Instituto Tecnológico de Aeronáutica

MOQ-13:
Probabilidade e Estatística

Prof. Mauri Aparecido de Oliveira

mauri@ita.br

2019

Estimação por Intervalos

- Estimadores pontuais não fornecem informação a respeito da precisão e confiabilidade da estimativa.
- A acurácia da estimativa pontual aumenta com o tamanho da amostra, mas é pouco provável que mesmo o estimador pontual não-tendencioso mais eficiente seja capaz de estimar o parâmetro populacional com exatidão.

Estimação por Intervalos

Consiste na construção de intervalos em torno da estimativa pontual contendo

valores plausíveis (com probabilidade conhecida)

para o parâmetro populacional de interesse.

$$P[\widehat{\Theta}_{inf} < \theta < \widehat{\Theta}_{sup}] = 1 - \alpha, \quad 0 < \alpha < 1$$

$\widehat{\Theta}_{inf}$; $\widehat{\Theta}_{sup}$: v.a.'s, variam com a amostra (distribuição amostral) $\rightarrow$ limites de confiança

$1 - \alpha$: nível de confiança (é um valor de probabilidade)

Intervalo de Confiança

O intervalo $\hat{\theta}_{inf} < \theta < \hat{\theta}_{sup}$, calculado para uma determinada amostra é chamado intervalo de confiança $100(1 - \alpha)\%$.

Exemplo:

$\alpha = 0,05 \Rightarrow$ CI 95%

$\alpha = 0,01 \Rightarrow$ CI 99%

... destes, qual é o mais largo?

Definição

O nível de confiança $(1-\alpha)$ de um IC corresponde à probabilidade de o intervalo aleatório $(\widehat{\Theta}_{inf}, \widehat{\Theta}_{sup})$ conter o parâmetro populacional de interesse (antes que a amostra tenha sido observada).

Interpretação:
Um nível de confiança de $100(1-\alpha)\%$ implica que $100(1-\alpha)\%$ de todas as amostras colhidas originariam um i.c. que incluiria o parâmetro populacional θ. Apenas $100\alpha\%$ dos intervalos construídos não incluiriam o valor real de θ.

- Valores comumente utilizados: $90\%, 95\%$ e 99%
- Quanto menor o valor de α, mais largo o intervalo.
- Idealmente, é desejável ter o intervalo mais curto (maior precisão) com o maior nível de confiança possível .

Seja a a.a. $X_1, X_2, \ldots, X_n$.

1. Definir a estatística
$$W = g(X_1, X_2, \ldots, X_n),$$
cuja distribuição seja conhecida e não dependa de θ.
2. Com base na distribuição amostral de W, calcular $\hat{\theta}_{inf}$ e $\hat{\theta}_{sup}$ tais que
$$P[\hat{\theta}_{inf} < \theta < \hat{\theta}_{sup}] = 1 - \alpha$$
3. Definir
$$\{\theta : \hat{\theta}_{inf} < w(x_1, x_2, \ldots, x_n) < \hat{\theta}_{sup}\}$$
como o intervalo de confiança $100(1-\alpha)\%$ para θ.

Veremos os seguintes casos:

CASO 1: IC para μ com σ^2 conhecida

CASO 2: IC para μ com σ^2 desconhecida

2.1. pop. Normal

2.2. pop. Normal, n grande

CASO 3: IC para p (proporção)

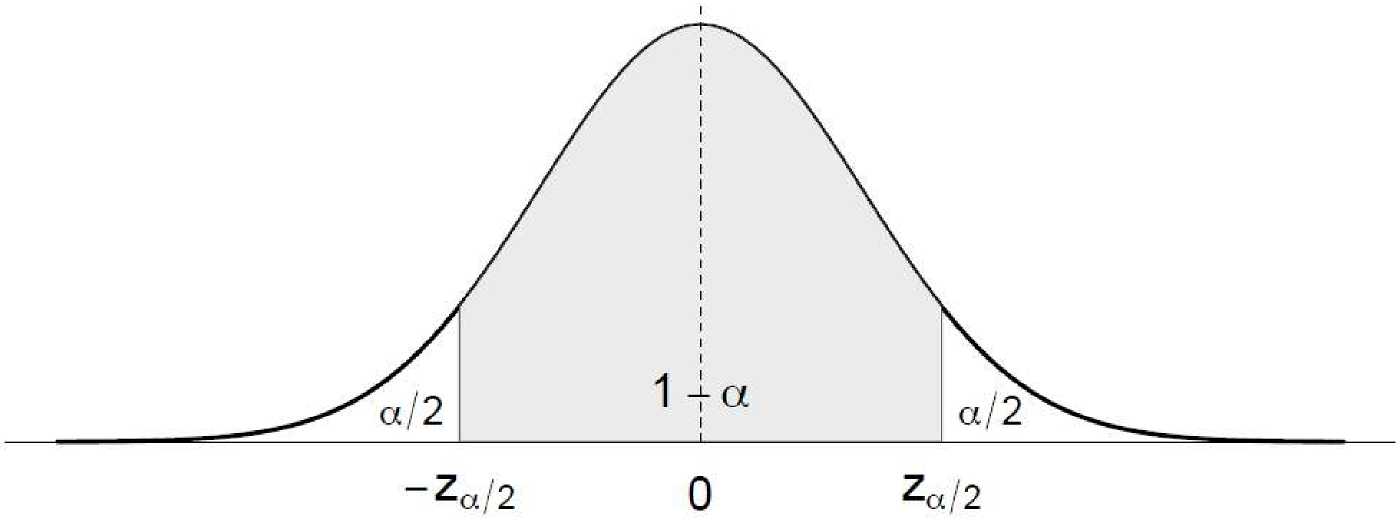
CASO 4: IC para σ^2 (pop. Normal)

CASO 1: IC para μ com σ^2 conhecida

Seja $X \sim N(\mu, \sigma^2)$, σ^2 conhecida.
Seja $\overline{x}_n$ a média de uma a.a. de tamanho n.

O IC $100(1-\alpha)\%$ para a μ é dado por:

$$\overline{x}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

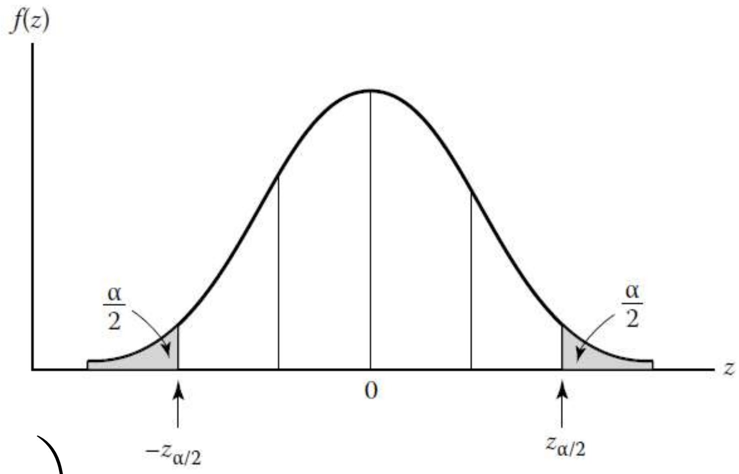


Nível de Confiança	$z_{\alpha/2}$
90%	1,645
95%	1,960
99%	2,575

OBS: O intervalo simétrico é o que tem menor comprimento possível: $L = 2z_{\alpha/2}\sigma/\sqrt{n}$

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

$$P\left(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$



$$P\left(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}\right) = P\left(-z_{\frac{\alpha}{2}} \le \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \le z_{\frac{\alpha}{2}}\right)$$

$$= P\left(-z_{\frac{\alpha}{2}}\sigma_{\bar{x}} \le \bar{x} - \mu \le z_{\frac{\alpha}{2}}\sigma_{\bar{x}}\right) = P\left(-z_{\frac{\alpha}{2}} \le \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \le z_{\frac{\alpha}{2}}\right)$$

$$= P\left(-\bar{x} - z_{\frac{\alpha}{2}}\sigma_{\bar{x}} \le -\mu \le -\bar{x} + z_{\frac{\alpha}{2}}\sigma_{\bar{x}}\right), \quad \text{como} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \quad \text{então}$$

$$= P\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Estimação por Intervalos

Considere uma amostra de tamanho n para uma distribuição Normal com média μ (desconhecida) e variância σ^2 conhecida. Encontre um IC de 95% para a média desconhecida, μ.

Nós sabemos que o melhor estimador de μ é $\overline{X}$ e a distribuição amostral de $\overline{X}$ é N~(μ, σ^2 /n). Então, da normal padrão temos que

$$P\left(\frac{\left|\overline{X}-\mu\right|}{\frac{\sigma}{\sqrt{n}}}<1{,}96\right)=0{,}95$$

o qual é equivalente ao evento

$$\overline{X}-\frac{1{,}96\sigma}{\sqrt{n}}<\mu<\overline{X}+\frac{1{,}96\sigma}{\sqrt{n}}.$$

Estimação por Intervalos

Consequentemente,

$$P\left(\bar{X}-\frac{1{,}96\sigma}{\sqrt{n}}<\mu<\bar{X}+\frac{1{,}96\sigma}{\sqrt{n}}\right)=0{,}95$$

As duas estatísticas

$$\bar{X}-\frac{1{,}96\sigma}{\sqrt{n}} \quad \text{e} \quad \bar{X}+\frac{1{,}96\sigma}{\sqrt{n}}$$

são os limites de um intervalo de confiança de 95% para μ.

Estimação por Intervalos

O intervalo de confiança de 95% para μ é

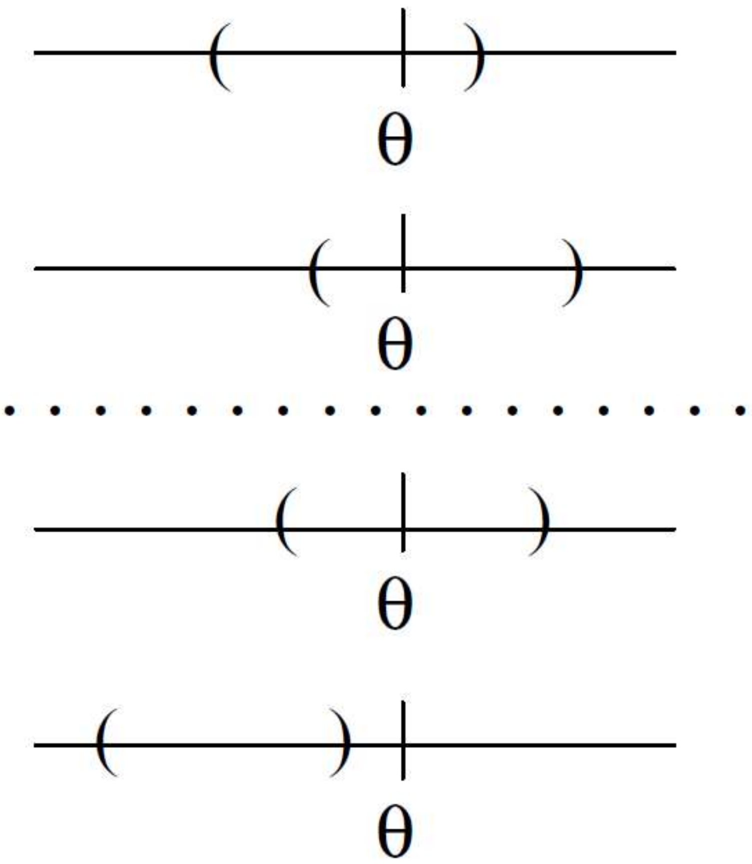
$$\left(\overline{X} - \frac{1{,}96\sigma}{\sqrt{n}}; \overline{X} + \frac{1{,}96\sigma}{\sqrt{n}}\right).$$

<u>Observação</u>

Uma afirmação do tipo $P(5{,}3 < \mu < 9{,}4) = 0{,}95$ é incorreta e deve ser substituída por: Um IC de 95% para μ é (5,3; 9,4).

Estimação por Intervalos

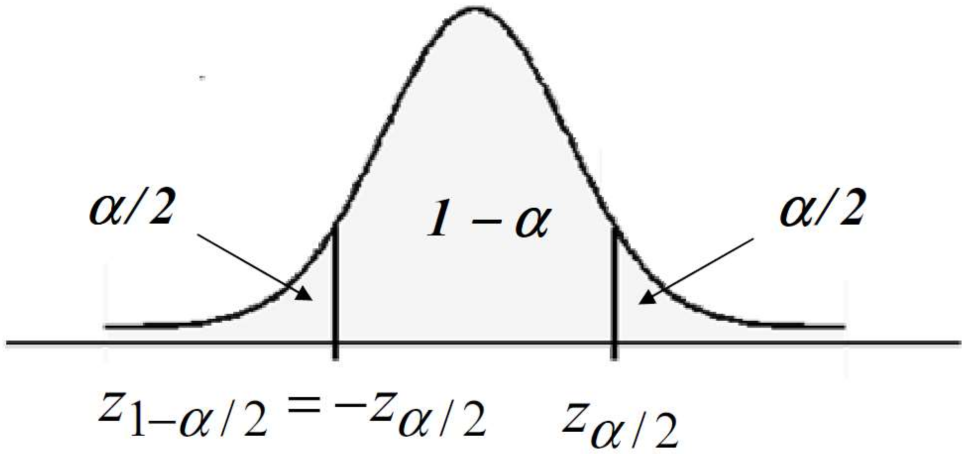
De cada 100 intervalos construídos a partir de 100 amostras, $100(1 - \alpha)\%$ deveriam conter o valor verdadeiro do parâmetro.



Distribuição da Média Amostral

$$\bar{X} \to N\left(\mu; \frac{\sigma^2}{n}\right) \Leftrightarrow Z = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \to N(0;1)$$

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



Estimação por Intervalos

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

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

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

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

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

IC para μ, com variância conhecida, ao nível de confiança $1 - \alpha$.

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Notas:

- Utilizamos o TLC $\Rightarrow$ grau de confiança é acurado para $n \geq 30$.
- Amostras diferentes produzem valores distintos de $\hat{\theta}_{inf}$ e $\hat{\theta}_{sup}$.
- A largura de cada intervalo é a mesma (depende apenas da escolha de $z_{\alpha/2}$), uma vez que $\bar{x}$ foi determinada.
  (Simulação)

Consequências:

Se $\overline{X}$ for utilizado como estimador de μ:

- podemos ter $100(1-\alpha)\%$ de certeza de que o erro $= |\bar{X} - \mu|$ não excederá $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$:
$$\text{erro} < e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
- podemos ter $100(1-\alpha)\%$ de certeza de que o erro não excederá um valor específico e se o tamanho da amostra for $n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$.

Exemplo 1:

A concentração média de zinco em 36 diferentes poços de água foi calculada como sendo $2,6$mg/l. Assuma que o desvio-padrão populacional vale 0,3.

1. Determine os IC 95% e 99% para a concentração média de zinco nas águas distribuídas para a população.
2. Qual deve ser o menor tamanho de amostra para que tenhamos 95% de confiança de que o erro na estimativa encontrada para μ seja no máximo 0,05?

Exemplo 1 – Solução:

Temos: $n = 36$; $\bar{x} = 2,6$mg/l; $\sigma = 0,3$.

