Sumário:

Intervalos de confiança

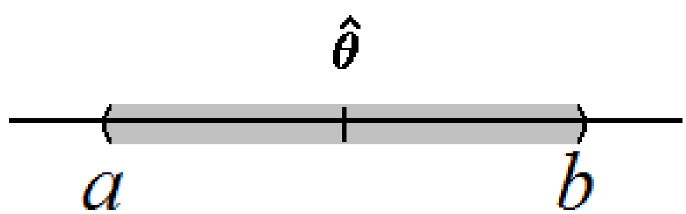
6.1. A estimação por intervalo

Normalmente, no processo de investigação de um parâmetro θ, necessitamos ir além da sua estimativa pontual $\hat{\theta}$. O fato de não se conhecer o valor de θ pode causar uma "insegurança" e levar a um questionamento:

Quão próximo estamos do valor real de θ quando obtemos sua estimativa?

A resposta depende da precisão (ou variância) do estimador e, também, do valor real do parâmetro.

Uma maneira de contornar esse problema consiste em se encontrar um intervalo em torno de $\hat{\theta}$ que tenha alta probabilidade de englobar θ.



$$P(\text{ do intervalo } [a,b] \text{ englobar } \theta\ ) = \gamma$$

O intervalo $[a,b]$, na prática, será construído com a amostra, ou seja, a partir dos dados e da distribuição amostral associada a $\hat{\theta}$.

Logo, os valores ***a*** e ***b*** serão *aleatórios*, variando de uma amostra para outra.

6.2. Intervalo de confiança para a média

6.2.1. Intervalo de confiança para a média com variância conhecida

Seja uma *aa* $X_1, X_2, \ldots, X_n$, com média μ e variância σ^2 conhecida. Para construir um intervalo de confiança para a média deve-se considerar a distribuição da média amostral $\bar{X}$,

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

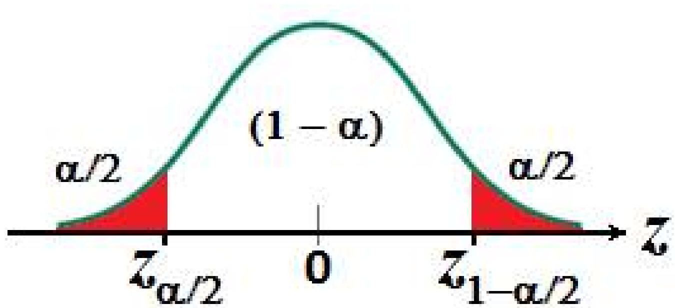
Intervalo de confiança **$(1-\alpha)\times 100\%$** para **μ**

Para construir um *I.C.* para a μ temos que obter constantes a e b tal que $P(a \leq \mu \leq b) = (1-\alpha)$.

A probabilidade $(1-\alpha)$ é chamada de *nível de confiança* do intervalo e α de *nível de significância*.

Então, da distribuição de $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, temos:

$$P\left(z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right) = (1-\alpha)$$



$$P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \geq \mu - \overline{X} \geq -z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = (1-\alpha)$$

$$P\left(\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = (1-\alpha)$$

Como $z_{1-\alpha/2} = -z_{\alpha/2}$, teremos:

$$P\left(\underbrace{\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{a} \leq \mu \leq \underbrace{\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{b}\right) = (1-\alpha)$$

$$a = \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{e} \quad b = \overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Nota: observe que, nessa notação, $z_{\alpha/2} < 0$.

Portanto, um *intervalo de confiança* **(1 – α)×100%** para μ, com **σ²** *conhecido*, é dado por:

$$\left[\overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \; ; \; \overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

Se α = 0,05, $\alpha/2 = 0.025$ e $z_{0.025} = -1.96$, logo, um *I.C.* 95% para μ, com σ² conhecido, é dado por:

$$\left[\overline{x} - 1.96\frac{\sigma}{\sqrt{n}} \; ; \; \overline{x} + 1.96\frac{\sigma}{\sqrt{n}}\right].$$

Exemplo 1: Testes de compressão foram aplicados na marca *A* de cimento para avaliar sua resistência em concretos. Foram produzidos 13 *corpos de prova* e os testes foram aplicados no Laboratório de testes do Departamento de Engenharia Civil da UFSCar.
(O corpo de prova padrão brasileiro, normatizado pela ABNT, é o cilíndrico, com 15 cm de diâmetro, 30 cm de altura e a idade de referência é 28 dias)

Foi registrada a resistência à compressão simples (f_c), para cada corpo de prova com o intuito de calcular a resistência característica do concreto à compressão (f_{ck}).

Um concreto concreto classe C30, por exemplo, corresponde a um concreto com f_{ck} = 30 Mpa (Mpa = 10^6Pa).

Pascal (unidade)

O Pascal (símbolo: Pa) é a unidade padrão de pressão e tensão no SI. Equivale a força de 1N aplicada uniformemente sobre uma superfície de 1m^2 (fonte: Wikipédia).

Dados (MPa):

31.04	31.11	39.56	24.83	36.97	34.86	29.44	$\bar{x}_{\mathbf{A}} = 33.76$
39.15	27.82	34.96	35.19	39.68	34.27		$s_{\mathbf{A}} = 4.665$

A empresa afirma que o processo tem variabilidade $\sigma^2_{\mathbf{A}}$ = 25MPa^2. Construir um intervalo de confiança 95% (nível de significância α = 0.05) para a resistência à compressão média.

Estatística: $\dfrac{\bar{X}_{\mathbf{A}} - \mu_{\mathbf{A}}}{\sigma_{\mathbf{A}} / \sqrt{n_{\mathbf{A}}}} \sim N(0\,;1)$

Encontrar *a* e *b* tais que: $P(a < \mu_{\mathbf{A}} < b) = 0.95$

$$P\left(-1.96 \le \frac{\overline{X}_{\mathbf{A}} - \mu_{\mathbf{A}}}{\sigma_{\mathbf{A}} / \sqrt{n_{\mathbf{A}}}} \le 1.96\right) = 0.95$$

$$P\left(-1.96\frac{\sigma_{\mathbf{A}}}{\sqrt{n_{\mathbf{A}}}} \le \overline{X}_{\mathbf{A}} - \mu_{\mathbf{A}} \le 1.96\frac{\sigma_{\mathbf{A}}}{\sqrt{n_{\mathbf{A}}}}\right) = 0.95$$

$$P\left(\overline{X}_{\mathbf{A}} - 1.96\frac{\sigma_{\mathbf{A}}}{\sqrt{n_{\mathbf{A}}}} \le \mu_{\mathbf{A}} \le \overline{X}_{\mathbf{A}} + 1.96\frac{\sigma_{\mathbf{A}}}{\sqrt{n_{\mathbf{A}}}}\right) = 0.95$$

Substituindo os valores da média amostral e tamanho da amostra

$$P\left(33.76 - 1.96\frac{5}{\sqrt{13}} \le \mu_{\mathbf{A}} \le 33.76 + 1.96\frac{5}{\sqrt{13}}\right) = 0.95$$

$$P(31.04 \le \mu_{\mathbf{A}} \le 36.48) = 0.95$$

Ou seja:

$$a = 33.76 - 1.96\frac{5}{\sqrt{13}} = 31.04\,\text{MPa}$$

$$b = 33.76 + 1.96\frac{5}{\sqrt{13}} = 36.48\,\text{MPa}$$

Logo, **(31.04, 36.48)** é um *I.C.* 95% para μ_A.

Interpretação: o intervalo (31.04 ; 36.48) tem probabilidade 0.95 (95%) de englobar o real valor da média μ_A.

6.2.2. Intervalo de confiança para a média com variância desconhecida

Seja uma *aa* $X_1, X_2, \ldots, X_n$, com média μ e variância σ^2 desconhecida. No caso da variância ser desconhecida devemos utilizar sua estimativa dada pela variância amostral s^2, porém, nesse caso a distribuição associada à média amostral $\bar{X}$ não será mais a normal.

Resultado: a estatística $\dfrac{\bar{X}-\mu}{s/\sqrt{n}}$ tem distribuição ***t – Student*** com $(n-1)$ graus de liberdade, ou seja

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

Notas:

1) A razão $\dfrac{\bar{X}-\mu}{s/\sqrt{n}}$ pode ser escrita como:

$$\frac{\bar{X}-\mu}{s/\sqrt{n}}\frac{\sigma}{\sigma} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\frac{\sigma}{s}$$

$$= \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)s^2/\sigma^2}{n-1}}} = \frac{N(0,1)}{\sqrt{\dfrac{\chi^2_{n-1}}{n-1}}}$$

Ou seja, a distribuição *t-Student* é dada pela razão de uma $N(0,1)$ por χ^2 uma dividida pelos seus graus de liberdade.

