Snow666

EXEMPLO 8.2 Verifique se a correspondência dada por

$$f(x) = \frac{x+3}{15} \quad \text{para} \quad = 1, 2 \text{ e } 3$$

pode ser a distribuição de probabilidade de alguma variável aleatória.

Solução Substituindo $x = 1$, 2 e 3 em $\frac{x+3}{15}$, obtemos $f(1) = \frac{4}{15}$, $f(2) = \frac{5}{15}$ e $f(3) = \frac{6}{15}$. Como nenhum desses valores é negativo ou maior do que 1, e como sua soma é

$$\frac{4}{15} + \frac{5}{15} + \frac{6}{15} = 1$$

a função dada pode ser a distribuição de probabilidade de alguma variável aleatória.

8.3 A DISTRIBUIÇÃO BINOMIAL

Existem muitos problemas aplicados em que estamos interessados na probabilidade de que um evento vá ocorrer x vezes a cada n tentativas. Por exemplo, podemos estar interessados na probabilidade de obter 45 respostas a 400 questionários distribuídos como parte de um estudo sociológico, a probabilidade de 5 em 12 ratos sobreviverem por determinado prazo após a injeção de uma substância cancerígena, a probabilidade de 45 em 300 motoristas retidos numa barreira de trânsito estarem usando seus cintos de segurança, ou a probabilidade de 66 em 200 telespectadores entrevistados por um serviço de medição de audiência lembrarem quais produtos foram anunciados num determinado programa. Utilizando a linguagem dos jogos de azar, poderíamos dizer, em cada um desses exemplos, que estamos interessados na probabilidade de obter "x sucessos em n provas" ou, em outras palavras, "x sucessos e $n - x$ fracassos em n tentativas."

Nos problemas que estudaremos nesta seção, faremos sempre as seguintes hipóteses:

Há um número fixo de provas.

A probabilidade de sucesso é a mesma em cada prova.

As provas são todas independentes.

Assim, a teoria que desenvolveremos não se aplica, por exemplo, ao número de vestidos que uma mulher pode experimentar antes de comprar um (em que o número de provas não é fixo), ou se verificarmos a cada hora se o tráfego está congestionado em certo cruzamento (em que a probabilidade de "sucesso" não é constante), ou se estivermos interessados no número de vezes que uma pessoa votou no candidato de um certo partido nas cinco últimas eleições presidenciais (em que as provas não são independentes).

No que segue, vamos conseguir estabelecer uma fórmula para resolver os problemas que verificam as condições indicadas anteriormente. Se p e $1 - p$ são as probabilidades de um sucesso e de um fracasso numa dada prova qualquer, então a probabilidade de obter x sucessos e $n - x$ fracassos *numa determinada ordem* é $p^x(1-p)^{n-x}$. Claramente, nesse produto de p com $(1 - p)$ há um fator p para cada sucesso e um fator $1 - p$ para cada fracasso e os x fatores p e os $n - x$ fatores $1 - p$ são todos multiplicados entre si em decorrência da generalização da regra especial de multiplicação para dois ou mais eventos independentes. Como essa probabilidade se aplica a qualquer ponto do espaço amostral que represente x sucessos e $n - x$ fracassos (em alguma ordem determinada), é suficiente contar quantos desses pontos há e multiplicar esse número por $p^x(1-p)^{n-x}$.

Snow666

Claramente, o número de maneiras em que podemos escolher as x provas em que devem ocorrer os sucessos é $\binom{n}{x}$, e assim chegamos ao resultado seguinte:

DISTRIBUIÇÃO BINOMIAL

A probabilidade de obter x sucessos em n provas independentes é

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \textit{para} \quad x = 0, 1, 2, \ldots, \textit{ou } n$$

onde p é a probabilidade constante de sucesso em cada prova.

Costuma-se dizer aqui que o número de sucessos em n provas é uma variável aleatória com a **distribuição binomial de probabilidade** ou, simplesmente, com a **distribuição binomial**. A distribuição binomial é assim denominada porque, para $x = 0, 1, 2, \ldots$, e n, os valores das probabilidades são os termos sucessivos da expansão binomial de $[(1-p)+p]^n$.

EXEMPLO 8.3 Verifique que a fórmula dada no Exemplo 8.1 para a probabilidade de obter x caras em quatro lançamentos de uma moeda equilibrada é, de fato, a fórmula da distribuição binomial com $n = 4$ e $p = \frac{1}{2}$.

Solução Fazendo $n = 4$ e $p = \frac{1}{2}$ na fórmula da distribuição binomial, obtemos

$$f(x) = \binom{4}{x}\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{4-x} = \binom{4}{x}\left(\frac{1}{2}\right)^4 = \frac{\binom{4}{x}}{16}$$

para $x = 0, 1, 2, 3$ e 4. Isso é precisamente a fórmula dada no Exemplo 8.1.

EXEMPLO 8.4 Se a probabilidade de um eleitor qualquer (escolhido aleatoriamente na relação oficial) votar em determinada eleição for de 0,70, qual é a probabilidade de dois dentre cinco eleitores da lista votarem na eleição?

Solução Substituindo $x = 2$, $n = 5$, $p = 0{,}70$ e $\binom{5}{2} = 10$ na fórmula da distribuição binomial, obtemos

$$f(2) = \binom{5}{2}(0{,}70)^2(1-0{,}70)^{5-2} = 10(0{,}70)^2(0{,}30)^3 = 0{,}132$$

Damos, a seguir, um exemplo em que calculamos todas as probabilidades de uma distribuição binomial.