1. (i) IC 95%: $\alpha = 1 - 0,95 = 0,05 \overset{tabela}{\Rightarrow} -z_{0,025} = -1,96$; $z_{0,975} = 1,96$.
   Portanto,
   IC 95%: $2,6 - (1,96)(0,3/\sqrt{36}) < \mu < 2,6 + (1,96)(0,3/\sqrt{36}) \Rightarrow 2,50 < \mu < 2,70$.
   Temos 95% de certeza que a média amostral $\bar{x} = 2,6$ difere da média populacional μ em no máximo 0,1 (erro máximo): $1,96(0,3/6) = 1,96 \cdot 0,05 = 0,1$

   (ii) IC 99%: $\alpha = 1 - 0,99 = 0,01 \overset{tabela}{\Rightarrow} -z_{0,005} \approx -2,57$; $z_{0,995} \approx 2,57$.
   Portanto,
   IC 99%: $2,6 - (2,57)(0,3/\sqrt{36}) < \mu < 2,6 + (2,57)(0,3/\sqrt{36}) \Rightarrow 2,47 < \mu < 2,73$.
   Temos 99% de certeza que o erro ao estimar μ como $\bar{x} = 2,6$ é menor que 0,13:
   $2,57(0,3/6) = 2,57 \cdot 0,05 \approx 0,13$

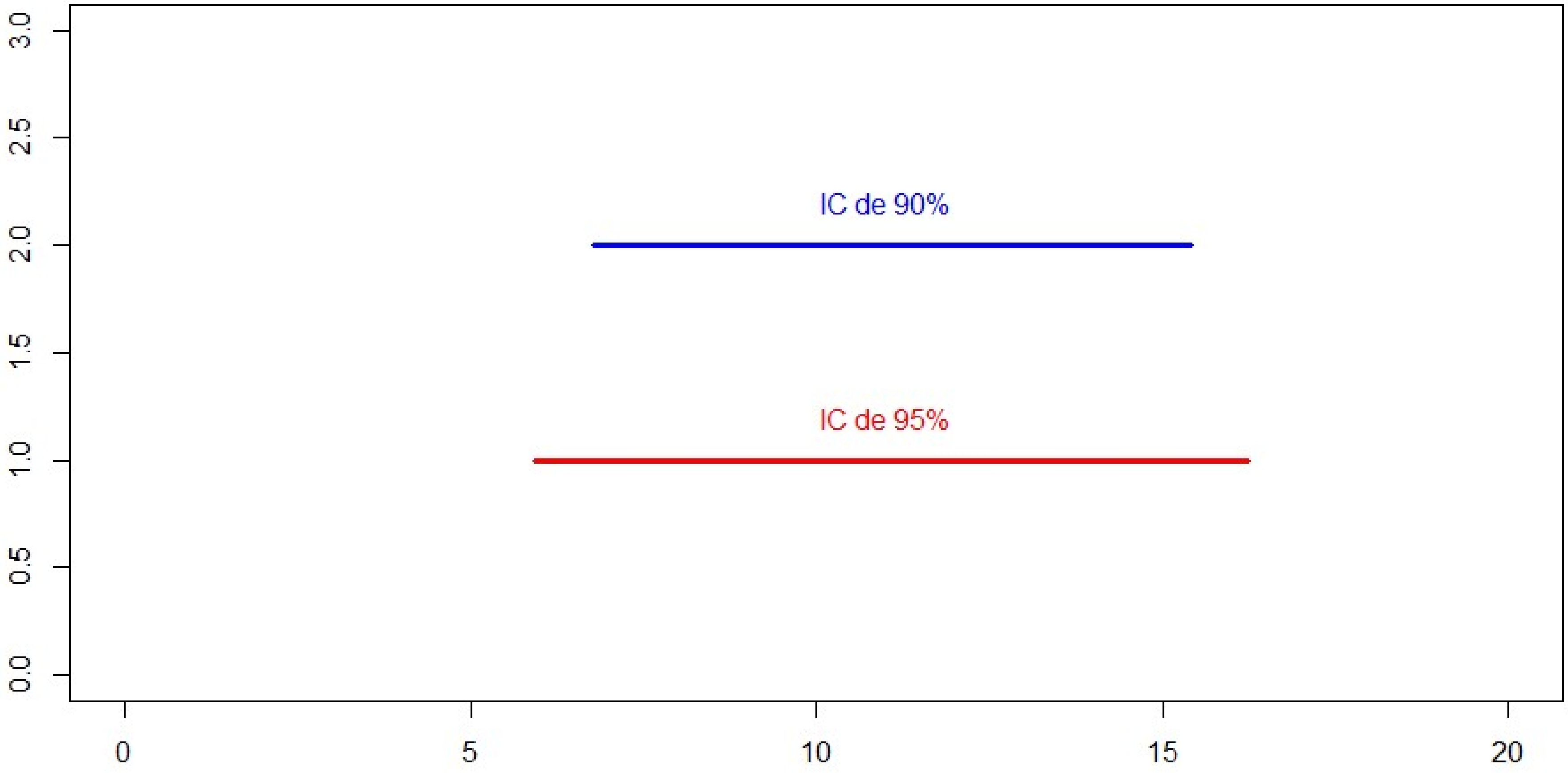
2. n=? tq. erro $< 0,05$ com 95% de confiança:
   erro $< e = z_{\alpha/2}\sigma/\sqrt{n} = 0,05 \Rightarrow n = (z_{\alpha/2}\sigma/0,05)^2 = 138,3 \quad \therefore n = 139$.

```r
# Exemplo Intervalo de Confiança para Média

dev.off(dev.list()["RStudioGD"])
rm(list=ls())
cat("\f")

# Os dados a seguir representam uma amostra aleatória dos ativos (em milhões de dólares)
# de 30 cooperativas de crédito de um estado do país. Suponha que o desvio padrão da
# população seja 14,405. Encontre o intervalo de confiança de 90% da média.

x <- c(12.23,16.56,4.39,2.89,1.24,2.17,13.19,9.16,1.42,73.25,1.91,14.64,11.59,6.69,1.06,
       8.74,3.17,18.13,7.92,4.78,16.85,40.22,2.42,21.58,5.01,1.47,12.24,2.27,12.77,2.76)

# Objetivo: estimar a média de todos os dados.
# Encontrar intervalos de confiança de:
# a) 90% e
# b) 95% para a média.

# Suposição: considere que os dados são oriundos de uma distribuição normal
#            com média desconhecida e desvio-padrão 14,405

# Sabemos que: um IC (1 - alfa)100% é
# Xbar +- z(alfa/2)*sigma/sqrt(n)

n = 30
sigma = 14.405
```

```r
# a) IC de 90%
Xbar <- mean(x)
alfa = 0.10
z <- qnorm(0.95)     # z(alfa/2) = z(0.05)
z

valores <- seq(from=-4, to=4, length=100)
normal = dnorm(valores, 0, 1)

plot(valores, normal, type="n", ylab="f(x)", las=1)
lines(valores, normal, lty=1, lwd=2)
abline(v=c(-z,z),col="red", lwd=3, lty=2)

# Margem de erro
me <- z*sigma/sqrt(n)
me

LI <- Xbar-me
LS <- Xbar+me

IC <- c(LI,LS)
IC


# Intrepretação: Por definição, estamos 90% confiantes de que a verdadeira
#                média de X esteja no intervalo (LI,LS).

# Intrepretação: Portanto, pode-se ter 90% de confiança de que a média da
#                população dos ativos de todas as cooperativas de crédito
#                está entre US$ 6,752 milhões e US$ 15,430 milhões, com base
#                em uma amostra de 30 cooperativas de crédito.

# Incorreto: Há uma probabilidade de 90% de que o intervalo de confiança contenha µ.
```

```r
# b) IC de 95%
Xbar <- mean(x)
alfa = 0.05
z <- qnorm(0.975)      # z(alfa/2) = z(0.025)
z

valores <- seq(from=-4, to=4, length=100)
normal = dnorm(valores, 0, 1)

plot(valores, normal, type="n", ylab="f(x)", las=1)
lines(valores, normal, lty=1, lwd=2)
abline(v=c(-z,z),col="blue", lwd=3, lty=2)

# Margem de erro
me <- z*sigma/sqrt(n)
me

LI <- Xbar-me
LS <- Xbar+me

IC <- c(LI,LS)
IC
```

```r
# Plotando od dois intervalos
ICs <- data.frame(x0 = c(5.935998, 6.764733),    # Cria um data frame com os intervalos
                  y0 = c(1,2),
                  x1 = c(16.245335, 15.416601),
                  y1 = c(1,2))

ICs
ICs[1, ]
ICs[2, ]

plot(c(0,20), c(0,3), col = "white", xlab = "", ylab = "")    # Plota a estrutura básica
segments(x0 = ICs$x0,
         y0 = ICs$y0,
         x1 = ICs$x1,
         y1 = ICs$y1, col=c("red","blue"), lwd=3)

text(c(11,11), c(1.2,2.2),
     c(expression("IC de 95%","IC de 90%")), col=c("red","blue"))
```



IC de 90%
IC de 95%
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
5
10
15
20

CASO 2: IC para μ com σ^2 desconhecida

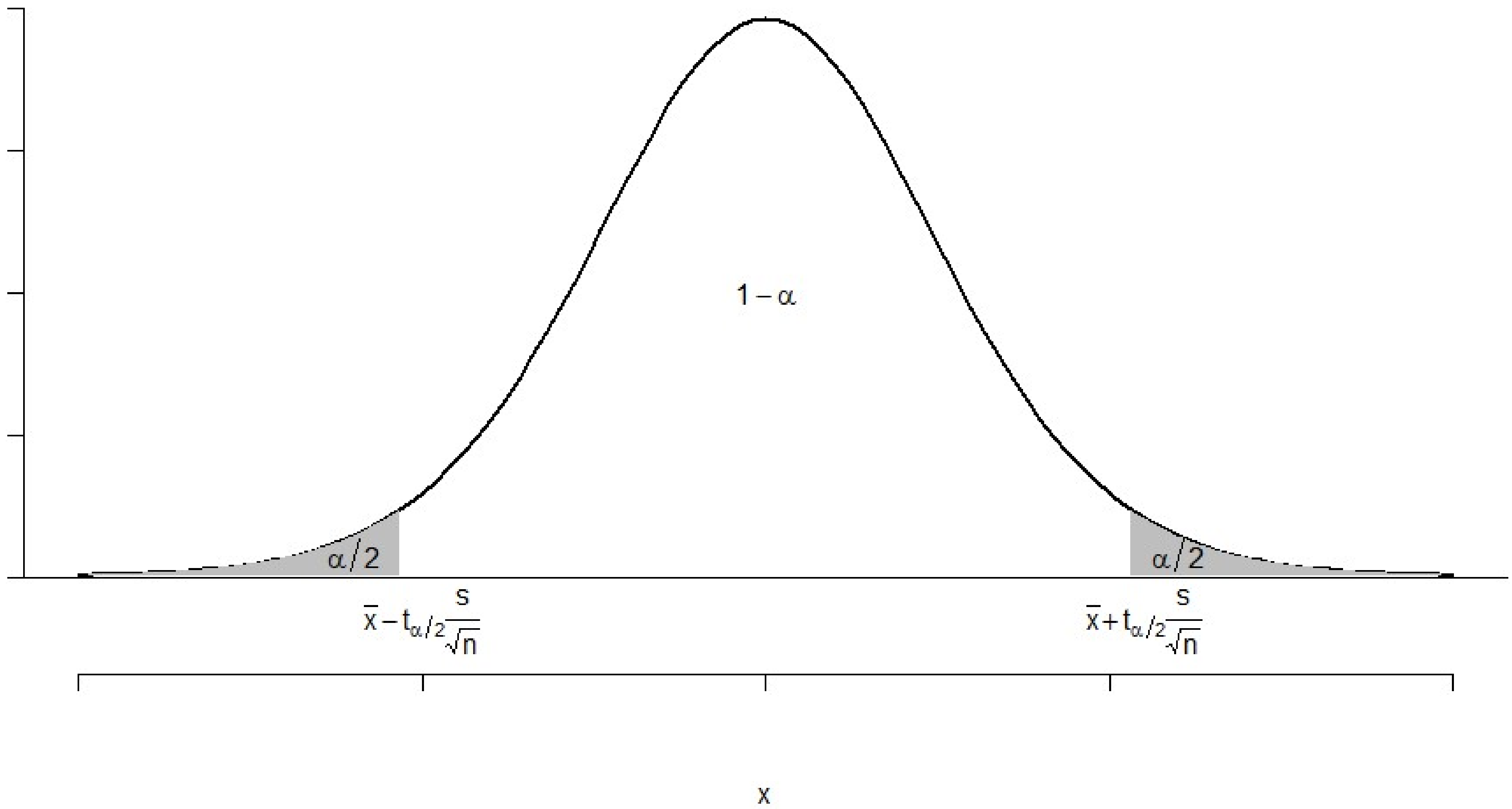
Distribuição t de Student

Em situações em que a população é normal, mas o desvio padrão σ é desconhecido, a **distribuição t de Student** deve ser usada no lugar da distribuição Normal. Isso é particularmente importante quando a amostra é pequena. Quando σ é desconhecido, a fórmula para um intervalo de confiança se parece com a fórmula para σ conhecido, exceto que a estatística t substitui z e s substitui σ.

$$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$ em que $\frac{s}{\sqrt{n}}$ é o erro padrão estimado da média

A interpretação do intervalo de confiança é a mesma dada σ conhecido, conforme mostrado na Figura a seguir. Entretanto, os intervalos de confiança serão maiores (sendo mantidos os outros fatores) porque $t_{\frac{\alpha}{2}}$ será sempre maior do que $z_{\frac{\alpha}{2}}$. Intuitivamente, nosso intervalo de confiança será maior pelo fato de nos depararmos com uma incerteza maior ao utilizarmos o desvio padrão amostral s para estimar o desvio desconhecido da população σ.

William Sealy Gosset (13 de julho de 1876 – 16 de outubro de 1937) foi um químico e estatístico inglês.



$1-\alpha$
$\alpha/2$
$\alpha/2$
$\bar{x}-t_{\alpha/2}\frac{s}{\sqrt{n}}$
$\bar{x}+t_{\alpha/2}\frac{s}{\sqrt{n}}$
x

```r
par(bty="n")

# Plotando a curva T
x = seq(from=-4, to=4, length=100)
gl = 15      # Graus de Liberdade
alpha = 0.05
curva.t = dt(x, df=gl)

plot(0,0,col="white",xlab="",ylab="",xlim=c(-4,4),ylim=c(0,0.4))
lines(x, curva.t, lty=1, lwd=2)

t.1 <- qt(alpha/2, gl, lower.tail = TRUE)
t.2 <- qt(1-alpha/2, gl, lower.tail = TRUE)
abline(v=c(t.1,t.2), lty=2, lwd=2, col="red")
abline(h=0)

# Curva e áreas sombreadas
curve(dt(x, gl), from = -4, to = 4,lwd=2, ylab="",ylim=c(-0.05,0.4), labels = FALSE)
coord.x1 <- seq(-4, t.1, len = 100)
coord.y1 <- dt(coord.x1, 15)
polygon(c(coord.x1[1], coord.x1, coord.x1[100]), c(dt(-4, 15), coord.y1, dt(4, 15)),
        col = "grey", border = NA)
coord.x2 <- seq(t.2,4, len = 100)
coord.y2 <- dt(coord.x2, 15)
polygon(c(coord.x2[1], coord.x2, coord.x2[100]), c(dt(-4, 15), coord.y2, dt(4, 15)),
        col = "grey", border = NA)
abline(h=0)


text(c(-2.4,0,2.4), c(0.015,0.2,0.015), c(expression(alpha/2,1-alpha,alpha/2)))
text(c(-2,2.2), c(-0.03,-0.03), c(expression(bar(x)-t[alpha/2]*frac(s,sqrt(n))),
                                  expression(bar(x)+t[alpha/2]*frac(s,sqrt(n)))))
```

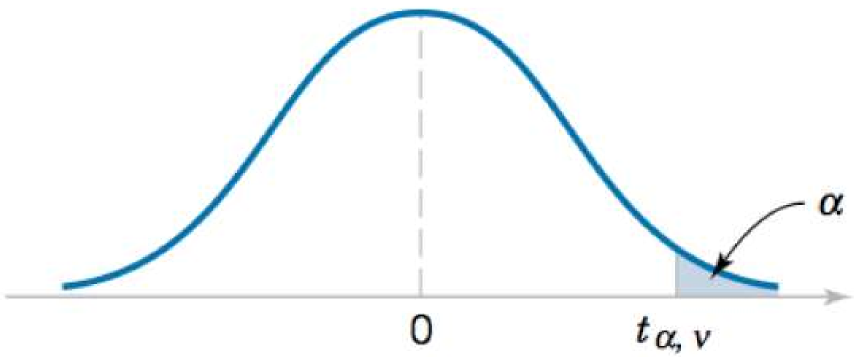
Distribuição $t-$Student

Sejam:
$X \sim N(\mu, \sigma^2)$, com σ desconhecido (em geral, este é o caso):
$X_1, X_2, \ldots, X_n$ a.a.

Teorema

Então:

$$T = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim T_\nu \qquad (\nu = n - 1 g.d.l.)$$



(ver Tabela $t-$Student)

Teorema: $t_{1-\alpha}(\nu) = -t_\alpha$ (Simetria)

(Simulação)

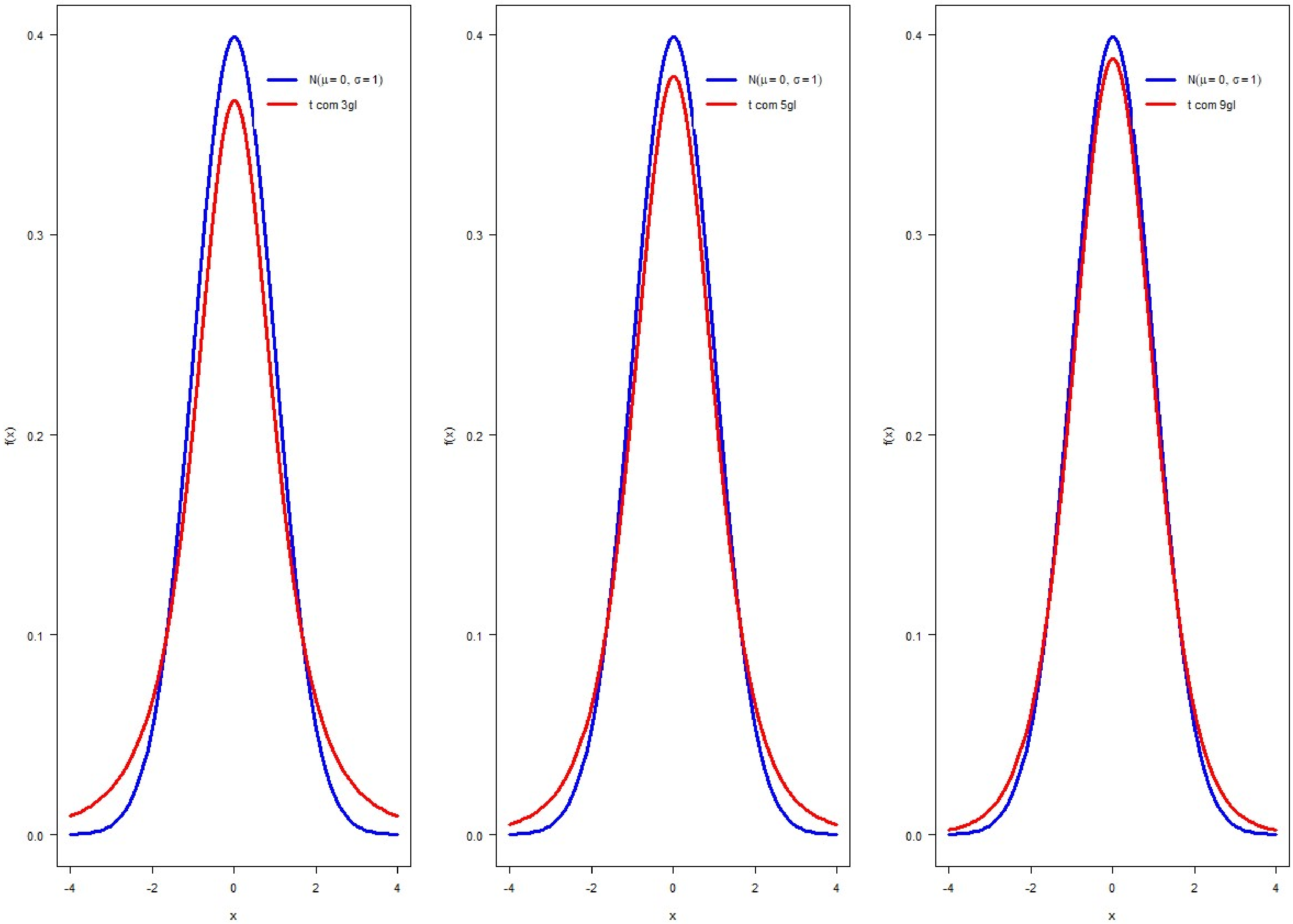
Graus de Liberdade

Conhecendo o tamanho da amostra podemos calcular um parâmetro chamado graus de liberdade (algumas vezes abreviado por g.l.). Esse parâmetro é usado para determinar o valor da estatística t usada na fórmula do intervalo de confiança. Os graus de liberdade refletem o número de observações que usamos para calcular s, o desvio padrão amostral, menos o número de estimativas intermediárias usadas em nosso cálculo. Lembre-se que a fórmula para s usa todos os n valores individuais da amostra e também $\bar{x}$, a média amostral. Assim sendo, os graus de liberdade são iguais ao tamanho da amostra menos 1.

