ estatísticas tais que:

$$P(\hat{\theta}_0 < \theta < \hat{\theta}_1) = 1 - \alpha$$

Então $[\hat{\theta}_0; \hat{\theta}_1]$ é chamado *intervalo de confiança* de nível $100(1 - \alpha)\%$ para o parâmetro θ . Usualmente toma-se $1 - \alpha = 0.95$ ou 0.99 .

$1 - \alpha \rightarrow$ é o nível ou grau de confiança e fornece a probabilidade de conter o verdadeiro parâmetro.

$\alpha \rightarrow$ é o nível de significância, representa o erro que se está cometendo ao afirmar que a probabilidade do intervalo $[\theta_i \leq \theta \leq \theta_s]$ conter o verdadeiro valor do parâmetro populacional θ é $(1 - \alpha)$.

Valores Críticos de Z para α e $\alpha / 2$.

α	10%	5%	1%	0,5%	0,2%
Z_{α}	1,28	1,64	2,36	2,58	2,88
$Z_{\alpha / 2}$	1,64	1,96	2,58	2,81	3,06

Muitos estatísticos consideram a construção de intervalos de confiança como o principal método de estudo de um parâmetro populacional através de uma amostra.

Como o **IC** é construído com base na estimativa por ponto, é aleatório, ao passo que o parâmetro é suposto uma constante da população.

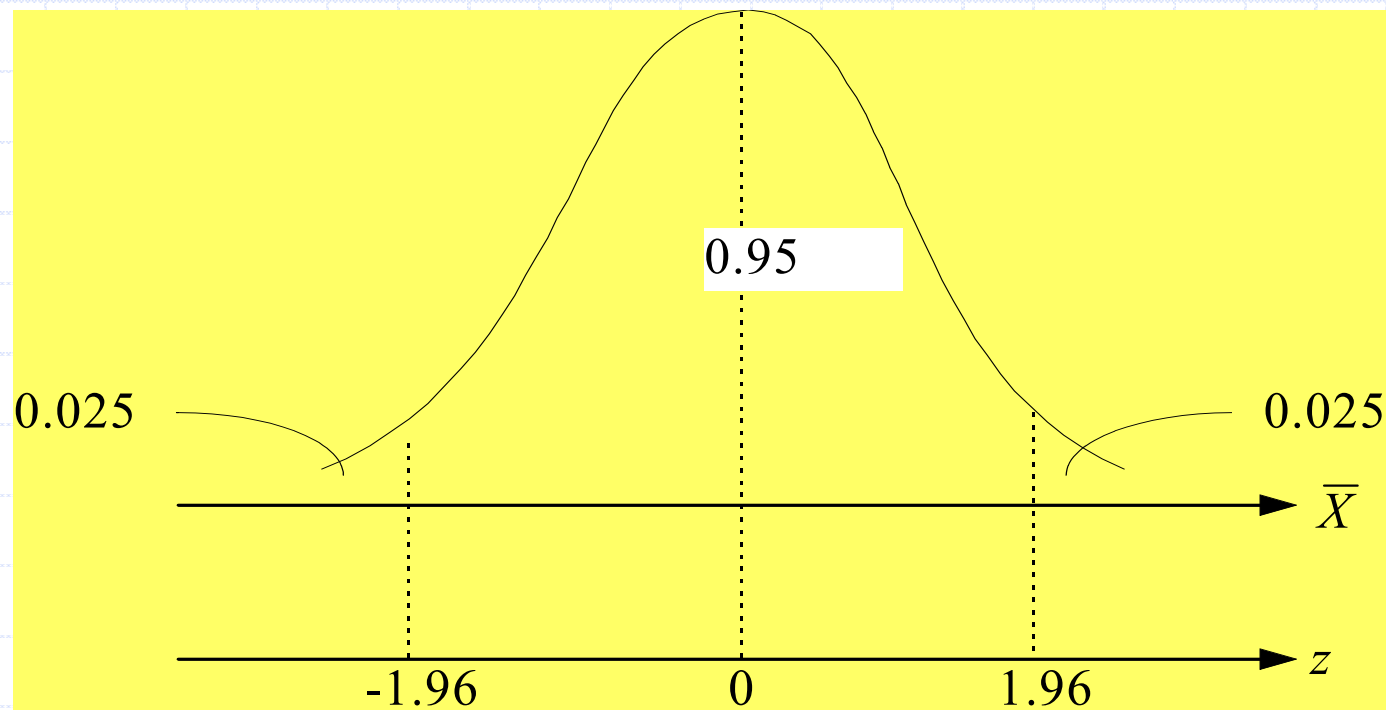
Assim, o **IC** conterá ou não o parâmetro, com probabilidades $1 - \alpha$ e α .

