Capítulo 9 | Problemas de estimação em uma e duas amostras

9.3 Métodos clássicos de estimação

Definição 9.1

Uma estatística $\hat{\Theta}$ é considerada um estimador *não viciado* do parâmetro θ se

$$\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta.$$

Definição 9.2

Se considerarmos todos os possíveis estimadores não viciados de algum parâmetro θ, aquele com a menor variância é chamado de *estimador mais eficiente* de θ.

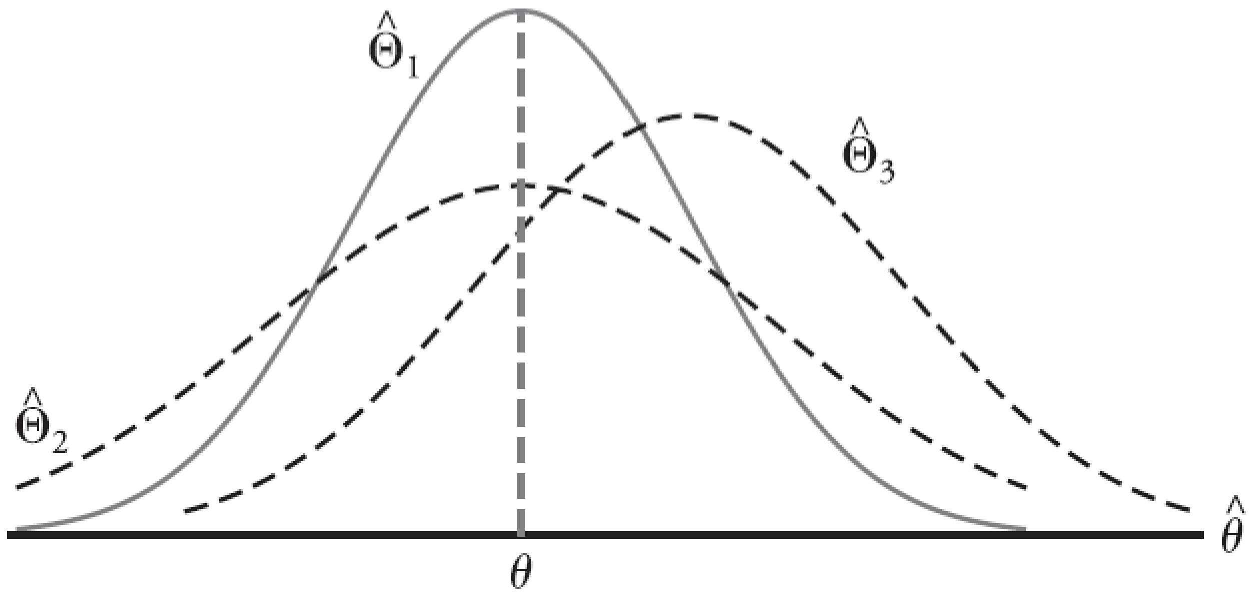

Figura 9.1 Distribuições amostrais de diferentes estimadores de θ.