$g.l. = n - 1$ (graus de liberdade para um intervalo de confiança para μ)

É costume se referir a S^2 como tendo n – 1 graus de liberdade ($g.l.$). Esta terminologia resulta do fato de que, embora S^2 seja baseada nas n quantidades: $x_1 - \bar{x}$, $x_2 - \bar{x}$, ..., $x_n - \bar{x}$, esta soma é zero, assim especificar os valores de qualquer n – 1 das quantidades determina o valor que resta. Por exemplo, se $n = 4$ e $x_1 - \bar{x} = 8$, $x_2 - \bar{x} = -6$ e $x_4 - \bar{x} = -6$, então automaticamente $x_3 - \bar{x} = 2$, então apenas três dos quatro valores de $x_i - \bar{x}$ são determinados livremente.

Para graus de liberdade grandes, a distribuição t se aproxima da forma da distribuição normal, como mostrado na Figura a seguir. Contudo, em pequenas amostras, a diferença é relevante, considerando a Figura a seguir, a escala do eixo inferior se estende além de ±4, enquanto um intervalo de ±3, cobriria a maior parte da área para uma distribuição normal padrão. Temos de ir um pouco mais longe nas caudas da distribuição t para cobrir uma mesma área, e assim, para um dado coeficiente de confiança, t é sempre maior que z, então o intervalo de confiança é sempre mais longo do que se z fosse usado.



Comparação da Normal e t de Student.

```r
par(mfrow=c(1,3))

x = seq(from=-4, to=4, length=100)
curva.normal = dnorm(x, 0, 1)
gl = 3
curva.t = dt(x, df=gl)

plot(x, curva.normal, type="n", ylab="f(x)", las=1)
lines(x, curva.normal, lty=1, lwd=3,col="blue")
lines(x, curva.t, lty=1, lwd=3,col="red")
legend(0.5, .39, lty=c(1,1), lwd=c(3,3),col=c("blue","red"),box.col = "white",
       legend=c(expression(N(mu == 0,sigma == 1)),paste("t com ", gl,"gl", sep="")))


x = seq(from=-4, to=4, length=100)
curva.normal = dnorm(x, 0, 1)
gl = 5
curva.t = dt(x, df=gl)

plot(x, curva.normal, type="n", ylab="f(x)", las=1)
lines(x, curva.normal, lty=1, lwd=3,col="blue")
lines(x, curva.t, lty=1, lwd=3,col="red")
legend(0.5, .39, lty=c(1,1), lwd=c(3,3),col=c("blue","red"),box.col = "white",
       legend=c(expression(N(mu == 0,sigma == 1)),paste("t com ", gl,"gl", sep="")))


x = seq(from=-4, to=4, length=100)
curva.normal = dnorm(x, 0, 1)
gl = 9
curva.t = dt(x, df=gl)

plot(x, curva.normal, type="n", ylab="f(x)", las=1)
lines(x, curva.normal, lty=1, lwd=3,col="blue")
lines(x, curva.t, lty=1, lwd=3,col="red")
legend(0.5, .39, lty=c(1,1), lwd=c(3,3),col=c("blue","red"),box.col = "white",
       legend=c(expression(N(mu == 0,sigma == 1)),paste("t com ", gl,"gl", sep="")))
```

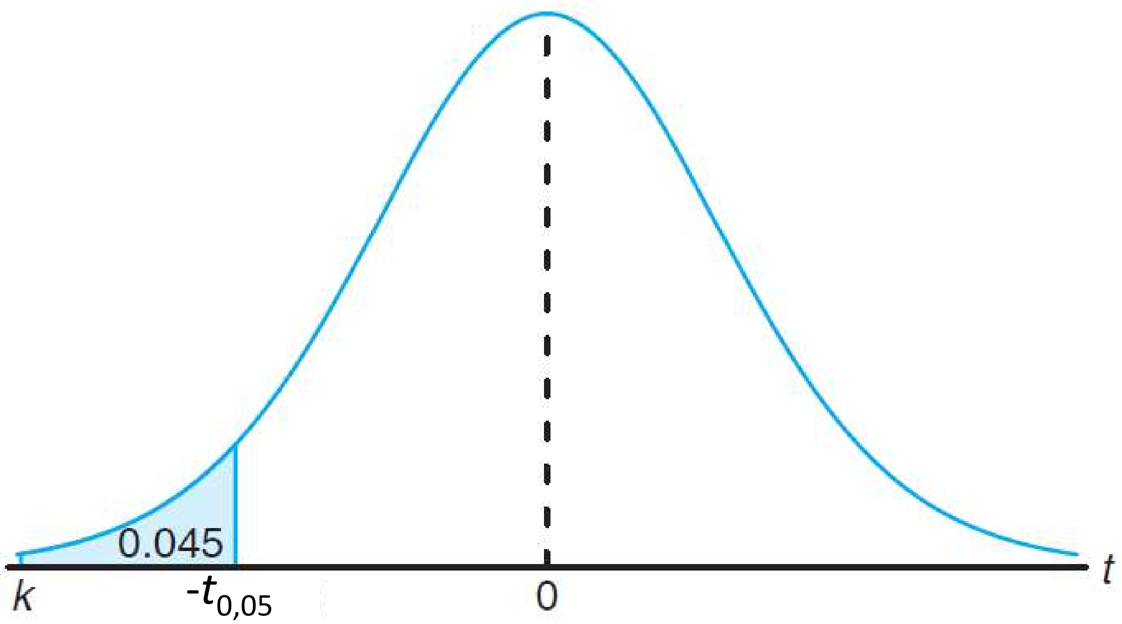
A distribuição $t-$Student é útil para realizar inferência a respeito da média populacional (ou entre diferenças de duas médias populacionais) quando σ é desconhecido.

É importante observar:

1. Para $n \downarrow$: S_n^2 flutua muito de amostra para amostra e $T \neq N(0,1)$
2. Para $n \geq 30$: $T \approx N(0,1)$
3. Para $n < 30$: usar a distr. exata T
4. Populações *não-normais* com forma de sino têm $T \approx T_\nu$

Determine o valor de k de forma que $P[k < T < -1,761] = 0,045$ para uma a.a. $X_1, \ldots, X_n$ de tamanho $n = 15$ selecionada a partir de uma distribuição normal e

$$T = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}.$$



Solução

$n = 15 \Rightarrow v = n - 14$ g.l.

Temos que 1,761 corresponde a $t_{0,05}$ quando $v = 14$. Portanto, $-t_{0,05} = -1,761$.
Como k na declaração de probabilidade original está à esquerda de $-t_{0,05} = -1,761$, fazemos $k = -t_\alpha$. Então, da Figura, temos

$$0,045 = 0,05 - \alpha, \text{ ou } \alpha = 0,005.$$

Consequentemente, com $v = 14$.

$$k = -t_{0,005} = -2,977 \text{ e } P(-2,977 < T < -1,761) = 0,045.$$

```
> pt(-1.761,14, lower.tail = TRUE)
[1] 0.05002709
>
> qt(0.005,14, lower.tail = TRUE)
[1] -2.976843
```

Um engenheiro afirma que a média populacional da concentração de um determinado produto químico é de 500mg/ml de solução. Considere que a distribuição da concentração deste produto é aproximadamente normal.

A fim de verificar esta hipótese, ele toma amostras de 25 lotes de produto todo mês. Se o valor t calculado estiver entre $-t_{0,05}$ e $t_{0,05}$, a hipótese é mantida.

A que conclusão pode-se chegar a partir de uma amostra cuja média vale $\overline{x} = 518$mg/ml e desvio padrão vale $s = 40$mg/ml?

Temos: $\mu = 500$mg/ml; n = 25; $\Rightarrow \nu = 25 - 1 = 24$;
Para a amostra obtida: $\overline{x} = 518$mg/ml e $s = 40$mg/ml.

O engenheiro está satisfeito se $-t_{0,05;24} < T < t_{0,05;24}$. Da tabela, $t_{0,05;24} = 1,711$.

$$T = \frac{518 - 500}{40/\sqrt{25}} = 2,25 > t_{0,05}!$$

Portanto, a hipótese de que $\mu = 500$ml/mg não parece ser válida.
É mais provável que $\mu > 500$ml/mg.

Distribuição t-Student

Função Densidade de Probabilidade da distribuição t de Student

$$f(x;m) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\,\Gamma\left(\frac{m}{2}\right)} \frac{1}{\left(1+\frac{x^2}{m}\right)^{\frac{m+1}{2}}}$$

Distribuição t-Student

```
dev.off(dev.list()["RStudioGD"])
rm(list=ls())
cat("\f")


# FUNÇÃO DENSIDADE DE PROBABILIDADE
df <- 15;
t <- function(x) (gamma((df+1)/2))/((sqrt(df*pi))*(gamma(df/2))*((1+((x^2)/df))^((df+1)/2)))
int <- integrate(t,-Inf,1.753)
alpha <- 1 - int$value
plot(t,-5,5)
int
alpha

polygon(x=c(1.753,seq(1.753,10,l=20),10), y=c(0,t(seq(1.753,10,l=20)), 0), col="gray")

text(c(2.2, 0), c(0.02, 0.2), c(expression(alpha,1-alpha)))
```

t de Student

$$f(x) = \frac{\Gamma\left[\frac{m+1}{2}\right]}{\sqrt{m\pi}\ \Gamma\left[\frac{m}{2}\right]} * \frac{1}{\left(1+\frac{x^2}{m}\right)^{\frac{m+1}{2}}}$$

In[30]:= $\int_{-\infty}^{\infty} \frac{\text{Gamma}\left[\frac{7+1}{2}\right]}{\sqrt{7\,\pi}\ \text{Gamma}\left[\frac{7}{2}\right]} * \frac{1}{\left(1+\frac{x^2}{7}\right)^{\frac{7+1}{2}}}\, dx$

Out[30]=

1

In[32]:= $\int_{-\infty}^{1.89} \frac{\text{Gamma}\left[\frac{7+1}{2}\right]}{\sqrt{7\,\pi}\ \text{Gamma}\left[\frac{7}{2}\right]} * \frac{1}{\left(1+\frac{x^2}{7}\right)^{\frac{7+1}{2}}}\, dx$

Out[32]=

0.949662


WolframAlpha.com
WolframCloud.com
Todos os Sites & Recursos para o público...
WOLFRAM
COMPUTATION MEETS KNOWLEDGE
Produtos & Serviços
Tecnologias
Soluções
Aprendizagem & Suporte
Empresa
Busca
23 de novembro de 2019
INSCREVA-SE AGORA
BRAZILIAN WOLFRAM
TECHNOLOGY CONFERENCE
23 de novembro de 2019 | das 9:00 às 17:00 | São Paulo
VISÃO GERAL
CRONOGRAMA
LOGÍSTICA
SUBMISSÃO
Blockchain
Cloud Computing
Numerical Computation
Audio
Natural Language
IoT
Image Processing
Visualization
User Interfaces
Machine Learning
Curated Repositories
Artificial Intelligence
Neural Networks
Enterprise Deployment
Algorithmic Startups
Data Science
Web Apps
Coding
Automation

TABLE 2 Critical Values for Two-Sided and One-Sided Tests Using the Student *t* Distribution

Degrees of Freedom	Significance Level				
	20% (2-Sided) 10% (1-Sided)	10% (2-Sided) 5% (1-Sided)	5% (2-Sided) 2.5% (1-Sided)	2% (2-Sided) 1% (1-Sided)	1% (2-Sided) 0.5% (1-Sided)
1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.41	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.05

Exemplo 9

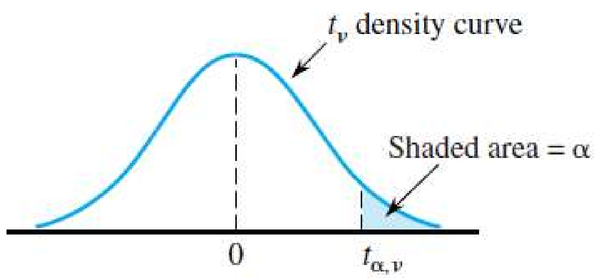
Determine

a) $t_{0,025}$ quando $\nu = 14$

b) $-t_{0,10}$ quando $\nu = 10$

c) $t_{0,995}$ quando $\nu = 7$

Table A.5 Critical Values for *t* Distributions



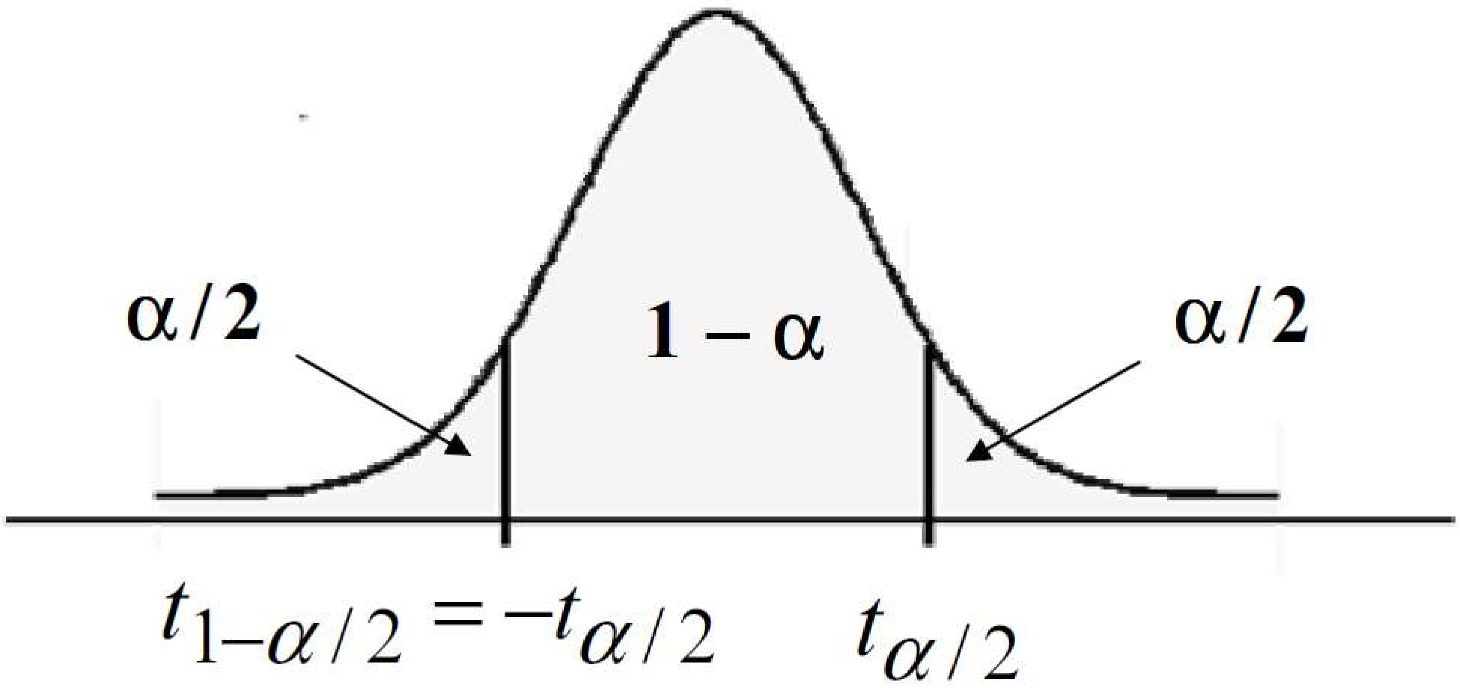
ν \ α	.10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850

Distribuição Amostral

$$T = \frac{\left(\bar{X} - \mu\right)}{S / \sqrt{n}} \rightarrow t_{n-1}$$

Estimação por Intervalos

$$S = \sqrt{\frac{1}{n-1}\sum_{i}\left(X_i - \bar{X}\right)^2}$$



Estimação por Intervalos

$$P\left[-t_{\alpha/2;n-1} \le T \le t_{\alpha/2;n-1}\right] = 1-\alpha$$

$$P\left[-t_{\alpha/2;n-1} \le \frac{\left(\bar{X}-\mu\right)}{S/\sqrt{n}} \le t_{\alpha/2;n-1}\right] = 1-\alpha$$

$$P\left[-t_{\alpha/2;n-1} \le \frac{\left(\bar{X}-\mu\right)}{S/\sqrt{n}} \le t_{\alpha/2;n-1}\right] = 1-\alpha$$

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

Estimação por Intervalos

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

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

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

IC para μ, com variância desconhecida, ao nível de confiança $1 - \alpha$.

$$\left[\bar{X} - t_{\alpha/2;n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2;n-1} \frac{S}{\sqrt{n}} \right]$$

Estimação por Intervalos

CASO 2.1: IC para μ com σ^2 desconhecida

Seja $X \dot{\sim} N(\mu, \sigma^2)$, σ^2 desconhecida.
Sejam $\overline{x}_n$ e s_n^2 a média a variância de uma a.a. de tamanho n.

