# Capítulo 7 - Estimação por intervalos

## 7.1 Intervalos de confiança

Para além duma estimativa pontual de um parâmetro é, em muitas situações, importante dispôr de alguma forma de intervalo que indique a confiança que se pode depositar na estimativa pontual.

Um intervalo de confiança (I.C.) para um parâmetro desconhecido $\theta$ é do tipo

$$l \leq \theta \leq u,$$

onde $l$ e $u$ dependem do valor observado $\hat{\theta}$ (estimativa pontual), e da distribuição por amostragem da estatística $\hat{\Theta}$, usada para estimar $\theta$.





Nota: o I. C. $l \leq \theta \leq u$ é chamado bilateral. Em alguns casos pode ser mais importante um intervalo unilateral (superior ou inferior) do tipo

$l \leq \theta$, intervalo inferior

$\theta \leq u$ intervalo superior

## 7.2 Intervalo de confiança para a média, variância conhecida

$X$ população tal que:

$E(X) = \mu$ (desconhecido)

$V(X) = \sigma^2$ (conhecido)

$(X_1, \ldots, X_n)$ a. a. de dimensão $n$

$\hat{\mu} = \overline{X}$ (estimador pontual de $\mu$)



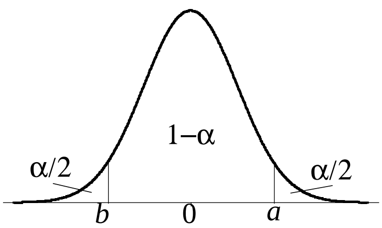


Sabemos já que

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \left\langle \begin{array}{ll} \sim & N(0,1) \text{ se } X \sim N(\mu, \sigma^2) \\ \underset{a}{\sim} & N(0,1) \text{ se } X \text{ qq e } n \text{ grande} \end{array} \right.$$

onde o caso $X$ qq é justificado pelo T. L. C.

Distribuição de $Z$:




Podemos determinar pontos, $a$ e $b$, tais que

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

onde $\alpha \approx 0$ e $\gamma \approx 1$.





Para minimizar $a - b$ deve ter-se $b = -a$.

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

$$= P\left( -a \frac{\sigma}{\sqrt{n}} \leq \overline{X} - \mu \leq a \frac{\sigma}{\sqrt{n}} \right) =$$

$$= P\left( -\overline{X} - a \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\overline{X} + a \frac{\sigma}{\sqrt{n}} \right) =$$

$$= P\left( \overline{X} - a \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + a \frac{\sigma}{\sqrt{n}} \right) =$$

$$= P\left( \mu \in \left[ \overline{X} - a \frac{\sigma}{\sqrt{n}} ; \overline{X} + a \frac{\sigma}{\sqrt{n}} \right] \right) = 1 - \alpha$$

com $a$: $P(Z > a) = \frac{\alpha}{2}$.

**Nota:** a $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ chama-se variável aleatória fulcral (deve ter distribuição conhecida e depender apenas do parâmetro desconhecido).

$\left[\overline{X} - a\frac{\sigma}{\sqrt{n}};\overline{X} + a\frac{\sigma}{\sqrt{n}}\right]$ é um intervalo aleatório, quando substituímos $\overline{X}$ por $\bar{x}$ (valor observado da média duma amostra aleatória) passamos a ter um intervalo concreto chamado intervalo de confiança.

**Definição** (Intervalo de confiança para a média com variância conhecida): Se $\bar{x}$ for a média observada duma amostra aleatória de dimensão $n$ duma população normal (ou duma população qualquer desde que $n$ grande, mas nesse caso o intervalo é apenas aproximado) com variância conhecida $\sigma^2$, um intervalo de confiança a $100 \times (1-\alpha)\%$ para $\mu$ é dado por

$$\bar{x} - a\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + a\frac{\sigma}{\sqrt{n}} \quad \text{com } a\text{: } P(Z > a) = \frac{\alpha}{2}$$

(ou $a = z_{\alpha/2}$, na notação do Montgomery).