EXEMPLO 8.5 A probabilidade de que uma pessoa fazendo compras num certo supermercado aproveite uma promoção especial de sorvete é de 0,30. Determine as probabilidades de que dentre seis pessoas fazendo compras nesse supermercado, haja 0, 1, 2, 3, 4, 5 ou as 6 aproveitando a promoção. Também esboce um histograma desta distribuição de probabilidade.

Solução Admitindo que a escolha seja aleatória, substituímos $n = 6$, $p = 0{,}30$ e, respectivamente, $x = 0, 1, 2, 3, 4, 5$ e 6 na fórmula da distribuição binomial, obtendo:

Snow666

$$f(0) = \binom{6}{0}(0{,}30)^0(0{,}70)^6 = 0{,}118$$

$$f(1) = \binom{6}{1}(0{,}30)^1(0{,}70)^5 = 0{,}303$$

$$f(2) = \binom{6}{2}(0{,}30)^2(0{,}70)^4 = 0{,}324$$

$$f(3) = \binom{6}{3}(0{,}30)^3(0{,}70)^3 = 0{,}185$$

$$f(4) = \binom{6}{4}(0{,}30)^4(0{,}70)^2 = 0{,}060$$

$$f(5) = \binom{6}{5}(0{,}30)^5(0{,}70)^1 = 0{,}010$$

$$f(6) = \binom{6}{6}(0{,}30)^6(0{,}70)^0 = 0{,}001$$

todos valores arredondados até a terceira casa decimal. A Figura 8.2 mostra o histograma dessa distribuição.

No caso de o leitor não gostar muito de sorvete e não se interessar particularmente nos hábitos de compra pessoais, vamos salientar a importância da distribuição binomial como um **modelo estatístico**. Os resultados do exemplo precedente aplicam-se também ao caso de ser de 0,30 a probabilidade de uma pilha de relógio durar dois anos sob condições normais e quisermos saber as probabilidades de que, dentre seis dessas pilhas, 0,1, 2, 3, 4, 5 ou as 6 durarem dois anos sob condições normais; se houver uma probabilidade de 0,30 de um ladrão ser preso e levado a julga-

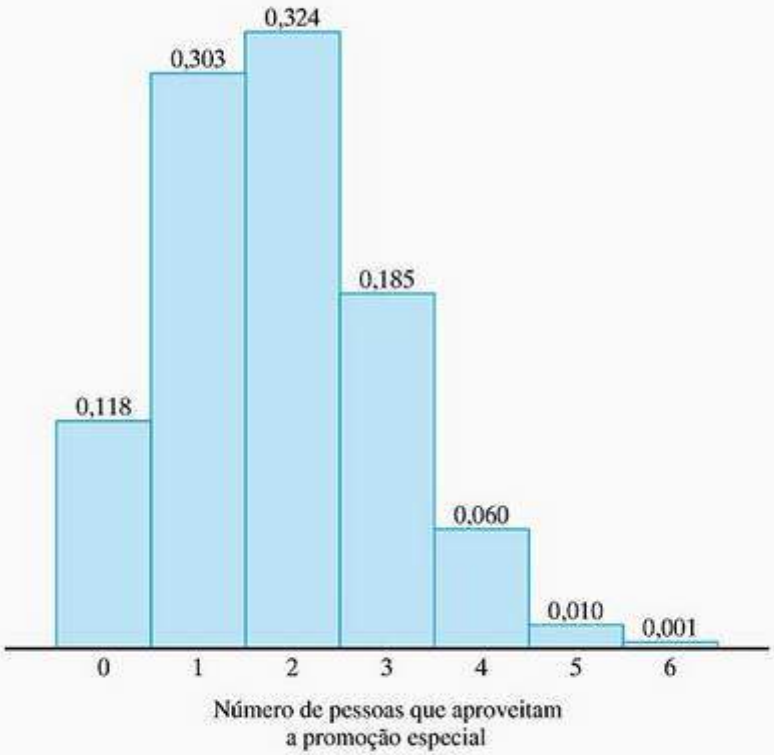

Figura 8.2
Histograma da distribuição binomial com $n = 6$ e $p = 0{,}30$.

Snow666

mento e quisermos saber as probabilidades de que, dentre seis ladrões, 0, 1, 2, 3, 4, 5 ou os 6 serem presos e levados a julgamento; se houver uma probabilidade de 0,30 de um chefe de família possuir pelo menos uma apólice de seguro de vida e quisermos saber as probabilidades de que, dentre seis chefes de família, 0, 1, 2, 3, 4, 5 ou os 6 possuírem pelo menos uma apólice de seguro de vida; ou se houver uma probabilidade de 0,30 de uma pessoa com uma certa doença viver por mais dez anos e quisermos saber as probabilidades de que, dentre seis pessoas com essa doença, 0, 1, 2, 3, 4, 5 ou as 6 viverem por mais dez anos. O argumento que apresentamos aqui é exatamente como o utilizado na Seção 1.2, em que tentamos incutir no leitor o senso da generalidade das técnicas estatísticas.

Na prática, raramente se calculam as probabilidades binomiais por substituição direta na fórmula. Às vezes utilizamos aproximações como as que serão estudadas mais adiante neste capítulo e no Capítulo 9, e às vezes utilizamos tabelas especiais como a Tabela V ao final deste livro. Atualmente, a fonte mais comum de probabilidades binomiais são os aplicativos de computador. No final deste capítulo, relacionamos algumas referências para tabelas binomiais detalhadas em formato de livro (embora estejam caindo em desuso).

A Tabela V está limitada às probabilidades binomiais para $n = 2$ até $n = 20$ e $p = 0{,}05$; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9 e 0,95, todas arredondadas até a terceira casa decimal. Os valores omitidos na Tabela V são todos menores do que 0,0005 e, portanto, arredondados para 0,000 até a terceira casa decimal.