O IC $100(1-\alpha)\%$ para a μ é dado por:

$$\overline{x}_n \pm t_{\alpha/2} \frac{s_n}{\sqrt{n}}$$

em que $t_{\alpha/2}$ é o valor-t com $\nu = n - 1$ graus de liberdade.

Estimação por Intervalos

CASO 2.2: IC para μ com σ^2 desconhecida (amostras grandes)

Seja $X \dot\sim N(\mu, \sigma^2)$, σ^2 desconhecida.
Sejam $\overline{x}_n$ e s_n^2 a média a variância de uma a.a. grande ($n \geq 30$).

Recomenda-se utilizar o seguinte IC $100(1-\alpha)\%$ para a μ:

$$\overline{x}_n \pm z_{\alpha/2} \frac{s_n}{\sqrt{n}}$$

Exemplo 2:

Um fabricante de computadores deseja avaliar o sistema de memória do disco rígido. Uma medida de desempenho corresponde ao tempo médio entre falhas do disco. A fim de estimar este valor, um engenheiro registrou o tempo entre falhas para uma a.a. de 45 computadores.
As seguintes estatísticas foram calculadas:

$$\overline{x} = 1.762\text{h} \quad \text{e} \quad s = 215\text{h}$$

1. Estime o valor do tempo médio entre falhas com um IC 90%.
2. Sabe-se que se o sistema de memória do disco rígido estiver funcionando adequadamente, o tempo médio real entre falhas deverá ser de pelo menos 1.700h. Com base no intervalo calculado no item anterior, o que se pode inferir a respeito do sistema de memória?

Exemplo 2 – Solução:

Temos: $\overline{x} = 1.762$h; $s = 215$h; n = 45
Usando a distribuição t: $\nu = 44$

1. IC 90%: $\alpha = 1 - 0,90 = 0,10; \alpha/2 = 0,05 \quad \overset{\text{(tabela t: } \nu=44)}{\Rightarrow} \quad t_{0,05;44} = 1,684.$
   Portanto,
   IC 90%: $\bar{x} \pm t_{\alpha/2;\nu} s_n/\sqrt{n}$: $\quad 1.762 \pm 1,684(215/\sqrt{45}) \quad \Rightarrow \quad 1.762 \pm 54h.$
   Temos 90% de certeza que o intervalo (1.708, 1.816) contém a o tempo médio entre falhas.
2. Sabe-se que se o sistema de memória do disco rígido estiver funcionando adequadamente, o tempo médio real entre falhas deverá ser de pelo menos 1.700h. Com base no intervalo calculado no item anterior, o que se pode inferir a respeito do sistema de memória?
   Como todos os valores no IC excedem 1.700h, podemos inferir que o sistema funcionará adequadamente com 90% de certeza.

Exemplo 2 – Solução (2):

Temos: $\overline{x} = 1.762$h; $s = 215$h; n $= 45$
Usando a aproximação para a distribuição normal:

1. IC 90%: $\alpha = 1 - 0,90 = 0,10; \alpha/2 = 0,05 \quad \overset{\text{(tabela)}}{\Rightarrow} \quad z_{0,05} = 1,645$.
   Portanto,
   IC 90%: $\bar{x} \pm z_{\alpha/2} s_n/\sqrt{n}: \quad 1.762 \pm 1,645(215/\sqrt{45}) \quad \Rightarrow \quad 1.762 \pm 52,7h$.

   O IC para amostras grandes é menor (mais preciso): válido por causa do TLC.

Exemplo – 3

As companhias de seguros assumem riscos. Quando eles seguram uma propriedade ou uma vida, eles devem precificar a apólice de tal forma que seu lucro esperado permita que eles continuem operando. Eles podem basear suas projeções em tabelas atuariais, mas a realidade do negócio de seguros muitas vezes exige que eles descontem as apólices para uma variedade de clientes e situações. Gerenciar esse risco se torna ainda mais difícil pelo fato de que, até que a apólice expire, a empresa não saberá se obteve lucro, independentemente do prêmio cobrado.

Um gerente queria ver o desempenho de um de seus representantes de vendas, por isso selecionou 30 apólices vencidas que haviam sido vendidas pelo representante de vendas e calculou o lucro líquido (prêmio cobrado menos sinistros pagos), para cada uma das 30 políticas.

O gerente gostaria que você, como consultor, construísse um intervalo de confiança de 95% para o lucro médio das apólices vendidas por esse representante de vendas.

Configuração da Análise

- Estabeleça o que queremos saber.
- Identifique as variáveis e seu contexto.
- Faça gráficos.
- Verifique o formato da distribuição e procure por assimetria, múltiplas modas e outliers.

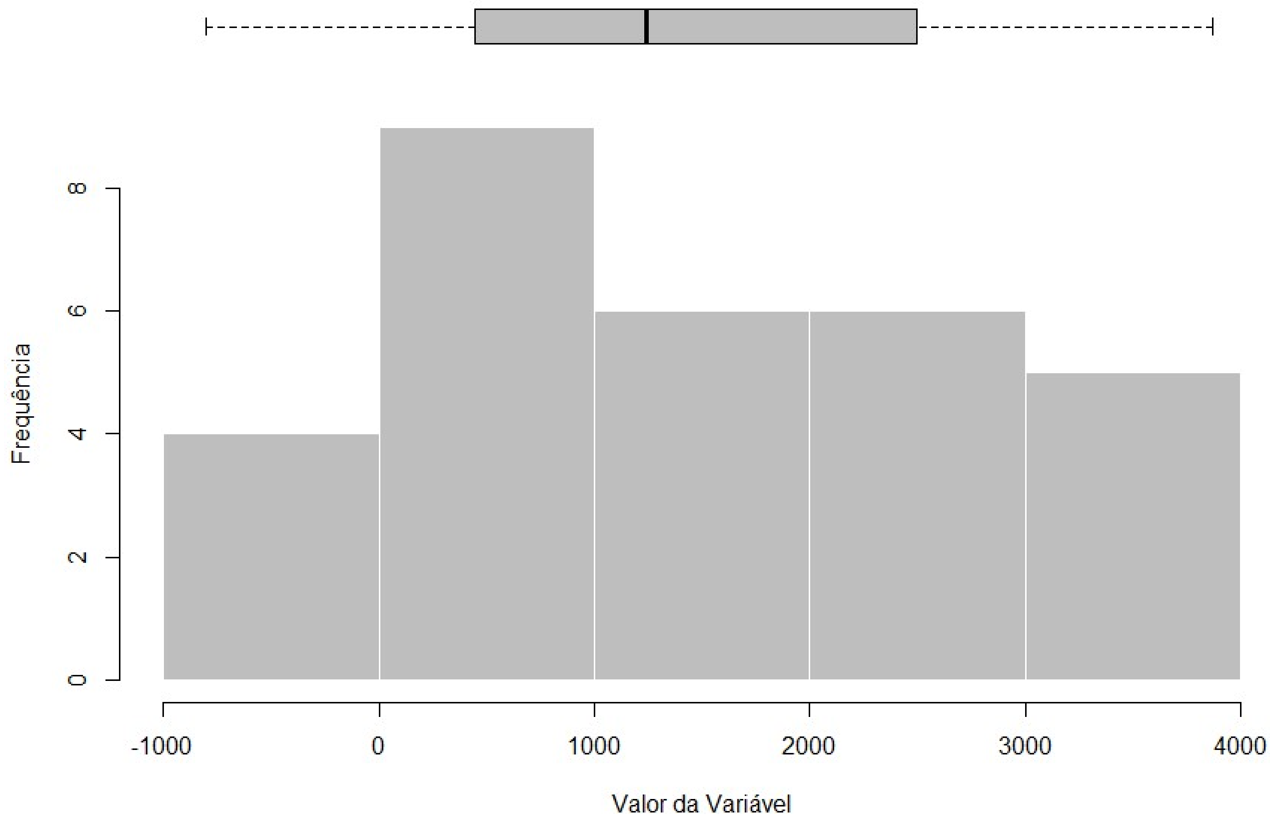
Modelo

- Pense nas suposições e verifique as condições.

Indique o modelo de distribuição de amostragem para a estatística.

Desejamos encontrar um intervalo de confiança de 95% para o lucro médio das apólices vendidas por esse representante de vendas. Temos dados de 30 apólices vencidas.

A seguir são mostrados o boxplot e o histograma desses valores.




```r
lucro <- c(222.80,463.35,2089.40,
           1756.23,-66.20,2692.75,
           1100.85,57.90,2495.70,
           3340.66,833.95,2172.70,
           1006.50,1390.70,3249.65,
           445.50,2447.50,-397.10,
           3255.60,1847.50,-397.31,
           3701.85,865.40,186.25,
           -803.35,1415.65,590.85,
           3865.90,2756.94,578.95)

# Layout para dividir a tela
layout(mat = matrix(c(1,2),2,1, byrow=TRUE),  height = c(1,8))

# Plotar o boxplot e o histograma
par(mar=c(0, 4.1, 0, 2.1))
boxplot(lucro , horizontal=TRUE , ylim=c(-1000,4000), xaxt="n" ,
        col="grey" , frame=F)
par(mar=c(4, 4.1, 1.1, 2.1))
hist(lucro , breaks=4 , col="grey" , border=F , main="" ,
     xlab="Valor da Variável", ylab="Frequência", xlim=c(-1000,4000))
```

A amostra parece ser unimodal e simétrica, com valores de lucro entre o valor – \$1000 e \$4000 e nenhum valor discrepante.

❖ Suposição de independência

Esta é uma amostra aleatória, portanto as observações devem ser independentes.

❖ Condição de Randomização

Essa amostra foi selecionada aleatoriamente das apólices vencidas vendidas pelo representante de vendas da empresa.

❖ Condição de Aproximadamente Normal

A distribuição dos lucros é unimodal e bastante simétrica, sem forte assimetria.

Vamos usar o modelo t de Student com $n - 1 = 30 - 1 = 29$ graus de liberdade e encontrar um intervalo t de uma amostra para a média.

Calcule estatísticas básicas e construa o intervalo de confiança.

```r
t1 <- summary(lucro)
t2 <- var(lucro)
t3 <- sd(lucro)

estats <- as.matrix(c(t1,t2,t3))
rownames(estats) <- c("Min.","1st Qu.","Median", "Mean", "3rd Qu.","Max.","Var.","SD")
colnames(estats) <- c("Estatísticas")
estats
```

```
> estats
         Estatísticas
Min.         -803.3500
1st Qu.       449.9625
Median       1245.7750
Mean         1438.9023
3rd Qu.      2483.6500
Max.         3865.9000
Var.      1767824.1534
SD           1329.5955
```

Lembre-se de que o erro padrão da média é igual ao desvio padrão dividido pela raiz quadrada de n.
O valor crítico que precisamos para fazer um intervalo de confiança de 95% vem da tabela t de um aluno, de um programa de computador ou de uma calculadora.

Nós temos 30 – 1 = 29 graus de liberdade.

Então procuramos o valor t correspondente em uma Tabela, ou obtemos a partir de um software.

O erro padrão da média é:

$$DP\left[\bar{X}\right] = \sqrt{Var\left[\bar{X}\right]} = \frac{s}{\sqrt{n}} = \frac{1329{,}5955}{\sqrt{30}} = 242{,}75$$

Existem 30 – 1 = 29 graus de liberdade. O gerente especificou um nível de confiança de 95%, portanto, o valor crítico (da Tabela) é 2,045.

A margem de erro é:

$$\begin{aligned} \text{Margem de Erro} &= 2{,}045 \times DP\left[\bar{X}\right] \\ &= 2{,}045 \times 242{,}75 \\ \text{Margem de Erro} &= \$496{,}42 \end{aligned}$$

O intervalo de confiança de 95% para o lucro médio é:

$$IC = \$1438{,}90 \pm \$496{,}42$$

$$IC = \left(\$942{,}48;\ \$1935{,}32\right)$$

Conclusão

Interprete o intervalo de confiança no contexto adequado.

Quando construímos intervalos de confiança dessa maneira, esperamos que 95% deles cubram a verdadeira média e 5% falhem em cobrir valor verdadeiro.

Isso é o que "95% de confiança" significa.

Da nossa análise das apólices selecionadas, estamos 95% confiantes de que o verdadeiro lucro médio das apólices vendidas por esse representante de vendas está contido no intervalo de $942,48 a $1.935,32.

Advertência:
Perdas de seguro são notoriamente sujeitas a outliers.
Uma perda muito grande poderia influenciar substancialmente o lucro médio. No entanto, não houve tais casos neste conjunto de dados.

```r
X.barra <- mean(lucro)
s <- sd(lucro)
n <- 30
alfa <- 0.05
gl <- n-1
t <- qt(alfa/2,gl,lower.tail = FALSE)

LI <- X.barra - t*(s/(n^0.5))   # Limite Inferior do IC
LS <- X.barra + t*(s/(n^0.5))   # Limite Superior do IC

IC <- c(LI,LS)
IC
```

```
> IC <- c(LI,LS)
> IC
[1]  942.4232 1935.3814
```

Cuidados com Interpretação de Intervalos de Confiança

Intervalos de confiança para médias oferecem interpretações novas, tentadoras e erradas. Aqui estão algumas maneiras de evitar comentários equivocados:

• Não diga: “95% de todas as apólices vendidas por esse representante de vendas têm lucros entre $942,48 e $1.935,32.” O intervalo de confiança é sobre a média, não sobre as medidas de apólices individuais.

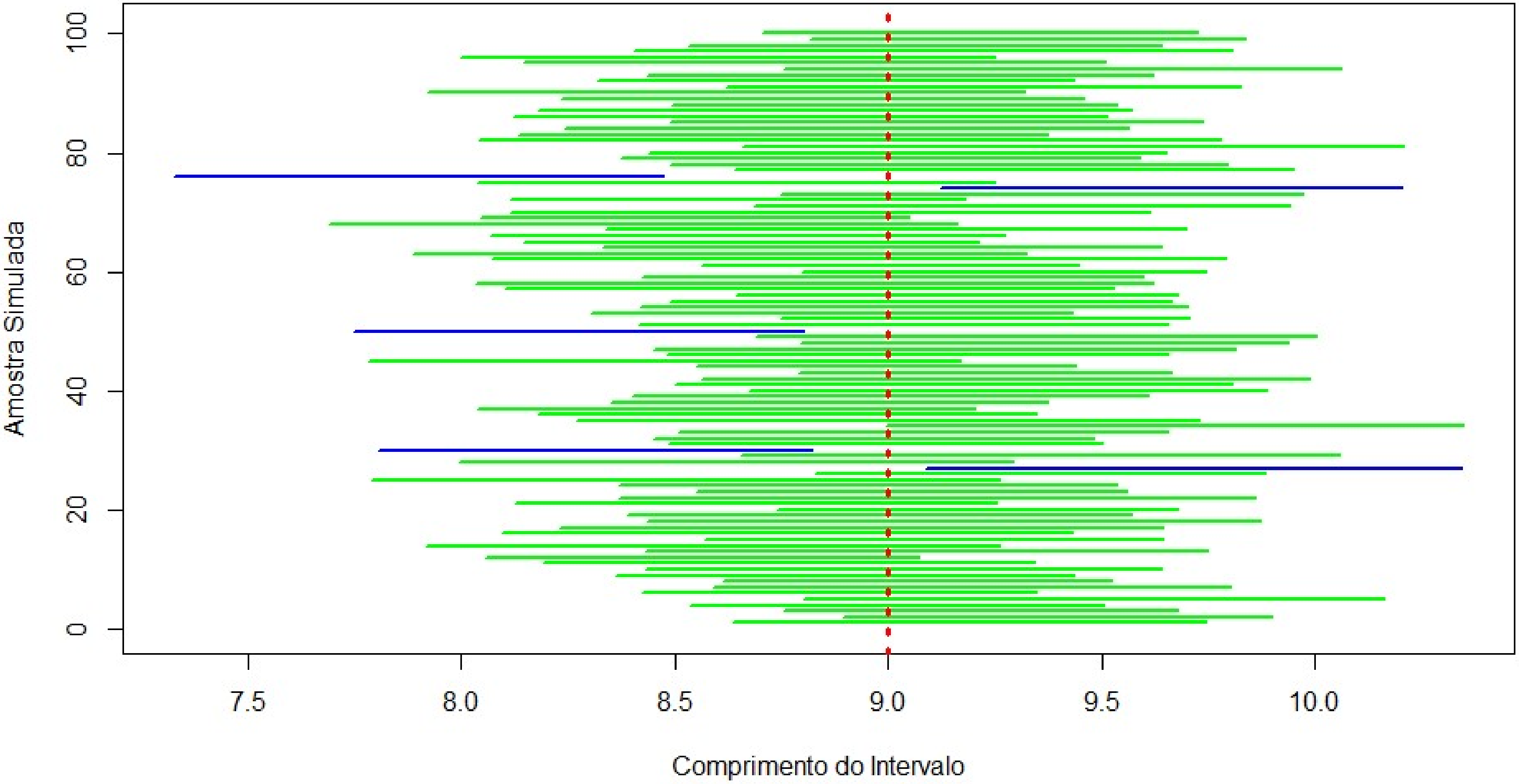
• Não diga: "Temos 95% de certeza de que uma apólice selecionada aleatoriamente terá um lucro líquido entre $942,48 e $1.935,32". Essa falsa interpretação também diz respeito a apólices individuais e não à média das apólices. Estamos 95% confiantes de que o lucro médio de todas as apólices (similares) vendidas por esse representante de vendas está entre $ 942,48 e $ 1.935,32.

