7

Distribuições Amostrais e Estimação de Parâmetros

Este capítulo apresenta a base para aprendermos a estatística indutiva, a qual fornece procedimentos formais para tirar conclusões sobre uma população, a partir dos dados de uma amostra. Para isso, veremos como se relacionam *estatísticas* (características dos elementos de uma amostra) com *parâmetros* (características dos elementos de uma população).

7.1 PARÂMETROS E ESTATÍSTICAS

Relembremos alguns conceitos básicos:

População: conjunto de elementos que formam o universo de nosso estudo e que são passíveis de ser observados, sob as mesmas condições.

Amostra: uma parte dos elementos de uma população.

Amostragem: o processo de seleção da amostra.

Amostragem aleatória simples: o processo de seleção é feito por sorteio, fazendo com que todos os elementos da população tenham a mesma chance de serem escolhidos e, além disso, todo subconjunto de n elementos tenha a mesma chance de fazer parte da amostra.

Quando a amostragem é aleatória e, em especial, quando é aleatória simples, podemos fazer inferências sobre a população, com base no estudo de uma amostra (ver a Figura 7.1).

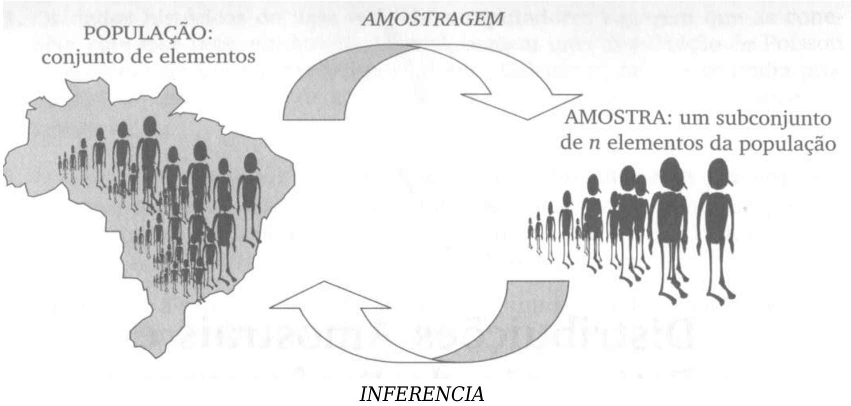


Figura 7.1 *Ilustração de conceitos básicos da estatística.*

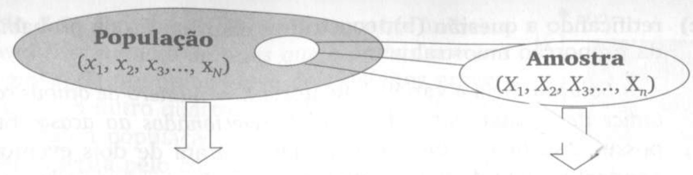
Em geral, estamos pesquisando uma ou mais variáveis associadas aos elementos da população ou da amostra. Nesse contexto, também podemos caracterizar a população e a amostra em termos da variável em estudo. Por exemplo, ao verificar se cada consumidor potencial de uma revendedora de automóveis planeja (X = 1) ou não (X = 0) comprar um carro novo no próximo ano, podemos representar a *população* pelo conjunto $\{x_1, x_2, x_3, ..., x_N\}$, onde   = 0 ou 1, dependendo se o j-ésimo indivíduo da população pretende ou não comprar um carro novo no próximo ano. Da mesma forma, a *amostra,* que será formada por n indivíduos a serem selecionados da população, pode ser representada por $\{X_l$ $X_2, ..., X_n\}$, onde X_i *é* a variável aleatória que corresponde ao valor de X (0 ou 1) na *i-é*sima observação (í = 1, 2, …n).

No presente contexto, definimos:

Parâmetro: alguma medida descritiva (média, variância, proporção etc.) dos valores $x_1, x_2, x_3, ...,$ associados à população.

Amostra aleatória simples: conjunto de n variáveis aleatórias independentes $\{X_1, X_2, ..., X_n\}$, cada uma com a mesma distribuição de probabilidades de certa variável aleatória X. Essa distribuição de probabilidades deve corresponder à distribuição de freqüências dos valores da população (X_1, X_2, X_3, …).

Estatística: alguma medida descritiva (média, variância, proporção etc.) das variáveis aleatóriaX_1, X_2,…,X_n, associadas à amostra (ver a Figura 7.2).



	Parâmetros	Estatísticas
Proporção	$p = \dfrac{n^{\text{o}} \text{ de elementos com o atributo}}{N}$	$\hat{P} = \dfrac{n^{\text{o}} \text{ de elementos com o atributo}}{n}$
Média	$\mu = \dfrac{1}{N}\sum_{i=1}^{N} x_i$	$\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$
Variância	$\sigma^2 = \dfrac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$	$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Figura 7.2 *Alguns parâmetros e estatísticas.*

Estatísticas e variáveis aleatórias

Vamos considerar uma amostragem aleatória simples, o que faz com que qualquer medida associada à amostra (*estatística*) seja uma variável aleatória. Isso ocorre por causa da aleatoriedade introduzida pelo sorteio, na amostragem. A fim de entender melhor esse conceito, vamos acompanhar os exemplos seguintes.

Exemplo 7.1 Em um estudo sobre emissões de $C0_2$, definiu-se uma população como sendo composta por quatro ônibus de uma pequena companhia de transporte urbano. Dos quatro ônibus, um deles apresentava alto índice de emissão, enquanto os outros três estavam dentro dos padrões. Assim, a população pode ser descrita por {1, 0, 0, 0}. O parâmetro de interesse é a *proporção de veículos fora do padrão.* Considere as seguintes questões acerca dessa população:

a) calcular a proporção populacional (p);

b) se for retirada uma amostra aleatória simples, com reposição, de tamanho $n = 2$, qual será a proporção P de veículos fora dos padrões na amostra?

Solução: Para a questão (a), a resposta é trivial: $p = {}^1/_4$i Já a questão (b) não pode ser respondida, pois a proporção amostrai ($\dot{P}$) é uma variável aleatória. Assim, não podemos dizer o que *vai* ocorrer, mas tão-somente o que *pode* ocorrer.

c) retificando a questão (b), construir a distribuição de probabilidades da proporção amostrai.