9.4 Amostra única: estimação da média

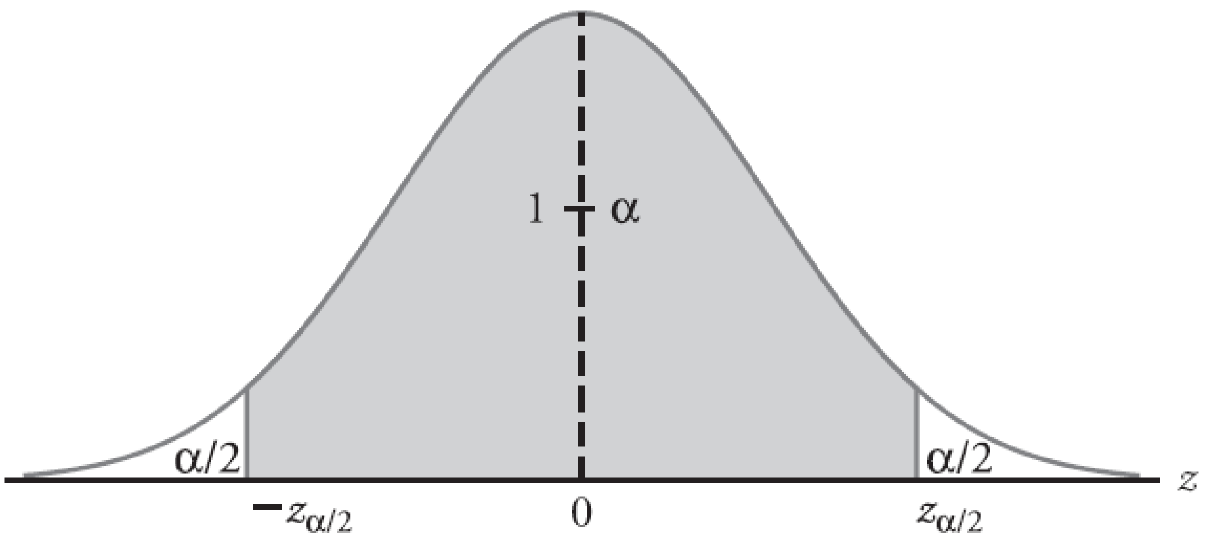

Figura 9.2 $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$.

Intervalo de confiança para μ; σ conhecido

Se $\bar{x}$ é a média de uma amostra aleatória de tamanho n de uma população com variância conhecida σ^2, um intervalo de confiança de $100(1-\alpha)\%$ para μ é dado por

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$

onde $z_{\alpha/2}$ é o valor z que deixa uma área de $\alpha/2$ à direita.

Capítulo 9 | Problemas de estimação em uma e duas amostras

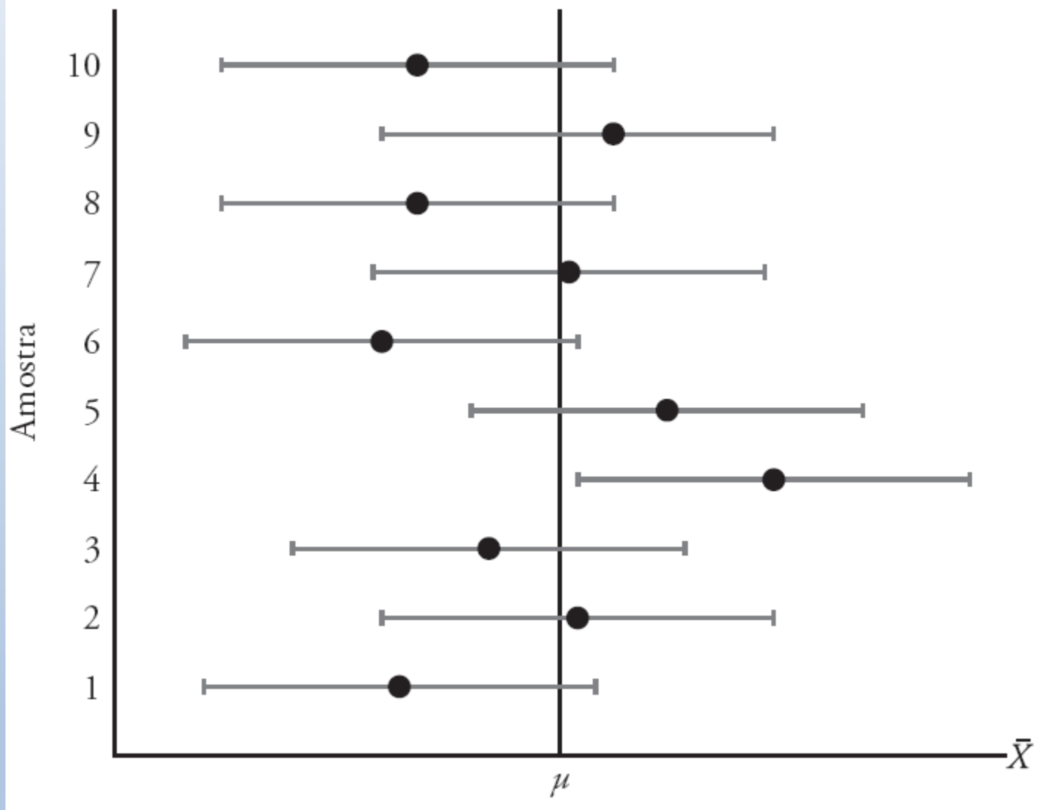

Figura 9.3 Estimativas intervalares de μ para diferentes amostras.