**Interpretação:**

- Não podemos dizer que $\mu$ pertence ao intervalo de confiança com probabilidade $1-\alpha$;
- o que podemos dizer é que se fizermos um grande número de intervalos nestas condições, aproximadamente $100 \times (1-\alpha)\%$ desses intervalos conterão de facto o verdadeiro valor de $\mu$ (que permanece desconhecido), é esta ideia que é traduzida por "confiança".

**Notas:**

**1)** O comprimento do intervalo de confiança está associado à precisão, quanto menor for o comprimento mais precisa é a média.





**2)** Se diminuímos $\alpha$, isto é, aumentarmos $1-\alpha$ (grau de confiança), mantendo $n$ fixo, $a$ vai aumentar e consequentemente o comprimento do intervalo. Não é possível fazer $\alpha = 0$ pois nesse caso $a = +\infty$.

**3)** Quando aumentamos $n$, mantendo $\alpha$ fixo, diminui o comprimento do intervalo.

**4)** Qual deve ser a dimensão da amostra para que se tenha um dado erro máximo (com confiança fixa)?
Para termos uma confiança de $100 \times (1-\alpha)\%$ em que $|\bar{x} - \mu| \le E$ deve ter-se

$$n = \left(\frac{a\sigma}{E}\right)^2, \quad \left(\text{visto que } \frac{a\sigma}{\sqrt{n}} = E\right).$$

(quando $n$ não der inteiro arredonda-se por excesso).





**Exemplo:** $X \sim N(\mu, \sigma^2)$ com $\sigma = 1$.

**a)** uma amostra aleatória de dimensão 10 conduziu a $\bar{x} = 10.1$. Calcular um intervalo de confiança a 95% para $\mu$:

$$\alpha = 0.05 \Leftrightarrow P(Z > a) = 0.025 \Leftrightarrow$$
$$\Leftrightarrow P(Z < a) = 0.975 \Leftrightarrow a = \Phi^{-1}(0.975) = 1.96$$

Logo o intervalo pedido é

$$\left[10.1 - 1.96 \times \frac{1}{\sqrt{10}};10.1 + 1.96 \times \frac{1}{\sqrt{10}}\right] = [9.48;10.72]$$

**b)** Qual deve ser $n$ para que se possa garantir com 95% de confiança que $|\bar{x} - \mu| \le 0.25$?

$$n = \left(\frac{1.96 \times 1}{0.25}\right)^2 = 61.47 \qquad \text{Resposta: } n = 62.$$

**7.3 Intervalo de confiança para a diferença de duas médias, variâncias conhecidas**

$X_1$, população 1, com $E(X_1)=\mu_1$ e $V(X_1)=\sigma_1^2$ (conhecida)

$X_2$, população 2, com $E(X_2)=\mu_2$ e $V(X_2)=\sigma_2^2$ (conhecida)

( $X_1$ e $X_2$ independentes)

a. a. da população 1 $\left(X_{11}, X_{12}, \ldots, X_{1n_1}\right)$ com média

$$\overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_{1i}}{n_1}$$

a. a. da população 2 $\left(X_{21}, X_{22}, \ldots, X_{2n_2}\right)$ com média

$$\overline{X}_2 = \frac{\sum_{i=1}^{n_2} X_{2i}}{n_2}$$





(e a a.a. $\left(X_{11}, X_{12}, \ldots, X_{1n_1}\right)$ é independente da a.a. $\left(X_{21}, X_{22}, \ldots, X_{2n_2}\right)$ )

Como construir um I.C. a $100\times(1-\alpha)\%$ para $\mu_1-\mu_2$ ?

**Nota:** um intervalo deste tipo é útil para comparar duas experiências ou dois métodos.

O estimador pontual de $\mu_1-\mu_2$ é $\overline{X}_1-\overline{X}_2$.