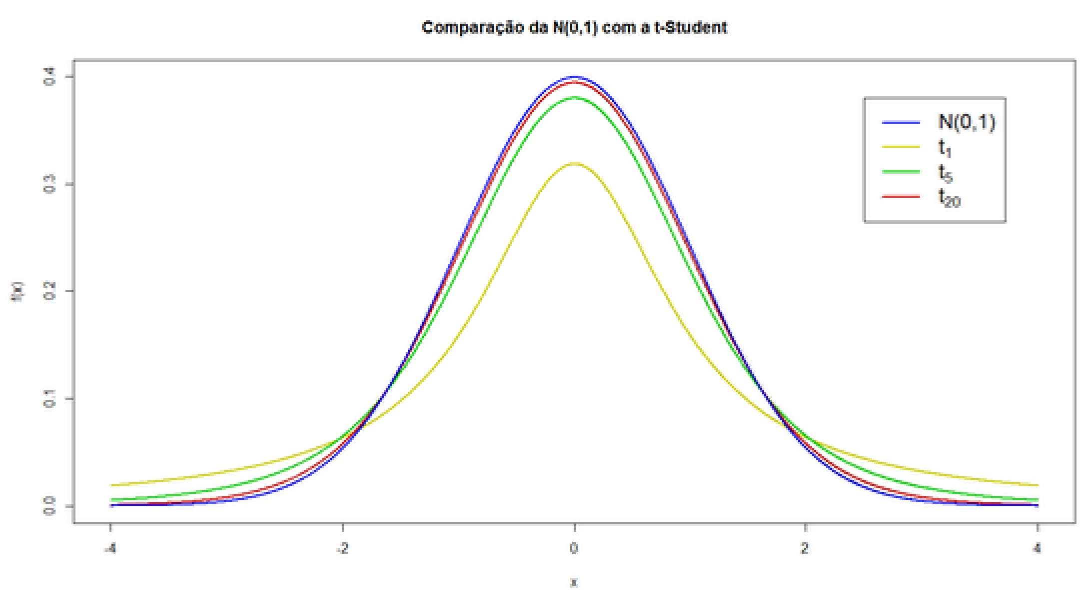
2) Assim com a normal padronizada a distribuição ***t – Student*** tem formato de sino, ou seja, é simétrica em torno do zero, porém, para graus de liberdade pequenos a moderados suas caudas são mais "pesadas".

3) Se uma $va\ T$ tem distribuição $t - Student$ com k graus de liberdade, então:

$$E(T) = 0 \qquad \text{e} \qquad Var(T) = \frac{k}{k-2}$$

4) Quando os graus de liberdade crescem, a distribuição $t - Student$ se aproxima da $N(0,1)$.

5) A distribuição $t - Student$ com 1 grau de liberdade é conhecida como ***distribuição de Cauchy***.



Para construir um $I.C.$ para a μ quando σ é desconhecida, devemos proceder como nos casos anteriores, porém substituindo a distribuição normal padrão pela *t-Student*, ou seja:

$$P\left(t_{(n-1);\alpha/2} \leq \frac{\overline{X}-\mu}{s/\sqrt{n}} \leq t_{(n-1);1-\alpha/2} \right) = (1-\alpha)$$

$$P\left(-t_{(n-1);\alpha/2}\frac{s}{\sqrt{n}} \geq \mu - \overline{X} \geq -t_{(n-1);1-\alpha/2}\frac{s}{\sqrt{n}} \right) = (1-\alpha)$$

$$P\left(\overline{X} - t_{(n-1);1-\alpha/2}\frac{s}{\sqrt{n}} \ \leq \ \mu \ \leq \ \overline{X} - t_{(n-1);\alpha/2}\frac{s}{\sqrt{n}} \right) = (1-\alpha)$$

Como $t_{(n-1);1-\alpha/2} = -t_{(n-1);\alpha/2}$, temos:

$$P\left(\overline{X} + t_{(n-1);\alpha/2}\frac{s}{\sqrt{n}} \ \leq \ \mu \ \leq \ \overline{X} - t_{(n-1);\alpha/2}\frac{s}{\sqrt{n}} \right) = (1-\alpha)$$

Logo, um *intervalo de confiança* **(1 – α)×100%** para μ, com **σ²** *desconhecido*, é dado por

$$\left[\overline{x} + t_{(n-1);\alpha/2}\frac{s}{\sqrt{n}} \ ; \ \overline{x} - t_{(n-1);\alpha/2}\frac{s}{\sqrt{n}} \right]$$

Exemplo 2: No caso dos testes de compressão em amostras de concreto, o gerente da companhia, desconfiando de que a informação a respeito da variância não seja verdadeira, refez os cálculos estimando a variância do processo por s^2.

Como o procedimento de cálculo é o mesmo, basta substituir o valor do quantil da normal ($Z_{0.025} = 1.96$) pelo quantil das distribuição t – *Student* com $(n - 1) = 12$ graus de liberdade.

Como $n_{\mathbf{A}} = 13$, então $t_{(n-1);\alpha/2} = t_{12;0.025} = 2.1788$

Com $\overline{x}_{\mathbf{A}} = 33.76$ e $s_{\mathbf{A}} = 4.665$ refazendo os cálculos temos que

$$\bar{x}_{\mathbf{A}} - t_{(n_{\mathbf{A}}-1);0.025}\frac{s_{\mathbf{A}}}{\sqrt{n_{\mathbf{A}}}} = 33.76 - 2.1788\frac{4.665}{\sqrt{13}} = 30.94\,\text{MPa}$$

$$\bar{x}_{\mathbf{A}} + t_{(n_{\mathbf{A}}-1);0.025}\frac{s_{\mathbf{A}}}{\sqrt{n_{\mathbf{A}}}} = 33.76 + 2.1788\frac{4.665}{\sqrt{13}} = 36.58\,\text{MPa}$$

Portanto, **(30.94 , 36.58)** é um IC 95% para μ_A para o caso em que a variância é desconhecida

***Interpretação*:** é mesma do caso anterior, porém, agora a variância é desconhecida.

6.2.3. Intervalo de confiança para a proporção

Como a proporção p é de fato a média amostral de uma *aa* cuja *va* tem distribuição de *Bernoulli*(p), para se construir intervalos de confiança para p devemos seguir os mesmos procedimentos anteriores.

Considerando que o estimador da proporção $\hat{p}$ tem valor esperado p e variância $\dfrac{p(1-p)}{n}$, dada a distribuição

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

um ***I.C.* (1 – α)×100% *para a proporção*** é dado por:

$$\left[\hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \;\; ; \;\; \hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right].$$

Exemplo 3: Nos testes de compressão em amostras de concreto, se a empresa afirma que 90% da produção atende ao valor do f_{ck} = 30Mpa, construir um *I.C.* de 95% (α = 0,05) para a proporção de corpos de provas com f_c abaixo de f_{ck}.

Dos 13 corpos de prova os valores 24.83, 29.44 e 27.82 são menores do que o f_{ck} de 30Mpa. Então, $\hat{p} = \dfrac{3}{13} = 0.231$

Considerando que $p = 0.10$:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.231 - 1.96\sqrt{\frac{0.10 \cdot 0.90}{13}} = 0.0679$$

$$\hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.231 + 1.96\sqrt{\frac{0.10 \cdot 0.90}{13}} = 0.3941$$

Ou seja: $P\left(0.0679 \le p \le 0.3941\right) = 0.95$

Portanto, **(0.0679 ; 0.3941)** é um *I.C.* 95% para p.

Interpretação: o intervalo (0,0679 ; 0,3941) tem probabilidade 0,95 (95%) de englobar o real valor do parâmetro p.

Nota: Como normalmente não conhecemos p, podemos construir intervalos de confiança para a proporção substituindo p e $(1-p)$ por $\hat{p}$ e $(1-\hat{p})$, respectivamente. Neste caso o intervalo fica:

$$\left[\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;;\; \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right].$$

Outra possibilidade seria considerar o fato de $p(1-p) \leq 1/4$ e construir um *intervalo conservador* para p assumindo p = ½.

Neste caso:

$$\frac{p(1-p)}{n} = \frac{1}{4n}$$

Logo, o intervalo de confiança *conservador* para p será

$$\left[\hat{p} + \frac{z_{\alpha/2}}{\sqrt{4n}} \;\; ; \;\; \hat{p} - \frac{z_{\alpha/2}}{\sqrt{4n}}\right].$$

Considerando α = 0.05, então, $I.C.$'s 95% para p, nos casos acima serão dados por:

i) utilizando $\hat{p}$: $$\left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;\; ; \;\; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

ii) conservador $p = ½$: $$\left[\hat{p} - \frac{1.96}{\sqrt{4n}} \;\; ; \;\; \hat{p} + \frac{1.96}{\sqrt{4n}}\right]$$

O procedimento em (ii) fornece intervalos de confiança excessivamente grandes quando p se distancia de ½ ($p \approx 0$ ou $p \approx 1$) (Bussab & Moretin, 2002). Para a utilização do intervalo conservador, portanto, devemos ter algum conhecimento do valor p, garantindo que seu valor esteja próximo de ½.