Teorema 9.1

Se $\bar{x}$ é usado como uma estimativa de μ, podemos então estar $100(1 - \alpha)\%$ confiantes de que o erro não excederá $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Teorema 9.2

Se $\bar{x}$ é usado como uma estimativa de μ, podemos estar $100(1-\alpha)\%$ confiantes de que o erro não excederá um valor específico e quando o tamanho da amostra for

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2.$$

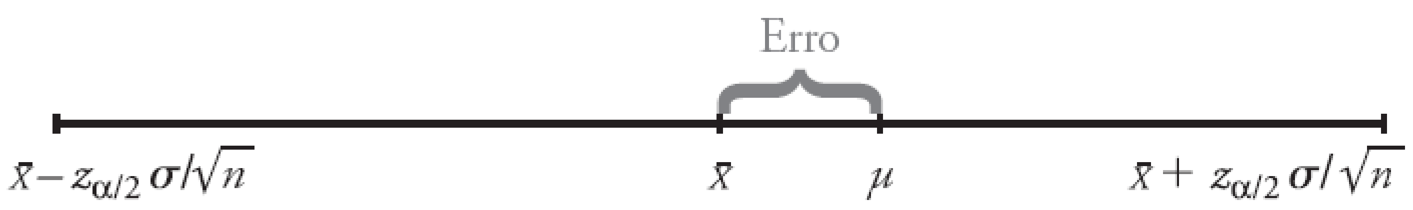

Figura 9.4 Erro ao estimar μ por $\bar{x}$.

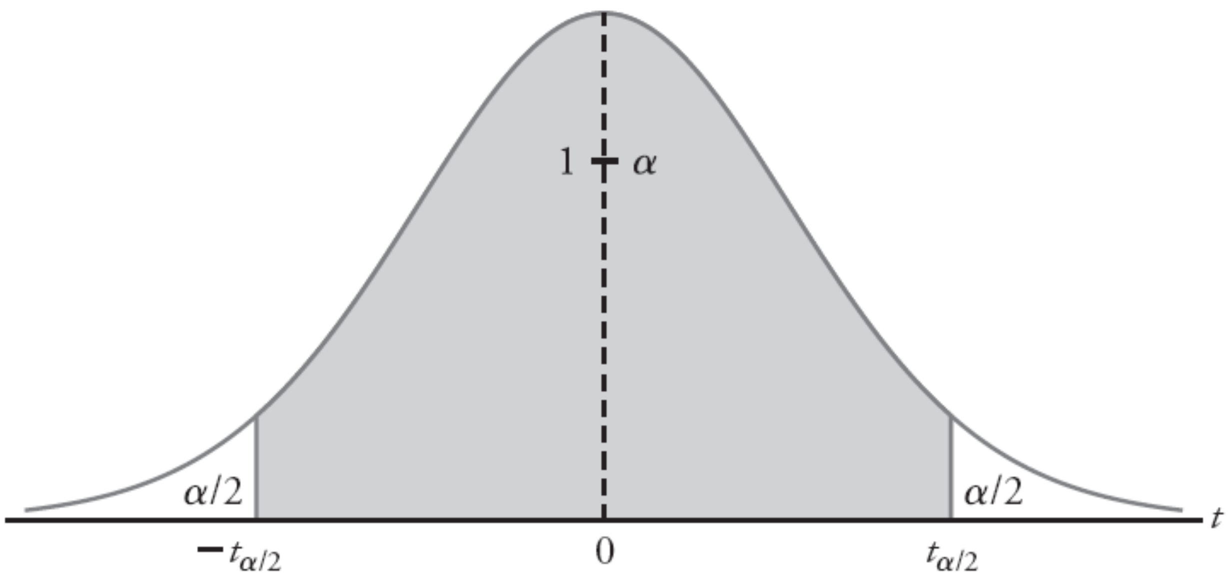

Figura 9.5 $P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$.

Intervalo de confiança para μ; σ desconhecido

Se $\bar{x}$ e s são a média e o desvio-padrão de uma amostra aleatória de tamanho n de uma população normal com variância desconhecida σ^2, um intervalo de confiança de $100(1 - \alpha)\%$ para μ é

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

onde $t_{\alpha/2}$ é um valor t com $v = n - 1$ graus de liberdade que deixa uma área de $\alpha/2$ à direita.