Já sabemos que se $X_1 \sim N(\mu_1,\sigma_1^2)$ e se $X_2 \sim N(\mu_2,\sigma_2^2)$ então

$$\overline{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ e } \overline{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right) \text{ pelo que}$$

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





logo a v.a. fulcral é:

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

**Notas:** a distribuição anterior também é válida, aproximadamente, para $X_1$ e $X_2$ com outra qualquer distribuição, não necessariamente normal, desde que $n_1$ e $n_2$ sejam elevados (pelo T.L.C.)

Fazendo, do mesmo modo que na secção anterior,

$$P(-a\le Z\le a)=1-\alpha, \quad \text{com } P(Z>a)=\frac{\alpha}{2}$$

obtém-se um I.C. a $100\times(1-\alpha)\%$ para $\mu_1-\mu_2$:

$$\left[\bar{x}_1-\bar{x}_2-a\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}; \bar{x}_1-\bar{x}_2+a\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$





**Notas:** Mantêm-se as notas 1) 2) e 3) da secção 7.2, relativas à interpretação e variação do comprimento do intervalo.

**4)** Dimensão da amostra tal que

$\left|\left(\overline{X}_1-\overline{X}_2\right)-\left(\mu_1-\mu_2\right)\right|\le E$ com $100\times(1-\alpha)\%$ de confiança ?

Não tem solução única para $n_1$ e $n_2$ gerais, mas se quisermos $n_1=n_2=n$ obtém-se

$$n=\left(\frac{a}{E}\right)^2\left(\sigma_1^2+\sigma_2^2\right)$$

### 7.4 Intervalo de confiança para a média de uma população normal, variância desconhecida

$X$ população tal que:

$E(X) = \mu$ (desconhecido)

$V(X) = \sigma^2$ (desconhecida)

$(X_1, \ldots, X_n)$ a. a. de dimensão $n$

$\hat{\mu} = \overline{X}$ (estimador pontual de $\mu$)

Não se pode usar $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ porque $\sigma$ é desconhecido. Um procedimento lógico consiste em substituir $\sigma$ por $S$ (desvio padrão amostral), ou seja, em considerar a v. a. fulcral $\dfrac{\overline{X} - \mu}{S / \sqrt{n}}$.



©Ana Pires, IST, Outubro de 2000

Mas qual será o efeito de fazer isto? Não é a mesma v. a.! Qual a sua distribuição?

Se $n$ for grande ($n > 30$, em geral) pode mostrar-se que o efeito é pequeno e tem-se

$$Z = \frac{\overline{X} - \mu}{S / \sqrt{n}} \underset{a}{\sim} N(0,1),$$

quer para $X \sim N(\mu, \sigma^2)$, quer para $X$ qualquer (com $E(X) = \mu$ e $V(X) = \sigma^2$). Ou seja, o I.C. calcula-se exactamente como na Secção 7.2 substituindo $\sigma$ por $s$ (desvio padrão amostral observado).

Se $n \leq 30$ o problema não é solúvel no caso geral (isto é, desconhecendo o tipo de distribuição da população). Se $X \sim N(\mu, \sigma^2)$ o teorema seguinte fornece o resultado que se pretende.



©Ana Pires, IST, Outubro de 2000

**Teorema:** Seja $(X_1, \ldots, X_n)$ uma a.a. duma população $X \sim N(\mu, \sigma^2)$. A variável aleatória

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

tem distribuição $t$ com $n-1$ graus de liberdade ($T \sim t_{n-1}$).

**Notas:**

**1)** Uma v.a. com distribuição $t$ tem função de densidade de probabilidade

$$f(x) = \frac{\Gamma[(k+1)/2]}{\sqrt{\pi k}\,\Gamma(k/2)} \cdot \frac{1}{\left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}}$$

onde $k > 0$ é o número de graus de liberdade. Pode mostrar-se que se $T \sim t_k$ então

$$E(T) = 0 \quad (k > 1) \text{ e } V(T) = k/(k-2) \quad (k > 2)$$



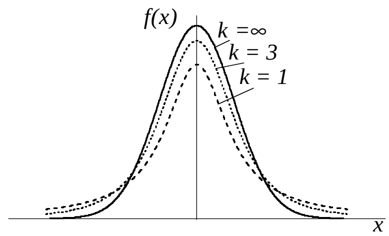
©Ana Pires, IST, Outubro de 2000

**2)** Função Gama

$$\Gamma(r) = \int_0^{+\infty} x^{r-1} e^{-x} dx \,, \quad r > 0$$

se $r$ inteiro $\Gamma(r) = (r-1)!$. Esta função pode ser vista como uma generalização do factorial (definido só para inteiros) aos reais positivos.