EXEMPLO 8.6 A probabilidade de um eclipse lunar ser ocultado por nuvens num observatório perto de São Paulo é de 0,60. Use a Tabela V para encontrar as probabilidades de

(a) no máximo três dentre dez eclipses lunares serem ocultados por nuvens naquele local;

(b) no mínimo sete dentre dez eclipses lunares serem ocultados por nuvens naquele local.

Solução **(a)** Para $n = 10$ e $p = 0{,}60$, as entradas na Tabela V correspondentes a $x = 0$, 1, 2 e 3 são 0,000; 0,002; 0,011 e 0,042. Assim, a probabilidade de no máximo três dentre dez eclipses lunares serem ocultados por nuvens naquele local é de aproximadamente

$$0{,}000 + 0{,}002 + 0{,}011 + 0{,}042 = 0{,}055$$

O resultado é somente aproximado porque as entradas na Tabela V estão todas arredondadas até a terceira casa decimal.

(b) Para $n = 10$ e $p = 0{,}60$, as entradas na Tabela V correspondentes a $x = 7$, 8, 9 e 10 são 0,215; 0,121; 0,040 e 0,006. Assim, a probabilidade de no mínimo sete dentre dez eclipses lunares serem ocultados por nuvens naquele local é de aproximadamente

$$0{,}215 + 0{,}121 + 0{,}040 + 0{,}006 = 0{,}382$$

Foi possível utilizar a Tabela V nesse exemplo pois $p = 0{,}60$ é um dos poucos valores para os quais a Tabela V fornece probabilidades binomiais. Se a probabilidade tivesse sido, digamos, 0,51 ou 0,63 de um eclipse lunar ser ocultado por nuvens num determinado local, precisaríamos usar uma das tabelas mais detalhadas listadas à página 213, ou um computador, como no exemplo a seguir ou, como último recurso, a fórmula da distribuição binomial.

EXEMPLO 8.7 Com referência ao Exemplo 8.6, suponha que a probabilidade de um eclipse lunar ser ocultado por nuvens num certo observatório é de 0,63. Use a tabela fornecida por computador na Figura 8.6 para refazer o Exemplo 8.6 com $p = 0{,}63$ no lugar de 0,60.

Snow666

Solução **(a)** Para $n = 10$ e $p = 0{,}63$, as entradas na tabela de computador na Figura 8.3 correspondentes a $x = 0, 1, 2$ e 3 são 0,0000; 0,0008; 0,0063 e 0,0285. Assim, a probabilidade de no máximo três dentre dez eclipses lunares serem ocultados por nuvens naquele local é de aproximadamente

$$0{,}0000 + 0{,}0008 + 0{,}0063 + 0{,}0285 = 0{,}0356$$

Observe que essa resposta também aparece na parte inferior do material impresso na Figura 8.3. Ela é a probabilidade acumulada debaixo da legenda $P(X <= x)$ correspondente a $x = 3{,}00$.

(b) Para $n = 10$ e $p = 0{,}63$, as entradas na tabela de computador na Figura 8.3 correspondentes a $x = 7, 8, 9$ e 10 são 0,2394; 0,1529; 0,0578 e 0,0098. Assim, a probabilidade de no mínimo sete dentre dez eclipses lunares serem ocultados por nuvens naquele local é de

$$0{,}2394 + 0{,}1529 + 0{,}0578 + 0{,}0098 = 0{,}4599$$

Novamente, a resposta pode ser lida na parte inferior do material impresso. Ela é 1 menos a entrada correspondente a $x = 6$ na coluna $P(X <= x)$, a saber, $1 - 0{,}5400 = 0{,}4600$. É claro que a pequena diferença entre 0,4599 e 0,4600 é devida ao arredondamento.

Quando observamos um valor de uma variável aleatória com distribuição binomial como, por exemplo, quando observamos o número de caras em 25 lançamentos de uma moeda, ou o número de sementes (de um pacote com 24 sementes) que germinam, ou o número de estudantes (den-

Função Densidade de Probabilidade

```
Binomial with n = 10 and p = 0.630000

          x        P( X = x)
       0.00           0.0000
       1.00           0.0008
       2.00           0.0063
       3.00           0.0285
       4.00           0.0849
       5.00           0.1734
       6.00           0.2461
       7.00           0.2394
       8.00           0.1529
       9.00           0.0578
      10.00           0.0098
```

Função Distribuição Cumulativa

```
Binomial with n = 10 and p = 0.630000

          x       P( X <= x)
       0.00           0.0000
       1.00           0.0009
       2.00           0.0071
       3.00           0.0356
       4.00           0.1205
       5.00           0.2939
       6.00           0.5400
       7.00           0.7794
       8.00           0.9323
       9.00           0.9902
      10.00           1.0000
```

Figura 8.3 Impressão de computador para o Exemplo 8.7.

Snow666

tre 200 entrevistados) que são contra uma alteração nas taxas escolares, ou o número de acidentes automobilísticos (dentre 20 investigados) causados por embriagues do motorista, dizemos que estamos extraindo uma **amostra de uma população binomial**. Essa terminologia é largamente utilizada em Estatística.

EXERCÍCIOS

8.1 Em cada caso, determine se os valores dados podem ser valores de uma distribuição de probabilidade de alguma variável aleatória que pode tomar os valores 1, 2 e 3, explicando suas respostas:

(a) $f(1) = 0{,}52$, $f(2) = 0{,}26$ e $f(3) = 0{,}32$;
(b) $f(1) = 0{,}18$, $f(2) = 0{,}02$ e $f(3) = 1{,}00$;
(c) $f(1) = \frac{10}{33}$, $f(2) = \frac{1}{3}$ e $f(3) = \frac{12}{33}$.