Solução: Seja a variável aleatória X = *número de ônibus com alto índice de emissão entre dois ônibus selecionados ao acaso*. Então, X possui distribuição binomial, já que se trata de dois eventos independentes, com duas possibilidades cada (*sucesso = ônibus com alto índice de emissão; fracasso = ônibus dentro dos padrões*). Os parâmetros da distribuição são $n = 2$ e $p = 0{,}25$. Assim, pela distribuição binomial (ver a Tabela I do apêndice), temos:

x	$p(x)$
0	0,5625
1	0,3750
2	0,0625
Total	1

A proporção amostral é dada por:

$$\hat{P} = \frac{X}{n} = \frac{X}{2}$$

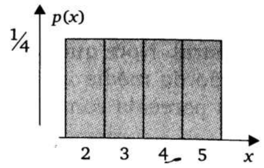
Assim, sua função de probabilidade é dada por:

$\hat{p}$	$p(\hat{p})$
0	0,5625
½	0,3750
1	0,0625
Total	1

A função de probabilidade de P também é chamada de *distribuição da proporção amostrai* ou *distribuição amostrai da proporção*, pois apresenta os possíveis resultados de uma proporção, que é calculada sobre os elementos de uma amostra a ser extraída da população em estudo. De maneira geral, temos:

Uma *estatística é* uma variável aleatória e sua distribuição de probabilidades é chamada de *distribuição amostra*l

Exemplo 7.2 Seja a população dos quatro ônibus e a variável X = *número de vezes que o ônibus teve um defeito grave.* Se um ônibus teve dois defeitos graves, o outro três, o outro quatro e o último cinco defeitos graves, então a população, em termos da variável X, pode ser descrita pelo conjunto {2, 3, 4, 5}. A população também pode ser descrita pela função de probabilidade ao lado - a *distribuição da população.* Essa distribuição tem os parâmetros valor esperado (média) e variância dados por:



$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{4}(2+3+4+5) = 3{,}5 \text{ e}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{1}{4}\left[(2-3{,}5)^2 + (3-3{,}5)^2 + (4-3{,}5)^2 + (5-3{,}5)^2\right] = 1{,}25$$

A Tabela 7.1 mostra a construção da distribuição da média amostrai, considerando uma amostragem aleatória simples com n = 2 elementos, extraída com reposição.

Tabela 7.1 *Construção da distribuição de X (Exemplo 7.2).*

Amostras possíveis	Valor de $\bar{X}$	Probabilidade
(2, 2)	2,0	$1/16$
(2, 3), (3, 2)	2,5	$2/16$
(2, 4), (3, 3), (4, 2)	3,0	$3/16$
(2, 5), (3, 4), (4, 3), (5, 2)	3,5	$4/16$
(3, 5), (4, 4), (5, 3)	4,0	$3/16$
(4, 5), (5, 4)	4,5	$2/16$
(5, 5)	5,0	$1/16$

O valor esperado e a variância da distribuição de X são:

$$E(\bar{X}) = 2\left(\frac{1}{16}\right) + 2{,}5\left(\frac{2}{16}\right) + 3\left(\frac{3}{16}\right) + 3{,}5\left(\frac{4}{16}\right) + 4\left(\frac{3}{16}\right) + 4{,}5\left(\frac{2}{16}\right) +$$

$$+5\left(\frac{1}{16}\right) = 3{,}5$$

$$V(\bar{X}) = (2-3{,}5)^2\,\frac{1}{16} + (2{,}5-3{,}5)^2\,\frac{2}{16} + \ldots + (5-3{,}5)^2\,\frac{1}{16} = 0{,}625$$

A Figura 7.3 mostra a distribuição da população e a distribuição da média amostrai. Note que ambas têm a mesma média (valor esperado), mas a distribuição da média amostrai é mais concentrada (menor variância) e tem forma mais parecida com a distribuição normal.

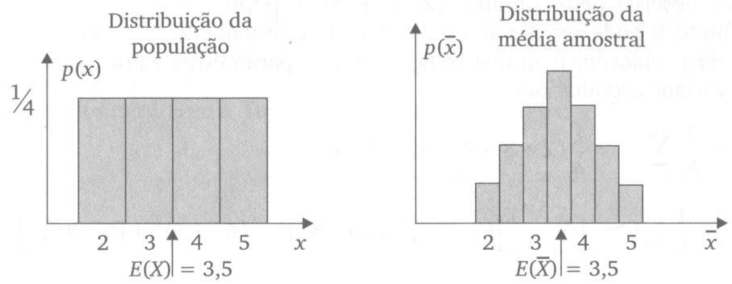


Figura 7.3 *Distribuição da população do Exemplo 7.2 e a distribuição da média amostral, considerando amostragem aleatória simples com* n = *2 elementos, extraídos com reposição.*

EXERCÍCIOS

1. Refaça o Exemplo 7.1c, considerando que a amostra seja retirada sem reposição.

2. Em um estudo sobre consumo de combustível, definiu-se uma população composta por quatro ônibus de uma pequena companhia de transporte urbano. Os consumos dos ônibus (km/l), em condições padrões de teste, eram 3,8, 3,9, 4,0 e 4,1. Uma amostra de dois elementos será sorteada, com reposição. Verifique todas as amostras possíveis e, em seguida, construa a distribuição amostrai para o *consumo médio* da amostra e calcule o valor esperado e a variância.

3. Refaça o exercício anterior considerando amostragem sem reposição.

7.2 DISTRIBUIÇÕES AMOSTRAIS

Quando a amostragem é aleatória simples, várias estatísticas apresentam distribuições amostrais que se aproximam de distribuições contínuas conhecidas, à medida que o tamanho da amostra cresce. É o caso da média e da proporção que apresentam distribuições amostrais aproximadamente normal.

7.2.1 Distribuição amostrai da média

Considere o esquema ilustrado na Figura 7.4.

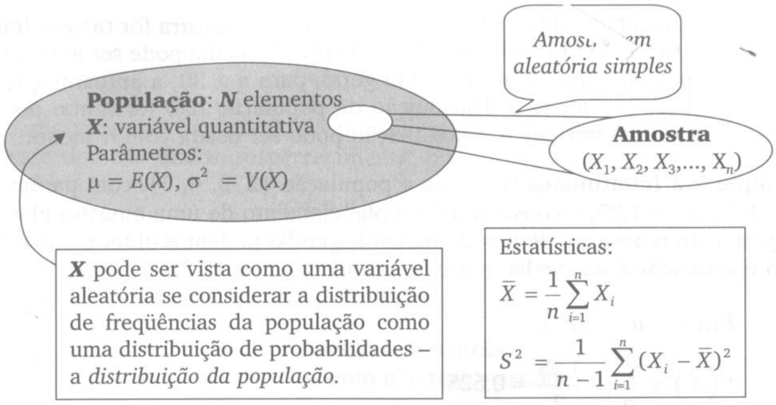


Figura 7.4 *Esquema geral de uma amostragem aleatória simples na observação de uma variável quantitativa* X.

Seja uma amostra aleatória simples $\{X_1, X_2, \ldots, X_n\}$ e a estatística X. *A* distribuição de *X (distribuição da média amostrai)* apresenta as seguintes propriedades:

a) O valor esperado da média amostrai é igual à média da população, ou seja:[1]

$$\boxed{E(\bar{X}) = \mu} \tag{7.1}$$

b) A variância da média amostral é inferior à variância populacional (σ^2) e a relação é dada por

$$\boxed{V(\bar{X}) = \frac{\sigma^2}{n}}$$ (se a amostragem for *com* reposição, ou *N* muito grande ou infinito).[2] (7.2)

---

1 $E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu$

2 Com as condicionantes, podemos supor independência entre $X_1, \ldots, X_n$, donde:
$V(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

ou

$$\boxed{V(\bar{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}}\quad \text{(se a amostragem for } \textit{sem} \text{ reposição e } N \text{ não muito grande, } N < 20n) \qquad (7.3)$$

c) (*Teorema limite central)* Se o tamanho da amostra for razoavelmente *grande,* então a distribuição amostrai da média pode ser aproximada pela *distribuição normal* Em geral, para $n \geq 30$, a aproximação já é boa, porém, se a distribuição da população não for muito distante de uma normal, a aproximação pode ser usada com n menor.