**3)** Representação gráfica

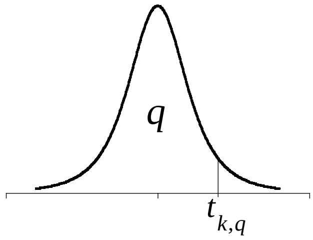


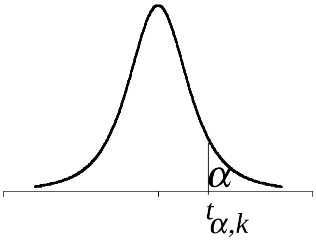

**4)** É fácil ver que quando $k \to +\infty$

$$f(x) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (\text{f.d.p. de } N(0,1))$$

**5)** Os percentis da distribuição $t_k$ encontram-se tabelados

Tabelas da disciplina: No livro:




$$P\left(T < t_{k,q}\right) = q \qquad P\left(T > t_{\alpha,k}\right) = \alpha$$

$$t_{k,1-q} = -t_{k,q} \qquad t_{1-\alpha,k} = -t_{\alpha,k}$$

---

Voltando à construção do I.C. e procedendo da forma habitual:

$$P\left(-a \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq a\right) = 1-\alpha \Leftrightarrow$$





$$\Leftrightarrow P\left(\mu \in \left[\overline{X} - a\frac{S}{\sqrt{n}} ; \overline{X} + a\frac{S}{\sqrt{n}}\right]\right) = 1-\alpha$$

com $a$: $P(T > a) = \dfrac{\alpha}{2}$.

Obtém-se então o I.C. para a média de uma população normal com variância desconhecida

$$\bar{x} - a\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + a\frac{s}{\sqrt{n}}$$

**Observações:**

**1)** A interpretação é semelhante à que foi feita em 7.2.

**2)** Se $\alpha$ diminuir, com $n$ fixo, aumenta o comprimento do intervalo.





**3)** Se $n$ aumentar, com $\alpha$ fixo, espera-se que diminua o comprimento do intervalo, mas não há certeza, pois $s$ varia de amostra para amostra.

**4)** Determinação de $n$ para um dado erro (com $\alpha$ fixo):

$$|\bar{x}-\mu| \leq E \Leftrightarrow n = \left(\frac{as}{E}\right)^2$$

Dificuldades:

*i)* $a$ também depende de $n$; Solução: resolução por tentativa-erro.

*ii)* $s$ é desconhecido antes de se ter a amostra; Solução: obter uma amostra preliminar para ter uma ideia do valor que $s$ pode vir a ter.





**Exemplo:** Considere-se uma v.a. $X \sim N\left(\mu, \sigma^2\right)$. Uma amostra aleatória de dimensão 10 conduziu a $\bar{x} = 10.1$ e $s = 1.2$. Calcular um intervalo de confiança a 95% para $\mu$.

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1=9}$$

$$\alpha = 0.05 \Leftrightarrow P(T > a) = 0.025 \Leftrightarrow$$
$$\Leftrightarrow P(T < a) = 0.975 \Leftrightarrow a = t_{9,0.975} = 2.262$$

$$I.C._{95\%}(\mu) = \left[\bar{x} - a\frac{s}{\sqrt{n}} ; \bar{x} + a\frac{s}{\sqrt{n}}\right]$$

Logo o intervalo pedido é

$$I.C._{95\%}(\mu) = \left[10.1 - 2.262\frac{1.2}{\sqrt{10}} ; 10.1 + 2.262\frac{1.2}{\sqrt{10}}\right] = $$
$$= [9.48; 10.72]$$