8.2 Em cada caso, determine se os valores dados podem ser valores de uma distribuição de probabilidade de alguma variável aleatória que pode tomar os valores 1, 2, 3 e 4, explicando suas respostas:

(a) $f(1) = 0{,}20$, $f(2) = 0{,}80$, $f(3) = 0{,}20$ e $f(4) = -0{,}20$;
(b) $f(1) = 0{,}25$, $f(2) = 0{,}17$, $f(3) = 0{,}39$ e $f(4) = 0{,}19$;
(c) $f(1) = \frac{1}{17}$, $f(2) = \frac{7}{17}$, $f(3) = \frac{6}{17}$ e $f(4) = \frac{2}{17}$.

8.3 Em cada caso, determine se pode servir como a distribuição de probabilidade de alguma variável aleatória:

(a) $f(x) = \frac{1}{7}$, para $x = 1, 2, 3, 4, 5, 6, 7$;
(b) $g(y) = \frac{1}{9}$, para $y = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$;
(c) $f(x) = \frac{x+2}{18}$, para $x = 1, 2, 3, 4$.

8.4 Em cada caso, determine se pode servir como a distribuição de probabilidade de alguma variável aleatória:

(a) $f(x) = \frac{x-1}{10}$, para $x = 0, 1, 2, 3, 4, 5$;
(b) $h(z) = \frac{z^2}{30}$, para $z = 0, 1, 2, 3, 4$;
(c) $f(y) = \frac{y+4}{y-4}$, para $y = 1, 2, 3, 4, 5$.

8.5 Em determinada cidade, as despesas médicas são consideradas como responsáveis por 75% de todas as falências pessoais. Use a fórmula para a distribuição binomial para calcular a probabilidade de as despesas médicas serem apontadas como responsáveis por duas das próximas três falências pessoais naquela cidade.

8.6 Use a fórmula da distribuição binomial para calcular a probabilidade de quatro de seis tomateiros morrerem uma geada se a probabilidade de qualquer uma dessas plantas sobreviver a uma geada é de 0,30. Também confira a sua resposta na Tabela V.

8.7 Um médico sabe por experiência que 10% dos pacientes para os quais ele prescreve um certo medicamento contra pressão alta terá efeitos colaterais indesejáveis. Use a fórmula da distribuição binomial para calcular a probabilidade de nenhum de quatro pacientes para os quais ele prescreve o medicamento ter efeitos colaterais indesejáveis. Também confira a sua resposta na Tabela V.

8.8 Tem sido alegado que 80% de todos os acidentes industriais podem ser evitados dando estrita atenção às normas de segurança. Se for assim, encontre a probabilidade de que quatro dentre seis acidentes industriais possam ser assim evitados, usando

(a) a fórmula da distribuição binomial;
(b) a Tabela V.

Snow666

8.9 A experiência mostra que 30% dos lançamentos de foguete de uma base da NASA foram adiados em virtude do mau tempo. Use a Tabela V para determinar as probabilidades de que em dez lançamentos de foguete daquela base

(a) no máximo três sejam adiados em virtude do mau tempo;
(b) no mínimo seis sejam adiados em virtude do mau tempo.

8.10 Um estudo mostra que, de todas as melancias embaladas por uma cooperativa agrícola, 95% estão maduras e prontas para comer. Encontre as probabilidades de que entre 20 melancias embaladas pela cooperativa

(a) no máximo 16 estejam maduras e prontas para comer;
(b) todas elas estejam maduras e prontas para comer.

8.11 Um estudo mostra que em 60% dos casos de divórcio requeridos num certo município, a incompatibilidade é apontada como causa. Encontre as probabilidades de que entre 15 casos de divórcios requeridos naquele município

(a) no máximo cinco apontem a incompatibilidade como causa;
(b) de oito a onze apontem a incompatibilidade como causa;
(c) no mínimo onze apontem a incompatibilidade como causa.

8.12 Um distribuidor de comida congelada alega que 80% de seus pratos prontos contêm pelo menos 100 gramas de carne de galinha. Para verificar a alegação, um serviço de proteção ao consumidor decide examinar dez desses pratos prontos e rejeitar a alegação a menos que em pelo menos sete deles forem encontradas, pelo menos, 100 gramas de carne de galinha. Encontre as probabilidades de o serviço de proteção ao consumidor cometer o erro de

(a) rejeitar a alegação mesmo que ela seja verdadeira;
(b) não rejeitar a alegação quando na realidade apenas 70% dos pratos prontos contêm pelo menos 100 gramas de carne de frango.

8.13 Uma engenheira de controle de qualidade deseja verificar se, de acordo com as especificações, 90% dos produtos embarcados estão em perfeitas condições de funcionamento. Para tanto, ela seleciona 12 itens ao acaso de cada lote pronto para o embarque e aprova o lote somente se todos os 12 estão em perfeitas condições de funcionamento. Se um ou mais itens não estão em perfeitas condições de funcionamento, ela submete o lote inteiro a uma inspeção completa. Encontre as probabilidades de ela cometer o erro de

(a) reter um lote para inspeção completa, mesmo que 90% dos itens estejam em perfeitas condições de funcionamento;
(b) liberar um lote, mesmo que apenas 80% dos itens estejam em perfeitas condições de funcionamento;
(c) liberar um lote, mesmo que apenas 70% dos itens estejam em perfeitas condições de funcionamento.

8.14 Um estudo mostra que 70% de todos os pacientes que chegam numa certa clínica precisam esperar pelo menos 15 minutos para ver seu médico. Encontre as probabilidades de que, dentre dez pacientes que chegam nessa clínica, 0, 1, 2, 3,..., ou 10 precisam esperar pelo menos 15 minutos para ver seu médico, e esboce um histograma dessa distribuição de probabilidade.