É incorreto dizer "probabilidade do parâmetro **CAIR** no intervalo".

Consideremos uma população normal com média μ e desvio padrão σ e uma amostra dessa população. Sabemos pelos resultados do Teorema Central do Limite que a média $\bar{X}$ desta amostra tem distribuição normal com média μ e desvio padrão $\sigma/\sqrt{n}$, ou seja:

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N: (0,1)$$

Fixando α em 0.05, ou seja, $1 - \alpha = 0.95$, vemos pela tabela de distribuição normal padronizada z , que:



$$P(-1.96 < Z < 1.96) = 0.95$$

isto é:

$$P \left[-1.96 < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < 1.96 \right] = 0.95$$

Reescrevendo as desigualdades entre parênteses, temos:

$$P \left[\bar{X} - 1.96(\sigma / \sqrt{n}) < \mu < \bar{X} + 1.96(\sigma / \sqrt{n}) \right] = 0.95$$

Neste caso: $\theta = \mu$ é o parâmetro

$$[\hat{\theta}_0; \hat{\theta}_1] = [\bar{X} - 1.96(\sigma / \sqrt{n}), \bar{X} + 1.96(\sigma / \sqrt{n})]$$

É o **IC** de 95% para μ , $IC(\mu:95\%)$

É importante observar que o *nível de confiança* ($1 - \alpha$) se aplica ao “processo” de construção de intervalos, e não a um intervalo específico.

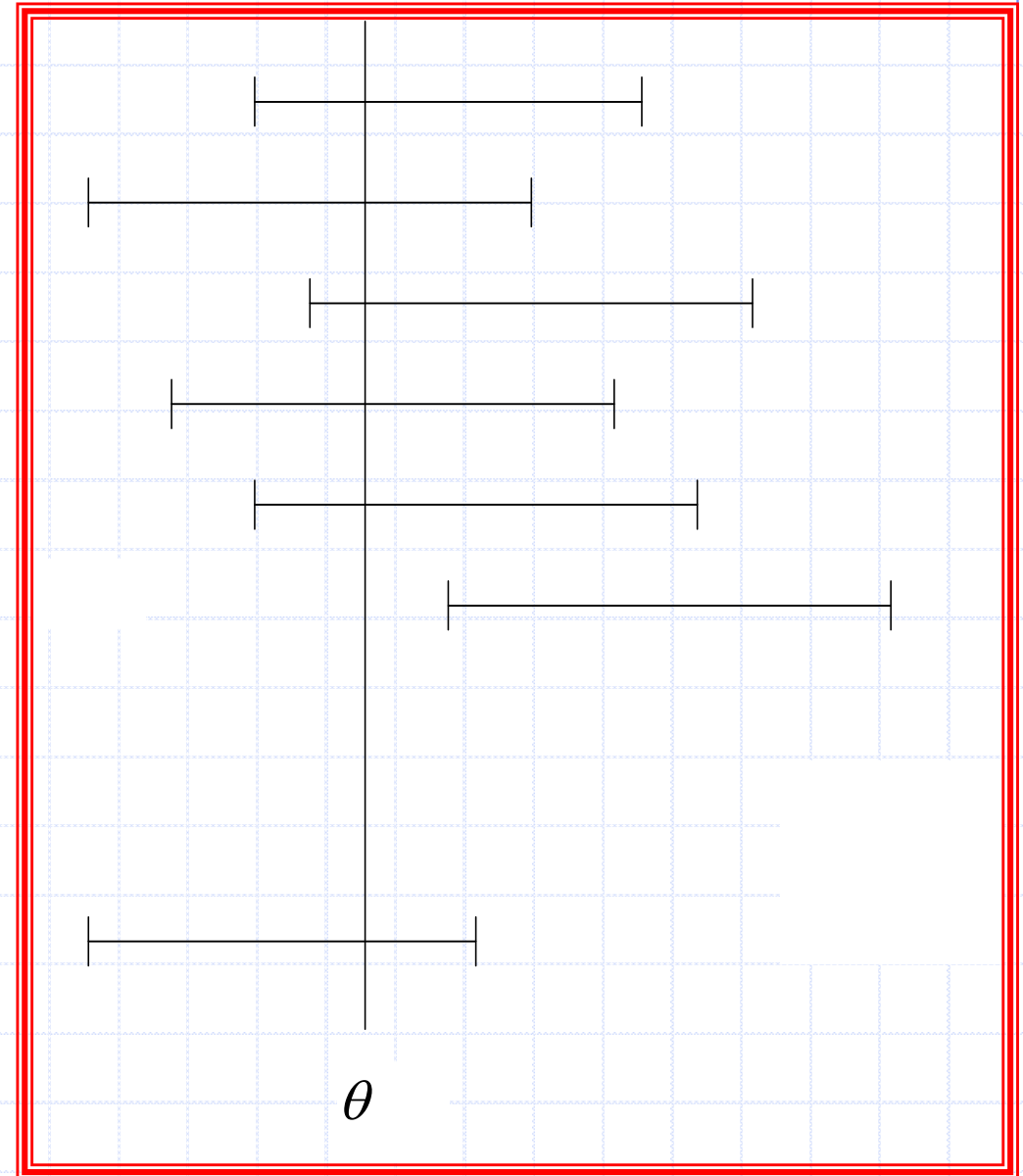
Para explicitar o conceito de **IC**, suponha que retiremos um grande número de amostras de tamanho n , fixo, da população em estudo e, para cada amostra, construamos um intervalo.