Exemplo 7.2 (continuação) Dada a população {2, 3, 4, 5}, com parâmetros $\mu = 3{,}5$ e $\sigma^2 = 1{,}25$, e considerando o planejamento de uma amostra aleatória simples, com reposição, de $n = 2$ elementos, então podemos obter o valor esperado e a variância da média amostrai usando (7.1) e (7.2):

$$E(\bar{X}) = \mu = 3{,}5 \text{ e}$$

$$V(\bar{X}) = \frac{\sigma^2}{n} = \frac{1{,}25}{2} = 0{,}625$$

Observe que são os mesmos valores encontrados anteriormente.

7.2.2 Distribuição amostrai da proporção

Quando o interesse é estudar uma proporção, tal como a *proporção dos elementos que têm certo atributo A,* a população pode ser vista como dividida em dois subgrupos:

1. o subgrupo dos elementos que *têm* o atributo *A;* e
2. o subgrupo dos elementos que *não têm* o atributo *A,* como mostra a Figura 7.5.

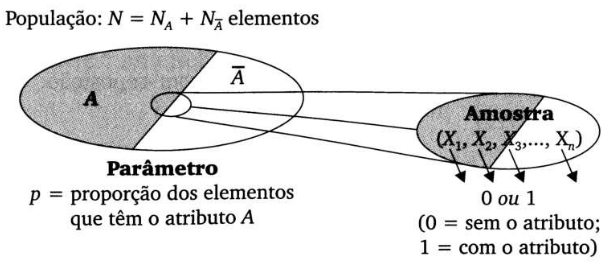


Figura 7.5 *Esquema geral de uma amostragem aleatória simples quando se observa a proporção de certo atributo* A.

A distribuição da população pode ser representada por uma variável aleatória de Bernoulli (tipo "0-1"), com função de probabilidade:

x	$P(x)$
0	$1-p$
1	p

Como vimos no Capítulo 5, o valor esperado e a variância de uma distribuição desse tipo são dados, respectivamente, por:

$$\mu = p \text{ e} \tag{7.4}$$

$$\sigma^2 = p(1-p) \tag{7.5}$$

Representando as observações amostradas por

$$X_i = \begin{cases} 1 & \text{se o } i \text{ - ésimo elemento tem o atributo } A \\ 0 & \text{se o } i \text{ - ésimo elemento não tem o atributo } A \end{cases}$$

verificamos que

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\textit{número de elementos com o atributo } A}{n} = \hat{P} \tag{7.6}$$

ou seja, a proporção eqüivale a uma média aritmética para dados de variáveis do tipo "0-1". Assim, as propriedades da distribuição amostrai da média também são aplicadas à distribuição amostrai da proporção. Usando as notações próprias da proporção, temos:

a) O valor esperado da proporção amostrai é igual à proporção da população:

$$\boxed{E(\hat{P}) = p} \tag{7.7}$$

b) A variância da proporção amostrai é dada por

$$\boxed{V(\hat{P}) = \frac{p(1-p)}{n}} \tag{7.8}$$

(se a amostragem for *com* reposição, ou N muito grande ou infinito).

ou

$$\boxed{V(\hat{P}) = \frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}} \tag{7.9}$$

(se a amostragem for *sem* reposição e N não muito grande, $N < 20n$)

c) Se o tamanho da amostra for razoavelmente *grande,* então a distribuição amostrai da proporção pode ser aproximada pela *distribuição normal.*[3]

As aplicações da distribuição da proporção amostrai podem ser feitas no contexto da aproximação da distribuição normal à binomial, inclusive com a correção de continuidade (Seção 6.3.1).

EXERCÍCIOS

4. Uma fundição produz blocos para motor de caminhões. Os furos para as camisas devem ter diâmetro de 100 mm, com tolerância de 5 mm. Para verificar qual é o diâmetro *médio* no processo, a empresa vai retirar uma amostra com 36 blocos e medir os diâmetros de 36 furos (1 a cada bloco). Suponha que o desvio padrão (populacional) dos diâmetros seja conhecido e igual a 3 mm.

a) Qual é o desvio padrão da distribuição da média amostrai?

b) Qual é a probabilidade de a média amostrai diferir da média populacional (desconhecida) em mais do que 0,5 mm (para mais ou para menos)?

c) Qual é a probabilidade de a média amostrai diferir da média populacional (desconhecida) em mais do que 1 mm (para mais ou para menos)?

d) Se alguém afirmar que a média amostrai não se distanciará da média populacional em mais do que 0,98 mm, qual é a probabilidade de essa pessoa acertar?

e) Se alguém afirmar que a média amostrai não se distanciará da média populacional em mais do que 1,085 mm, qual é a probabilidade de essa pessoa errar?

5. Uma empresa fabricante de pastilhas para freios efetua um teste para controle de qualidade de seus produtos. Supondo que 1% das pastilhas fabricadas pelo processo atual apresenta desempenho deficiente quanto ao nível de desgaste, qual é a probabilidade, em uma amostra aleatória simples com 10.000 pastilhas, de serem encontradas 85 ou menos pastilhas com problemas?

6. Sabe-se que 50% dos edifícios construídos em uma grande cidade apresentam problemas estéticos relevantes em menos de cinco anos após a entrega

3 Observamos que a distribuição exata é a binomial (ou a hipergeométrica se a amostragem for feita de população pequena e sem reposição).

da obra. Considerando a seleção de uma amostra aleatória simples com 200 edifícios com cinco anos, qual é a probabilidade de menos de 90 deles apresentarem problemas estéticos relevantes (considerar que não tenha havido obras de reparo nos edifícios selecionados)?

7. Existem vários algoritmos computacionais que permitem gerar números aleatórios (ou, mais apropriadamente, *pseudo-aleatórios)* no intervalo [0, 1], com distribuição uniforme. Considere a geração de 100 números (X_1, X_2, ..., X_{100}) desta forma e seja X a média aritmética simples desses 100 números.

a) Qual é o valor esperado e a variância de X_1

b) Qual é a probabilidade de X_1 assumir um valor no intervalo [0,47, 0,53]?

c) Qual é o valor esperado e a variância de X'?

d) Qual é a distribuição de probabilidade de X'?

e) Qual é a probabilidade de X' assumir um valor no intervalo [0,47, 0,53]?

8. Um profissional de Computação observou que seu sistema gasta entre 20 e 24 segundos para realizar determinada tarefa. Além disso, o tempo gasto, *X,* pode ser razoavelmente representado pela seguinte função de densidade:

$$f(x) = \begin{cases} \dfrac{x}{4} - 5, & \text{para } 20 \le x < 22 \\ 6 - \dfrac{x}{4}, & \text{para } 22 \le x < 24 \\ 0, & \text{para } x \notin [20, 24] \end{cases}$$

a) Numa particular rodada, qual é a probabilidade de o sistema gastar mais que 22,4 segundos?

b) Em 30 rodadas, qual é a probabilidade de o sistema gastar, em média, mais que 22,4 segundos por rodada?