• Não diga: "O lucro médio é de $1.438,90, 95% do tempo. “Essa afirmação é sobre média, mas ainda está errado. Isso implica que a média verdadeira varia, quando na verdade é o intervalo de confiança que teria sido diferente se tivéssemos obtido uma amostra diferente.

• Finalmente, não diga: “95% de todas as amostras terão lucros médios entre $ 942,48 e $1935,32”. Essa declaração sugere que esse intervalo de alguma forma define um padrão para todos os outros intervalos. De fato, esse intervalo não é mais (ou menos) provável de estar correto do que qualquer outro. Você poderia dizer que 95% de todas as amostras possíveis produziriam intervalos que contêm o verdadeiro lucro médio. (O problema é que, como nunca saberemos qual é o verdadeiro lucro médio, não podemos saber se nossa amostra foi um desses 95%.)

```r
dev.off(dev.list()["RStudioGD"])
rm(list=ls())
cat("\f")

# Verifica como os IC são aleaórios

amostra = c(6.2,6.6,7.1,7.4,7.6,7.9,8,8.3,8.4,8.5,8.6,8.8,8.8,9.1,9.2,9.4,
            9.4,9.7,9.9,10.2,10.4,10.8,11.3,11.9)

amostra
mean(amostra)

# O Teste-t fornece um IC baseado na amostra em que não conhececemos a variância.

t.test(amostra)

# Com base nos dados da amostra, extraímos 100 amostras aleatórias com média igual a 9
# e desvio-padrão igual ao da amostra.

simulacoes <- 100
mi <- 9;
n <- 24;
desvpad.amostra = sd(amostra);
desvpad.amostra

m <- matrix(rnorm(n*simulacoes, mi, desvpad.amostra), n)
```

```r
# O Teste-t pode ser usado para ter o IC de um amostra de cada vez.

f.IC <- function(x) t.test(x)$conf.int
IC <- apply(m, 2, f.IC)
IC
#IC[1, ]
#IC[2, ]
sum(IC[1, ] <= mi & IC[2, ] >= mi)

plot(range(IC), c(0, 1 + simulacoes), type = "n", xlab = "Comprimento do Intervalo",
     ylab = "Amostra Simulada")

for (i in 1:simulacoes){
  if (IC[1,i]<mi & IC[2,i]>mi){
    lines(IC[, i], rep(i, 2), lwd = 2,col="green")
  }else{
    lines(IC[, i], rep(i, 2), lwd = 2,col='blue')
  }
}

abline(v = 9, lwd = 3, lty = 3, col="red")
```


Amostra Simulada
Comprimento do Intervalo
0
20
40
60
80
100
7.5
8.0
8.5
9.0
9.5
10.0

CASO 3: IC para p (proporção)

p : proporção de sucessos na população

X : número de sucessos em n realizações independentes

n : tamanho da amostra (conhecido)

$$X \longrightarrow B\,(n;\,p)$$

Estimador pontual de p: $\hat{p} = X/n$

Distribuição assintótica ($n \to \infty$):

$$X \xrightarrow{n \to \infty} N\left(np; np(1-p)\right)$$

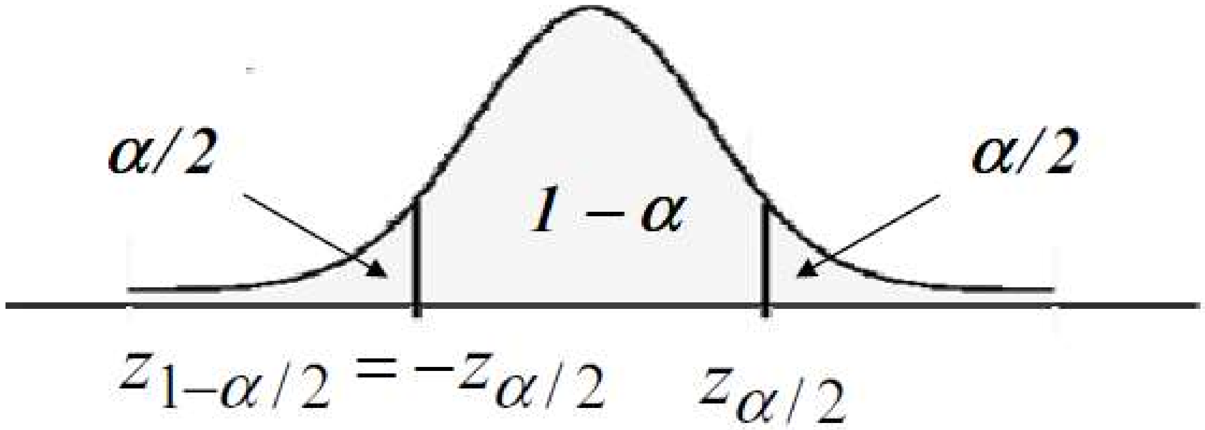
Estimação por Intervalos

$$\hat{p} = \frac{X}{n} \xrightarrow{n \to \infty} N\left(p; \frac{p(1-p)}{n}\right)$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \xrightarrow{n \to \infty} N(0; 1)$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \xrightarrow{n \to \infty} N(0;1)$$

Estimação por Intervalos



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

$$P[-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq z_{\alpha/2}] = 1-\alpha$$

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

IC para o parâmetro proporção ao nível de confiança $1 - \alpha$.

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

Conhecido como IC Clássico de Wald.

Estimação por Intervalos – Wilson Score

$$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n} \rightarrow \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}},\ \text{ou seja, } \hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Temos que:

$$\hat{p} - p = z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}},\ \text{elevando ao quadrado permite escrever}$$

$$\frac{(\hat{p}-p)^2}{\sigma_{\hat{p}}^2} = z_{\frac{\alpha}{2}}^2.$$

Estimação por Intervalos – Wilson Score

Substituindo $\sigma_{\hat{p}}^2$ na equação anterior

$$\frac{\left(\hat{p}-p\right)^2}{\dfrac{p\left(1-p\right)}{n}} = z_{\frac{\alpha}{2}}^2 \rightarrow \left(\hat{p}-p\right)^2 = \frac{z_{\alpha/2}^2}{n}\left(p-p^2\right)$$

$$\hat{p}^2 - 2\hat{p}p + p^2 = \frac{z_{\alpha/2}^2}{n}\left(p-p^2\right)$$

$$\hat{p}^2 - 2\hat{p}p + p^2 = \frac{z_{\alpha/2}^2}{n}p - \frac{z_{\alpha/2}^2}{n}p^2$$

$$\hat{p}^2 - 2\hat{p}p + p^2 - \frac{z_{\alpha/2}^2}{2}p + \frac{z_{\alpha/2}^2}{n}p^2 = 0$$

Estimação por Intervalos – Wilson Score

$$\left(1+\frac{z_{\alpha/2}^{2}}{n}\right)p^{2}-2\left(\hat{p}+\frac{z_{\alpha/2}^{2}}{2n}\right)p+\hat{p}^{2}=0.$$

Resolvendo em p, produz

$$p=\frac{\hat{p}+\frac{z_{\alpha/2}^{2}}{2n}\mp\sqrt{\left(\hat{p}+\frac{z_{\alpha/2}^{2}}{2n}\right)^{2}-\left(1+\frac{z_{\alpha/2}^{2}}{n}\right)\hat{p}^{2}}}{1+\frac{z_{\alpha/2}^{2}}{n}}.$$

Estimação por Intervalos – Wilson Score

Simplificando o radical,

$$p = \frac{\hat{p} + \dfrac{z_{\alpha/2}^2}{2n} \mp z_{\alpha/2}\sqrt{\dfrac{\hat{p}\left(1-\hat{p}\right)}{n} + \dfrac{z_{\alpha/2}^2}{4n^2}}}{1 + \dfrac{z_{\alpha/2}^2}{n}}.$$

Conhecido como IC de Escores de Wilson.

Estimação por Intervalos
CASO 3: IC para p (proporção)

1. Estatística: $\widehat{P} = \frac{X}{n}$ (proporção amostral),
   $X =$ no. de sucessos no experimento binomial.
   Propriedades: $E[\widehat{P}] = p$; $\quad Var[\widehat{P}] = \frac{pq}{n}$
   Padronizando $\widehat{P}$ (para $n\widehat{p} \geq 5$ ou $n\widehat{q} \geq 5$), temos:

$$Z = \frac{\widehat{P} - p}{\sqrt{pq/n}} \dot{\sim} N(0, 1)$$

2. O IC $100(1-\alpha)\%$ para p é dado por:

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

OBS:

– Quando $n \downarrow$ e $p \rightarrow 0$ (ou $p \rightarrow 1$), este procedimento não é válido.

– Regra de bolso: deve-se ter $n\hat{p} \geq 5$ e $n\hat{q} \geq 5$.

Exemplo 3:

Em uma a.a. de tamanho $n = 500$ de famílias assinantes de TV a cabo em SJC, um total de $x = 340$ possuíam assinatura de um canal de filmes.

1. Qual o IC 95% para a proporção na população de assinantes de TV a cabo em SJC que têm acesso ao canal?
2. Qual o erro máximo que se comete com esta estimativa?
3. Qual o tamanho da amostra mínimo que garanta confiança de 95% de que o erro máximo desta estimativa seja de 0,02?

Exemplo 4 – Solução:

Temos $n = 500$, $x = 340$.

1. IC 95%: $\alpha = 1 - 0,95 = 0,05 \overset{tabela}{\Rightarrow} z_{0,025} = 1,96$.
   $\hat{p} = 340/500 = 0,68$ $\hat{q} = 1 - \hat{p} = 0,32$
   Portanto,
   IC 95%: $0,68 \pm (1,96)\sqrt{(0,68 \cdot 0,32)/500} \Rightarrow 0,68 \pm 0,04$.
   Temos 95% de certeza que a proporção de assinantes que possuem o canal está no intervalo $(0,64; 0,72)$.

2. erro $< e = z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} = 0,04$, com 95% de confiança.
3. $n = (z_{\alpha/2}/e)^2\hat{p}\hat{q} = (1,96/0,02)^2 0,68 \cdot 0,32 = 2089,8 \quad \therefore n = 2090$.
   Garante 95% de confiança de cometer erro máximo de 0,02, ao estimar que a proporção de assinantes que possuam o canal seja 0,68.

CASO 4: IC para σ^2 (pop. Normal)

Estimador $$S^2 = \frac{1}{n-1}\sum_i \left(X_i - \bar{X}\right)^2$$

Distribuição Amostral

$$\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \rightarrow \chi^2_{n-1}$$

Distribuição Amostral de S_n^2

$X_1, X_2, \ldots, X_n \sim f_X(x)$: a.a.

Variância amostral:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \sim \ ?$$

Atenção

Não existe um teorema análogo ao TLC que garanta uma aproximação para a distribuição amostral de S_n^2 (mesmo quando $n \to \infty$), para uma distribuição $f_X(x)$ arbitrária...

Distribuição Amostral de S_n^2

A distribuição exata para S_n^2 pode ser obtida para o caso normal:

Teorema

$X_1, X_2, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$

Variância amostral:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Então:

$$Q^2 = \frac{(n-1)S_n^2}{\sigma^2} = \sum_{i=1}^{n}\frac{(X_i - \bar{X}_n)^2}{\sigma^2} \sim \chi_\nu^2$$

$\nu = n - 1$ graus de liberdade

Observação:

- O valor esperado (média) da distribuição χ^2 é igual ao seu número de graus de liberdade.

- A variância da distribuição χ^2 é igual a duas vezes o seu número de graus de liberdade.

Se uma amostra aleatória de tamanho n é tirada de uma população normalmente distribuída, então

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right]=(n-1)$$

O que resulta em

$$\frac{(n-1)}{\sigma^2}E\left[S^2\right]=(n-1)\Rightarrow E\left[S^2\right]=(n-1)\frac{\sigma^2}{(n-1)}\Rightarrow E\left[S^2\right]=\sigma^2.$$

De maneira similar, podemos obter a variância de S^2 da seguinte forma

$$Var\left[\frac{(n-1)S^2}{\sigma^2}\right]=2(n-1),$$

$$\left(\frac{n-1}{\sigma^2}\right)^2 Var\left[S^2\right]=2(n-1)\Rightarrow Var\left[S^2\right]=2(n-1)\left(\frac{\sigma^2}{n-1}\right)^2,$$

$$Var\left[S^2\right]=\frac{2\sigma^4}{n-1}.$$

Distribuição Amostral de S_n^2

Interpretação dos g.d.l. como medida de informação:

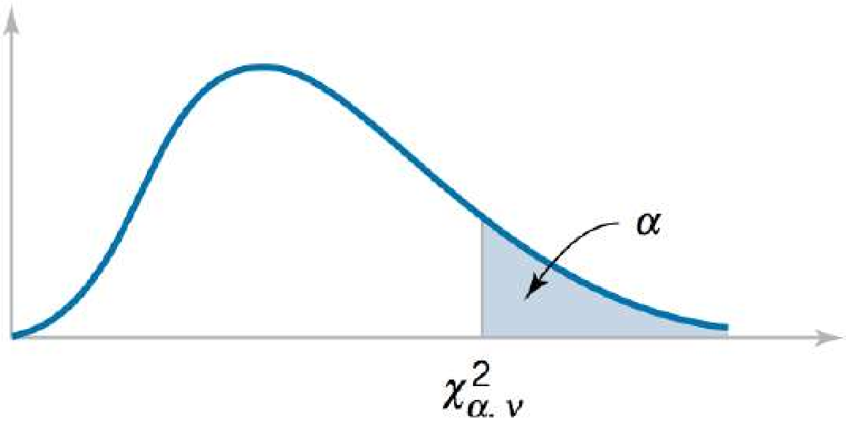
– $\exists$ n g.d.l. na a.a. $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$
– quando não conhecemos o parâmetro populacional μ, temos que usar $\bar{X}_n$ para estimá-lo, ou seja, perdemos 1 g.d.l. na estimação de μ.
– portanto, quando queremos estimar σ_2 usando S_n^2, temos 1 g.d.l. a menos (n-1).

Distribuição Amostral de S_n^2

Relembrando...

Distribuição Chi-quadrado

$\chi^2_\nu = Gamma(\nu/2, 2)$, em que:
$\nu \in \mathcal{Z}^+ =$ graus de liberdade



(ver Tabela χ^2)

Distribuição Amostral de S_n^2

Exemplo

Uma companhia produz medicamentos que contém 8g de um determinado composto em um frasco do medicamento. Engenheiros da qualidade determinaram que o processo estará operando em conformidade com o especificado se a variabilidade real σ^2 do composto por frasco for menos que 0,0025.

Uma a.a. de 10 frascos é selecionada e a quantidade do composto em cada frasco é medida. Assuma que a quantidade do composto seja normalmente distribuída. Estamos interessados na variância amostral, Sn^2.

Se, de fato, $\sigma^2 = 0,001$, qual a probabilidade de que S_n^2 exceda 0,0025?

Distribuição Amostral de S_n^2

Temos: $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$
Queremos: $P[S_n^2 > 0,0025]$:

(i) $\chi^2 = (n-1)S_n^2/\sigma^2 \sim \chi^2_{\nu=n-1}$

(ii)

$$P[S_n^2 > 0,0025] = P\left[(n-1)\frac{S_n^2}{\sigma^2} > (n-1)\frac{0,0025}{\sigma^2}\right]; \quad n = 10; \sigma^2 = 0,001$$

$$= P\left[\chi^2 > 9\frac{0,0025}{0,001}\right] = P[\chi^2 > 22,5]$$

Da tabela, $\nu = 9$, $\chi^2_\alpha = 22,5$.
Na linha de $\nu = 9$, temos $\chi^2_{0,01} = 21,67$ e $\chi^2_{0,005} = 23,59$.
Portanto:
$0,005 < P[\chi^2 > 22,5] < 0,01$, se a variância populacional for, de fato, $\sigma^2 = 0,001$.

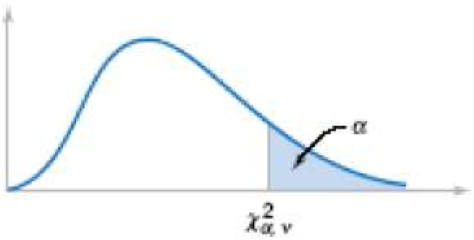
```
> pchisq(22.5,9, lower.tail = FALSE)
[1] 0.007422449
```

Distribuição amostral de s^2

Table III Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution

ν \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	$\boxed{21.67}$	$\boxed{23.59}$
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

ν = degrees of freedom.