Os limites dos intervalos resultantes serão diferentes.

O verdadeiro valor do parâmetro estará contido, em média, em $100(1 - \alpha)\%$ desses intervalos.

$100(1 - \alpha)\%$ dos intervalos construídos abrangerão o verdadeiro valor do parâmetro.

No caso μ , conforme ilustrado na figura, mas cada valor contém, ou não contém, o parâmetro.



A expressão

$$P \left[\bar{X} - 1.96 \left(\sigma / \sqrt{n} \right) < \mu < \bar{X} + 1.96 \left(\sigma / \sqrt{n} \right) \right] = 0.95$$

deve ser interpretada muito cuidadosamente.

Ela não significa que a probabilidade do parâmetro μ cair dentro de um intervalo especificado seja igual a 0.95.

μ sendo o parâmetro, está ou não está dentro do intervalo acima. De preferência a expressão acima deve ser interpretada assim:

“0.95 é a probabilidade de que um intervalo aleatório contenha μ .”

$$\left[\bar{X} - 1.96 \left(\sigma / \sqrt{n} \right); \bar{X} + 1.96 \left(\sigma / \sqrt{n} \right) \right]$$

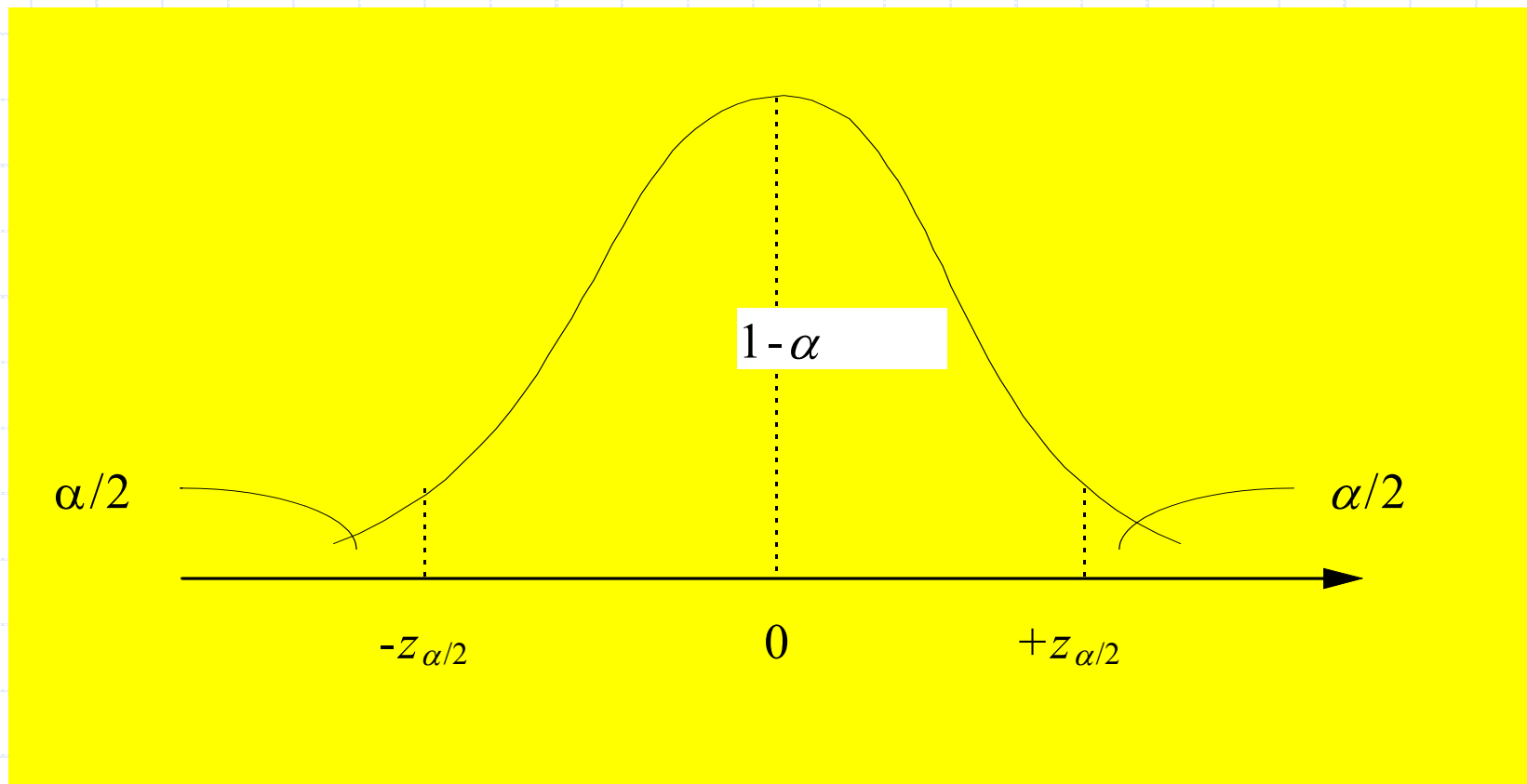
Para uma amostra de 50 observações de uma população normal com média desconhecida μ e desvio padrão $\sigma = 6$, seja 20,5 a média amostral $\bar{X}$. Construir um intervalo de 95% de confiança para a média populacional.

Temos, de imediato que:

$$P\left[\bar{X} - 1.96\left(\sigma/\sqrt{n}\right) < \mu < \bar{X} + 1.96\left(\sigma/\sqrt{n}\right)\right] = 0.95$$

Assim, tal intervalo é [18.84; 22.16].

A figura seguinte representa a curva $N(0,1)$ e a notação que iremos utilizar, onde $P[|Z| < Z_{\alpha/2}] = 1 - \alpha$



Na situação aqui apresentada, $N(\mu; \sigma)$ com σ conhecido, sabemos que:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0;1)$$

logo o intervalo de confiança para μ será:

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

$$P\left[\bar{X} - Z_{\alpha/2} \left(\sigma/\sqrt{n}\right) < \mu < \bar{X} + Z_{\alpha/2} \left(\sigma/\sqrt{n}\right)\right] = 1 - \alpha$$

$$IC(\mu; (1 - \alpha)100) = \left[\bar{X} - Z_{\alpha/2} \left(\sigma/\sqrt{n}\right); \bar{X} + Z_{\alpha/2} \left(\sigma/\sqrt{n}\right)\right]$$

Intervalo de Confiança para a média da população σ é conhecido

μ = média da população

$\bar{x}$ = média da amostra

σ = desvio-padrão da população

S = desvio-padrão da amostra

n = tamanho da amostra

e_o = Semi-amplitude do intervalo de confiança

Seja uma população $X \sim N(\mu, \sigma^2)$, sabe-se que:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad e \quad Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = N(0,1)$$