7.3 ESTIMAÇÃO DE PARÂMETROS

Nesta seção, estudaremos o problema de avaliar parâmetros populacionais, a partir de operações com os dados de uma amostra. É um raciocínio tipicamente indutivo, em que se generalizam resultados *da parte* (amostra) para o *todo* (população), conforme ilustra a Figura 7.6.

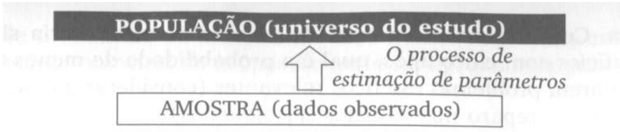


Figura 7.6 O *raciocínio indutivo da estimação.*

Por exemplo, podemos ter interesse em avaliar a resistência mecânica X de um novo material. Contudo, X não é um número, mas uma variável aleatória, porque há uma infinidade de fatores não controláveis que provocarão variações nas possíveis medidas de resistência mecânica do material. É até razoável admitir que a distribuição de X seja aproximadamente normal, por se tratar de medidas físicas. E o interesse pode estar na avaliação dos *parâmetros populacionais* $\mu = E(X)$ e $\sigma^2 = V(X)$.

Medidas de resistência mecânica $(X_1, X_2, \quad X_n)$, a serem realizadas de forma independente e sob as mesmas condições, constituem uma *amostra aleatória simples* de X. Cálculos podem ser feitos sobre essas medidas para *estimar* os parâmetros de interesse. Exemplos desses cálculos podem ser

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{e} \qquad \textbf{(7.10)}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \qquad (7.11)$$

que são *estimadores* dos parâmetros μ e σ^2, respectivamente.

De forma genérica, considere uma população caracterizada pela distribuição de certa variável aleatória X, com parâmetro θ. E seja $(X_1, X_2, \quad X_n)$ uma amostra aleatória simples de X.

Uma *estatística T é* uma função dos elementos da amostra, isto é $T = f(X_1, X_2, \quad X_n)$. Quando ela é usada para avaliar certo parâmetro θ, é também chamada de **estimador** de θ.

Observe que um estimador é uma variável aleatória, pois depende da amostra a ser selecionada. Realizada a amostragem, o estimador assume determinado valor (o resultado do cálculo), o qual denominamos de *estimativa.*[4]

4 Neste texto, as estimativas serão representadas por letras minúsculas, contrastando com os estimadores, os quais serão representados por letras maiúsculas.

Algumas propriedades desejáveis de um estimador serão discutidas, o que permite, pelo menos em tese, escolher o *melhor* estimador para cada situação prática. Usando as distribuições amostrais, é possível avaliar probabilisticamente o erro que se está cometendo por se usar uma amostra e não toda a população - o *erro amostral*. Isso será feito com estimativas em forma de *intervalos de confiança* (Seções 7.3.2 e 7.3.3).

7.3.1 Propriedades de um estimador

Um estimador, por ser uma variável aleatória, pode assumir valores, segundo uma distribuição de probabilidades. Contudo, é desejável que, em média, ele seja igual ao parâmetro que se deseja estimar. Mais formalmente,

> T é um estimador **não-viesado** (ou **não-tendencioso**) de um parâmetro θ se e só se $E(T) = \theta$.

Por exemplo, $\bar{X}$ e $\hat{P}$ são estimadores *não viesados* dos parâmetros μ e p, respectivamente, porque $E(\bar{X}) = \mu$ e $E(\hat{P}) = p$, conforme foi visto na Seção 7.2. Já o estimador

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \tag{7.12}$$

é um estimador *viesado* do parâmetro σ^2, pois $E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$. Por isso, definimos a variância amostral, S^2, com denominador $(n - 1)$ no lugar de n. A diferença

$$E(\hat{\sigma}^2) - \sigma^2 = -\frac{1}{n}\sigma^2 \tag{7.13}$$

é chamada de ***viés* do estimador** $\hat{\sigma}^2$.

Na prática, retiramos só uma amostra, produzindo um único valor para o estimador - uma *estimativa*. Mesmo que o estimador seja não viesado, o valor da estimativa pode estar longe do valor do parâmetro. Outra propriedade desejável é que o estimador tenha variância pequena, porque isso reduz a chance de a estimativa acusar um valor distante do parâmetro.

Dados dois estimadores não viesados T_1 e r_2, sendo $V(T_1) < V(T_2)$, então T_1 é dito mais *eficiente* do que T_2. E a *eficiência relativa* de T_1 em relação a T_2 é dada por:

$$ef(T_1, T_2) = \frac{V(T_2)}{V(T_1)} \quad (6.14)$$

A Figura 7.7 ilustra os conceitos de viés e eficiência, fazendo analogia com tiros ao alvo, realizados por três rifles. Os rifles T_1 e T_2 são *não viesados*, porque, em média, acertam o alvo; enquanto o rifle T_3 é viesado. Embora T_1 e T_2 sejam não viesados, T_1 é mais *eficiente* do que T_2, pois a variância entre os tiros é menor.

Figura 7.7 *Tiros ao alvo com três rifles.*

Considere a média amostrai (X') e a mediana amostrai (M_d) como estimadores do parâmetro Sabemos que $V(X') = \sigma^2/n$. Se supusermos a população com distribuição normal e a amostra grande, é possível mostrar que $V(M_d) \approx$

/2) * (σ^2/n) . Como $V(X') < V(M_d)$, então X' é um estimador *mais eficiente* do que M_d

na estimação de nas condições estabelecidas. E a *eficiência relativa* de X' em relação à M_d é dada por

$$\frac{V(M_d)}{V(\bar{X})} \approx \frac{\pi\sigma^2/2n}{\sigma^2/n} \approx 1{,}57 \quad (7.15)$$

Assim, em amostras grandes de populações normais, a média amostrai é cerca de 57% mais eficiente do que a mediana amostrai. Isso significa que se formos usar M_d no lugar de X', precisamos ter uma amostra 57% maior, para garantir a mesma eficiência na estimação de μ.

De modo geral, a qualidade de um estimador *T,* na estimação de um parâmetro θ, pode ser avaliada em função de seu *erro quadrático médio,* o qual é definido por

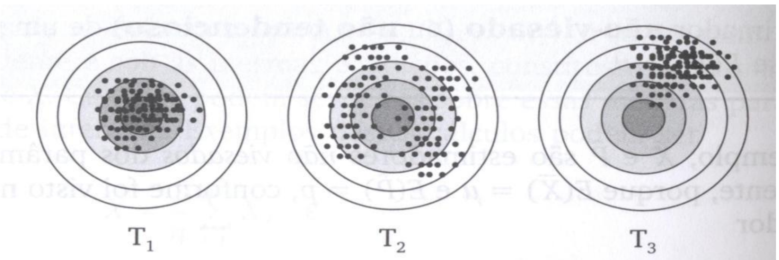
$$EQM(T) = E(T-\theta)^2 \qquad (7.16)$$

que pode ser escrito como[5]

$$EQM(T) = V(T) + (\text{viés})^2 \qquad (7.17)$$

Então, para um estimador *T não viesado,* temos *EQM(T) = V(T).*

Para dois estimadores e T_1 T_2 quaisquer (não necessariamente não viesados), definimos a *eficiência relativa* de T_1 em relação à T_2 por:



7.3.2 Intervalo de confiança para proporção

Em muitas situações, o principal parâmetro de interesse é alguma proporção p. Por exemplo:

- a proporção de itens defeituosos em uma linha de produção;
- a proporção de consumidores que vão comprar certo produto;
- a proporção de mensagens que chegam adequadamente a seu destino etc.