Exemplo: No exemplo do teste de compressão em concretos temos $\hat{p} = \dfrac{3}{13} = 0.231$, logo

i) utilizando $\hat{p}$:

$$\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.231 - 1.96\sqrt{\frac{0.231 \cdot 0.769}{13}} = 0.0019$$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.231 + 1.96\sqrt{\frac{0.231 \cdot 0.769}{13}} = 0.4601$$

Portanto, **(0.0019 ; 0.4601)** é um *I.C.* 95% para p.

ii) conservador $p =$ ½:

$$\hat{p} + \frac{z_{\alpha/2}}{\sqrt{4n}} = 0.231 - \frac{1.96}{\sqrt{52}} = -0.0408 \quad (< 0\ !!)$$

$$\hat{p} - \frac{z_{\alpha/2}}{\sqrt{4n}} = 0.231 + \frac{1.96}{\sqrt{52}} = 0.5028$$

Portanto, **(– 0.0408 ; 0.5028)** é um *I.C.* 95% conservador para p.

Note que no intervalo acima o limite inferior é negativo, consequência da utilização da máxima variância de p e do fato de que a proporção a ser estimada está longe do valor ½.

Nota: usualmente, nestes casos, arredondamos o limite inferior para 0 (zero), porém, o mais indicado é a utilização da estimativa $\hat{p}$.

Forma simplificada de representação:

i) Média com variância conhecida: $\bar{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

ii) Média com variância desconhecida: $\bar{x} \pm t_{(n-1);\alpha/2} \dfrac{s}{\sqrt{n}}$

iii) Proporção: $\hat{p} \pm z_{\alpha/2} \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Exemplos:

1) Um provedor de acesso à internet deseja implantar um plano sem limite de horas. Para isso, verificou numa amostra de $n = 25$ usuários os tempos de utilização mensal, obtendo: média amostral $\bar{x} = 26.8$ horas. Sabendo que $\sigma^2 = 6.25$ horas^2:

a) Encontre um intervalo de confiança 90% para a média.

b) De quanto deve ser aumentado o tamanho da amostra para que, mantidas as demais medidas, o comprimento do intervalo caia pela metade?

2) Observou-se a estatura de 20 recém-nascidos num hospital conforme dados abaixo. Pesquisas anteriores indicam que a estatura média das crianças nascidas neste hospital é de $\mu = 51$ cm.
Dados: $\Sigma x = 987$ e $\Sigma x^2 = 48845.25$

a) Qual a probabilidade de que a estatura média da amostra não ultrapasse 50.20 cm?

b) Construa um I.C. 99% para a média.

3) 10 corpos de provas foram submetidos a um teste de corrosão onde foram submersos em água salgada durante 60 segundos/dia. A corrosão foi medida pela perda de peso em miligramas/decímetro quadrado/dia (MDD). Os dados obtidos foram:

130.1 124.2 122.0 110.8 113.1 103.9 101.5 92.3 91.4 83.7

a) Encontre estimativas para a média e variância para a perda de peso em MDD.

b) Construa um intervalo de 95% de confiança para a média.

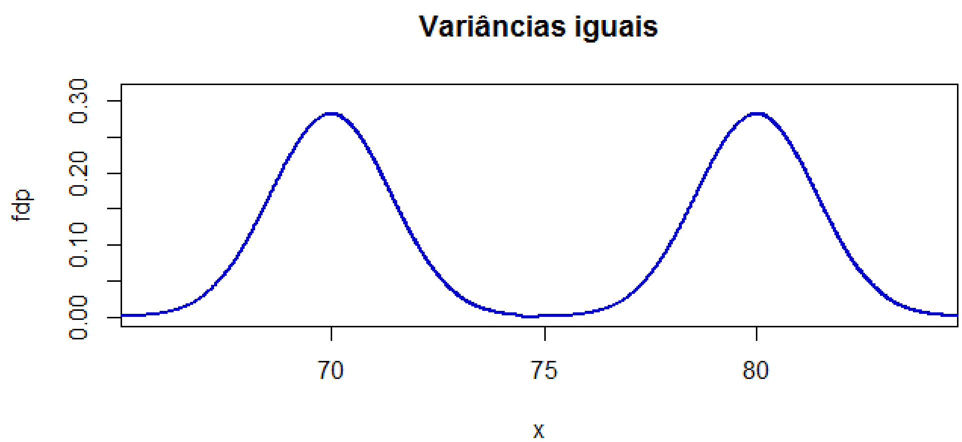
c) Supondo que a verdadeira média seja $\mu = 110$, calcule a probabilidade de que $\overline{X}$ seja superior ao máximo valor da amostra considerando:
i) desvio padrão conhecido $\sigma = 16$; **ii)** desvio padrão desconhecido.

6.3. Intervalo de confiança para a diferença entre as médias de duas populações independentes

Sejam duas populações $\mathbf{A}$ e $\mathbf{B}$ cujas médias são $\mu_\mathbf{A}$ e $\mu_\mathbf{B}$ e variâncias $\sigma^2_\mathbf{A}$ e $\sigma^2_\mathbf{B}$, respectivamente.

Um estimador não viciado para $(\mu_\mathbf{A} - \mu_\mathbf{B})$ é dado pela estatística $\left(\overline{X}_\mathbf{A} - \overline{X}_\mathbf{B}\right)$ e sua distribuição amostral é obtida conforme três diferentes situações:

i) Populações independentes com variâncias conhecidas;
ii) Populações independentes com variâncias desconhecidas, porém, iguais;
iii) Populações independentes com variâncias diferentes e desconhecidas.



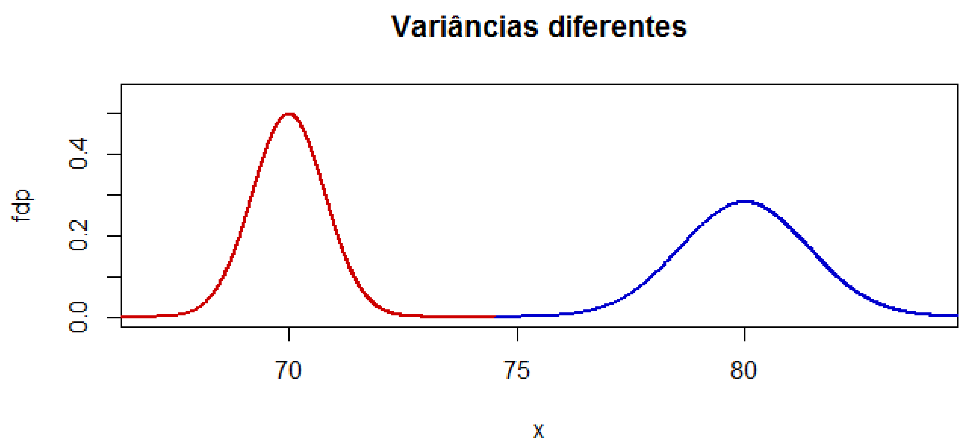


Figura: Populações normais.

6.3.1. Intervalo de confiança para a diferença entre as médias de duas populações independentes com variâncias conhecidas

Seja uma *aa* de tamanho $n_{\mathbf{A}}$, retirada da população **A** e uma *aa* de tamanho $n_{\mathbf{B}}$ retirada da população **B**, independentes. Considerando que as variâncias $\sigma_{\mathbf{A}}^2$ e $\sigma_{\mathbf{B}}^2$ sejam ambas conhecidas, temos que:

$$\frac{\overline{X}_{\mathbf{A}} - \mu_{\mathbf{A}}}{\sigma_{\mathbf{A}} / \sqrt{n_{\mathbf{A}}}} \quad \text{e} \quad \frac{\overline{X}_{\mathbf{B}} - \mu_{\mathbf{B}}}{\sigma_{\mathbf{B}} / \sqrt{n_{\mathbf{B}}}} \quad \text{são} \quad N(0\,,1)$$

Da teoria da probabilidade temos que

$$E\left(\overline{X}_{\mathbf{A}} - \overline{X}_{\mathbf{B}}\right) = \mu_{\mathbf{A}} - \mu_{\mathbf{B}} \quad \text{e} \quad Var\left(\overline{X}_{\mathbf{A}} - \overline{X}_{\mathbf{B}}\right) = \frac{\sigma_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \frac{\sigma_{\mathbf{B}}^2}{n_{\mathbf{B}}}$$