8.15 Use a Figura 8.3 para determinar a probabilidade de que uma variável aleatória de distribuição binomial com $n = 10$ e $p = 0{,}63$ tome um valor menor do que cinco usando

(a) as probabilidades binomiais;
(b) as probabilidades acumuladas.

Snow666

8.16 Use a Figura 8.3 para determinar a probabilidade de que uma variável aleatória de distribuição binomial com $n = 10$ e $p = 0{,}63$ tome um valor maior do que oito usando

(a) as probabilidades binomiais;

(b) as probabilidades acumuladas.

8.17 Em algumas situações em que se aplicaria a distribuição binomial, podemos estar interessados na probabilidade de o primeiro sucesso ocorrer em determinada tentativa. Para que isso ocorra na x-ésima tentativa, deve ser precedida por $x - 1$ fracassos, cuja probabilidade é de $(1 - p)^{x-1}$, e decorre que a probabilidade de o primeiro sucesso ocorrer na x-ésima tentativa é

$$f(x) = p(1 - p)^{x-1} \quad \text{para} \quad x = 1, 2, 3, 4, \ldots$$

Essa distribuição é denominada **distribuição geométrica** (porque seus valores sucessivos constituem uma progressão geométrica) e deve ser observado que há uma infinidade enumerável de possibilidades.* Usando a fórmula, constatamos, por exemplo, que para jogadas repetidas de um dado equilibrado, a probabilidade de o primeiro 6 ocorrer na quinta jogada é

$$\frac{1}{6}\left(\frac{5}{6}\right)^{5-1} = \frac{625}{7.776} \approx 0{,}080$$

(a) Ao gravar um comercial de televisão, a probabilidade de um ator infantil acertar sua fala em uma tomada arbitrária é 0,40. Qual é a probabilidade de que esse ator infantil acerte sua fala pela primeira vez na quarta tomada?

(b) Suponhamos que haja uma probabilidade de 0,25 de que uma pessoa qualquer acredite num rumor sobre a vida particular de certo político. Qual é a probabilidade de que a quinta pessoa a ouvir o rumor seja a primeira a acreditar nele?

(c) A probabilidade de uma criança contrair uma doença contagiosa à qual está exposta, é de 0,70. Qual é a probabilidade de a terceira criança exposta à doença ser a primeira a contraí-la?

8.4 A DISTRIBUIÇÃO HIPERGEOMÉTRICA

No Exercício 6.62, introduzimos os termos "amostragem com reposição" e "amostragem sem reposição" em relação a cartas extraídas de um baralho. Para levar essa distinção para mais além, ressaltamos que a distribuição binomial se aplica quando extraímos amostras com reposição e as provas são todas independentes, mas que não se aplica quando a amostragem é sem reposição. Para introduzir uma distribuição de probabilidade aplicável ao caso de amostragem sem reposição, consideremos o exemplo seguinte: uma floricultura envia limoeiros de três anos em lotes de 24 e, quando eles chegam ao destino, um inspetor seleciona ao acaso três de cada lote. Se essas três árvores estão saudáveis, todo o lote é aceito; caso contrário, as outras 21 árvores do lote também são inspecionadas. Como um lote pode ser aceito sem inspeção adicional, mesmo que haja muitas árvores em más condições, esse procedimento de inspeção envolve um risco considerável. Para ilustrar a magnitude do risco, vamos supor que, na realidade, 6 das 24 árvores estejam em más condições e determinemos a probabilidade de que um lote inteiro seja, mesmo assim, aceito

* De acordo com a formulação no Capítulo 6, os postulados de Probabilidade aplicam-se somente quando o espaço amostral é finito. Quando o espaço amostral é infinito enumerável, como ocorre aqui, o terceiro postulado deve ser modificado de acordo. Isso será explicado à página 201.

Snow666

sem inspeção adicional. Isso significa que devemos encontrar a probabilidade de três sucessos (árvores saudáveis) em três provas (árvores inspecionadas) e poderíamos ser tentados a argumentar que, como 18 das 24 árvores no lote estão saudáveis, a probabilidade é de $\frac{18}{24} = \frac{3}{4}$ que alguma delas esteja saudável e, portanto, a probabilidade procurada é

$$f(3) = \binom{3}{3}\left(\frac{3}{4}\right)^3\left(1 - \frac{3}{4}\right)^{3-3} = 0{,}42$$

Esse resultado, obtido com a fórmula da distribuição binomial, seria correto se a amostragem fosse com reposição, mas não é isso que ocorre em problemas reais de inspeção por amostragem. Para obtermos a resposta correta de nosso problema quando a amostragem é sem reposição, devemos raciocinar como segue: há um total de $\binom{24}{3} = 2.024$ maneiras de escolher três das 24 árvores, e todas elas são equiprováveis em virtude da hipótese de que a seleção é aleatória. Entre estas, há $\binom{18}{3} = 816$ maneiras de selecionar três das 18 árvores saudáveis e decorre, portanto, que a probabilidade procurada é $\frac{816}{2.024} = 0{,}40$.

Para generalizar o método aqui utilizado, suponha que devamos escolher n objetos de um conjunto em que a objetos são de um tipo (sucessos) e b objetos são de outro tipo (fracassos), que a amostragem seja sem reposição, e que estejamos interessados na probabilidade de obter x sucessos e $n - x$ fracassos. Argumentando como anteriormente, vemos que é possível escolher n objetos do conjunto total com $a + b$ objetos de $\binom{a+b}{n}$ maneiras, e que x dos a sucessos e $n - x$ dos b fracassos podem ser escolhidos de $\binom{a}{x}\cdot\binom{b}{n-x}$ maneiras. Decorre que, na amostragem sem reposição, a probabilidade de "x sucessos em n provas" é

DISTRIBUIÇÃO HIPERGEOMÉTRICA

$$f(x) = \frac{\binom{a}{x}\cdot\binom{b}{n-x}}{\binom{a+b}{n}} \quad \textit{para } x = 0, 1, 2, \ldots, \textit{ou } n$$

onde x não pode exceder a e $n - x$ não pode exceder b. Essa é a fórmula da **distribuição hipergeométrica**.