Seja uma população caracterizada por uma variável aleatória *X,* que assume o valor 0 ou 1, conforme o elemento tenha ou não certo atributo de interesse. Por exemplo, nas peças que saem de uma linha de produção, o código C pode identificar peça boa e o código 1 peça defeituosa. Para um elemento tomado ao acaso, seja $p = P(X = 1)$. Note que *p* representa a proporção de elementos com o atributo, na população.

Já vimos que a proporção amostrai *P* é um bom estimador da proporção populacional *p.* Dada uma amostra aleatória simples de tamanho *n,* o que se pode dizer sobre o *erro amostrai:* $|\hat{P} - p|$?

Conforme discutido na Seção 7.2, se *n* for grande, a distribuição amostra] de *P* é aproximadamente normal com[6]

5 $EQM(T) = E\{T - E(T) + E(T) - \theta\}^2 = E\{T - E(T)\}^2 + \{E(T) - \theta\}^2 + 2E\{T - E(T)\}\{E(T) - \theta\} = E\{T - E(T)\}^2 + \{E(T) - \theta\}^2 + 0 = V(T) + (\text{viés})^2.$

6 Estamos supondo população infinita ou bastante grande. Caso contrário, deveríamos usar a correção (7.9), como discutido na Seção 7.2.

$$E(\hat{P}) = p \text{ e} \tag{7.19}$$

$$V(\hat{P}) = \frac{p(1-p)}{n} \tag{7.20}$$

Denotaremos por

$$\sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{n}} \tag{7.21}$$

o desvio padrão da distribuição amostrai de P, que no presente contexto será chamado de **erro padrão** de P.

Seja uma variável aleatória normal padrão, Z. Como vale a relação[7]

$$P\{-1{,}96 \leq Z \leq 1{,}96\} = 0{,}95 \tag{7.22}$$

podemos escrever

$$P\left\{-1{,}96 \leq \frac{\hat{P}-p}{\sigma_{\hat{P}}} \leq 1{,}96\right\} = 0{,}95 \tag{7.23}$$

ou

$$P\left\{\hat{P} - (1{,}96)\sigma_{\hat{P}} \leq p \leq \hat{P} + (1{,}96)\sigma_{\hat{P}}\right\} = 0{,}95 \tag{7.24}$$

ou seja, com probabilidade de 95%, temos:

$$\left|\hat{P} - p\right| \leq (1{,}96)\sigma_{\hat{P}} \tag{7.25}$$

Observada efetivamente a amostra, e chamando de p a proporção obtida nesta amostra, podemos definir um *intervalo de confiança* para p, com *nível de confiança* de 95%, por

$$IC(p,\ 95\%) = \hat{p} \pm (1{,}96)\sigma_{\hat{P}} \tag{7.26}$$

Na prática, $\sigma_{\hat{P}}$ não pode ser calculado por (7.21), porque depende do parâmetro desconhecido p. Então, usamos em seu lugar a estimativa:

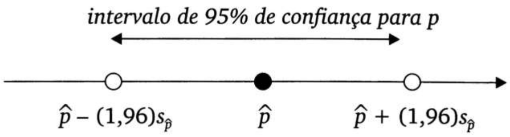
7 Ver Exemplo 6.5 (Capítulo 6).

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{7.27}$$

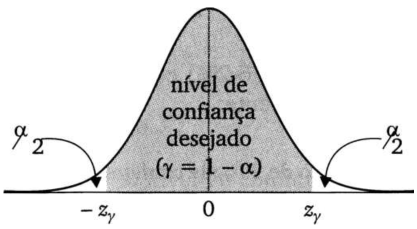
Desde que a amostra seja grande (p. ex., $n \geq 50$), a diferença entre s_p e σ_p pode ser considerada desprezível, e um *intervalo de confiança* para p, com *nível de confiança* de 95%, pode ser calculado por:

$$IC(p,\ 95\%) = \hat{p} \pm (1{,}96)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{7.28}$$

Esquematicamente,



Em suma, embora p seja um parâmetro populacional desconhecido, é possível, com base em uma amostra aleatória simples, construir um intervalo que deve conter p com alto nível de confiança. É bastante usual o nível de confiança de 95%, mas o intervalo pode ser construído com um nível y qualquer, bastando encontrar o valor de z_y na distribuição normal padrão, conforme mostra a Figura 7.8.



y	0,800	0,900	0,950	0,980	0,990	0,995	0,998
zy	1,282	1,645	1,960	2,326	2,576	2,807	3,090

Figura 7.8 *Valores de z_y para alguns níveis de confiança.*

Com Zy tomado adequadamente, conforme o esquema da Figura 7.8, calculamos o intervalo de confiança para p por:

$$\boxed{IC(p,\ \gamma) = \hat{p} \pm z_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \qquad (7.29)$$

Exemplo 7.3 Na avaliação de dois sistemas computacionais, *A* e *B,* foram selecionadas 400 cargas de trabalho (tarefas) - supostamente uma amostra aleatória da infinidade de cargas de trabalho que poderiam ser submetidas a esses sistemas. O sistema *A* foi melhor que o *B* em 60% dos casos. Construir intervalos de confiança para p (proporção de vezes que o sistema *A é* melhor que o sistema *B,* considerando todas as possíveis cargas de trabalho) usando níveis de confiança de 95% e 99%.

Para o nível de confiança de 95%, temos z_y = 1,96, resultando em

$$IC(p,\ 95\%) = \hat{p} \pm (1{,}96)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0{,}6 \pm (1{,}96)\sqrt{\frac{0{,}6(0{,}4)}{400}} =$$

$$= 0{,}600 \pm 0{,}048$$

ou, em porcentagens:

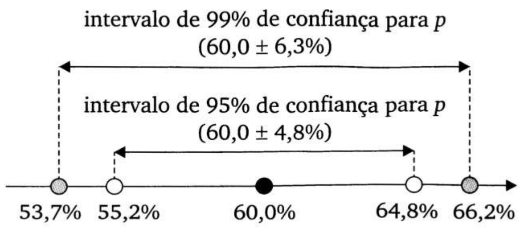
$$\text{IC}(p, 95\%) = 60{,}0\% \pm 4{,}8\%$$

Concluímos, então, que o intervalo (55,2%; 64,8%) contém o parâmetro p, com nível de confiança de 95%.

Para o nível de confiança de 99%, temos z_y = 2,576, resultando em

$$IC(p,\ 99\%) = 0{,}6 \pm (2{,}576)\sqrt{\frac{0{,}6(0{,}4)}{400}} = 0{,}600 \pm 0{,}063$$

Ou seja, o intervalo (53,7%; 66,2%) contém o parâmetro p, com nível de confiança de 99%. Esquematicamente:



Observe que, ao exigir maior nível de confiança, o intervalo de confiança **aumenta** em magnitude. Tente entender o porquê disso! Para *dado nível de con-*

plitude de seu intervalo de confiança. A forma natural de aumentar a precisão é aumentando o tamanho da amostra. Voltaremos a esse ponto na Seção 7.4.