Distribuição Amostral de S_n^2

Exemplo

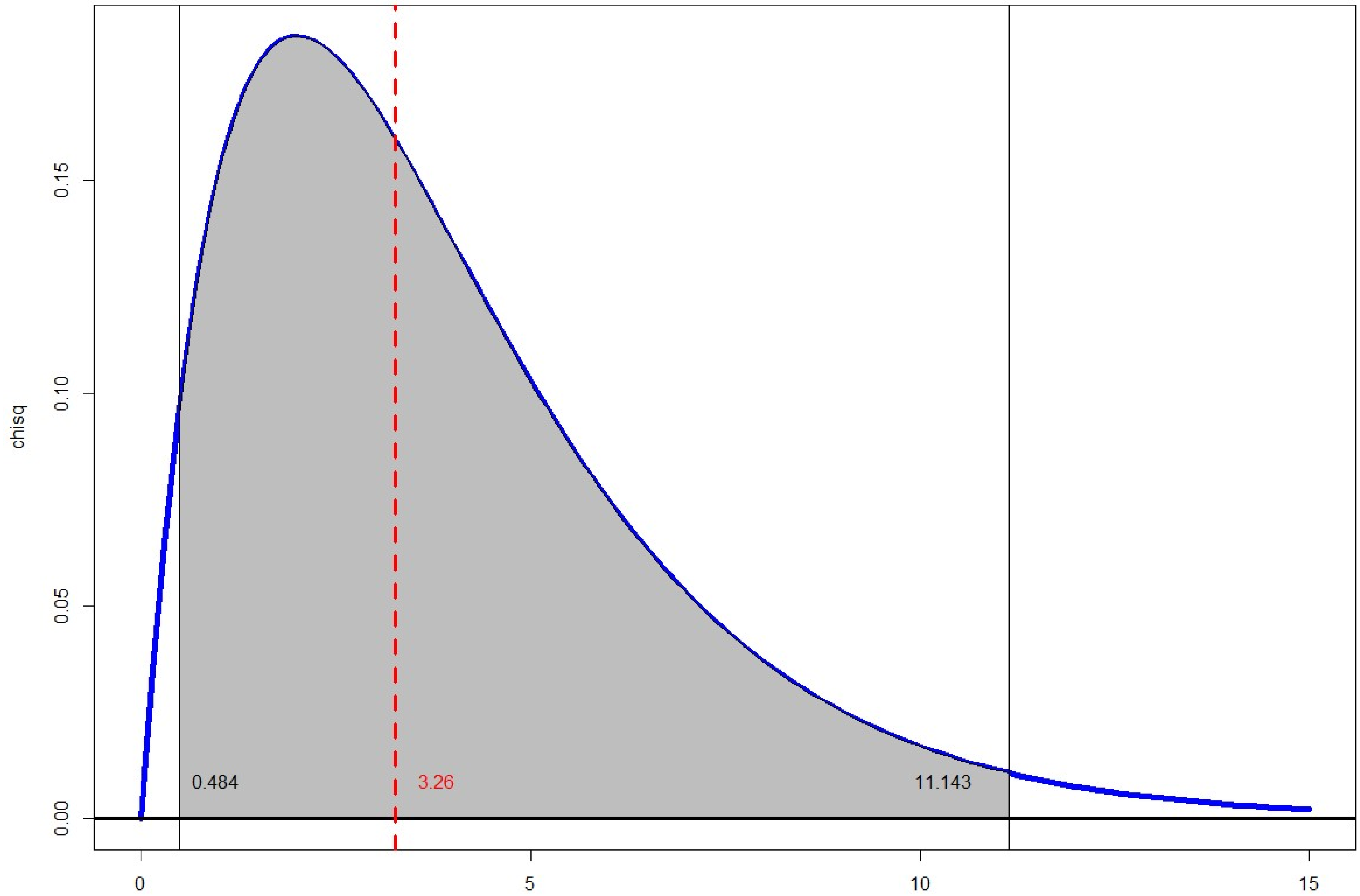
Um fabricante de baterias de carro garante que seus produtos duram, em média, 3 anos com desvio-padrão de 1 ano. Assuma que a vida útil da bateria segue a distribuição normal.

Se 5 destas baterias têm vida útil de 1,9, 2,4, 3,0, 3,5 e 4,2, deve-se acreditar que o desvio-padrão é de, realmente, 1 ano?

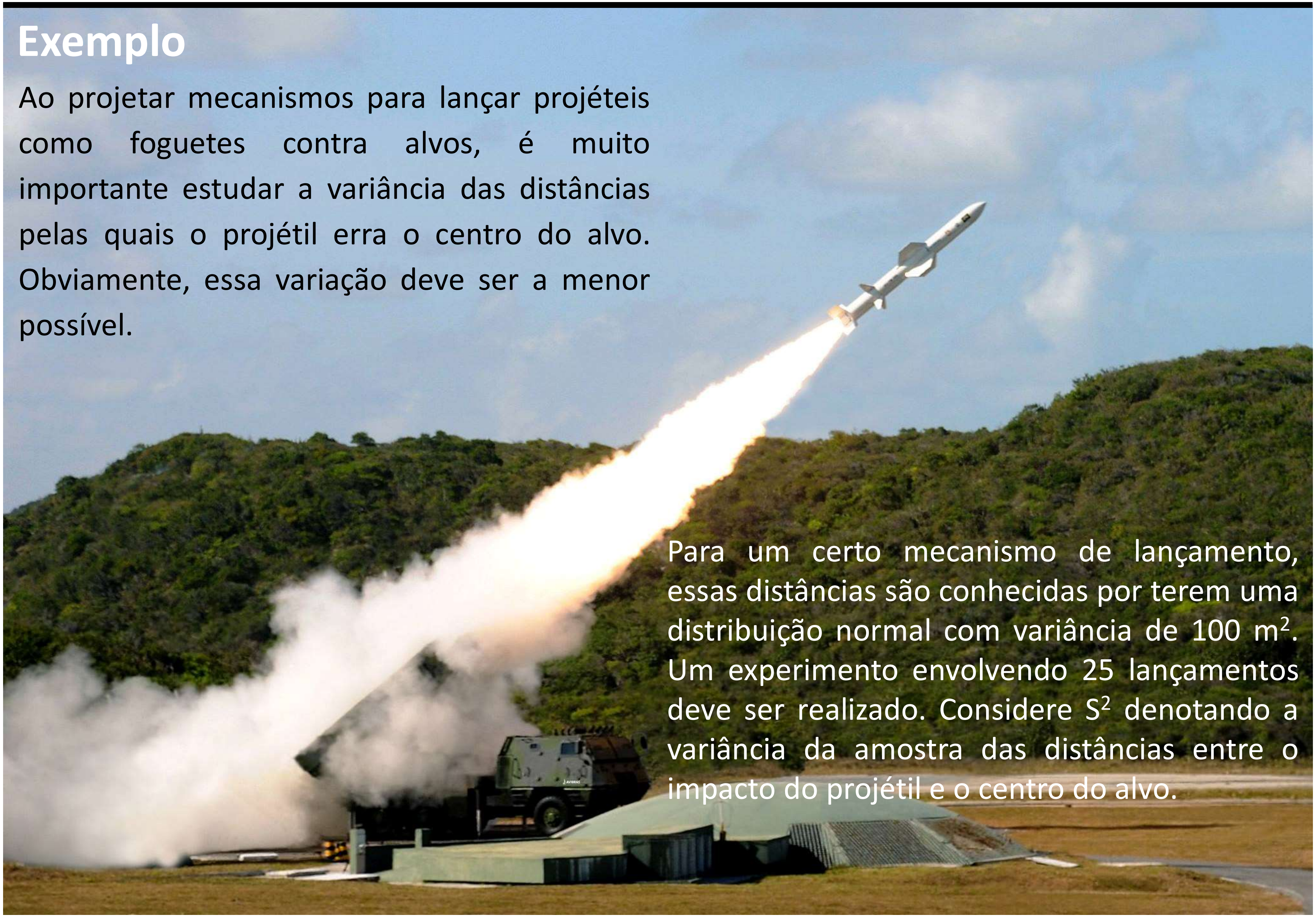
Solução:

i. Variância amostral: $S_n^2 = \sum(X_i - \bar{X})/(n-1) = \ldots = 0,815$
ii. $\chi^2 = (n-1)S_n^2/\sigma^2 = 5 \cdot 0,815/1 = 3,26$, com $\nu = n-1 = 4$ g.d.l.
iii. Para $\nu = 4$, temos $\chi^2_{0,975;4} = 0,48$ e $\chi^2_{0,025;4} = 11,14$

Portanto, o valor $\sigma^2 = 1$ parece razoável.


chisq
0.00
0.05
0.10
0.15
0
5
10
15
0.484
3.26
11.143

```
dev.off(dev.list()["RStudioGD"])
rm(list=ls())
cat("\f")

pchisq(0.484,4, lower.tail = FALSE)
pchisq(11.14,4, lower.tail = FALSE)

df <- 4;
chisq <- function(x) (1/((2^(df/2))*gamma(df/2)))*(x^((df/2)-1))*exp(-x/2)
plot(chisq,0,15,lwd=5, col="blue")
polygon(x=c(0.484,seq(0.484,11.143,l=100),11.143), y=c(0,chisq(seq(0.484,11.143,l=100)), 0), col="gray")
abline(h=0,lwd=3,col="black")
abline(v=c(0.484,3.26,11.143), col=c("black","red", "black"), lty=c(1,2,1), lwd=c(1,3,1))
text(c(0.95,3.8, 10.3), c(0.009,0.009, 0.009), c(expression(0.484,3.26,11.143)),
     col=c("black","red", "black"))
```



Exemplo

Ao projetar mecanismos para lançar projéteis como foguetes contra alvos, é muito importante estudar a variância das distâncias pelas quais o projétil erra o centro do alvo. Obviamente, essa variação deve ser a menor possível.

Para um certo mecanismo de lançamento, essas distâncias são conhecidas por terem uma distribuição normal com variância de 100 m^2. Um experimento envolvendo 25 lançamentos deve ser realizado. Considere S^2 denotando a variância da amostra das distâncias entre o impacto do projétil e o centro do alvo.

a) Determine $P(S^2 > 50)$.

b) Determine $P(S^2 > 150)$.

c) Encontre $E[S^2]$, o valor esperado da variância amostral.

d) Encontre $Var[S^2]$, a variância da variância amostral.

a) Determine $P(S^2 > 50)$.

Seja $Q = (n-1)S^2/\sigma^2$ tendo distribuição $\chi^2(24)$ para $n = 25$.

$P(S^2 > 50) = P((n-1)S^2/\sigma^2 > (25-1)50/100) = P(Q^2 > 12) = 0{,}9799$

b) Determine $P(S^2 > 150)$.

$P(S^2 > 150) = P((n-1)S^2/\sigma^2 > (25-1)150/100) = P(Q^2 > 36) = 0{,}0549$

c) Encontre $E[S^2]$, o valor esperado da variância amostral.

$E[S^2] = \sigma^2 = 100$

d) Encontre $Var[S^2]$, a variância da variância amostral.

$Var[S^2] = 2\,\sigma^4/(n-1) = (2(100)^2)/24 = (100)^2/12 = 833{,}3333$

De acordo com este resultado, o desvio padrão de S^2 é $\sqrt{833,3333} \approx 29$.

Então, pelo teorema de Tchebysheff, pelo menos 75% dos valores, em amostragens repetidas, devem situar-se entre 100 ± 2(29), ou seja, entre 42 e 158.

Como consideração prática, essa grande variação nas variâncias da distância pode fazer com que o engenheiro reavalie o projeto do mecanismo de lançamento.

Distribuição Qui-quadrado

Função Densidade de Probabilidade

$$f_{\chi^2_\nu}(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}, \qquad x > 0.$$

QUI - QUADRADO

$$f(x) = \frac{1}{2^{\frac{m}{2}}\,\Gamma\left(\frac{m}{2}\right)}\,x^{\frac{m}{2}-1}\,\mathrm{Exp}\left[-\frac{x}{2}\right]$$

In[8]:= $\int_0^{\infty} \frac{1}{2^{\frac{3}{2}}\,\mathrm{Gamma}\left[\frac{3}{2}\right]}\,x^{\frac{3}{2}-1}\,\mathrm{Exp}\left[-\frac{x}{2}\right]\,dx$

Out[8]= 1

In[9]:= $\int_0^{7.81} \frac{1}{2^{\frac{3}{2}}\,\mathrm{Gamma}\left[\frac{3}{2}\right]}\,x^{\frac{3}{2}-1}\,\mathrm{Exp}\left[-\frac{x}{2}\right]\,dx$

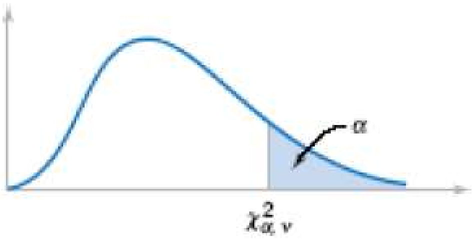
Out[9]= 0.949894

Distribuição amostral de s^2

Table III Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution

ν \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

ν = degrees of freedom.



Exemplo

Calcular

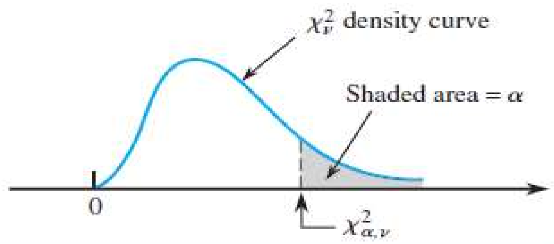
a) $\chi^2_{0,1;15}$

b) $\chi^2_{0,005;25}$

c) $\chi^2_{0,99;25}$

d) $\chi^2_{0,995;25}$

Table A.7 Critical Values for Chi-Squared Distributions

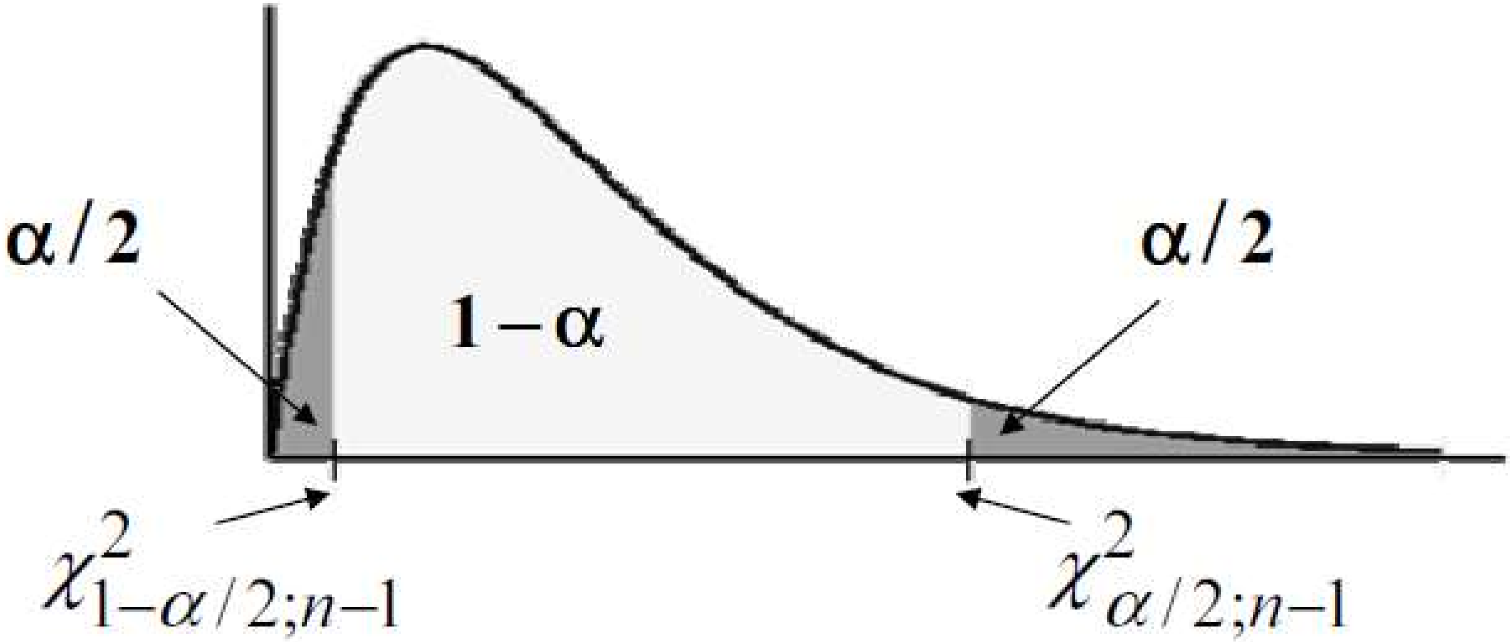


	α									
ν	**.995**	**.99**	**.975**	**.95**	**.90**	**.10**	**.05**	**.025**	**.01**	**.005**
1	0.000	0.000	0.001	0.004	0.016	2.706	3.843	5.025	6.637	7.882
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.344	12.837
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.085	16.748
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.440	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.012	18.474	20.276
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.534	20.090	21.954
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.022	21.665	23.587
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.724	26.755
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.735	27.687	29.817
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.600	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.577	32.799
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.407	7.564	8.682	10.085	24.769	27.587	30.190	33.408	35.716
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.843	7.632	8.906	10.117	11.651	27.203	30.143	32.852	36.190	38.580
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.033	8.897	10.283	11.591	13.240	29.615	32.670	35.478	38.930	41.399
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.195	11.688	13.090	14.848	32.007	35.172	38.075	41.637	44.179
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.519	11.523	13.120	14.611	16.473	34.381	37.652	40.646	44.313	46.925
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.807	12.878	14.573	16.151	18.114	36.741	40.113	43.194	46.962	49.642
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.120	14.256	16.147	17.708	19.768	39.087	42.557	45.772	49.586	52.333

Estimação por Intervalos

$$P\left[\chi^2_{1-\alpha/2;n-1} \leq \chi^2_{n-1} \leq \chi^2_{\alpha/2;n-1}\right] = 1-\alpha$$

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



Estimação por Intervalos

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

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

Estimação por Intervalos

$$P\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2;\,n-1}} \geq \sigma^2 \geq \frac{(n-1)S^2}{\chi^2_{\alpha/2;\,n-1}}\right] = 1-\alpha$$

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

IC para σ^2, com média populacional desconhecida, ao nível de confiança $1 - \alpha$.