Os dois próximos exemplos ilustram o uso da distribuição hipergeométrica em problemas em que tomamos uma amostragem sem reposição.

EXEMPLO 8.8 Um funcionário da expedição deveria remeter 6 de 15 pacotes por via expressa para a Europa, mas ele acaba misturando todos e aleatoriamente manda 6 dos pacotes por via expressa para a Europa. Qual é a probabilidade de que apenas três dos pacotes que deveriam ir por via expressa sigam realmente por via expressa?

Solução Substituindo $a = 6$, $b = 9$, $n = 6$ e $x = 3$ na fórmula da distribuição hipergeométrica, obtemos

$$f(3) = \frac{\binom{6}{3}\cdot\binom{9}{6-3}}{\binom{15}{6}} = \frac{20 \cdot 84}{5.005} \approx 0{,}336$$

Snow666

EXEMPLO 8.9 Das 16 ambulâncias de um serviço de ambulâncias, cinco emitem uma quantidade excessiva de poluentes. Se oito das ambulâncias são escolhidas ao acaso para inspeção, qual é a probabilidade de que essa amostra inclua pelo menos três das ambulâncias que emitem uma quantidade excessiva de poluentes?

Solução A probabilidade que devemos encontrar é $f(3) + f(4) + f(5)$, onde cada parcela dessa soma é um valor da distribuição hipergeométrica com $a = 5$, $b = 11$ e $n = 8$. Substituindo essas quantidades, juntamente com $x = 3$, 4 e 5 na fórmula da distribuição hipergeométrica, obtemos

$$f(3) = \frac{\binom{5}{3} \cdot \binom{11}{5}}{\binom{16}{8}} = \frac{10 \cdot 462}{12.870} = 0{,}359$$

$$f(4) = \frac{\binom{5}{4} \cdot \binom{11}{4}}{\binom{16}{8}} = \frac{5 \cdot 330}{12.870} = 0{,}128$$

$$f(5) = \frac{\binom{5}{5} \cdot \binom{11}{3}}{\binom{16}{8}} = \frac{1 \cdot 165}{12.870} \approx 0{,}013$$

e a probabilidade de que a amostra vá incluir pelo menos três das ambulâncias que emitem uma quantidade excessiva de poluentes é

$$0{,}359 + 0{,}128 + 0{,}013 = 0{,}500$$

Esse resultado sugere que a inspeção talvez devesse incluir mais do que oito das ambulâncias e deixamos a cargo do leitor mostrar no Exercício 8.18 que a probabilidade de pegar pelo menos três das ambulâncias que emitem uma quantidade excessiva de poluentes teria sido de 0,76 se a inspeção tivesse incluído dez das ambulâncias.

No início desta seção, demos um exemplo em que a distribuição binomial foi usada erroneamente em lugar da distribuição hipergeométrica. O erro, contudo, foi bastante pequeno (0,42 em vez de 0,40) e, na prática, a distribuição binomial é usada seguidamente para aproximar a distribuição hipergeométrica. De modo geral, essa aproximação é considerada satisfatória desde que n não exceda 5 por cento de $a + b$, isto é, se

$$n \leq (0{,}05)(a + b)$$

As vantagens principais da aproximação são que existem tabelas muito mais extensas da distribuição binomial do que da hipergeométrica e que, das duas fórmulas, a da distribuição binomial é mais fácil de usar, ou seja, os cálculos binomiais geralmente são menos complicados. Observe também que a distribuição binomial é descrita por dois parâmetros (n e p), enquanto que a distribuição hipergeométrica requer três (a, b e n).

EXEMPLO 8.10 Numa prisão federal, 120 dos 300 internos estão cumprindo pena por crimes relacionados com drogas. Se oito dos internos devem ser escolhidos ao acaso para comparecerem perante um comitê legislativo, qual é a probabilidade de que três dentre os oito escolhidos estejam cumprindo pena por crimes relacionados com drogas?

Snow666

Solução Como $n = 8$, $a + b = 300$ e 8 é menos do que $(0{,}05)(300) = 15$, podemos usar a aproximação binomial da distribuição hipergeométrica. Pela Tabela V, obtemos que para $n = 8$, $p = \frac{120}{300} = 0{,}40$ e $x = 3$, a probabilidade procurada é de 0,279. Com cálculos bastante extensos, poderíamos mostrar que o erro dessa aproximação é de apenas 0,003.

EXERCÍCIOS

8.18 No Exemplo 8.9, indicamos que, se dez das ambulâncias tivessem sido inspecionadas, então a probabilidade de pegar pelo menos três das que emitem uma quantidade excessiva de poluentes teria sido de 0,76. Verifique essa probabilidade.

8.19 Dentre os 14 mecânicos de uma revendedora, dez fizeram curso de treinamento na fábrica. Se três dos mecânicos são escolhidos ao acaso para um trabalho especial, encontre as probabilidades de

(a) todos três terem feito curso de treinamento na fábrica;
(b) somente dois deles terem feito curso de treinamento na fábrica.

8.20 Dentre 20 painéis solares apresentados numa exposição, 12 são do tipo de placa plana e os outros são do tipo concentrador. Se uma pessoa que visita a exposição escolher ao acaso seis painéis para verificar, qual é a probabilidade de três deles serem do tipo de placa plana?