Logo, para o caso em que as variâncias $\sigma_{\mathbf{A}}^2$ e $\sigma_{\mathbf{B}}^2$ são conhecidas, a distribuição amostral associada à estatística $\left(\overline{X}_{\mathbf{A}} - \overline{X}_{\mathbf{B}}\right)$ é dada por:

$$\frac{\left(\overline{X}_{\mathbf{A}} - \overline{X}_{\mathbf{B}}\right) - \left(\mu_{\mathbf{A}} - \mu_{\mathbf{B}}\right)}{\sqrt{\dfrac{\sigma_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \dfrac{\sigma_{\mathbf{B}}^2}{n_{\mathbf{B}}}}} \sim N(0\,,1)$$

Observe que a variável padronizada tem expressão similar aos casos anteriores, ou seja, a *diferença entre a va e sua média, dividida pelo seu desvio padrão.*

Podemos, assim, construir um $I.C.$ para $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$ a partir de

$$P\left[z_{\alpha/2} \le \frac{(\bar{X}_{\mathbf{A}} - \bar{X}_{\mathbf{B}}) - (\mu_{\mathbf{A}} - \mu_{\mathbf{B}})}{\sqrt{\dfrac{\sigma^2_{\mathbf{A}}}{n_{\mathbf{A}}} + \dfrac{\sigma^2_{\mathbf{B}}}{n_{\mathbf{B}}}}} \le z_{1-\alpha/2} \right] = (1-\alpha).$$

Ou seja, um $I.C.$ $(1-\alpha)\times 100\%$ para $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$ considerando *amostras independentes e variâncias conhecidas* é dado por:

$$\left[(\bar{x}_{\mathbf{A}} - \bar{x}_{\mathbf{B}}) + z_{\alpha/2}\sqrt{\frac{\sigma^2_{\mathbf{A}}}{n_{\mathbf{A}}} + \frac{\sigma^2_{\mathbf{B}}}{n_{\mathbf{B}}}} \;;\; (\bar{x}_{\mathbf{A}} - \bar{x}_{\mathbf{B}}) - z_{\alpha/2}\sqrt{\frac{\sigma^2_{\mathbf{A}}}{n_{\mathbf{A}}} + \frac{\sigma^2_{\mathbf{B}}}{n_{\mathbf{B}}}} \right].$$

Exemplo 4: Considere que no exemplo com os testes de compressão em amostras de concretos, além da *A* uma segunda marca *B* tenha sido avaliada com o intuito de que fossem comparadas.

Dados (MPa):

A	31.04	31.11	39.56	24.83	36.97	34.86	29.44	39.15	$\bar{x}_{\mathbf{A}} = 33.76$
	27.82	34.96	35.19	39.68	34.27				$s_{\mathbf{A}} = 4.665$
B	27.91	40.94	39.25	37.42	32.16	34.29	38.69	21.21	$\bar{x}_{\mathbf{B}} = 33.08$
	29.30	29.21	33.76	32.71	31.91	34.10	33.34		$s_{\mathbf{B}} = 5.017$

a) Sabendo que as empresas afirmam que ambos os processos têm variabilidade σ^2 = 25MPa^2, construir um $I.C.$ para a diferença entre as médias das duas marcas.

Solução:

a) Como $\sigma_{\mathbf{A}}^2 = \sigma_{\mathbf{B}}^2 = 25$ então:

$$\frac{(\bar{X}_{\mathbf{A}} - \bar{X}_{\mathbf{B}}) - (\mu_{\mathbf{A}} - \mu_{\mathbf{B}})}{\sqrt{\dfrac{\sigma_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \dfrac{\sigma_{\mathbf{B}}^2}{n_{\mathbf{B}}}}}$$

Logo, um $I.C.$ 95% para $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$ é dado por:

$$\left[(\bar{x}_{\mathbf{A}} - \bar{x}_{\mathbf{B}}) + z_{\alpha/2}\sqrt{\frac{\sigma_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \frac{\sigma_{\mathbf{B}}^2}{n_{\mathbf{B}}}} \;;\; (\bar{x}_{\mathbf{A}} - \bar{x}_{\mathbf{B}}) - z_{\alpha/2}\sqrt{\frac{\sigma_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \frac{\sigma_{\mathbf{B}}^2}{n_{\mathbf{B}}}} \right].$$

Ou seja:

$$(\bar{x}_{\mathbf{A}} - \bar{x}_{\mathbf{B}}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \frac{\sigma_{\mathbf{B}}^2}{n_{\mathbf{B}}}}$$

$$(33.76 - 33.08) \pm 1.96\sqrt{\frac{25}{13} + \frac{25}{15}}$$

Portanto **(–3.034 , 4.394)** é um $I.C.$ 95% para $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$.

6.3.2. Intervalo de confiança para a diferença entre as médias de duas populações independentes com variâncias iguais e desconhecidas

Sejam duas populações **A** e **B** cujas médias são $\mu_{\mathbf{A}}$ e $\mu_{\mathbf{B}}$ e variâncias desconhecidas, porém iguais, ou seja, $\sigma_{\mathbf{A}}^2 = \sigma_{\mathbf{B}}^2 = \sigma^2$

Nesse caso, contudo, tanto $s_{\mathbf{A}}^2$ como $s_{\mathbf{B}}^2$ estimam a variância comum, logo, *podemos utilizar as informações de ambas as amostras* para estimar a variância populacional.

O que se faz, na prática, é combinar as somas de quadrados das duas variâncias amostrais e dividir pelos graus de liberdade total, ou seja

$$\sum_{i=1}^{n_{\mathbf{A}}}(x_{\mathbf{A}i}-\bar{x}_{\mathbf{A}})^2=(n_{\mathbf{A}}-1)\,s_{\mathbf{A}}^2 \qquad \Rightarrow \qquad (n_{\mathbf{A}}-1) = g.l. \text{ de } s_{\mathbf{A}}^2$$

$$\sum_{i=1}^{n_{\mathbf{B}}}(x_{\mathbf{B}i}-\bar{x}_{\mathbf{B}})^2=(n_{\mathbf{B}}-1)\,s_{\mathbf{B}}^2 \qquad \Rightarrow \qquad (n_{\mathbf{B}}-1) = g.l. \text{ de } s_{\mathbf{B}}^2$$

que combinadas, resultam em

$$s_p^2 = \frac{\sum_{i=1}^{n_{\mathbf{A}}}(x_{\mathbf{A}i}-\bar{x}_{\mathbf{A}})^2+\sum_{i=1}^{n_{\mathbf{B}}}(x_{\mathbf{B}i}-\bar{x}_{\mathbf{B}})^2}{(n_{\mathbf{A}}-1)+(n_{\mathbf{B}}-1)}$$

$$s_p^2 = \frac{(n_{\mathbf{A}}-1)\,s_{\mathbf{A}}^2+(n_{\mathbf{B}}-1)\,s_{\mathbf{B}}^2}{n_{\mathbf{A}}+n_{\mathbf{B}}-2}$$

A *variância combinada* s_p^2 (ou *pooled*), nada mais é do que uma variância ponderada pelos graus de liberdade das duas amostras:

$$s_p^2 = \frac{(n_{\mathbf{A}}-1)\,s_{\mathbf{A}}^2+(n_{\mathbf{B}}-1)\,s_{\mathbf{B}}^2}{n_{\mathbf{A}}+n_{\mathbf{B}}-2},$$

Assim como $s_{\mathbf{A}}^2$ e $s_{\mathbf{B}}^2$, s_p^2 é um estimador não viesado para σ^2.

Prova:

$$E\left(s_p^2\right)=\frac{(n_{\mathbf{A}}-1)\,E\left(s_{\mathbf{A}}^2\right)+(n_{\mathbf{B}}-1)E\left(s_{\mathbf{B}}^2\right)}{n_{\mathbf{A}}+n_{\mathbf{B}}-2}$$

$$=\frac{(n_{\mathbf{A}}-1)\,\sigma^2+(n_{\mathbf{B}}-1)\sigma^2}{n_{\mathbf{A}}+n_{\mathbf{B}}-2}$$

$$=\frac{(n_{\mathbf{A}}-1+n_{\mathbf{B}}-1)}{n_{\mathbf{A}}+n_{\mathbf{B}}-2}\,\sigma^2$$

$$=\sigma^2$$

Pelo fato de σ^2 ser desconhecida, temos que

$$\frac{\left(\bar{X}_{\mathbf{A}}-\mu_{\mathbf{A}}\right)}{s_{\mathbf{A}}/\sqrt{n_{\mathbf{A}}}}\sim t_{n_{\mathbf{A}}-1} \qquad \text{e} \qquad \frac{\left(\bar{X}_{\mathbf{B}}-\mu_{\mathbf{B}}\right)}{s_{\mathbf{B}}/\sqrt{n_{\mathbf{B}}}}\sim t_{n_{\mathbf{B}}-1}.$$

Como temos um estimador comum para a variância populacional, podemos derivar uma distribuição de probabilidade para $\left(\bar{X}_{\mathbf{A}}-\bar{X}_{\mathbf{B}}\right)$.