Pela figura anterior temos:

$$P[-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}] = 1 - \alpha$$

$$P\left[-Z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq Z_{\alpha/2}\right] = 1 - \alpha$$

$$P\left[-Z_{\alpha/2} \leq \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

Multiplicando-se por (-1):

$$P\left[-\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

Ordenando, temos:

$$P\left[\bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

$$P\left[\bar{x} - \frac{\sigma}{\sqrt{n}} \cdot Z_{\alpha/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} \cdot Z_{\alpha/2}\right] = 1 - \alpha$$

✦ Como a distribuição normal é simétrica

$$Z_{\alpha/2} = Z_{1 - \alpha/2}$$

✦ Se σ for desconhecido e $n \geq 30$, pode-se usar S , resultando em um intervalo aproximado.

✦ $e_0 = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$, podendo o **IC** ser escrito como:

$$P(\bar{x} - e_0 \leq \mu \leq \bar{x} + e_0) = 1 - \alpha$$

Exemplo:

Feito um ensaio de corrosão com 64 peças de um lote de produção, verificou-se que o tempo que a peça suportou nesse teste apresentou uma média igual a 200 horas. Calcular o IC de 95% para a verdadeira média, sabendo que $\sigma = 16$ horas.

Sabe-se que os comprimentos das barras produzidas por uma siderúrgica tem uma distribuição normal de variância $1,69 \text{ m}^2$. Numa amostra de cinco barras encontrou-se: 20,1; 21,0; 21,4; 22,1; 23,3 m. Determinar o IC para a média, com:

$$\text{a) } \alpha = 0,10 \rightarrow Z_{\alpha/2} = 1,645$$

$$\text{b) } \alpha = 0,06 \rightarrow Z_{\alpha/2} = 1,881$$

σ é desconhecido

Em geral nos problemas práticos σ é desconhecido e devemos estimá-lo:

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

Quanto menor a amostra, mais necessária se torna a introdução de uma correção, a qual consiste em a variável " t " de *Student* ao invés de Z .

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

$$P[-t_{\alpha/2} \leq t \leq t_{\alpha/2}] = 1 - \alpha$$

$$P[-t_{\alpha/2} \leq \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \leq t_{\alpha/2}] = 1 - \alpha$$

.....

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

✦ "t" possui $n - 1$ graus de liberdade $\Rightarrow t_{(n-1); \alpha/2}$

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

$$S = \sqrt{\frac{1}{n-1} \left(\sum x_i^2 \cdot f_i - \frac{(\sum x_i \cdot f_i)^2}{n} \right)}$$

✦ Convém ressaltar que, quanto maior for o nível de confiança (isto é, quanto menor for o nível de significância); mais amplo será o intervalo

$$\star e_0 = t_{\alpha/2} \cdot$$

Exemplo:

① A seguinte amostra foi extraída de uma população normal: 6, 6, 7, 8, 9, 9, 9, 10, 11, 12. Construir um intervalo para μ , ao nível de significância de 10%.

① Construir um I.C. de 95%, para a média populacional, a partir da seguinte distribuição amostral:

Classe	0 — 5	5 — 10	10 — 15	15 — 20
Frequência	2	3	5	2

Exemplo:

① Suponha que x tenha uma distribuição $N(\mu, \sigma^2)$.

Uma amostra de tamanho 25 fornece os seguintes

valores: $\sum x_i = 70,8$; $\sum x_i^2 = 3546,8$. Determine um

IC de 95% para μ .

IC para a σ^2 de uma população normal

Seja X uma população com média μ e variância σ^2 . Sabe-se pelo Teorema de Fisher, que: $X_{n-1}^2 = \frac{S^2(n-1)}{\sigma^2}$;

$$P[\chi_1^2 \leq \chi^2 \leq \chi_2^2] = 1 - \alpha$$

$$P[\chi_1^2 \leq \frac{S^2(n-1)}{\sigma^2} \leq \chi_2^2] = 1 - \alpha$$

$$P \left[\frac{\chi_1^2}{(n-1).S^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi_2^2}{(n-1).S^2} \right] = 1 - \alpha$$

$$P \left[\frac{S^2 (n-1)}{\chi_{sup}^2} \leq \sigma^2 \leq \frac{S^2 (n-1)}{\chi_{inf}^2} \right] = 1 - \alpha$$

$$\chi_{inf}^2 = \chi_1^2 = \chi_{1 - \alpha/2}^2$$

$$\chi_{sup}^2 = \chi_2^2 = \chi_{\alpha/2}^2 \quad \text{ambos com } \delta = n - 1$$