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

Estimação por Intervalos

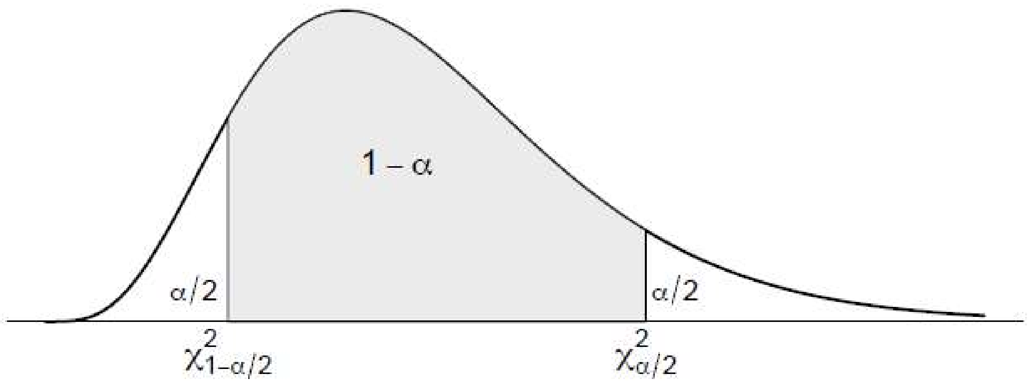
CASO 4: IC para σ^2

Seja $X \sim N(\mu, \sigma^2)$
Seja S_n a variância de uma a.a. de tamanho n.

$$Q^2 = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{\nu=n-1}$$

Portanto, o IC $100(1-\alpha)\%$ para σ^2 é dado por:

$$\frac{(n-1)s_n^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s_n^2}{\chi^2_{1-\alpha/2}}$$



Exemplo 5

Uma companhia produtora de sementes de vegetais coletou uma amostra de 10 pacotes de sementes, cujos pesos (em g) valem:

46,4 46,1 45,8 47,0 46,1 45,9 45,8 46,9 45,2 46,0

Qual o IC 95% para a variância dos pacotes de sementes?
Assuma população normalmente distribuída.

Exemplo 5 – Solução:

Temos $s_n^2 = 0,286$; $n = 10$
Vamos supor $X \sim$ Normal.

IC 95% (σ^2):
$\alpha = 1 - 0,95 = 0,05 \quad \Rightarrow \alpha/2 = 0,025$
$\nu = n - 1 = 9$ g.d.l.
$\chi_U^2 = \chi_{\alpha/2,\nu}^2 = \chi_{0,025,9}^2 \overset{tabela}{=} 19,023$
$\chi_L^2 = \chi_{1-\alpha/2,\nu}^2 = \chi_{0,975,9}^2 \overset{tabela}{=} 2,700$
Portanto,

$$(n-1)\frac{S_n^2}{\chi_{\alpha/2}^2} < \sigma^2 < (n-1)\frac{S_n^2}{\chi_{1-\alpha/2}^2}$$

$$9 \cdot \frac{0,286}{19,023} < \sigma^2 < 9 \cdot \frac{0,286}{2,700} \quad \Rightarrow \quad \text{IC } 95\%(\sigma^2) : (0,135; 0,953)$$

Estimação Envolvendo Duas Amostras

Veremos os seguintes casos:

CASO 1: IC para a diferença entre duas médias: $\mu_1 - \mu_2$

- 1.1. σ_1^2, σ_2^2 conhecidas
- 1.2. $\sigma_1^2 = \sigma_2^2 = \sigma^2$ desconhecida
- 1.3. $\sigma_1^2 \neq \sigma_2^2$ desconhecidas
- 1.4. observações emparelhadas (amostras dependentes)

CASO 2: IC para a diferença entre duas proporções $p_1 - p_2$

CASO 3: IC para a razão entre duas variâncias σ_1^2/σ_2^2

CASO 1: IC para a diferença entre duas médias: $\mu_1 - \mu_2$

1.1. σ_1^2, σ_2^2 conhecidas

Estimação por Intervalos Envolvendo Duas Amostras

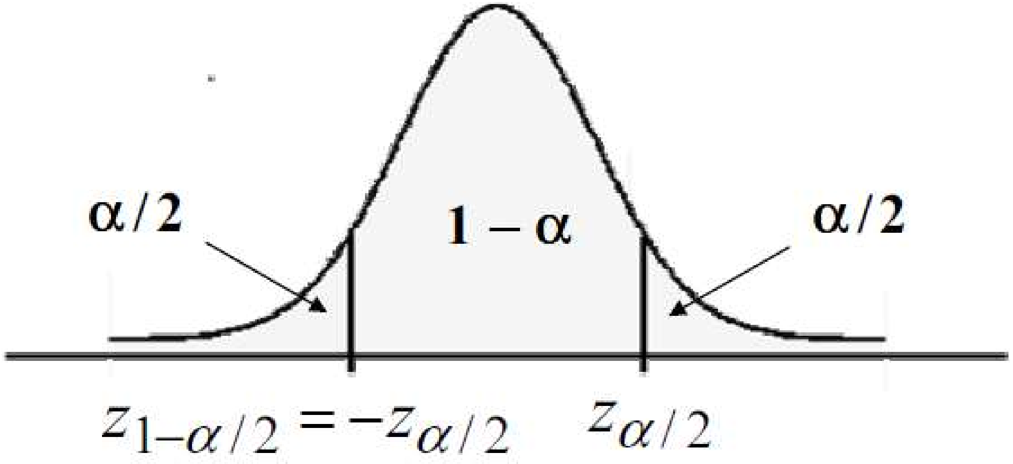
Estimador $T = \bar{X} - \bar{Y}$

Distribuição Amostral

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \rightarrow N(0;1)$$

Estimação por Intervalos Envolvendo Duas Amostras

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



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

$$P\left[-z_{\alpha/2} \le \frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}} \le z_{\alpha/2}\right] = 1-\alpha$$

Estimação por Intervalos Envolvendo Duas Amostras

$$P\left[\left(\bar{X}-\bar{Y}\right)-z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}} \le \mu_X-\mu_Y \le \right.$$

$$\left. \le \left(\bar{X}-\bar{Y}\right)+z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}\right]=1-\alpha$$

IC para $\mu_X - \mu_Y$, com variâncias populacionais conhecidas, ao nível de confiança $1 - \alpha$.

$$\left[(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \right]$$

Caso 1.1: $\mu_1 - \mu_2$ com σ_1^2 e σ_2^2 conhecidas

População 1: μ_1, σ_1^2, a.a. de tamanho n_1
População 2: μ_2, σ_2^2, a.a. de tamanho n_2

1. Estatística: $W = \overline{X}_1 - \overline{X}_2$
   Propriedades: $E[W] = \mu_1 - \mu_2$; $\quad Var[W] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
2. Padronizando W:

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

3. O IC $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ é dado por:

$$(\overline{x}_1 - \overline{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Estimação por Intervalos Envolvendo Duas Amostras

CASO 1: IC para $\mu_1 - \mu_2$

Hipóteses:

- As duas a.a. são selecionadas independentemente a partir de cada uma das populações (i.e., a escolha dos elementos de uma a.a. não afeta e não é afetada pela escolha dos elementos da outra a.a.)
- Ambas as amostras são suficientemente grandes para que o TLC seja válido ($n_1, n_2 \geq 30$)
- As variâncias populacionais são conhecidas.

Nota:

- O nível de confiança é exato quando $f_i(X) \sim N(\mu_i, \sigma_i^2), \ i = 1, 2$
- Para populações não normais, o TLC garante boa aproximação para a.a.'s suficientemente grandes.

Exemplo 6

Deseja-se comparar a eficiência de dois tipos de motores, A e B. Para tanto, foram testados 50 motores do tipo A e 75 motores do tipo B. O consumo médio amostral para o motor A foi de 36 mi/gal e para o motor B foi de 42 mi/gal. Assumindo que os desvios-padrão populacionais valham $\sigma_A = 6$ e $\sigma_B = 8$, construa um IC 96% para $\mu_A - \mu_B$.

Exemplo 6 – Solução:

Temos: $\begin{array}{lll} n_A = 50 & \bar{x}_A = 36 & \sigma_A = 6 \\ n_B = 75 & \bar{x}_B = 42 & \sigma_B = 8 \end{array}$

Portanto, $w = \bar{x}_B - \bar{x}_A = 42 - 36 = 6$
$\alpha = 1 - 0,96 = 0,04 \quad \Rightarrow \alpha/2 = 0,02$
Da tabela, $z_{\alpha/2} = 2,05$
Então,

$$6 \pm 2,05\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

$$6 \pm 2,05\sqrt{\frac{36}{50} + \frac{64}{75}} \quad \Rightarrow \quad \text{IC } 96\% : \quad 3,43 < \mu_B - \mu_A < 8,57$$

CASO 1: IC para a diferença entre duas médias: $\mu_1 - \mu_2$

1.2. $\sigma_1^2 = \sigma_2^2 = \sigma^2$ desconhecida

Estimador $T = \overline{X} - \overline{Y}$

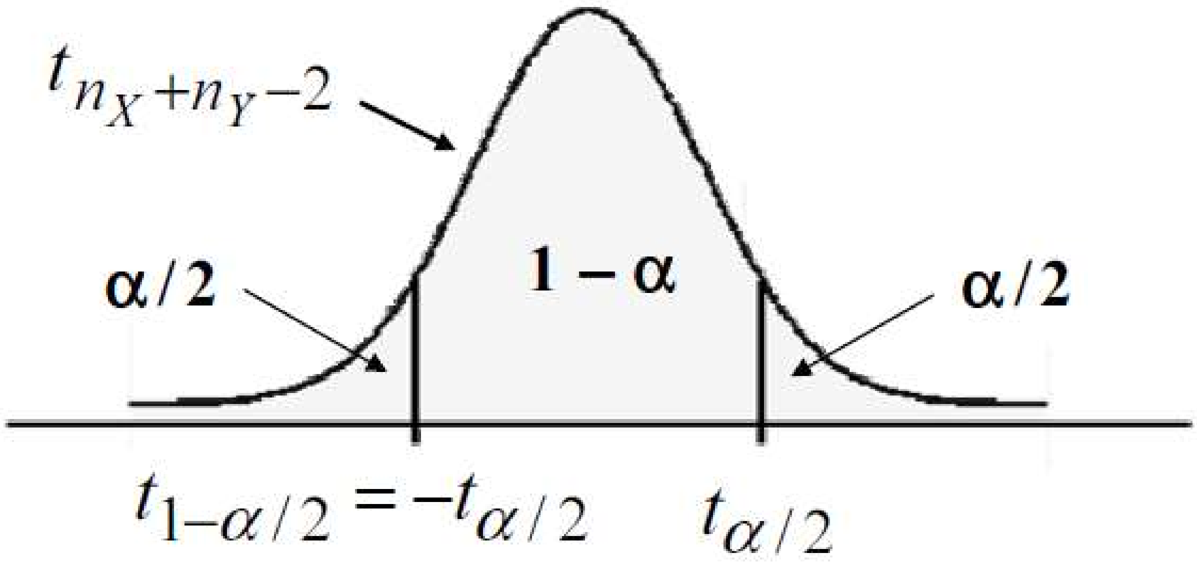
Distribuição Amostral

$$T = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_x - \mu_y\right)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \rightarrow t_{n_x + n_y - 2}$$

$$S_p = \sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}}$$

Estimação por Intervalos Envolvendo Duas Amostras

$$P\left[-t_{\alpha/2;n_X-n_Y-2} \leq t_{n_X-n_Y-2} \leq t_{\alpha/2;n_X-n_Y-2}\right] = 1-\alpha$$



Estimação por Intervalos Envolvendo Duas Amostras

$$P\left[-t_{\alpha/2;n_X-n_Y-2} \le t_{n_X-n_Y-2} \le t_{\alpha/2;n_X-n_Y-2}\right] = 1-\alpha$$

$$P\left[-t_{\alpha/2;\,n_X+n_Y-2} \le \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)}{S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}} \le t_{\alpha/2;\,n_X+n_Y-2}\right] =$$

$$P\left[\bar{X}-\bar{Y}-t_{\alpha/2;\,n_X+n_Y-2}\times S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}} \le \mu_X-\mu_Y \le\right.$$

$$\left.\le \bar{X}-\bar{Y}+t_{\alpha/2;\,n_X+n_Y-2}\times S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\right] = 1-\alpha$$

IC para $\mu_X - \mu_Y$, ao nível de confiança $1 - \alpha$, com variâncias populacionais desconhecidas porém iguais e amostras pequenas.

$$\left[\bar{X} - \bar{Y} \pm t_{\alpha/2;\, n_X + n_Y - 2} \times S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}\right]$$

Caso 1.2: $\mu_1 - \mu_2$ com $\sigma_1^2 = \sigma_2^2 = \sigma^2$ desconhecida

1. Estatística:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2[1/n_1 + 1/n_2]}} \dot{\sim} N(0,1)$$

2. Estimador para σ^2:

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

3. O IC $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ é dado por:

$$(\overline{x}_1 - \overline{x}_2) \pm t_{\alpha/2;\nu}\sqrt{s_p^2(1/n_1 + 1/n_2)}$$

em que $\nu = n_1 + n_2 - 2$ é o número de graus de liberdade.

Hipóteses:

- Ambas as populações tem distribuição aproximadamente normal
- As a.a.'s são selecionadas independentemente
- As variâncias populacionais são iguais (a estimativa de σ^2 é construída com base na informação contida em ambas as amostras)

Exemplo 7

Testes laboratoriais foram realizados a fim de analisar a permeabilidade de dois tipos de concreto. Quatro espécimes foram preparados para cada composição de concreto. As medidas de permeabilidade para as a.a.'s são dadas a seguir:

Tipo 1	1.189	840	1.020	980
Tipo 2	853	900	733	785

Construa um IC 95% para a diferença das permeabilidades médias para os dois tipos de concreto.

Exemplo 7 – Solução:

Temos: $n_1 = n_2 = 4$ $\quad \begin{aligned} \bar{x}_1 &= 1007,25 & \sigma_1 &= 143,66 \\ \bar{x}_2 &= 817,75 & \sigma_2 &= 73,63 \end{aligned}$

$\alpha = 1 - 0,95 = 0,05 \quad \Rightarrow \alpha/2 = 0,025$
$\nu = n_1 + n_2 - 2 = 4 + 4 - 2 = 6$. Da tabela, $t_{\alpha/2;\nu} = 2,447$.
$s_p^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2) = \ldots = 13.028,92$
Então,

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2;\nu}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$(1007,25 - 817,75) \pm 2,447\sqrt{13.028,92(1/4 + 1/4)}$$

$$\text{IC } 95\% : 189,5 \pm 197,5 \quad \Rightarrow \quad -8 < \mu_1 - \mu_2 < 387$$

Interpretação:

Temos 95% de certeza que o intervalo (-8, 387) contém a diferença real entre as permeabilidades médias para os dois tipos de concreto. Como intervalo contém 0, não é possível afirmar, com 95% de certeza, que as médias são estatisticamente diferentes.

CASO 1: IC para a diferença entre duas médias: $\mu_1 - \mu_2$

1.3. $\sigma_1^2 \neq \sigma_2^2$ desconhecidas

Caso 1.3: $\mu_1 - \mu_2$ com $\sigma_1^2 \neq \sigma_2^2$ desconhecidas

1. Estatística:

$$T' = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left[S_1^2/n_1 + S_2^2/n_2\right]}} \sim t_\nu$$

em que:

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left[\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}\right]}$$

Nota:

ν deve ser arredondado para o menor inteiro mais próximo.

Se $n_1 = n_2$, $\nu = n_1 + n_2 - 2 = 2(n-1)$

2. O IC $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ é dado por:

$$(\overline{x}_1 - \overline{x}_2) \pm t_{\alpha/2;\nu}\sqrt{s_1^2/n_1 + s_2^2/n_2}$$

Exemplo 8

Um estudo foi conduzido a fim de estimar a diferença na concentração (mg/l) de um determinado componente químico em duas estações ao longo do Rio Paraíba.

Foram coletadas 15 observações na estação 1 e 12, na estação 2. As concentrações médias de cada uma das amostras foi de, respectivamente, 3,84mg/l e 1,49 mg/l. Os desvios-padrão correspondentes para as duas amostras foram de 3,07mg/l e 0,80mg/l.

Assumindo que as concentrações sejam normalmente distribuídas com variâncias distintas, calcule um IC 95% para a diferença entre as concentrações médias reais.

Exemplo 8 – Solução:

Temos: $f_1 \sim N(\mu_1, \sigma_1^2)$; $f_2 \sim N(\mu_2, \sigma_2^2)$ $\begin{array}{llll} n_1 = 15 & \bar{x}_1 = 3,84 & s_1 = 3,07 \\ n_2 = 12 & \bar{x}_2 = 1,49 & s_2 = 0,80 \end{array}$

$\alpha = 1 - 0,95 = 0,05 \quad \Rightarrow \alpha/2 = 0,025$
$\nu = \ldots = 16,3 \approx 16$. Da tabela, $t_{\alpha/2;\nu} = 2,120$.
Então,

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2;\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$2,35 \pm 2,120\sqrt{\frac{30,7^2}{15} + \frac{0,80^2}{12}}$$

$$\text{IC } 95\% : 2,35 \pm 1,75 \quad \Rightarrow \quad 0,60 < \mu_1 - \mu_2 < 4,10$$

CASO 1: IC para a diferença entre duas médias: $\mu_1 - \mu_2$

1.4. observações emparelhadas (amostras dependentes)

Caso 1.4: $\mu_1 - \mu_2$ com observações emparelhadas

Hipóteses:

- Amostras dependentes (comparação de "tratamentos" para um mesmo indivíduo)
- Variâncias populacionais não necessariamente iguais

1. Estatística: $\bar{D} = \frac{1}{n}\sum_{i=1}^{n}(X_{1i} - X_{2i})$
   Propriedades: $E[\bar{D}] = \mu_D = \mu_1 - \mu_2, \quad Var[\bar{D}] = \frac{S_D^2}{n}$
2. Padronizando:
$$T = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} \sim t_{\nu=n-1}$$
3. O IC $100(1-\alpha)\%$ para $\mu_D = \mu_1 - \mu_2$ é dado por:
$$\bar{d} \pm t_{\alpha/2;\nu} s_D/\sqrt{n}$$

Nota: Se $n \geq 30$, temos: $\bar{d} \pm z_{\alpha/2} s_D/\sqrt{n}$

Exemplo 9

Um estudo foi conduzido a fim de investigar os níveis de cloreto na água distribuída na cidade. Para tanto, foram medidas durante 12 meses as concentrações de cloreto em duas estações distribuidoras de água que recebem a água de uma mesma estação de tratamento. Os seguintes resultados foram obtidos:

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Est. 1	2,0	2,0	2,1	1,9	1,7	1,8	1,7	1,9	2,0	2,0	2,1	2,1
Est. 2	2,2	2,2	2,1	2,0	1,9	1,9	1,8	1,7	1,9	1,9	1,8	2.0

Calcule um IC 95% para a diferença entre as concentrações médias nas duas estações.