8.21 Dentre 12 candidatos do sexo masculino a um emprego nos correios, nove têm esposas que trabalham. Se dois dos candidatos são escolhidos ao acaso para um exame mais detido, quais são as probabilidades de

(a) nenhum dos dois ter esposa que trabalha;
(b) apenas um ter esposa que trabalha;
(c) ambos terem esposas que trabalham?

8.22 Um inspetor alfandegário resolve inspecionar 3 de 16 carregamentos que chegam de Caracas por via aérea. Se a escolha é aleatória e se cinco dos carregamentos contêm contrabando, qual é a probabilidade de que o inspetor apanhe

(a) nenhum dos carregamentos com contrabando;
(b) somente um dos carregamentos com contrabando;
(c) dois dos carregamentos com contrabando;
(d) os três dos carregamentos com contrabando?

8.23 Em cada caso, verifique se é satisfeita a condição para aproximação da distribuição hipergeométrica pela distribuição binomial:

(a) $a = 140$, $b = 60$ e $n = 12$;
(b) $a = 220$, $b = 280$ e $n = 20$;
(c) $a = 250$, $b = 390$ e $n = 30$;
(d) $a = 220$, $b = 220$ e $n = 25$.

8.24 Um embarque de 250 bombas de piscina contém quatro bombas com pequenos defeitos. Se cinco delas forem escolhidas ao acaso para mandar para uma revenda de piscinas e componentes, use a aproximação binomial da distribuição hipergeométrica para encontrar a probabilidade de que elas incluirão uma bomba com pequeno defeito.

8.25 Com referência ao Exercício 8.24, encontre o erro dessa aproximação binomial da distribuição hipergeométrica.

8.26 Dentre os 200 empregados de uma firma, 120 são membros de um sindicato, ao passo que os outros não são. Se seis dos empregados são escolhidos por sorteio para servir no comi-

Snow666

tê administrativo do fundo de pensão, encontre a probabilidade de que três deles serem membros do sindicato, ao passo que os outros não são, usando

(a) a fórmula para a distribuição hipergeométrica;

(b) a distribuição binomial com $\frac{120}{200} = 0{,}60$ e $n = 6$ como uma aproximação.

8.5 A DISTRIBUIÇÃO DE POISSON

Quando n é grande e p é pequeno, as probabilidades binomiais costumam ser aproximadas por meio da fórmula

APROXIMAÇÃO DE POISSON PARA A DISTRIBUIÇÃO BINOMIAL

$$f(x) = \frac{(np)^x \cdot e^{-np}}{x!} \quad para \quad x = 0, 1, 2, 3, \ldots$$

que é uma forma especial da **distribuição de Poisson**, assim designada em homenagem ao matemático e físico francês S. D. Poisson (1781–1840). Nessa fórmula, o número irracional $e = 2{,}71828...$ é a base do sistema dos logaritmos naturais, e os valores necessários de e^{-np} podem ser obtidos da Tabela XII no final do livro. Observe também que, como no Exercício 8.17, nos deparamos aqui com uma variável aleatória que pode tomar uma infinidade enumerável de valores (a saber, tantos valores quantos são os números inteiros). Correspondentemente, o terceiro postulado de Probabilidade deve ser modificado de modo que, para qualquer seqüência de eventos mutuamente excludentes $A_1, A_2, A_3, ...$, a probabilidade de ocorrência de um deles é

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

É difícil dar condições precisas sob as quais podemos usar a aproximação de Poisson da distribuição binomial, ou seja, explicar precisamente o que queremos dizer com "quando n é grande e p é pequeno". Embora outros livros possam dar regras empíricas menos rigorosas, preferimos uma relativa segurança e usaremos a aproximação de Poisson para a distribuição binomial somente quando

$$n \geq 100 \quad \text{e} \quad np < 10$$

Para termos uma idéia sobre a precisão da aproximação binomial pela aproximação de Poisson, consideremos os impressos de computador da Figura 8.4, que mostram, uma ao lado da outra, as probabilidades binomiais com $n = 150$ e $p = 0{,}05$ e as probabilidades de Poisson com $np = 150(0{,}05) = 7{,}5$. [Comparando as probabilidades nas colunas encabeçadas por $P(X = x)$, vemos que a maior diferença, correspondente a $x = 8$, é $0{,}1410 - 0{,}1373 = 0{,}0037$.]

Na Figura 8.4, o programa MINITAB se refere às distribuições de probabilidade como funções de densidade probabilística (*probability density function*). Além disso, o parâmetro dessa forma especial da distribuição de Poisson é o produto np, que aqui é "mu" e denotado por μ, como será explicado na Seção 8.7.

Os dois exemplos a seguir ilustram a aproximação de Poisson da distribuição binomial.

EXEMPLO 8.11 Sabe-se por experiência que 2% dos livros encadernados em certa gráfica apresentam defeitos de encadernação. Use a aproximação de Poisson da distribuição binomial para encontrar a probabilidade de que cinco apresentem defeitos de encadernação, num lote de 400 livros encadernados nessa gráfica.