Padronizando a diferença entre as médias amostrais teremos:

$$\frac{\left(\bar{X}_{\mathbf{A}}-\bar{X}_{\mathbf{B}}\right)-\left(\mu_{\mathbf{A}}-\mu_{\mathbf{B}}\right)}{\sqrt{\dfrac{s_p^2}{n_{\mathbf{A}}}+\dfrac{s_p^2}{n_{\mathbf{B}}}}} = \frac{\left(\bar{X}_{\mathbf{A}}-\bar{X}_{\mathbf{B}}\right)-\left(\mu_{\mathbf{A}}-\mu_{\mathbf{B}}\right)}{s_p\sqrt{\dfrac{1}{n_{\mathbf{A}}}+\dfrac{1}{n_{\mathbf{B}}}}}$$

Resultado: $$\frac{(\bar{X}_{\mathbf{A}} - \bar{X}_{\mathbf{B}}) - (\mu_{\mathbf{A}} - \mu_{\mathbf{B}})}{s_p \sqrt{\dfrac{1}{n_{\mathbf{A}}} + \dfrac{1}{n_{\mathbf{B}}}}} \sim t_{n_{\mathbf{A}} + n_{\mathbf{B}} - 2}$$

Um $I.C.\ (1-\alpha)\times 100\%$ para $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$, quando as *variâncias são iguais e desconhecidas*, é dado por:

$$(\bar{x}_{\mathbf{A}} - \bar{x}_{\mathbf{B}}) \pm t_{(n_{\mathbf{A}} + n_{\mathbf{B}} - 2);\alpha/2}\, s_p \sqrt{\frac{1}{n_{\mathbf{A}}} + \frac{1}{n_{\mathbf{B}}}}$$

Exemplo 5: Construir um $I.C.\ 95\%$ para a diferença entre as resistências médias à compressão em concretos feitos com cimentos das marcas *A* e *B*, considerando variâncias iguais e desconhecidas.
(você acha válida a suposição de variâncias iguais?)

$$\bar{x}_{\mathbf{A}} = 33.76 \qquad \bar{x}_{\mathbf{B}} = 33.08$$
$$s_{\mathbf{A}} = 4.665 \qquad s_{\mathbf{B}} = 5.017$$
$$n_{\mathbf{A}} = 13 \qquad n_{\mathbf{B}} = 15$$

$$\Rightarrow s_p^2 = \frac{12(4.665)^2 + 14(5.017)^2}{13 + 15 - 2} = \frac{613.5307}{26} = 26.597$$

$$s_p = 4.8577$$

$$\Rightarrow t_{(n_{\mathbf{A}} + n_{\mathbf{B}} - 2);\alpha/2} = t_{26;0.025} = 2.0555$$

Logo, um *I.C.* 95% para $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$ é dado por:

$$(\bar{x}_{\mathbf{A}} - \bar{x}_{\mathbf{B}}) \pm t_{26;0.025}\, s_p \sqrt{\frac{1}{n_{\mathbf{A}}} + \frac{1}{n_{\mathbf{B}}}}$$

$$(33.76 - 33.08) \pm 2.0555 \times 4.8577\sqrt{\frac{1}{13} + \frac{1}{15}}$$

Portanto **(–3.105 , 4.465)** é um *I.C.* 95% para $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$ considerando variâncias iguais e desconhecidas.

6.3.3. Intervalo de confiança para a diferença entre as médias de duas populações independentes com variâncias diferentes e desconhecidas

Sejam duas populações **A** e **B** cujas médias são $\mu_{\mathbf{A}}$ e $\mu_{\mathbf{B}}$ e variâncias diferentes e desconhecidas, $\sigma^2_{\mathbf{A}}$ e $\sigma^2_{\mathbf{B}}$.

Com $\sigma^2_{\mathbf{A}}$ e $\sigma^2_{\mathbf{B}}$ diferentes e desconhecidas, devemos utilizar suas estimativas $s^2_{\mathbf{A}}$ e $s^2_{\mathbf{B}}$ individualmente e, nesse caso, a distribuição da estatística utilizada, apesar de continuar sendo a *t-Student*, não tem mais os graus de liberdade obtidos diretamente, como nos casos anteriores, isto é

$$\frac{(\bar{X}_{\mathbf{A}} - \bar{X}_{\mathbf{B}}) - (\mu_{\mathbf{A}} - \mu_{\mathbf{B}})}{\sqrt{\dfrac{s^2_{\mathbf{A}}}{n_{\mathbf{A}}} + \dfrac{s^2_{\mathbf{B}}}{n_{\mathbf{B}}}}} \sim t_{\nu} \,,$$

em que os graus de liberdade ν são dados por:

$$\nu = \frac{\left(\dfrac{s_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \dfrac{s_{\mathbf{B}}^2}{n_{\mathbf{B}}}\right)^2}{\dfrac{\left(s_{\mathbf{A}}^2/n_{\mathbf{A}}\right)^2}{(n_{\mathbf{A}}-1)} + \dfrac{\left(s_{\mathbf{B}}^2/n_{\mathbf{B}}\right)^2}{(n_{\mathbf{B}}-1)}}$$

Logo, um $I.C.\ (1-\alpha)\times 100\%$ para $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$, quando as *variâncias são diferentes e desconhecidas*, é dado por:

$$(\bar{x}_{\mathbf{A}} - \bar{x}_{\mathbf{B}}) \pm t_{\nu;\alpha/2}\sqrt{\frac{s_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \frac{s_{\mathbf{B}}^2}{n_{\mathbf{B}}}}\ .$$

Exemplo 6: Com os dados de resistências à compressão em concretos com cimentos das marcas *A* e *B*, considerando variâncias iguais e desconhecidas.

$$\bar{x}_{\mathbf{A}} = 33.76 \qquad \bar{x}_{\mathbf{B}} = 33.08$$

$$s_{\mathbf{A}} = 4.665 \qquad s_{\mathbf{B}} = 5.017$$

$$s_{\mathbf{A}}^2 = 21.759 \qquad s_{\mathbf{B}}^2 = 25.174$$

$$n_{\mathbf{A}} = 13 \qquad n_{\mathbf{B}} = 15$$

$$\nu = \frac{\left(\dfrac{21.759}{13} + \dfrac{25.174}{15}\right)^2}{\dfrac{(21.759/13)^2}{(13-1)} + \dfrac{(25.174/15)^2}{(15-1)}} = \frac{11.23614}{0.43464}$$

$$\nu = 25.86 \approx 26$$

Nota: Os graus de liberdade não precisam ser valores inteiros. De fato, $t_{25.86;0.025} = 2.056071$ (pelo R).

Enfim, um *I.C.* 95% para $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$ é dado por:

$$(\bar{x}_{\mathbf{A}} - \bar{x}_{\mathbf{B}}) \pm t_{26;0.025} \sqrt{\frac{s_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \frac{s_{\mathbf{B}}^2}{n_{\mathbf{B}}}}$$

$$(33.76 - 33.08) \pm 2.0555 \sqrt{\frac{21.759}{13} + \frac{25.174}{15}}$$

Portanto **(–3.084 , 4.444)** é um *I.C.* 95% para $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$ considerando variâncias diferentes e desconhecidas.

Resumindo:

Variâncias	**Estatística**	**I.C. 95% p/ $(\mu_{\mathbf{A}} - \mu_{\mathbf{B}})$**
Variâncias conhecidas	$\dfrac{(\bar{X}_{\mathbf{A}} - \bar{X}_{\mathbf{B}}) - (\mu_{\mathbf{A}} - \mu_{\mathbf{B}})}{\sqrt{\dfrac{\sigma_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \dfrac{\sigma_{\mathbf{B}}^2}{n_{\mathbf{B}}}}} \sim N(0,1)$	(–3.034 , 4.394)
Variâncias desconhecidas e iguais	$\dfrac{(\bar{X}_{\mathbf{A}} - \bar{X}_{\mathbf{B}}) - (\mu_{\mathbf{A}} - \mu_{\mathbf{B}})}{s_p \sqrt{\dfrac{1}{n_{\mathbf{A}}} + \dfrac{1}{n_{\mathbf{B}}}}} \sim t_{n_{\mathbf{A}} + n_{\mathbf{B}} - 2}$	(–3.105 , 4.465)
Variâncias desconhecidas e diferentes	$\dfrac{(\bar{X}_{\mathbf{A}} - \bar{X}_{\mathbf{B}}) - (\mu_{\mathbf{A}} - \mu_{\mathbf{B}})}{\sqrt{\dfrac{s_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \dfrac{s_{\mathbf{B}}^2}{n_{\mathbf{B}}}}} \sim t_{\nu}$	(–3.084 , 4.444)

6.3.4. Intervalo de confiança para a diferença entre duas proporções em populações independentes

Considere que se queira estimar a diferença entre duas proporções p_1 e p_2, associadas a duas populações independentes. Então, um estimador não viesado para a diferença $(p_1 - p_2)$ é dado por $(\hat{p}_1 - \hat{p}_2)$.