9.5 Erro-padrão de um estimador pontual

Limites de confiança em μ para σ desconhecido

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = \bar{x} \pm t_{\alpha/2}\,\text{e.p.}(\bar{x})$$

9.6 Intervalos de predição

Intervalo de predição de uma observação futura; σ conhecido

Para uma distribuição normal das medidas, com média desconhecida μ e variância conhecida σ^2, um *intervalo de predição* de $100(1-\alpha)\%$ de uma observação futura x_0 é

$$\bar{x} - z_{\alpha/2}\sigma\sqrt{1+1/n} < x_0 < \bar{x} + z_{\alpha/2}\sigma\sqrt{1+1/n}.$$

onde $z_{\alpha/2}$ é o valor z que deixa uma área de $\alpha/2$ à direita.

Intervalo de predição de uma observação futura; σ desconhecido

Para uma distribuição normal das medições, com média μ e variância σ^2 desconhecidas, um *intervalo de predição* de $100(1 - \alpha)\%$ de uma observação futura x_0 é

$$\bar{x} - t_{\alpha/2}s\sqrt{1 + 1/n} < x_0 < \bar{x} + t_{\alpha/2}s\sqrt{1 + 1/n},$$

onde $t_{\alpha/2}$ é o valor t, com $v = n - 1$ graus de liberdade, que deixa uma área de $\alpha/2$ à direita.

9.7 Limites de tolerância

Limites de tolerância

Para uma distribuição normal das medidas, com média μ e desvio-padrão σ desconhecidos, os *limites de tolerância* são dados por $\bar{x} \pm ks$, onde k é determinado de modo que podemos afirmar, com $100(1 - \gamma)\%$ de confiança, que os limites contêm, pelo menos, a proporção $1 - \alpha$ das medidas.

9.8 Duas amostras: estimando a diferença entre duas médias

Intervalo de confiança para $\mu_1 - \mu_2$; σ_1^2 e σ_2^2 conhecidos

Se $\bar{x}_1$ e $\bar{x}_2$ são médias de amostras aleatórias independentes de tamanhos n_1 e n_2 de populações com variâncias conhecidas σ_1^2 e σ_2^2, respectivamente, um intervalo de confiança de $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ é dado por

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

onde $z_{\alpha/2}$ é o valor de z que deixa uma área de $\alpha/2$ à direita.

Estimativa combinada da variância

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

Intervalo de confiança para $\mu_1 - \mu_2$; $\sigma_1^2 = \sigma_2^2$, mas desconhecidos

Se $\bar{x}_1$ e $\bar{x}_2$ são as médias de amostras aleatórias independentes de tamanhos n_1 e n_2, respectivamente, de populações aproximadamente normais com variâncias desconhecidas, mas iguais, um intervalo de confiança de

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 <$$

$$(\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

onde s_p é a estimativa combinada do desvio-padrão da população e $t_{\alpha/2}$ é o valor t com $v = n_1 + n_2 - 2$ graus de liberdade, que deixa uma área de $\alpha/2$ à direita.

Intervalo de confiança para $\mu_1 - \mu_2$; $\sigma_1^2 \neq \sigma_2^2$, e desconhecidos

Se $\bar{x}_1$ e s_1^2 e $\bar{x}_2$ e s_2^2 são as médias e as variâncias de amostras aleatórias independentes de tamanhos n_1 e n_2, respectivamente, de uma população aproximadamente normal, com variâncias diferentes e desconhecidas, um intervalo de confiança de $100(1 - \alpha)\%$ para $\mu_1 - \mu_2$ é dado por

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

onde $t_{\alpha/2}$ é o valor t com

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

graus de liberdade, que deixa uma área de $\alpha/2$ à direita.

9.9 Observações emparelhadas

Intervalo de confiança para $\mu_D = \mu_1 - \mu_2$ com observações emparelhadas

Se $\bar{d}$ e s_d são a média e o desvio-padrão, respectivamente, de diferenças normalmente distribuídas de n pares de medidas aleatórias, um intervalo de confiança de $100(1 - \alpha)\%$ para $\mu_D = \mu_1 - \mu_2$ é

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}},$$

onde $t_{\alpha/2}$ é o valor t, com $v = n - 1$ graus de liberdade, que deixa uma área de $\alpha/2$ à direita.

Tabela 9.1 Dados para o Exemplo 9.12

Fonte: Schecter, A. et al. "Partitioning of 2, 3, 7, 8-chlorinated dibenzo-p-dioxins and dibenzofurans between adipose tissue and plasma lipid of 20 Massachusetts Vietnan veterans", *Chemosphere*, v. 20, n. 7–9, 1990, p. 954–955 (tabelas I e II).