7.3.3 Intervalo de confiança para média

Seja uma população caracterizada pela distribuição de uma variável aleatória X com os seguintes parâmetros: $E(X) = \mu$ e $V(X) = \sigma^2$. Por exemplo, X pode representar a mensuração da resistência mecânica de um novo material. Devido ao erro experimental onipresente, X é uma variável aleatória; assim, o interesse recai em seu valor esperado µ.

Considere uma amostra aleatória simples $\{X_1, X_2, \ldots, X_n\}$ de X. Supondo X com distribuição aproximadamente normal, então

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (7.30)$$

é o estimador natural de µ. Vimos na Seção 7.2 que X' tem distribuição aproximadamente normal com média e variância dadas por[8]

$$E(\bar{X}) = \mu \qquad (7.31)$$

$$V(\bar{X}) = \frac{\sigma^2}{n} \qquad (7.32)$$

O desvio padrão da distribuição amostral de $\bar{X}$,

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \qquad (7.33)$$

será chamado de **erro padrão** de X'.

Escolhendo z_y em função do nível de confiança y desejado, tal que P{- $z_y \leq Z \leq z_y$} = y, podemos escrever

$$P\left\{-z_\gamma \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_\gamma\right\} = \gamma \qquad (7.34)$$

ou

8 Novamente estamos supondo a população infinit
(7.3), como visto na Seção 7.2.

$$P\left\{\bar{X} - z_\gamma \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_\gamma \frac{\sigma}{\sqrt{n}}\right\} = \gamma \qquad \textbf{(7.35)}$$

Observada efetivamente a amostra, e chamando de a média aritmética dos dados, podemos definir um intervalo de confiança para µ, com nível de confiança ϒ, por:

$$IC(\mu,\ \gamma) = \bar{x} \pm z_\gamma \frac{\sigma}{\sqrt{n}} \qquad (7.36)$$

Exemplo 7.4 Em uma indústria de cerveja, a quantidade de cerveja inserida em latas tem-se comportado como uma variável aleatória com média 350 ml e desvio padrão 3 ml. Após alguns problemas na linha de produção, suspeita-se que houve alteração na média. Uma amostra de 20 latas acusou média = 346 ml. Construa um intervalo de confiança para o novo valor da quantidade média // de cerveja inserida em latas, com nível de confiança 95%, supondo que não tenha ocorrido alteração no desvio padrão do processo.

Solução:

$$IC(\mu,\ 95\%) = \bar{x} \pm (1{,}96)\frac{\sigma}{\sqrt{n}} = 346 \pm (1{,}96)\frac{3}{\sqrt{20}} = 346 \pm 1{,}31\ ml$$

Interpretando: a quantidade média µ de cerveja inserida em latas, após os problemas na linha de produção, é 346 ml, tolerando, com 95% de confiança, uma margem de erro de até 1,31 ml. Assim, o intervalo (344,69; 347,31) contém, com 95% de confiança, o valor µ. Isso mostra que estatisticamente houve alteração na média do processo, pois o valor da média antiga (350 ml) não pertence ao intervalo.

Desvio padrão desconhecido

O intervalo de confiança descrito em (7.36) somente poderá ser usado nas situações em que conhecemos o desvio padrão a da população, o que não é comum na prática. Caso contrário, o procedimento usual é substituir a pelo desvio padrão calculado com os dados da amostra:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)} \qquad (7.37)$$

Duas situações a considerar:

1. quando a amostra for grande (digamos, $n \ge 50$), a diferença entre a e s pode ser desprezível, permitindo ainda o uso de (7.36);
2. quando a amostra for pequena, é necessário efetuar uma correção, como veremos a seguir.

A distribuiçko *t* de Student

Supondo a população com distribuição normal, a estatística

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \tag{7.38}$$

tem distribuição de probabilidades conhecida como *distribuição t de Student,* com $gl = n - 1$ graus de liberdade.

A *distribuição t de Student,* como mostra a Figura 7.9, tem forma parecida com a normal padrão, mas é um pouco mais dispersa. Essa dispersão varia com o tamanho da amostra. É bastante dispersa para amostras pequenas, mas se aproxima da normal padrão para amostras grandes.

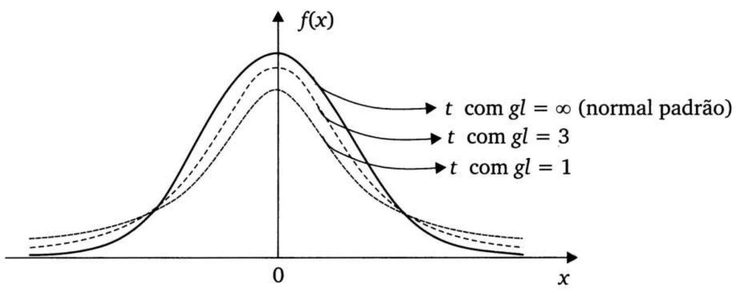


Figura 7.9 *Gráficos de distribuições* t *de Student e da normal padrão.*

Dado um nível de confiança y, podemos obter o valor t_y da distribuição t de Student, usando a Tabela 4 do apêndice, na linha correspondente a $gl = n - 1$. A Figura 7.10 ilustra esse processo, com $gl = 9$ e nível de confiança de 95%.

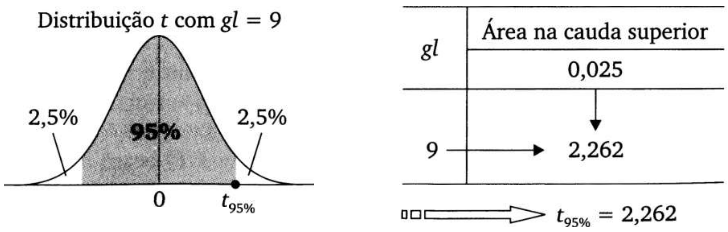


Figura 7.10 *Uso da tabela da distribuição* t *de Student: ilustração com* gl *- 9 e nível de confiança de 95%.*

Intervalo de confiança para μ, **com uso da distribuição** t **de** *Student*

Usando a Tabela 4 com $gl = n - 1$, podemos escolher o valor t_γ em função do nível de confiança y desejado, tal que

$$P\left\{-t_\gamma \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_\gamma\right\} = \gamma \tag{7.39}$$

ou

$$P\left\{\bar{X} - t_\gamma \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_\gamma \frac{S}{\sqrt{n}}\right\} = \gamma \tag{7.40}$$

Assim, o intervalo de confiança para µ com a amostra efetivamente observada, é dado por[9]

$$\boxed{IC(\mu,\ \gamma) = \bar{x} \pm t_\gamma \frac{s}{\sqrt{n}}} \tag{7.41}$$

Exemplo 7.5 Deseja-se avaliar a dureza esperada μ do aço produzido sob un novo processo de têmpera. Uma amostra de dez corpos de prova do aço produ ziu os seguintes resultados de dureza, em HRc:

36,4 35,7 37,2 36,5 34,9 35,2 36,3 35,8 36,6 36,9

Construir um intervalo de confiança para com nível de confiança de 95%.