Sabendo que

$$\hat{p}_1 \sim N\left[p_1, \frac{p_1(1-p_1)}{n_1}\right] \quad \text{e} \quad \hat{p}_2 \sim N\left[p_2, \frac{p_2(1-p_2)}{n_2}\right]$$

Então:

$$(\hat{p}_1 - \hat{p}_2) \sim N\left[(p_1 - p_2), \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right]$$

Desta forma, um $I.C.\ (1-\alpha)\times 100\%$ para $(p_1 - p_2)$ é dado por

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Exemplo 7: Um grupo de biólogos interessados em estudar populações de animais em regiões isoladas por longas distâncias estão avaliando o desenvolvimento de peixes de uma determinada espécie em duas lagoas separadas por uma grande distância geográfica. Numa amostra de 116 peixes da primeira lagoa, 84 são da espécie em questão, enquanto que, de uma amostra de 80 peixes da outra lagoa, 45 são da espécie estudada.

Estimar a diferença entre as proporções de peixes das duas lagoas e construir um *I.C.* 90% para a diferença.

As estimativas individuais para p_1 e p_2 são:

$$\hat{p}_1 = \frac{84}{116} = 0.724 \qquad \hat{p}_2 = \frac{46}{80} = 0.575$$

Então, uma estimativa para a diferença entre p_1 e p_2 é dada por

$$\hat{p}_1 - \hat{p}_2 = 0.724 - 0.575 = 0.149$$

e, a estimativa do desvio padrão da diferença

$$\sqrt{\frac{0.724 \times 0.276}{116} + \frac{0.575 \times 0.425}{80}} = \sqrt{0.04777}.$$

Logo, um *I.C. 90%* para a *diferença entre as proporções* é dado por

$$0.149 \pm 1.645\sqrt{0.04777},$$

ou seja, (0.0353 , 0.2627) é o *I.C. 90%* para $(p_1 - p_2)$.

O que se pode concluir?

6.4. Intervalo de confiança para a variância de uma população normal

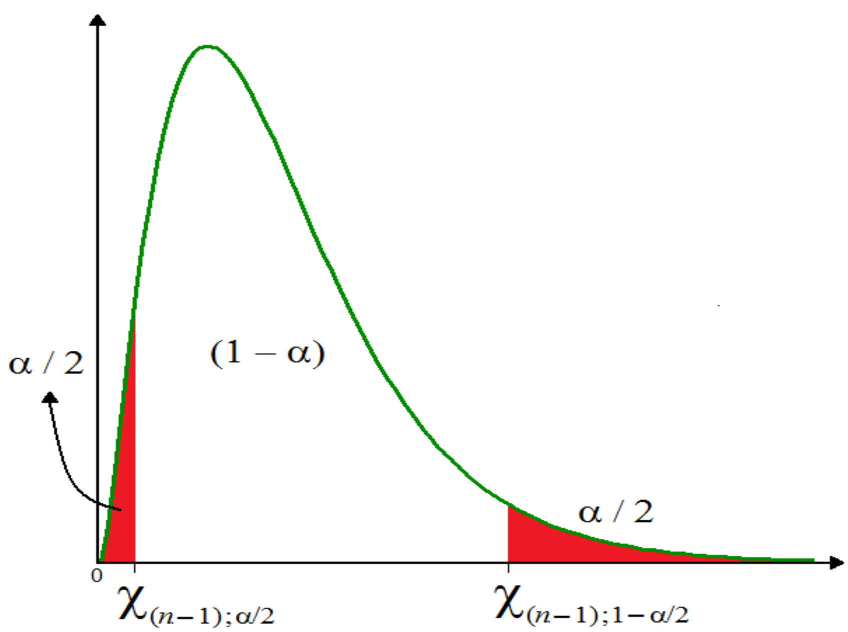
Considere uma população normal com média μ e variância σ^2, ambas desconhecidas. Em muitas aplicações práticas temos o interesse em avaliar a variabilidade dos fenômenos em estudo. Nessa situação, devemos estimar e, também, construir intervalos de confiança para a variância populacional.

Considerando que a população seja normal, temos que

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

Desta forma, a partir da distribuição χ^2_{n-1} podemos construir *I.C.'s* para σ^2 a partir de seus quantis:

$$P\left(\chi^2_{(n-1);\alpha/2} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{(n-1);1-\alpha/2}\right) = (1-\alpha)$$



$$P\left(\frac{\chi^2_{(n-1);\alpha/2}}{(n-1)s^2} \le \frac{1}{\sigma^2} \le \frac{\chi^2_{(n-1);1-\alpha/2}}{(n-1)s^2}\right) = (1-\alpha)$$

$$P\left(\underbrace{\frac{(n-1)s^2}{\chi^2_{(n-1);1-\alpha/2}}}_{a} \le \sigma^2 \le \underbrace{\frac{(n-1)s^2}{\chi^2_{(n-1);\alpha/2}}}_{b}\right) = (1-\alpha)$$

$$a = \frac{(n-1)s^2}{\chi^2_{(n-1);1-\alpha/2}} \quad \text{e} \quad b = \frac{(n-1)s^2}{\chi^2_{(n-1);\alpha/2}}.$$

Desta forma, um $I.C.\ (1-\alpha)\times 100\%$ para σ^2 é dado por:

$$\left[\frac{(n-1)s^2}{\chi^2_{(n-1);1-\alpha/2}} \ ; \ \frac{(n-1)s^2}{\chi^2_{(n-1);\alpha/2}}\right].$$

Exemplo 8: O peso de um componente mecânico é uma *va* com distribuição normal com média μ e variância σ^2, desconhecidos. Pretende-se estudar a variabilidade do processo de produção e, para isso, uma amostra com $n = 11$ componentes foi avaliada. Os pesos (g) são dados

98 97 102 100 98 101 102 105 95 102 100

$\sum x = 1100$ e $\sum x^2 = 110080$.

Portanto: $\bar{x} = \dfrac{1100}{11} = 100g$ $\qquad s^2 = \dfrac{110080 - 11(100)^2}{10} = 8g^2.$

Construir um $I.C.$ 95% para a variância populacional ($\alpha = 0.05$).

$\chi^2_{10;0.025} = 3.25$ e $\chi^2_{10;0.975} = 20.48$

$$a = \frac{(n-1)s^2}{\chi^2_{(n-1);1-\alpha/2}} = \frac{10\times 8}{20.48} = 3.906$$

$$b = \frac{(n-1)s^2}{\chi^2_{(n-1);\alpha/2}} = \frac{10\times 8}{3.25} = 24.615$$

Um $I.C.$ 95% para σ^2 é dado por (3.906 , 24.615).

6.5. Intervalo de confiança para a razão entre as variâncias de duas populações normais

É muito comum, em aplicações estatísticas, precisarmos comparar as variâncias de duas populações, como, por exemplo, quando comparamos a média dessas populações.

A comparação de duas variâncias não é feita pela diferença entre elas, mas sim pela *razão das mesmas*.

Resultado:

Seja $W_1 \sim \chi^2_{k_1}$ e $W_2 \sim \chi^2_{k_2}$, prova-se facilmente que a razão

$$F = \frac{W_1 / k_1}{W_2 / k_2} \sim F_{k_1;k_2}$$

A razão de duas *va* independentes, com distribuição quiquadrado, divididas pelos seus respectivos graus de liberdade (k_1 e k_2), tem *distribuição F de Snedecor*, em que k_1 são os graus de liberdade do numerados e k_2 os graus de liberdade do denominador.

Notas:

i) Se $X \sim t_k$, então $X^2 \sim F_{1,k}$.