**7.5 Intervalo de confiança para a diferença entre as médias de duas populações normais, variâncias desconhecidas**

$X_1$, população 1, com $E(X_1) = \mu_1$ e $V(X_1) = \sigma_1^2$ (desconhecida)

$X_2$, população 2, com $E(X_2) = \mu_2$ e $V(X_2) = \sigma_2^2$ (desconhecida)

( $X_1$ e $X_2$ independentes)

a. a. da população 1 $\left(X_{11}, X_{12}, \ldots, X_{1n_1}\right)$ com média

$$\overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_{1i}}{n_1} \text{ e variância } S_1^2 = \frac{\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)^2}{n_1 - 1}$$

a. a. da população 2 $\left(X_{21}, X_{22}, \ldots, X_{2n_2}\right)$ com média

$$\overline{X}_2 = \frac{\sum_{i=1}^{n_2} X_{2i}}{n_2} \text{ e variância } S_2^2 = \frac{\sum_{i=1}^{n_2}\left(X_{2i} - \overline{X}_2\right)^2}{n_2 - 1}$$





(e a a.a. $\left(X_{11}, X_{12}, \ldots, X_{1n_1}\right)$ é independente da a.a. $\left(X_{21}, X_{22}, \ldots, X_{2n_2}\right)$ )

Como construir um I.C. a $100 \times (1-\alpha)\%$ para $\mu_1 - \mu_2$ ?

Quando $\sigma_1^2$ e $\sigma_2^2$ eram conhecidas usava-se a v.a. fulcral

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Se $n_1 > 30$ e $n_2 > 30$ pode-se substituir $\sigma_1^2$ por $S_1^2$ e $\sigma_2^2$ por $S_2^2$ obtendo-se

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \underset{a}{\sim} N(0,1)$$





quer para $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$ e $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$ quer para $X_1$ e $X_2$ com outra qualquer distribuição.

Então um I.C. calcula-se exactamente como na Secção 7.4, apenas substituindo $\sigma_1^2$ por $s_1^2$ e $\sigma_2^2$ por $s_2^2$.

Quando $n_1 \le 30$ ou $n_2 \le 30$ o problema só tem solução no caso em que $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$ e $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$ e mesmo assim para se obter uma v.a. fulcral com distribuição exacta é necessário supor que (embora desconhecidas) $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (esta suposição é razoável em muitas situações reais, e além disso pode ser testada).

$\overline{X}_1 - \overline{X}_2$ continua a ser o estimador pontual de $\mu_1 - \mu_2$





e tem-se ainda que

$$V\left(\overline{X}_1 - \overline{X}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Vai ser necessário estimar $\sigma^2$. Um estimador natural obtém-se combinando as duas variâncias amostrais

$$\hat{\sigma}^2 = S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Note-se que quando $n_1 = n_2$ resulta $\hat{\sigma}^2 = \dfrac{S_1^2 + S_2^2}{2}$.

Sabemos que

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim N(0,1)$$

substituindo $\sigma$ por $S_p$ obtém-se

$$T = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

(sob as suposições, ou hipóteses de trabalho, consideradas). Então de

$P(-a \le T \le a) = 1 - \alpha$ com $a$: $P(T > a) = \frac{\alpha}{2}$

obtém-se um I.C. a $100 \times (1-\alpha)\%$ para $\mu_1 - \mu_2$:

$$\left[ \overline{x}_1 - \overline{x}_2 - a s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} ; \overline{x}_1 - \overline{x}_2 + a s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

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





**Observações:**

**1)** A determinação da dimensão da amostra é mais complicada (ver 7.3 e 7.4).

**2)** E se não for razoável admitir que $\sigma_1^2 = \sigma_2^2 = \sigma^2$? Este problema (conhecido por problema de Behrens-Fisher) não tem solução exacta. Há soluções aproximadas, o Montgomery (pág. 397) apresenta uma (não faz parte do programa).