Calculando as estatísticas para a amostra observada, temos:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 36{,}15$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)} = 0{,}7352$$

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = 0{,}2325$$

9 Para amostras pequenas, a validade do *IC(µ, y)* está condicionada à suposição de que os dados provenham de uma distribuição aproximadamente normal. Para n > 30, o teorema limite central garante a validade de *IC(µ, Υ)* e, além disso, tem-se para *N* grande: t_γ « z_r

Como vimos na Figura 7.10, usando nível de confiança Υ = 95%, temos, pela Tabela 4 com gl = 9, o valor $t_{95\%}$ = 2,262, resultando em:

$$IC(\mu,\ 95\%) = \bar{x} \pm t_{95\%} \frac{s}{\sqrt{n}} = 36{,}15 \pm 0{,}53$$

Ou seja, a resistência mecânica esperada do aço produzido pelo novo processo de têmpera é 36,15 HRc, tolerando, com 95% de confiança, uma margen] de erro de até 0,53 HRc.

EXERCÍCIOS

9. Sejam $X_1, X_2,$ X_7 uma amostra aleatória simples de uma população com média// e desvio padrão a. Considere os seguintes estimadores de μ :

$T_1 = (X_1, +X_2+X_3+X_4+X_5+X_6+X_7)/7$

$T_2 = (X_2 +X_3+X_4+X_5+X_6) / 5$

$T_3 = (X_2 + X_3 + X_4 + X_5 + X_6) / 7$

a) Quais estimadores são não viciados? Justifique.

b) Qual estimador é o mais eficiente entre T_1 e T_2? Justifique.

10. Em uma amostra aleatória simples com 200 edifícios com cinco anos, em certa cidade, 55% apresentaram problemas estéticos relevantes após a entrega da obra. Construir um intervalo de confiança para a proporção de edifícios da cidade que apresentam problemas estéticos relevantes nos cinco primeiros anos. Use nível de confiança de 95%.

11. Uma empresa fabricante de pastilhas para freios efetua um teste para controle de qualidade de seus produtos. Selecionou-se uma amostra de 600 pastilhas, das quais 18 apresentaram níveis de desgaste acima do tolerado. Construir um intervalo de confiança para a proporção de pastilhas com desgaste acima do tolerado, do atual processo industrial, com nível de confiança de 95%. Interpretar o resultado.

12. Uma fundição produz blocos para motor de caminhões. Os blocos têm furos para as camisas e deseja-se verificar qual é o diâmetro *médio* no processo do furo. A empresa retirou uma amostra de 36 blocos e mediu os diâmetros de 36 furos (1 a cada bloco). A amostra acusou média de 98,0 mm e desvio padrão de 4,0 mm. Construir um intervalo de confiança para a média do processo, com nível de confiança de 99%. Interpretar o resultado. Se o processo deveria ter média 100 mm, há evidência (com 99% de confiança) de que a média do processo não está no valor ideal? Explique.

7.4 TAMANHO DE AMOSTRA

Na seção anterior, aprendemos como estimar um parâmetro mediante observação de uma amostra aleatória simples de tamanho n. A estimação é feita com certa precisão, no sentido de que também avaliamos o *erro amostrai* que podemos estar cometendo.

Contudo, ainda na fase do planejamento da pesquisa, muitas vezes precisamos calcular o tamanho n da amostra, para garantir certa precisão desejada, que é descrita em termos do *erro amostrai máximo tolerado* (E_{0y}) e do nível de confiança (γ) a ser adotado no processo de estimação. No caso de estimação de μ, podemos exigir

$$|\bar{X} - \mu| \leq E_0 \tag{7.42}$$

Usando (7.34), temos:

$$z_\gamma \frac{\sigma}{\sqrt{n}} \leq E_0 \tag{7.43}$$

Isolando n, temos:

$$n \geq \frac{z_\gamma^2 \sigma^2}{E_0^2} \tag{7.44}$$

O tamanho mínimo da amostra é o menor n que satisfaz a inequação precedente.

A dificuldade operacional para calcularmos o tamanho da amostra é que o cálculo depende da variância populacional σ^2, que em geral é desconhecida. Em alguns problemas, σ^2 pode ser avaliada por meio de estudos anteriores ou pela experiência do engenheiro; em outras situações, σ^2 é obtida de uma *amostragem piloto*, isto é, alguns elementos da população são examinados e a variância encontrada nesta amostra piloto é usada no lugar de σ^2. Nesse caso, é melhor usar a expressão (7.44) com t_y no lugar de z_y.

Exemplo 7.5 (continuação) Considere que o pesquisador julgou o resultado encontrado, *IC*(μ, 95%) = 36,15 ± 0,53, pouco preciso. Ele tolera um erro amostrai máximo de 0,3 HRc. Além disso, ele quer realizar as estimações com nível de confiança de 99%. Qual deve ser o tamanho da amostra?

Solução: Para efetuar o cálculo, é necessário o conhecimento da variância populacional, σ^2. Usaremos, em seu lugar, a variância calculada sobre as dez observações, isto *é* $s^2 = (0{,}7352)^2 \approx 0{,}54$. Logo,

$$n \geq \frac{t_{99\%}^{2}\,\sigma^{2}}{E_0^{2}} = \frac{(3{,}250)^{2}(0{,}54)}{(0{,}3)^{2}} = 63{,}375$$

Portanto, precisamos de $n = 64$ corpos de prova para satisfazer a precisão desejada.

O Quadro 7.1 apresenta o formulário para o cálculo do tamanho da amostra, em função do parâmetro desejado. Quando a população é infinita ou muito grande, n_0 já é o tamanho da amostra. Porém, se N é conhecido e, especialmente, se não for muito grande, devemos dar seqüência ao cálculo, conforme indica a última linha do Quadro 7.1.

Quadro 7.1 *Tamanho mínimo de uma amostra aleatória simples.*

Parâmetro de interesse	Valor inicial do tamanho da amostra	
a) uma média (μ):	$n_0 = \dfrac{z_\gamma^2\,\sigma^2}{E_0^2}$	(7.45)
b) uma proporção (p):	$n_0 = \dfrac{z_\gamma^2\,p(1-p)}{E_0^2}$	(7.46)
c) várias proporções (p_1, p_2, ...):	$n_0 = \dfrac{z_\gamma^2}{4E_0^2}$	(7.47)
Tamanho da amostra		
População infinita:	$n = n_0$ (arredondamento para o inteiro superior)	(7.48)
População de tamanho N:	$n = \dfrac{N.n_0}{N + n_0 - 1}$ (arredondamento para o inteiro superior)	(7.49)

Quando o objetivo é estimar uma proporção p ($0 < p < 1$), podemos usar a seguinte relação (ver a Figura 7.11):

$$\sigma^2 = p \cdot (1-p) \leq \frac{1}{4} \qquad (7.50)$$

A primeira parte da relação (7.50) foi usada para construir a expressão de n_0 em (7.46) (Quadro 7.1b); e a segunda em (7.47) (Quadro 7.1c).