Prova: Sai direto do resultado (1) da distribuição *t-Student*.

ii) Existe uma relação entre os quantis das distribuições F, de forma que

$$F_{k_1;k_2;\alpha} = \frac{1}{F_{k_2;k_1;1-\alpha}}$$

Sejam duas populações normais com variâncias σ_1^2 e σ_2^2 e sejam s_1^2 e s_2^2 seus estimadores a partir de amostras de tamanho n_1 e n_2, então

$$F = \frac{\dfrac{(n_1 - 1)s_1^2/\sigma_1^2}{(n_1 - 1)}}{\dfrac{(n_2 - 1)s_2^2/\sigma_2^2}{(n_2 - 1)}} \sim F_{n_1 - 1\,;\,n_2 - 1}$$

Mas a razão F acima pode ser simplificada por:

$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} = \frac{\sigma_2^2\ s_1^2}{\sigma_1^2\ s_2^2} \sim F_{n_1 - 1\,;\,n_2 - 1}$$

Logo, um $I.C.$ para razão entre duas variâncias é construído a partir de:

$$P\left(f_1 \le \frac{\sigma_2^2\ s_1^2}{\sigma_1^2\ s_2^2} \le f_2 \right) = (1 - \alpha)$$

em que: $f_1 = F_{(n_1-1);(n_2-1);\alpha/2}$ e $f_2 = F_{(n_1-1);(n_2-1);1-\alpha/2}$,

$$P\left(f_1 \frac{s_2^2}{s_1^2} \le \frac{\sigma_2^2}{\sigma_1^2} \le f_2 \frac{s_2^2}{s_1^2} \right) = (1 - \alpha)$$

Portanto, escrevendo o resultado para $\frac{\sigma_1^2}{\sigma_2^2}$, um $I.C.$ $(1-\alpha)\times 100\%$ para a razão de variâncias é dado por:

$$P\left(\frac{s_1^2}{f_2\, s_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{f_1\, s_2^2}\right) = (1-\alpha)$$

Ou seja, o intervalo para a *razão entre duas variâncias de populações normais* é definido por:

$$\left[\frac{s_1^2}{s_2^2\, F_{(n_1-1);(n_2-1);1-\alpha/2}} \;;\; \frac{s_1^2}{s_2^2\, F_{(n_1-1);(n_2-1);\alpha/2}}\right].$$

Nota: O intervalo é construído de forma que $\frac{s_1^2}{s_2^2}$ seja maior do que 1.

Exemplo 9: Construir um $I.C.$ 95% para a razão entre as variâncias da resistência à compressão em concretos dos cimentos das marcas *A* e *B*.

$s_{\mathbf{A}}^2 = 21.759$ $\quad$ $s_{\mathbf{B}}^2 = 25.174$

$n_{\mathbf{A}} = 13$ $\quad$ $n_{\mathbf{B}} = 15$

Com $\frac{s_{\mathbf{B}}^2}{s_{\mathbf{A}}^2} > 1$, $\Rightarrow$ $F_{14;12;0.025} = 0.3279$ e $F_{14;12;0.975} = 3.2062$.

$$\frac{s_{\mathbf{B}}^2}{s_{\mathbf{A}}^2\, F_{14;12;0.975}} = \frac{25.174}{21.759 \times 3.2062} = 0.3608$$

$$\frac{s_{\mathbf{B}}^2}{s_{\mathbf{A}}^2\, F_{14;12;0.025}} = \frac{25.174}{21.759 \times 0.3279} = 3.5284$$

Assim, um $I.C.$ 95% para $\dfrac{\sigma_{\mathbf{B}}^2}{\sigma_{\mathbf{A}}^2}$ é dado por (0.3608 , 3.5284).

6.6. Exercícios

1. Deseja-se comparar a qualidade de um produto produzido por duas indústrias. Essa qualidade será definida pela uniformidade com que o produto é produzido. Tomaram-se duas amostras, uma de cada indústria, medindo-se o tamanho dos produtos (cm).

a) A qualidade das duas fábricas é a mesma? Caso a sua resposta seja negativa, dê um intervalo de confiança para indicar a intensidade dessa desigualdade.

b) Construir um *I.C.* 99% para a diferença entre as médias, $(\mu_A - \mu_B)$

Estatísticas	Indústria ***A***	Indústria ***B***
Tamanho da Amostra	21	17
Médias	21.15	21.12
Variâncias	0.0412	0.1734

2. Num estudo comparativo do tempo médio de adaptação dos empregados de um grande complexo bancário, uma amostra aleatória, de 50 homens e 50 mulheres, produziu os seguintes resultados:

Estatísticas	Homens	Mulheres
Tamanho da Amostra	50	50
Médias	3.2 anos	3.7 anos
Desvios-padrões	0.8 anos	0.9 anos

Que conclusões você pode tirar para a população de homens e mulheres desse banco? (Indique quais as suposições feitas)

3. Suponha que uma associação de defesa de consumidores deseja estimar o consumo médio um novo modelo de automóvel que será lançado no mercado. Para fazer esta verificação, a associação observa uma amostra de 10 veículos, conduzidos por motoristas treinados, num percurso de 100 milhas. O consumo, em galões, foi registrado com os seguintes resultados:

$$\sum x = 43.28 \quad \text{e} \quad \sum x^2 = 188.4886$$

Assumindo que estes valores representam uma amostra aleatória de uma variável normalmente distribuída com média μ e variância σ^2.

a) Calcule estimativas pontuais para μ e σ^2.

b) Calcule um intervalo de 75 % de confiança para σ^2.

4. Os dados abaixo são uma amostra aleatória para estimar a proporção estudantes de uma universidade que possuem automóvel.
Foi construído o intervalo conservador de 90% de confiança para p:

(0.5555 ; 0.8845)

Um segundo intervalo foi construído considerando a normalidade de $\hat{p}$:

(0.4887 ; 0.9513)

a) Qual é a estimativa pontual para $\hat{p}$?
b) Qual é o tamanho da amostra?
c) Qual o nível de confiança do segundo intervalo

5. Da população $X \sim$ *Normal*(50*;* 100) retirou-se uma *aa* de $n = 10$ elementos e da população $Y \sim$ *Normal*(60*;* 100) retirou-se uma *aa* de $m = 6$ elementos, independente da primeira, obtendo-se as variâncias amostrais s_1^2 e s_2^2, respectivamente.

a) Encontre o valor de *a*, tal que $P\left(s_1^2 / s_2^2 < a\right) = 0.95$

b) Encontre o valor de *b*, tal que $P\left(s_1^2 / s_2^2 > b\right) = 0.95$

6. Uma das maneiras de medir o grau de satisfação dos empregados de uma mesma categoria quanto à política salarial é por meio do desvio padrão de seus salários. A Fábrica *A* diz ser mais coerente na política salarial do que a Fábrica *B*. Para verificar essa afirmação, sorteou-se uma amostra de 10 funcionários não especializados de *A*, e 15 de *B*, obtendo-se os desvios padrões $s_A = 1000$ reais e $s_B = 1600$ reais. Qual seria a sua conclusão?

Resolução:

1. Deseja-se comparar a qualidade de um produto produzido por duas indústrias. Essa qualidade será definida pela uniformidade com que o produto é produzido. Tomaram-se duas amostras, uma de cada indústria, medindo-se o tamanho dos produtos (cm).

a) A qualidade das duas fábricas é a mesma? Caso a sua resposta seja negativa, dê um intervalo de confiança para indicar a intensidade dessa desigualdade.

$$n_A = 21 \qquad n_B = 17$$
$$\bar{x}_A = 21.15 \qquad \bar{x}_B = 21.12$$
$$s_A^2 = 0.0412 \qquad s_B^2 = 0.1734$$

I.C. 95% para $\dfrac{\sigma_B^2}{\sigma_A^2}$:

Limite inferior: $$\frac{0.1734}{0.0412\ F_{16;20;0.975}} = \frac{0.1734}{0.0412 \times 2.547} = 1.652$$

Como $$F_{16;20;0.025} = \frac{1}{F_{20;16;0.975}} = \frac{1}{2.68} = 0.3731$$

Limite superior: $$\frac{0.1734}{0.0412\ F_{16;20;0.025}} = \frac{0.1734}{0.0412 \times 0.3731} = 11.280$$

O intervalo (1.652 ; 11.283) é um intervalo de confiança 95% para $\dfrac{\sigma_B^2}{\sigma_A^2}$.

Como o intervalo não engloba o valor 1, então, *há evidência suficiente para afirmar* que $\sigma_A^2 \neq \sigma_B^2$.

Logo, a qualidade das duas indústrias não é a mesma. A indústria ***A***, com menor variabilidade, tem melhor qualidade.

b) *I.C.* 99% para $(\mu_A - \mu_B)$ considerando variâncias diferentes.

$$\nu = \frac{\left(\dfrac{s_{\mathbf{A}}^2}{n_{\mathbf{A}}} + \dfrac{s_{\mathbf{B}}^2}{n_{\mathbf{B}}}\right)^2}{\dfrac{\left(s_{\mathbf{A}}^2 / n_{\mathbf{A}}\right)^2}{\left(n_{\mathbf{A}} - 1\right)} + \dfrac{\left(s_{\mathbf{B}}^2 / n_{\mathbf{B}}\right)^2}{\left(n_{\mathbf{B}} - 1\right)}}$$

$$\nu = \frac{\left(\dfrac{0.0412}{21} + \dfrac{0.1734}{17}\right)^2}{\dfrac{(0.0412/21)^2}{20} + \dfrac{(0.1734/17)^2}{16}} = 22.1 \approx 22\ gl$$

$$\alpha/2 = 0.005 \quad \Rightarrow \quad t_{22;0.005} = 2.8188$$

$$(\bar{x}_A - \bar{x}_B) \pm t_{22;0.005}\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

$$0.03 \pm 2.8188\sqrt{\frac{0.0412}{21} + \frac{0.1734}{17}}$$

(–0.281 ; 0.341) é o *I.C.* 99% para a diferença entre as médias de tamanhos dos produtos das indústrias ***A*** e ***B***.