✦ Quando $\delta > 30$ graus de liberdade é comum usar a seguinte aproximação:

$$\chi^2_{\alpha} = \frac{1}{2} (\pm Z_{\alpha} \cdot \sqrt{2\delta - 1})^2$$

└ é a abscissa normal reduzida

✦ Como a tabela χ^2 é unicaudal a direita e o I.C. deve ser central, deve-se entrar na tabela com $\alpha/2$ e

$(1 - \alpha/2)$, para encontrar χ^2_{inf} e χ^2_{sup} .

Exemplo:

- ❶ Para 6, 6, 7, 8, 9, 9, 9, 10, 11, 12, calcular o IC para σ^2 , ao nível de 90%.
- ❷ De uma população normal foi retirada uma amostra de 15 elementos e calculou-se:
 $\sum x_i = 8,7$ e $\sum x_i^2 = 27,3$. Determinar um IC de 80% para a variância dessa população.
- ❸ Calcular um IC de 96% para a variância da distribuição mostrada a seguir (suposta como normal):

Classe	2,2	—	6,2	6,2	—	10,2	10,2	—	14,2	14,2	—	18,3
Freq.			3		4		5					3

IC para desvio-padrão σ de uma população normal

Se $S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ é estimador justo de σ^2 , mas S não é

estimador de σ , pois $E[S] = \sigma$, tem-se:

$V(S^2) = E(S^2) - E^2(S^2)$, logo $V(S^2) = \sigma^2 - \sigma^2 = 0$, o que

não tem sentido. O vício de S com estimador de σ ,

tende a zero quando $n \rightarrow \infty$.

Deveríamos adotar um coeficiente de correção, mas calcularemos um

IC aproximado para σ , bastando

extrair a raiz quadrada do IC da σ^2 .



Fazer EX.1, EX.2, EX.3 anteriores.

I.C. para a proporção populacional p

Sabemos que $f \cong N \left(p, \frac{pq}{n} \right)$ e que, para $n > 30$, a distribuição binomial tende a uma normal, então

$$Z = \frac{f - p}{\sqrt{pq / n}}$$

$$P[-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}] = 1 - \alpha$$

$$P\left[-Z_{\alpha/2} \leq \frac{f - p}{\sqrt{pq/n}} \leq Z_{\alpha/2}\right] = 1 - \alpha$$

.....

$$P[f - Z_{\alpha/2} \sqrt{pq/n} \leq p \leq f + Z_{\alpha/2} \sqrt{pq/n}] = 1 - \alpha$$

$$f = \frac{x}{n} = \frac{\text{casos de interesse}}{\text{total da amostra}}$$

$$P\left[f - Z_{\alpha/2} \cdot \sqrt{\frac{f(1-f)}{n}} \leq p \leq f + Z_{\alpha/2} \cdot \sqrt{\frac{f(1-f)}{n}} \right] = 1 - \alpha$$

Exemplo:

- ❶ Retirada uma amostra de 1.000 peças da produção de uma máquina, verificou-se que 35 eram defeituosas. Construir um IC, ao nível de 95%, para a proporção real de peças defeituosas fornecidas por essa máquina.
- ❷ Uma a.a. de 400 domicílios de uma cidade mostra que 25% são casas de aluguel. Qual o IC que podemos supor que seja o número de casas de aluguel dessa cidade, usando $\alpha = 2\%$, supondo que tal cidade tem 20.000 casas?

IC para a soma e para a diferença entre duas médias (μ_1 e μ_2) de duas populações normais, conhecidas suas variâncias σ_1^2 e σ_2^2 .