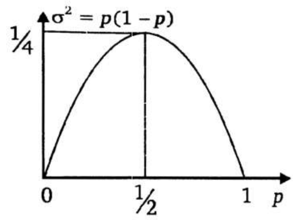


Figura 7.11 *O parâmetro σ^2 em função do valor da proporção p.*

A expressão (7.47) (Quadro 7.1c) também é usada com o objetivo de estimar uma única proporção *p*, quando não temos informação *a priori* sobre o valor de p. Contudo, deve resultar em um valor de n maior do que o necessário.

No caso de estarmos usando nível de confiança de 95%, temos $z_y = 1{,}96 \approx 2$, fazendo com que (7.47) resulte em:

$$n_0 = \frac{1}{E_0^{\ 2}} \tag{7.51}$$

A expressão (7.51) é muito usada no planejamento de pesquisas de levantamento, com o objetivo de estimar várias proporções, como nos exemplos seguintes:

- numa pesquisa eleitoral, em que é comum a necessidade de avaliar a proporção de cada candidato;
- em pesquisas de mercado, em que normalmente desejam-se avaliar as proporções de vários atributos nos consumidores;
- no levantamento de arquivos que trafegam em uma rede, em que é comum o interesse em verificar a proporção de cada tipo de arquivo.

EXERCÍCIOS

13. Um pesquisador precisa determinar o tempo médio gasto para perfurar três orifícios em uma peça de metal. Qual deve ser o tamanho da amostra para que a média amostrai esteja a menos de 15 s da média populacional? Por experiência prévia, pode-se supor o desvio padrão em torno de 40 s. Considere também, que a estimação será realizada com nível de confiança de 95%.

14. Seja a construção de um plano para garantir a qualidade dos parafusos vendidos em caixas com 100 unidades. Um dos requisitos é controlar o comprimento médio dos parafusos. Quer-se saber quantos parafusos de-

ve-se examinar em cada caixa, para garantir que a média da amostra ( ) não difira do comprimento médio dos parafusos da caixa (μ) em mais que 0,8 mm. Considere que a estimação seja realizada com nível de confiança de 95%. Análises feitas na linha de produção indicam variância em torno de 2 mm^2.

15. Considerando o Exercício 14, mas supondo a caixa com 1.000 parafusos, qual é o tamanho da amostra necessário?

16. Com o objetivo de avaliar a confiabilidade de um novo sistema de transmissão de dados, torna-se necessário verificar a proporção de *bits* transmitidos com erro em cada lote de 100 Mb. Considere que seja tolerável um erro amostrai máximo de 2% e que em sistemas similares a taxa de erro na transmissão é de 10%. Qual deve ser o tamanho da amostra?

a) Use $\gamma = 0{,}95$.

b) Use $\gamma = 0{,}99$.

EXERCÍCIOS COMPLEMENTARES

17. Sob condições normais, realizaram-se dez observações sobre o tempo de resposta de uma consulta a certo banco de dados. Os resultados, em segundos, foram:

28 35 43 23 62 38 34 27 32 37

Construa um intervalo de confiança para o tempo médio de uma consulta, sob condições normais. Use $\gamma = 0{,}99$.

18. Fixados certos parâmetros de entrada, o tempo de execução de um algoritmo foi medido 12 vezes, obtendo-se os seguintes resultados, em minutos:

15 12 14 15 16 14 16 13 14 11 15 13

a) Apresente um intervalo de 95% de confiança para o tempo médio de execução do algoritmo.

b) Considerando as 12 mensurações como uma amostra piloto, avalie o número de mensurações (tamanho da amostra) necessário para garantir um erro máximo de 15 segundos (0,25 minutos). Use $\gamma = 0{,}95$.

19. Uma empresa tem 2.400 empregados. Deseja-se extrair uma amostra de empregados para verificar o grau de satisfação em relação à qualidade da comida no refeitório. Em uma amostra piloto, numa escala de 0 a 10, o grau de satisfação recebeu nota média 6,5 e desvio padrão 2,0.

a) Determine o tamanho mínimo da amostra, supondo amostragem aleatória simples, com erro máximo de 0,5 unidade e nível de confiança de 99%.

b) Considere que a amostra planejada no item anterior tenha sido realizada e obteve-se média 5,3 e desvio padrão 1,8 ponto. Construa um intervalo de 99% de confiança para o parâmetro μ.

c) Considerando o resultado do item anterior, você diria, com nível de confiança de 99%, que a nota média seria superior a cinco se a pesquisa fosse aplicada a todos os 2.400 funcionários? Justifique.

d) Realizada a amostra planejada no item (a), suponha que 70 atribuíram notas iguais ou superiores a cinco. Apresente um intervalo de 90% de confiança para a porcentagem de indivíduos da população que atribuiriam notas iguais ou superiores a cinco.

20. Com os dados históricos sobre a temperatura do pasteurizador de um laticínio, sabe-se que a variância é aproximadamente 1,8 $(°C)^2$. Planeja-se fazer uma amostragem para avaliar o valor médio da temperatura do pasteurizador. Suponha que as observações sejam feitas sob as mesmas condições e de forma independente. Qual deve ser o tamanho da amostra, para garantir um erro máximo de 0,3°C, com nível de confiança de 95%?

21. Planeja-se extrair uma amostra aleatória simples dos 2.000 funcionários de uma empresa, para avaliar a satisfação com o trabalho. A satisfação será avaliada através de um questionário com vários itens numa escala de 1 a 5. Pretende-se avaliar o valor médio de cada item. Qual deve ser o tamanho da amostra para garantir um erro máximo de 0,2 unidade, com nível de confiança de 95%?

Nota: Use como variância o valor teórico que se obtém ao supor probabilidade igual para cada um dos cinco níveis da escala. Observe que dificilmente algum item terá variância maior do que esta que você está calculando, pois, na prática, a tendência é que as respostas se concentrem em torno de algum nível.

22. Numa pesquisa para estudar a preferência do eleitorado a uma semana da eleição presidencial, qual deve ser o tamanho de uma amostra aleatória simples para garantir, com nível de confiança de 95%, um erro amostrai não superior a 2%?

23. Um analista de sistemas está avaliando o desempenho de um novo programa de análise numérica. Forneceu como entrada do programa 14 operações similares e obteve os seguintes tempos de processamento (em milissegundos):

12,0	13,5	16,0	15,7	15,8	16,5	15,0	13,1
15,2	18,1	18,5	12,3	17,5	17,0		

a) Calcule a média e o desvio padrão da amostra do tempo de processamento.

b) Construir um intervalo de confiança para o tempo médio de processamento, com nível de confiança de 95%.

c) Qual deve ser o tamanho da amostra para garantir um erro amostrai máximo de 0,5 milissegundo, na estimação do tempo médio de processamento, com nível de confiança de 99%?

24. Uma unidade fabril da Intel produziu 500.000 *chips Pentium IV* em certo período. São selecionados, aleatoriamente, 400 *chips* para testes.

a) Supondo que 20 *chips* não tenham a velocidade de processamento adequada, construir o intervalo de confiança para a proporção de *chips* adequados. Use nível de confiança de 95%.

b) Verificar se essa amostra é suficiente para obter um intervalo de 99% de confiança, com erro amostrai máximo de 0,5%, para a proporção de *chips* adequados. Caso contrário, qual deveria ser o tamanho da amostra?