2. Num estudo comparativo do tempo médio de adaptação dos empregados de um grande complexo bancário, uma amostra aleatória de 50 homens e 50 mulheres produziu os seguintes resultados:

A qualidade das duas fábricas é a mesma? (*comparar as variâncias*)

$n_H = 50$	$n_M = 50$
$\bar{x}_H = 3.2\ anos$	$\bar{x}_M = 3.7\ anos$
$s_H = 0.8\ anos$	$s_M = 0.9\ anos$

Que conclusões você pode tirar com relação ao tempo de adaptação de homens e mulheres desse banco? (Indique quais suposições foram feitas)

I.C. 95% para $\dfrac{\sigma_M^2}{\sigma_H^2}$:

Limite inferior: $\dfrac{(0.9)^2}{(0.8)^2\ F_{49;49;0.975}} = \dfrac{0.81}{0.64 \times 1.7622} = 0.7182$

Limite superior: $\dfrac{(0.9)^2}{(0.8)^2\ F_{49;49;0.025}} = \dfrac{0.81}{0.64 \times 0.5675} = 2.2302$

O intervalo (0.7182 ; 2.2302) é um intervalo 95% para $\dfrac{\sigma_M^2}{\sigma_H^2}$.

Como o intervalo engloba o valor 1, então, *não há evidência suficiente para afirmar que as variâncias são diferentes*.

I.C. 90% para a diferença entre os tempos médios de adaptação entre homens e mulheres, com variâncias iguais.

$$s_p^2 = \frac{(n_M - 1)s_M^2 + (n_H - 1)s_H^2}{n_M + n_H - 2} = \frac{49 \times 0.64 + 49 \times 0.81}{98} = 0.725$$

$$\alpha / 2 = 0.05 \quad \Rightarrow \quad t_{98;0.05} = 1.6606$$

$$(\bar{x}_M - \bar{x}_H) \pm t_{98;0.05} \sqrt{\frac{s_p^2}{n_M} + \frac{s_p^2}{n_H}}$$

$$(3.7 - 3.2) \pm 1.6606 \sqrt{\frac{0.725}{50} + \frac{0.725}{50}}$$

(0.2172 ; 0.7828) é o *I.C.* 90% para a diferença entre os tempos médios de adaptação de entre mulheres e homens.

O intervalo não engloba o zero, portanto, *há evidências suficientes* para afirmar que os homens têm um tempo de adaptação menor do que as mulheres.

Suposições: Normalidade dos tempos de adaptação de homens e mulheres

3. Uma associação de defesa de consumidores deseja estimar o consumo médio um novo modelo de automóvel que será lançado no mercado. Para fazer esta verificação, a associação observa uma amostra de 10 veículos, conduzidos por motoristas treinados, num percurso de 100 milhas. O consumo, em galões, foi registrado com os seguintes resultados:

$$\sum x = 43.28 \text{ e } \sum x^2 = 188.4886$$

Assumindo que estes valores representam uma amostra aleatória de uma variável normalmente distribuída com média μ e variância σ^2.

a) Calcule estimativas pontuais para μ e σ^2.

$$\hat{\mu} = \bar{x} = \frac{43.28}{10} = 4.328$$

$$\hat{\sigma}^2 = s^2 = \frac{188.4886 - 10(4.328)^2}{(10-1)} = \frac{1.17276}{9} = 0.13031$$

b) Calcule um intervalo de 75% de confiança para σ^2. $(\alpha = 0.25)$

$$\chi_{9;0.125} = 4.507 \quad \text{e} \quad \chi_{9;0.875} = 13.926$$

Limite inferior: $\dfrac{(n-1)s^2}{\chi_{9;0.875}} = \dfrac{1.17276}{13.926} = 0.0842$

Limite superior: $\dfrac{(n-1)s^2}{\chi_{9;0.125}} = \dfrac{1.17276}{4.507} = 0.2602$

I.C. 75% para a variância σ^2 é dado por: (0,0842 ; 0,2602)

4. Os dados abaixo são uma amostra aleatória para estimar a proporção estudantes de uma universidade que possuem automóvel.

Intervalo conservador de 90% de confiança para p: (0.5555 ; 0.8845)

Intervalo considerando a normalidade de $\hat{p}$: (0.4887 ; 0.9513)

a) E pontual para $\hat{p}$? *(ponto médio dos intervalos)*

$$\hat{p} = \frac{0.5555 + 0.8845}{2} = \frac{0.4887 + 0.9513}{2} = 0.72$$

b) Qual é o tamanho da amostra?

Sabe-se que o tamanho do *I.C.* 90% conservador é dado por:

$$Z_{0.05}\sqrt{\frac{1}{4n}} = 0.5555 - 0.72$$

$$\frac{1.645}{\sqrt{4n}} = 0.1645$$

$$\sqrt{4n} = 10a \qquad \Rightarrow \qquad n = 25$$

c) Qual o nível de confiança do segundo intervalo

$$0.72 + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{25}} = 0.9513$$

$$Z_{\alpha/2} = \frac{5 \times (0.9513 - 0.72)}{\sqrt{0.72 \times 0.28}} = 2.576$$

O nível de confiança do *I.C.* é $(1-\alpha) = 0.99$ ou 99%.

5. Da população $X \sim$ *Normal*(50;100) retirou-se uma *aa* de $n = 10$ elementos e da população $Y \sim$ *Normal*(60;100) retirou-se uma *aa* de $m = 6$ elementos, independente da primeira, obtendo-se as variâncias amostrais s_1^2 e s_2^2, respectivamente.

a) Encontre o valor de *a*, tal que $P\left(s_1^2 / s_2^2 < a\right) = 0.95$

Obs: $\sigma_1^2 = \sigma_2^2 = 100$

$$\frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \sim F_{n-1;m-1} \qquad \Rightarrow \qquad \frac{s_1^2}{s_2^2} \sim F_{9;5}$$

$$P\left(s_1^2 / s_2^2 < a\right) = 0.95 \qquad \Rightarrow \qquad a = F_{9;5;0.95} = 4.772$$

b) Encontre o valor de *b*, tal que $P\left(s_1^2 / s_2^2 > b\right) = 0.95$

Da relação entre as distribuições *F*'s

$$F_{9;5;0.05} = \frac{1}{F_{5;9;0.95}} = \frac{1}{3.482} = 0.2872$$

$$P\left(s_1^2 / s_2^2 > b\right) = 0.95 \qquad \Rightarrow \qquad b = F_{9;5;0.05} = 0.2872$$

6. Uma das maneiras de medir o grau de satisfação dos empregados de uma mesma categoria quanto à política salarial é por meio do desvio padrão de seus salários. A Fábrica *A* diz ser mais coerente na política salarial do que a Fábrica *B*. Para verificar essa afirmação, sorteou-se uma amostra de 10 funcionários não especializados de *A*, e 15 de *B*, obtendo-se os desvios padrões $s_A = 1000$ reais e $s_B = 1600$ reais. Qual seria a sua conclusão?

O grau de satisfação com o salário é o mesmo nas duas fábricas?
(comparar as variâncias)

$$n_A = 10 \qquad\qquad n_B = 15$$
$$s_A = 1000 \qquad\qquad s_B = 1600$$
$$s_A^2 = 1\times 10^6 \qquad\qquad s_B^2 = 2.56\times 10^6$$

Construir um *I.C.* 95% para $\dfrac{\sigma_B^2}{\sigma_A^2}$ e verificar se engloba o valor 1:

Limite inferior: $\dfrac{s_B^2}{s_A^2\ F_{14;9;0.975}} = \dfrac{2.56\times 10^6}{1\times 10^6\times 3.7980} = 0.6740$

Como $F_{14;9;0.025} = \dfrac{1}{F_{9;14;0.975}} = \dfrac{1}{3.209} = 0.3116$

Limite superior: $\dfrac{s_B^2}{s_A^2\ F_{14;9;0.025}} = \dfrac{2.56\times 10^6}{1\times 10^6\times 0.3116} = 8.2157$

I.C. 95% para $\dfrac{\sigma_B^2}{\sigma_A^2}$: (0.6740 ; 8.2157)

O *I.C.* engloba o valor 1, portanto, *não há evidência* para sustentar a afirmação da Fábrica *A*.