**Exemplo:** Um mesmo tipo de material pode ser adquirido a dois fabricantes. As variáveis de interesse são a resistência mecânica do material (em unidades convenientes) para cada fabricante. Para comparar os seus valores médios obteve-se (por amostragem aleatória) uma amostra de cada:





| Fabricante 1 | Fabricante 2 |
|---|---|
| $\overline{x}_1 = 8.73$ | $\overline{x}_2 = 8.68$ |
| $s_1^2 = 0.35$ | $s_2^2 = 0.40$ |
| $n_1 = 15$ | $n_2 = 18$ |

Com o objectivo de ajudar a decidir qual dos dois é melhor pretende-se calcular um intervalo de confiança a 95% para a diferença dos valores médios.

Sejam:

$X_1$ - v.a. que representa a resistência do material produzido pelo fabricante 1

$X_2$ - v.a. que representa a resistência do material produzido pelo fabricante 2





Admitimos que (hipóteses de trabalho):

- $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$ e $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$
- $X_1$ e $X_2$ são independentes;
- $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (parece razoável porque $s_1^2$ e $s_2^2$ são da mesma ordem de grandeza).

A estimativa de $\sigma$ é

$$s_p = \sqrt{\frac{14 \times 0.35 + 17 \times 0.40}{15 + 18 - 2}} = 0.614$$

para grau de confiança = 95% vem $\alpha = 0.05$ e

$$a = t_{31;0.975} = 2.04 \quad (n_1 + n_2 - 2 = 31)$$

$$I.C._{95\%}\left(\mu_1 - \mu_2\right) =$$
$$= \left[ \overline{x}_1 - \overline{x}_2 - a s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} ; \overline{x}_1 - \overline{x}_2 + a s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right] =$$
$$= [-0.388; 0.488]$$

Podemos então afirmar (com 95% de confiança) que não existe grande diferença entre a resistência média do material produzido pelos dois fabricantes.

## 7.6 Intervalo de confiança para a variância de uma população normal

$X \sim N(\mu, \sigma^2)$ $(X_1, \ldots, X_n)$ a.a.

Queremos um I.C. a $100 \times (1-\alpha)\%$ para $\sigma^2$.

O estimador pontual de $\sigma^2$ é $S^2$. A v.a. fulcral obtém-se do teorema seguinte.

**Teorema:** Seja $(X_1, \ldots, X_n)$ uma a.a. duma população $X \sim N(\mu, \sigma^2)$. A v.a.

$$Q = \frac{(n-1)S^2}{\sigma^2}$$

tem distribuição do chi-quadrado com $n-1$ graus de liberdade ($Q \sim \chi^2_{n-1}$).





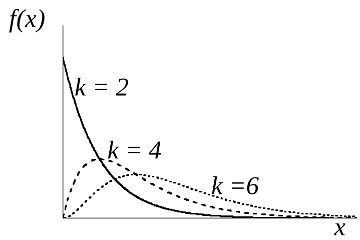
**Notas:**

**1)** Uma v.a. com distribuição do chi-quadrado tem função de densidade de probabilidade

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}, x > 0$$

onde $k > 0$ é o número de graus de liberdade. Pode mostrar-se que se $Q \sim \chi^2_k$ então $E(Q) = k$ e $V(Q) = 2k$.

**2)** Representação gráfica

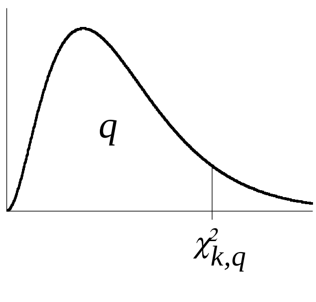


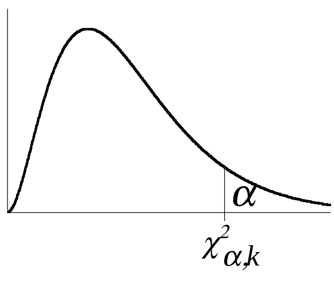





**3)** Os percentis da distribuição $\chi^2_k$ encontram-se tabelados

Tabelas da disciplina: No livro:




$P(Q < \chi^2_{k,q}) = q$ $P(Q > \chi^2_{\alpha,k}) = \alpha$

Para construir um I.C. a $100 \times (1-\alpha)\%$ para $\sigma^2$ parte-se de

$$P(a \le Q \le b) = 1 - \alpha$$





Dado que a distribuição de $Q$ não é simétrica é preciso determinar *a* e *b*. Faz-se então

$a$: $P(Q < a) = \frac{\alpha}{2}$ e $b$: $P(Q > b) = \frac{\alpha}{2}$.

$$P\left(a \le \frac{(n-1)S^2}{\sigma^2} \le b\right) = 1 - \alpha \Leftrightarrow$$

$$\Leftrightarrow P\left(\sigma^2 \in \left[\frac{(n-1)S^2}{b}; \frac{(n-1)S^2}{a}\right]\right) = 1 - \alpha$$

Um I.C. a $100 \times (1-\alpha)\%$ para $\sigma^2$ é então

$$\left[\frac{(n-1)s^2}{b}; \frac{(n-1)s^2}{a}\right]$$

**Exemplo:** Considere-se uma v.a. $X \sim N(\mu, \sigma^2)$. Uma amostra aleatória de dimensão 10 conduziu a $\sum_{i=1}^{10} x_i = 101$ e $\sum_{i=1}^{10} x_i^2 = 1040$. Calcular um intervalo de confiança a 99% para $\sigma$.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) =$$
$$= \frac{1}{9}\left(1040 - 10 \times \left(\frac{101}{10}\right)^2\right) = \frac{19.9}{9}$$

Da tabela: $a = \chi^2_{9;0.005} = 1.735$
$b = \chi^2_{9;0.995} = 23.59$

$$I.C._{99\%}(\sigma^2) = \left[\frac{19.9}{23.59};\frac{19.9}{1.735}\right] = [0.844;11.47]$$

e

$$I.C._{99\%}(\sigma) = \left[\sqrt{0.844};\sqrt{11.47}\right] = [0.918;3.39]$$





**7.7 Intervalo de confiança para uma proporção**

$(X_1, \ldots, X_n)$ amostra aleatória de uma população muito grande ou infinita.

Seja $Y (\leq n)$ o número de observações desta amostra que pertencem a uma dada categoria de interesse.

Seja $p$ a proporção de indivíduos na população que pertencem a essa categoria de interesse.

**Exemplos:**

| População | Categoria |
|---|---|
| Peças | ser defeituosa |
| Eleitores | vota no partido X |

O estimador pontual de $p$ é $\hat{P} = \frac{Y}{n}$.





Sabemos que $Y \sim Bin(n, p)$ e que se $n$ for grande

$$Y \underset{a}{\sim} N(np, np(1-p))$$

ou ainda $\hat{P} = \frac{Y}{n} \underset{a}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$ pois

$$E(\hat{P}) = E\left(\frac{Y}{n}\right) = \frac{E(Y)}{n} \quad \text{e} \quad V(\hat{P}) = V\left(\frac{Y}{n}\right) = \frac{V(Y)}{n^2}$$

Então tem-se $Z = \dfrac{\hat{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \underset{a}{\sim} N(0,1)$

$$P\left(-a \leq \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq a\right) \approx 1 - \alpha,\ a\text{: } P(Z > a) = \frac{\alpha}{2}$$

Resolver em ordem a $p$? Aproximação que dá resultados satisfatórios





$$P\left(-a \leq \frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \leq a\right) \approx 1 - \alpha$$

O I.C. (aproximado) a $100 \times (1-\alpha)\%$ para $p$ é:

$$\left[\hat{p} - a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}};\hat{p} + a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

**Exemplo:** População de eleitores portugueses. Sondagem (aleatória) a 1200 eleitores revelou que 683 tencionam votar no partido ABC. Determinar a e.m.v. de $p$ (proporção de eleitores na população que tencionam votar no partido ABC) e um I.C. aproximado a 95% para $p$.

$\hat{p} = 683/1200 = 0.569$

$a = 1.96$ $\qquad I.C._{95\%}(p) = [0.541;0.597]$