Exemplo 9 – Solução:

Temos:

$\mu_1 =$ média mensal do nível de cloreto na estação 1

$\mu_2 =$ média mensal do nível de cloreto na estação 2

$d = \bar{x}_1 - \bar{x}_2$: -0,2 -0,2 0 -0,1 -0,2 -0,1 -0,1 0,2 0,1 0,1 0,3 0,1

Então: $\bar{d} = -0,0083$; $s_d = 0,1676$; $n = 12\ (< 30) \Rightarrow \nu = n - 1 = 12$

Precisamos assumir $d \dot{\sim} N$.

$\alpha = 1 - 0,95 = 0,05 \quad \Rightarrow \alpha/2 = 0,025$

Da tabela, $t_{\alpha/2;\nu} = 2,201$.

Então,

$$\bar{d} \pm t_{\alpha/2;\nu}\frac{s_d}{\sqrt{n}} : \qquad -0,0083 \pm 2,201\frac{0,1676}{\sqrt{12}}$$

$$\text{IC } 95\% : -0,0083 \pm 0,1065 \quad \Rightarrow \quad -0,115 < \mu_1 - \mu_2 < 0,098$$

Interpretação:

Não é possível descartar, com 95% de certeza, que $\mu_1 = \mu_2$ (pois o intervalo obtido contém o 0).

CASO 2: IC para a diferença entre duas proporções $p_1 - p_2$

$p_1 - p_2$: diferença de proporções de sucessos na população

X : número de sucessos em n realizações independentes

Y : número de sucessos em m realizações independentes

n_X e n_Y : conhecidos

Parâmetro a ser estimado $p_1 - p_2$

Estimador pontual de $p_1 - p_2$: $\widehat{p_1} - \widehat{p_2} = X/n_X - Y/n_Y$

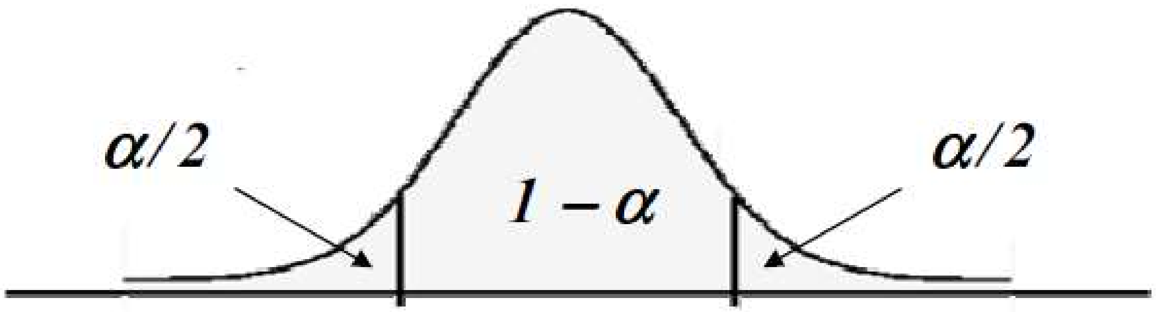
Distribuição assintótica $\left(n_X,\ n_Y \ \rightarrow \ \infty\right)$:

$$\widehat{p_1} - \widehat{p_2} = \frac{X}{n_X} - \frac{Y}{n_Y} \quad \xrightarrow{\; n_X,\, n_Y \ \rightarrow \ \infty \;}$$

$$N\left(p_1 - p_2 ; \frac{\widehat{p_1}\left(1-\widehat{p_1}\right)}{n_X} + \frac{\widehat{p_2}\left(1-\widehat{p_2}\right)}{n_Y} \right)$$

Estimação por Intervalos Envolvendo Duas Amostras

$$Z=\frac{\left(\widehat{p_1}-\widehat{p_2}\right)-\left(p_1-p_2\right)}{\sqrt{\frac{\widehat{p_1}\left(1-\widehat{p_1}\right)}{n_X}+\frac{\widehat{p_2}\left(1-\widehat{p_2}\right)}{n_Y}}}\xrightarrow{n_X,\,n_Y\;\to\;\infty} N(0;1)$$



$z_{1-\alpha/2} = -z_{\alpha/2} \qquad z_{\alpha/2}$

Estimação por Intervalos Envolvendo Duas Amostras

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

$$P\left(-z_{\alpha/2} \leq \frac{\left(\widehat{p_1} - \widehat{p_2}\right) - \left(p_1 - p_2\right)}{\sqrt{\frac{\widehat{p_1}\left(1 - \widehat{p_1}\right)}{n_X} + \frac{\widehat{p_2}\left(1 - \widehat{p_2}\right)}{n_Y}}} \leq z_{\alpha/2} \right) = 1 - \alpha$$

Estimação por Intervalos Envolvendo Duas Amostras

$$P\left[-z_{\alpha/2} \leq \frac{\left(\widehat{p_1}-\widehat{p_2}\right)-\left(p_1-p_2\right)}{\sqrt{\frac{\widehat{p_1}\left(1-\widehat{p_1}\right)}{n_X}+\frac{\widehat{p_2}\left(1-\widehat{p_2}\right)}{n_Y}}} \leq \right]=1-\alpha$$

$$P\left[\left(\widehat{p_1}-\widehat{p_2}\right)-z_{\alpha/2}\sqrt{\frac{\widehat{p_1}\left(1-\widehat{p_1}\right)}{n_X}+\frac{\widehat{p_2}\left(1-\widehat{p_2}\right)}{n_Y}} \leq \left(p_1-p_2\right) \leq \right.$$

$$\left. \leq \left(\widehat{p_1}-\widehat{p_2}\right)+z_{\alpha/2}\sqrt{\frac{\widehat{p_1}\left(1-\widehat{p_1}\right)}{n_X}+\frac{\widehat{p_2}\left(1-\widehat{p_2}\right)}{n_Y}}\right]=1-\alpha$$

IC para a diferença de proporções ao nível de confiança $1 - \alpha$.

$$\left[\left(\widehat{p_1} - \widehat{p_2} \right) \pm z_{\alpha/2} \sqrt{\frac{\widehat{p_1}\left(1 - \widehat{p_1}\right)}{n_X} + \frac{\widehat{p_2}\left(1 - \widehat{p_2}\right)}{n_Y}} \right]$$

CASO 2: $p_1 - p_2$ (amostras grandes)

$X_1 \sim Bin(n_1, p_1)$; $X_2 \sim Bin(n_2, p_2)$

1. Estatística: $W = \widehat{P_1} - \widehat{P_2}$; $\quad \widehat{P_1} = \frac{X_1}{n_1}$; $\widehat{P_2} = \frac{X_2}{n_2}$
   Propriedades: $E[W] = p_1 - p_2$, $\quad Var[W] = p_1 q_1/n_1 + p_2 q_2/n_2$
2. Padronizando:
$$Z = \frac{(\widehat{P_1} - \widehat{P_2}) - (p_1 - p_2)}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} \dot{\sim} N(0,1)$$
3. O IC $100(1-\alpha)\%$ para $p_1 - p_2$ é dado por:
$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\hat{p}_1 \hat{q}_1/n_1 + \hat{p}_2 \hat{q}_2/n_2}$$

Nota: $n_1\hat{p}_1,\ n_1\hat{q}_1,\ n_2\hat{p}_2,\ n_2\hat{q}_2 \geq 5$

Exemplo 10

Engenheiros consideram implantar uma mofificação no processo de fabricação de um determinado produto. Antes disso, porém, precisam constatar que a modificação resulta em melhoria do processo. A fim de comparar os processos, amostras do produto produzido através do processo atual e do processo alternativo são coletadas. Para o processo atual, em 75 itens de 1500 observados foram detectadas falhas. Para o processo modificado, 80 itens dos 2000 observados apresentaram defeitos.

Construa um IC 90% para a diferença real entre as frações de itens defeituosos produzidos através dos dois processos.

Exemplo 10 – Solução:

Temos:
$$\begin{array}{llll} n_1 = 1500 & x_1 = 75 & \Rightarrow & \hat{p}_1 = x_1/n_1 = 0,05 \\ n_2 = 2000 & x_2 = 80 & \Rightarrow & \hat{p}_2 = x_2/n_2 = 0,04 \end{array}$$

$\alpha = 1 - 0,90 = 0,10 \quad \Rightarrow \alpha/2 = 0,05$
Da tabela, $z_{\alpha/2} = 1,645$.
Então,

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

$$0,01 \pm 1,645\sqrt{\frac{0,05 \cdot 0,95}{1500} + \frac{0,04 \cdot 0,96}{2000}}$$

$$\text{IC } 90\% : 0,01 \pm 0,0117 \quad \Rightarrow \quad -0,0017 < p_1 - p_2 < 0,0217$$

Interpretação:
Este IC contém o valor 0. Não existe evidência suficiente para crer, com 90% de confiança, que o novo processo apresenta redução significativa na proporção de itens produzidos com defeito.

CASO 3: IC para a razão entre duas variâncias σ_1^2/σ_2^2

Estimador

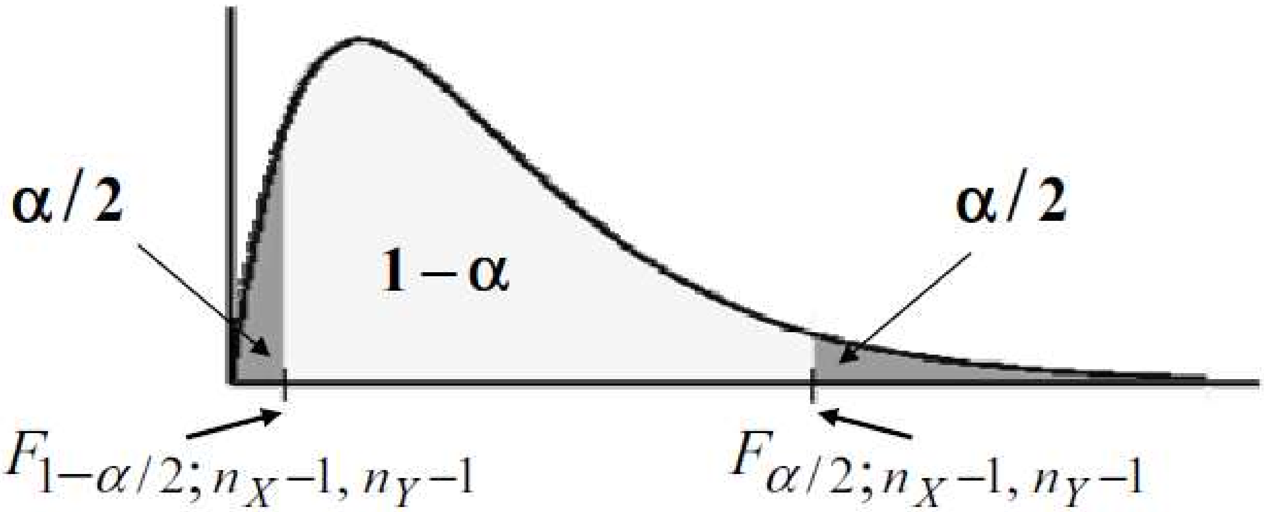
$$T = \frac{S_X^2}{S_Y^2} = \frac{\frac{1}{n_X - 1}\sum_{i=1}^{n_X}\left(X_i - \bar{X}\right)^2}{\frac{1}{n_Y - 1}\sum_{i=1}^{n_Y}\left(Y_i - \bar{Y}\right)^2}$$

Distribuição Amostral

$$F = \frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2} = \frac{S_X^2 \sigma_Y^2}{S_Y^2 \sigma_X^2} \rightarrow F_{n_X - 1,\, n_Y - 1}$$

Estimação por Intervalos Envolvendo Duas Amostras

$$P\left[F_{1-\alpha/2;n_X-1,n_Y-1} \leq F_{n_X-1,n_Y-1} \leq F_{\alpha/2;n_X-1,n_Y-1}\right] = 1-\alpha$$



Estimação por Intervalos Envolvendo Duas Amostras

$$P\left[F_{1-\alpha/2;n_X-1,n_Y-1} \le F_{n_X-1,n_Y-1} \le F_{\alpha/2;n_X-1,n_Y-1}\right] = 1-\alpha$$

$$P\left[F_{1-\alpha/2;n_X-1,n_Y-1} \le \frac{\sigma_Y^2}{\sigma_X^2}\frac{S_X^2}{S_Y^2} \le F_{\alpha/2;n_X-1,n_Y-1}\right] = 1-\alpha$$

$$P\left[F_{1-\alpha/2;n_X-1,n_Y-1}\frac{S_Y^2}{S_X^2} \le \frac{\sigma_Y^2}{\sigma_X^2} \le F_{\alpha/2;n_X-1,n_Y-1}\frac{S_Y^2}{S_X^2}\right] = 1-\alpha$$

$$P\left[\frac{1}{F_{1-\alpha/2;n_X-1,n_Y-1}}\frac{S_X^2}{S_Y^2} \ge \frac{\sigma_X^2}{\sigma_Y^2} \ge \frac{S_X^2}{S_Y^2}\frac{1}{F_{\alpha/2;n_X-1,n_Y-1}}\right] = 1-\alpha$$

Estimação por Intervalos Envolvendo Duas Amostras

$$P\left[\frac{1}{F_{\alpha/2;n_X-1,n_Y-1}}\frac{S_X^2}{S_Y^2}\leq\frac{\sigma_X^2}{\sigma_Y^2}\leq\frac{S_X^2}{S_Y^2}\frac{1}{F_{1-\alpha/2;n_X-1,n_Y-1}}\right]=1-\alpha$$

$$P\left[\frac{1}{F_{\alpha/2;n_X-1,n_Y-1}}\frac{S_X^2}{S_Y^2}\leq\frac{\sigma_X^2}{\sigma_Y^2}\leq\frac{S_X^2}{S_Y^2}\frac{1}{F_{1-\alpha/2;n_X-1,n_Y-1}}\right]=1-\alpha$$

$$\frac{1}{F_{1-\alpha/2;n_X-1,n_Y-1}}=F_{\alpha/2;n_Y-1,n_X-1}$$

$$P\left[\frac{1}{F_{1-\alpha/2;n_X-1,n_Y-1}}\frac{S_X^2}{S_Y^2}\leq\frac{\sigma_X^2}{\sigma_Y^2}\leq F_{\alpha/2;n_Y-1,n_X-1}\frac{S_X^2}{S_Y^2}\right]=1-\alpha$$

IC para $\dfrac{\sigma_X^2}{\sigma_Y^2}$, com médias populacionais desconhecidas, ao nível de confiança $1 - \alpha$.

$$\left[\frac{1}{F_{\alpha/2;n_X-1,n_Y-1}} \frac{S_X^2}{S_Y^2} \, , \, F_{\alpha/2\,;\,n_Y-1\,,\,n_X-1} \frac{S_X^2}{S_Y^2} \right]$$

CASO 3: σ_1^2/σ_2^2

$X_1 \sim N(\mu_1, \sigma_1^2)$; $X_2 \sim N(\mu_2, \sigma_2^2)$

1. Estatística: $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{\nu_1,\nu_2}$
2. O IC $100(1-\alpha)\%$ para σ_1^2/σ_2^2 é dado por:

$$\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(\nu_2,\nu_1)$$

Exemplo 11

No experimento realizado para análise das águas do Rio Paraíba, foi considerado que as variâncias nas duas estações eram distintas. Justifique esta hipótese através da construção de um IC 98% para σ_1/σ_2.

Exemplo 11 – Solução:

Temos: $\begin{array}{llll} n_1 = 15 & s_1 = 3,07 & \Rightarrow & \nu_1 = 14 \\ n_2 = 12 & s_2 = 0,80 & \Rightarrow & \nu_2 = 11 \end{array}$

$\alpha = 1 - 0,98 = 0,02 \quad \Rightarrow \alpha/2 = 0,01$
Interpolando valores da tabela: $f_{\alpha/2}(14,11) \approx 4,30$; $f_{\alpha/2}(11,14) \approx 3,87$.
Então,

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \cdot f_{\alpha/2}(\nu_2,\nu_1)$$

$$\left(\frac{3,07}{0,80}\right)^2 \cdot \frac{1}{4,30} < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{3,07}{0,80}\right)^2 \cdot 3,87 \quad \Rightarrow \quad 3,425 < \frac{\sigma_1^2}{\sigma_2^2} < 56,991$$

$$\text{IC98\%}(\sigma_1/\sigma_2) : 1,851 < \frac{\sigma_1}{\sigma_2} < 7,549$$

Interpretação:

Este IC não inclui o valor 1. Podemos garantir, com 98% de certeza, que $\sigma_1 \neq \sigma_2$.