Veterano	TCDD Níveis em plasma	TCDD Níveis em tecido adiposo	d_i
1	2,5	4,9	– 2,4
2	3,1	5,9	– 2,8
3	2,1	4,4	– 2,3
4	3,5	6,9	– 3,4
5	3,1	7,0	– 3,9
6	1,8	4,2	– 2,4
7	6,0	10,0	– 4,0
8	3,0	5,5	– 2,5
9	36,0	41,0	– 5,0
10	4,7	4,4	– 0,3
11	6,9	7,0	– 0,1
12	3,3	2,9	0,4
13	4,6	4,6	0,0
14	1,6	1,4	0,2
15	7,2	7,7	– 0,5
16	1,8	1,1	0,7
17	20,0	11,0	9,0
18	2,0	2,5	– 0,5
19	2,5	2,3	0,2
20	4,1	2,5	1,6

9.10 Amostra única: estimando uma proporção

Intervalo de confiança para *p* em amostra grande

Se $\hat{p}$ é a proporção de sucessos em uma amostra aleatória de tamanho n, e $\hat{q} = 1 - \hat{p}$, um intervalo de confiança de $100(1 - \alpha)\%$ aproximado para o parâmetro binomial p é dado por

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}},$$

onde $z_{\alpha/2}$ é o valor z que deixa uma área de $\alpha/2$ à direita.

Teorema 9.3

Se $\hat{p}$ é usado como estimativa de p, podemos estar $100(1 - \alpha)\%$ confiantes de que o erro não excederá $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.

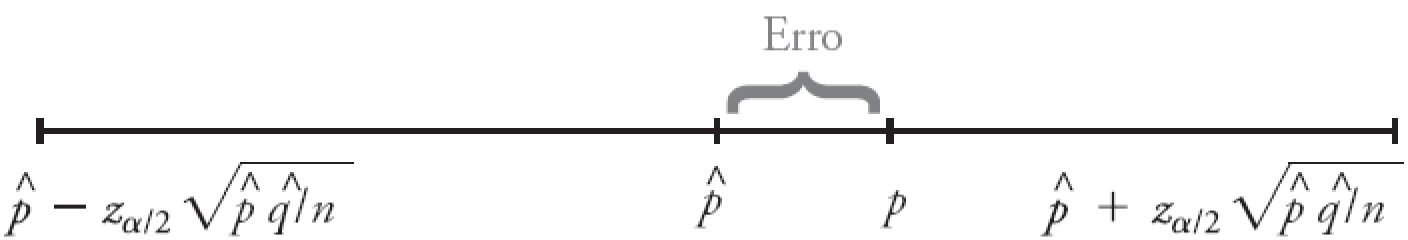

Figura 9.6 Erro ao estimar p por $\hat{p}$.

Teorema 9.4

Se $\hat{p}$ é usado como estimativa de p, podemos estar $100(1-\alpha)\%$ confiantes de que o erro será menor do que a quantidade especificada e quando o tamanho da amostra for aproximadamente

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{e^2}.$$

Teorema 9.5

Se $\hat{p}$ é usado como uma estimativa de p, podemos estar *pelo menos* $100(1-\alpha)\%$ confiantes de que o erro não excederá uma quantidade específica e quando o tamanho da amostra é

$$n = \frac{z_{\alpha/2}^2}{4e^2}.$$

9.11 Duas amostras: estimando a diferença entre duas proporções

Intervalo de confiança para $p_1 - p_2$ em amostras grandes

Se $\hat{p}_1$ e $\hat{p}_2$ são as proporções de sucessos em amostras aleatórias de tamanho n_1 e n_2, respectivamente, $\hat{q}_1 = 1 - \hat{p}_1$ e $\hat{q}_2 = 1 - \hat{p}_2$, um intervalo de confiança de $100(1 - \alpha)\%$ aproximado para a diferença dos dois parâmetros binomiais $p_1 - p_2$ é dado por

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2$$

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

onde $z_{\alpha/2}$ é o valor z que deixa uma área de $\alpha/2$ à direita.