Se $X_1 = N(\mu_1, \sigma_1^2)$ e $X_2 = N(\mu_2, \sigma_2^2)$, sendo X_1 e X_2 independentes. Como $\bar{x}_1 = N(\mu_1, \frac{\sigma_1^2}{n_1})$
 $\bar{x}_2 = N(\mu_2, \frac{\sigma_2^2}{n_2})$

$(\bar{x}_1 \pm \bar{x}_2) = N[(\mu_1 \pm \mu_2); \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}]$ para amostras a.a. independentes, temos:

$$Z = \frac{(\bar{x}_1 \pm \bar{x}_2) - (\mu_1 \pm \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$P[(\bar{x}_1 \pm \bar{x}_2) - Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 \pm \mu_2 \leq (\bar{x}_1 \pm \bar{x}_2) + Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}] = 1 - \alpha$$

❗ Uma empresa tem duas filiais (A e B), para os quais os desvios-padrões de vendas diárias são de 5 e 3 peças, respectivamente. Uma amostra de 20 dias fornecem uma venda média diária de 40 peças para a filial A e 30 peças para a filial B. Supondo que a distribuição diária de vendas seja normal, construir um I.C. de 92% para a venda diária das duas filiais.

I.C. para a Soma e para a Diferença entre duas médias (μ_1 e μ_2) de duas populações normais, de mesma variância σ^2 desconhecida.

Sendo $X_1 = N(\mu_1, \sigma^2)$ e $X_2 = N(\mu_2, \sigma^2)$, com X_1 e X_2

independentes, logo $\bar{x}_1 = N(\mu_1, \frac{\sigma^2}{n_1})$ e $\bar{x}_2 = N(\mu_2, \frac{\sigma^2}{n_2})$,

tem-se que $(\bar{x}_1 \pm \bar{x}_2) = N[(\mu_1 \pm \mu_2); \sigma^2 \cdot (\frac{1}{n_1} + \frac{1}{n_2})]$ e:

$$Z = \frac{(\bar{x}_1 \pm \bar{x}_2) - (\mu_1 \pm \mu_2)}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Como não se conhece σ^2 , deve-se estimá-lo por S'^2 , onde:

$$S'^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}$$

$\delta = (n_1 + n_2 - 2)$ graus de liberdade

$$t(n_1 + n_2 - 2), = t_{\delta, \alpha}$$

$$P[(\bar{x}_1 \pm \bar{x}_2) - t_{\alpha/2} \cdot S' \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 \pm \mu_2 \leq (\bar{x}_1 \pm \bar{x}_2) + t_{\alpha/2} \cdot S' \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}] = 1 - \alpha$$

① Duas populações normais: X_1 e X_2 tem a mesma variância. Da população 1 foi extraída uma amostra de tamanho 10, obtendo-se $\bar{X}=15$ e $S=8$. Da população 2 foi extraída uma amostra de 12 elementos, obtendo-se $\bar{X}=12$ e $S=2$. Construir o I.C. de 95% para a diferença de médias.

I.C. para a Soma e para a Diferença entre duas médias (μ_1 e μ_2) de duas populações normais, de variâncias desconhecidas e $\sigma_1^2 \neq \sigma_2^2$ —

Se $X_1 = N(\mu_1, \sigma_1^2)$ e $X_2 = (\mu_2, \sigma_2^2)$, com X_1 e X_2 independentes, então

$(\bar{x}_1 \pm \bar{x}_2) = N[(\mu_1 \pm \mu_2); (\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})]$ Como σ_1^2 e σ_2^2 são

desconhecidos, temos que estimá-las.

$$P[(\bar{x}_1 \pm \bar{x}_2) - t_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 \pm \mu_2 \leq (\bar{x}_1 \pm \bar{x}_2) + t_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 1 - \alpha$$

- $t_{\delta, \alpha/2}$ onde δ é o gl. dado pelo método de Aspin-Welch, com arredondamento para menos.

$$\delta = \frac{(V_1 + V_2)^2}{\frac{V_1^2}{n_1 + 1} + \frac{V_2^2}{n_2 + 1}} - 2 \quad \text{onde } V_1 = \frac{S_1^2}{n_1} \quad \text{e} \quad V_2 = \frac{S_2^2}{n_2}$$

❗ Duas máquinas de embalar de embalar arroz estão sendo usadas por uma empresa, sendo uma nova e outra velha; pegas duas amostras de sacos embalados, encontramos os seguintes pesos, em Kg.

Máquina Nova: 82, 83, 79, 81, 81, 80

Máquina Velha: 79, 82, 78, 74, 80, 77, 75, 84, 78

Construir o I.C. para a diferença dos pesos médios populacionais, ao nível de significância de 5%.