9.12 Amostra única: estimando a variância

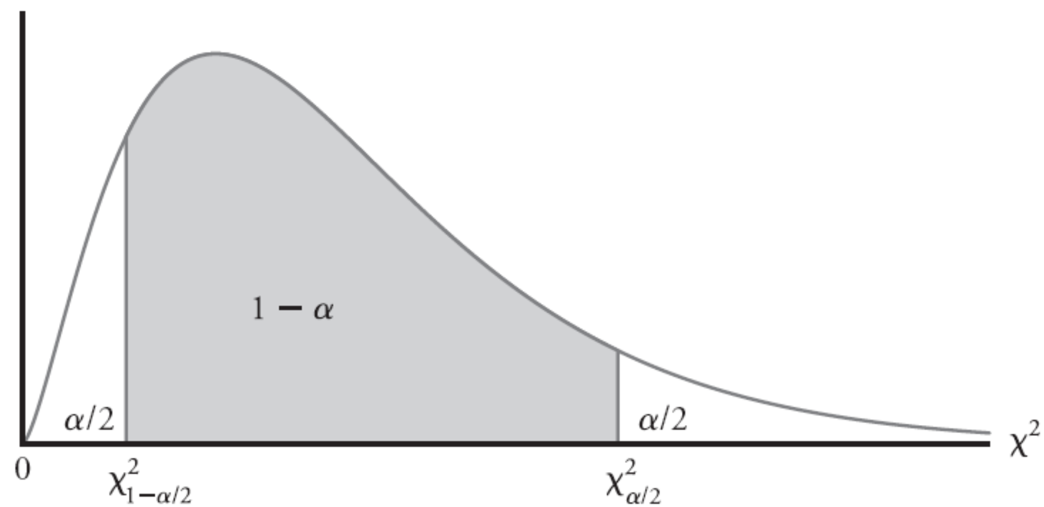

Figura 9.7 $P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha$.

Intervalo de confiança para σ^2

Se s^2 é a variância da amostra aleatória de tamanho n de uma população normal, um intervalo de confiança de $100(1-\alpha)\%$ para σ^2 é dado por

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}},$$

onde $\chi^2_{\alpha/2}$ e $\chi^2_{1-\alpha/2}$ são os valores χ^2 com $v = n - 1$ graus de liberdade, que deixam uma área de $\alpha/2$ e $1 - \alpha/2$ à direita, respectivamente.

9.13 Duas amostras: estimando a razão de duas variâncias

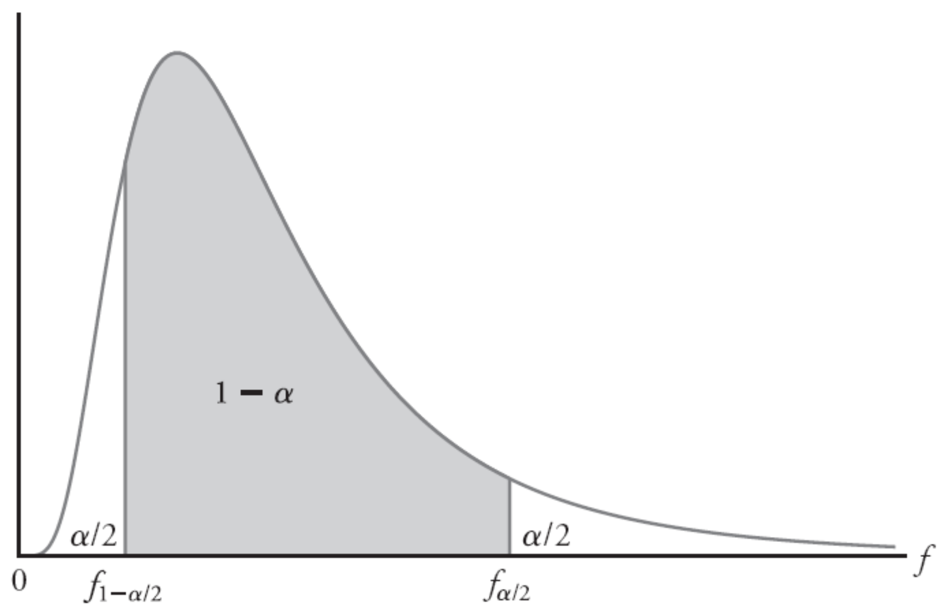

Figura 9.8 $P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha$.

9.14 Estimação de máxima verossimilhança (opcional)

Definição 9.3

Dadas as observações independentes $x_1, x_2, \ldots, x_n$ de uma função de densidade de probabilidade (caso contínuo) ou função de massa de probabilidade (caso discreto) $f(x; \theta)$, o estimador de máxima verossimilhança $\hat{\theta}$ é aquele que maximiza a função de verossimilhança

$$L(x_1, x_2, \ldots, x_n; \theta) = f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta).$$