I.C. para a Soma e para a Diferença de duas proporções populacionais p_1 e p_2 .

Se $f_1 = N \left(p_1, \frac{p_1 \cdot q_1}{n_1} \right)$ e $f_2 = N \left(p_2, \frac{p_2 \cdot q_2}{n_2} \right)$; logo,

$$(f_1 \pm f_2) = N \left[(p_1 \pm p_2); \frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2} \right] \text{ e,}$$

$$Z = \frac{(f_1 \pm f_2) - (p_1 \pm p_2)}{\sqrt{\frac{p_1 \cdot (1 - p_1)}{n_1} + \frac{p_2 \cdot (1 - p_2)}{n_2}}} \text{ logo o intervalo de confiança}$$

será:

$$P \left[-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2} \right] = 1 - \alpha$$

$$P \left[-Z_{\alpha/2} \leq \frac{(f_1 \pm f_2) - (p_1 \pm p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \leq Z_{\alpha/2} \right] = 1 - \alpha$$

$$P \left[(f_1 \pm f_2) - Z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \leq p_1 \pm p_2 \leq (f_1 \pm f_2) + Z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right] = 1 - \alpha$$

Como p_1 e p_2 são valores populacionais desconhecidos, eles podem ser estimados por f_1 e f_2 , desde que os tamanhos das amostras sejam maiores do que 30, daí temos:

$$P \left[(f_1 \pm f_2) - Z_{\alpha/2} \sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}} \leq p_1 \pm p_2 \right. \\ \left. \leq (f_1 \pm f_2) + Z_{\alpha/2} \sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}} \right] = 1 - \alpha$$

Exemplo:

- ① Um levantamento estatístico mostrou que 80 pessoas, das 200 consultadas, numa cidade, vão votar no candidato A na próxima eleição; uma outra amostra de 500 pessoas, dessa mesma cidade, mostrou que 150 delas vão votar no candidato B. Construir um IC de 93% para a diferença das proporções de pessoas que vão votar em A e B.

IC para o quociente das variâncias populacionais

Seja: $F(\delta_1, \delta_2) = \frac{\frac{\chi^2_{\delta_1}}{\delta_1}}{\frac{\chi^2_{\delta_2}}{\delta_2}} = \frac{\chi^2_{\delta_1}}{\chi^2_{\delta_2}} \cdot \frac{\delta_2}{\delta_1} \rightarrow$

para duas populações normais de variâncias descontinuas,

pelo Teorema de Fischer: $\chi^2_{n-1} = \frac{(n-1) \cdot S^2}{\sigma^2}$ ou

$$\chi^2_{\delta} = \frac{\delta \cdot S^2}{\sigma^2}$$

Portanto $F(\delta_1, \delta_2) = \frac{\frac{\delta_1 \cdot S_1^2}{\sigma_1^2}}{\frac{\delta_2 \cdot S_2^2}{\sigma_2^2}} \cdot \frac{\delta_2}{\delta_1} \therefore F(\delta_1, \delta_2) = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$

logo o intervalo será:

$$P[F_{1-\alpha/2} \leq F \leq F_{\alpha/2}] = 1 - \alpha$$

$$P[F_{1-\alpha/2} \leq \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leq F_{\alpha/2}] = 1 - \alpha$$

$$P\left[\frac{S_2^2}{S_1^2} \cdot \frac{1}{F_{\frac{\alpha}{2}}(\delta_2, \delta_1)} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \frac{S_2^2}{S_1^2} \cdot F_{\frac{\alpha}{2}}(\delta_1, \delta_2)\right] = 1 - \alpha$$

ou:

$$P\left[\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{\frac{\alpha}{2}}(\delta_1, \delta_2)} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \frac{S_1^2}{S_2^2} \cdot F_{\frac{\alpha}{2}}(\delta_2, \delta_1)\right] = 1 - \alpha$$

Exemplo:

- ① Construir um IC, para $\alpha = 2\%$, para o quociente de variâncias de duas populações normais, das quais foram extraídas as amostras seguintes: 41 elementos da primeira, obtendo-se $S_1^2 = 43,8$ e 31 elementos da segunda, obtendo-se $S_2^2 = 29,5$.