Snow666

Função Densidade de Probabilidade — Binomial with n = 150 and p = 0.0500000

x	P(X = x)
0.00	0.0005
1.00	0.0036
2.00	0.0141
3.00	0.0366
4.00	0.0708
5.00	0.1088
6.00	0.1384
7.00	0.1499
8.00	0.1410
9.00	0.1171
10.00	0.0869
11.00	0.0582
12.00	0.0355
13.00	0.0198
14.00	0.0102
15.00	0.0049
16.00	0.0022
17.00	0.0009
18.00	0.0003
19.00	0.0001
20.00	0.0000
21.00	0.0000

Probabilidade Densidade de Probabilidade — Poisson with mu = 7.50000

x	P(X = x)
0.00	0.0006
1.00	0.0041
2.00	0.0156
3.00	0.0389
4.00	0.0729
5.00	0.1094
6.00	0.1367
7.00	0.1465
8.00	0.1373
9.00	0.1144
10.00	0.0858
11.00	0.0585
12.00	0.0366
13.00	0.0211
14.00	0.0113
15.00	0.0057
16.00	0.0026
17.00	0.0012
18.00	0.0005
19.00	0.0002
20.00	0.0001
21.00	0.0000

Figura 8.4
A aproximação de Poisson da distribuição binomial.

Solução Como $n = 400 \geq 100$ e $np = 400(0{,}02) = 8 < 10$, as condições para a aproximação estão satisfeitas. Assim, substituindo $np = 8$, $e^{-8} = 0{,}00033546$ (da Tabela XII) e $x = 5$ na fórmula da distribuição de Poisson, obtemos

$$f(5) = \frac{8^5 \cdot e^{-8}}{5!} = \frac{(32.768)(0{,}00033546)}{120} \approx 0{,}0916$$

EXEMPLO 8.12 Os registros mostram que há uma probabilidade de 0,00006 de um pneu de um carro furar durante a travessia de um certo túnel. Utilize a aproximação de Poisson para a distribuição binomial para encontrar a probabilidade de que ao menos 2 dentre 10.000 carros tenham um pneu furado durante a travessia daquele túnel.

Solução Como $n = 10.000 \geq 100$ e $np = 10.000(0{,}00006) = 0{,}6 < 10$, as condições para a aproximação estão satisfeitas. Em vez de somar as probabilidades para $x = 2, 3, 4, \ldots$, subtrairemos de 1 as probabilidades para $x = 0$ e $x = 1$. Assim, substituindo $np = 0{,}6$, $e^{-0{,}6}$ (da Tabela XII) e, respectivamente, $x = 0$ e $x = 1$ na fórmula da distribuição de Poisson, obtemos

$$f(0) = \frac{(0{,}6)^0 \cdot e^{-0{,}6}}{0!} = \frac{1(0{,}5488)}{1} = 0{,}5488$$

$$f(1) = \frac{(0{,}6)^1 \cdot e^{-0{,}6}}{1!} \approx \frac{(0{,}6)(0{,}5488)}{1} = 0{,}3293$$

e, finalmente, $1 - (0{,}5488 + 0{,}3293) = 0{,}1219$.

Na prática, as probabilidades de Poisson (os valores da distribuição de Poisson) raramente são calculadas por substituição direta na fórmula da distribuição de Poisson ou usando uma tabela especial. Essas tarefas podem ser simplificadas enormemente usando um computador. Por exemplo, se tivéssemos utilizado os resultados de computador da Figura 8.5 no exemplo precedente, teríamos obtido diretamente $1 - 0{,}8781 = 0{,}1219$, onde 0,8781 é o valor correspondente a $x = 1{,}00$ na coluna de cabeçalho $P(X <= x)$.

Snow666

Figura 8.5
Impressão de computador para a distribuição de Poisson com $np = 0{,}60$

Função Densidade de Probabilidade

Poisson with mu = 0.600000

x	P(X = x)
0.00	0.5488
1.00	0.3293
2.00	0.0988
3.00	0.0198
4.00	0.0030
5.00	0.0004
6.00	0.0000

Função Distribuição Cumulativa

Poisson with mu = 0.600000

x	P(X <= x)
0.00	0.5488
1.00	0.8781
2.00	0.9769
3.00	0.9966
4.00	0.9996
5.00	1.0000
6.00	1.0000

Em alguns casos, as distribuições hipergeométricas podem ser aproximadas por uma distribuição binomial que pode, por sua vez, ser aproximada por uma distribuição de Poisson. Consideremos o exemplo seguinte.

EXEMPLO 8.13 Uma auditoria em 2002 examinou 4.000 faturas de vendas de uma loja de departamentos, encontrando 28 faturas com erros. Agora um perito-contador quer investigar a auditoria voltando a conferir uma amostra aleatória de 150 das 4.000 faturas. Em particular, ele gostaria de saber a probabilidade de encontrar duas faturas com erros nessas 150 faturas.

Solução Esse problema pede a utilização da probabilidade hipergeométrica, com $x = 2$, $a = 28$, $b = 4.000 - 28 = 3.972$ e $n = 150$, mas como $150 \leq 0{,}05(4.000) = 200$, podemos usar a aproximação binomial da distribuição hipergeométrica. Além disso, como esta é a distribuição binomial com $n = 150$ e $p = 28/4.000 = 0{,}007$, para os quais $n \geq 100$ e $np = 150\,(0{,}007) = 1{,}05 < 10$, podemos aproximá-la pela distribuição de Poisson com $np = 1{,}05$. Assim, podemos aproximar a probabilidade hipergeométrica original pela probabilidade de Poisson com $x = 2$ e $np = 1{,}05$. Logo

$$f(2) = \frac{1{,}05^2 \cdot e^{-1{,}05}}{2!} = \frac{1{,}1025(0{,}349938)}{2} = 0{,}1929$$

onde o valor de $e^{-1{,}05}$ foi obtido com uma calculadora estatística.

Como há muitas situações em que n ou $a + b$ são grandes, a distribuição de Poisson fornece aproximações muito úteis. Também observe que a distribuição de Poisson envolve apenas um parâmetro, np, enquanto a distribuição binomial envolve dois, n e p, e a distribuição hipergeométrica envolve três, a, b e n.

A distribuição de Poisson tem muitas aplicações importantes que não apresentam ligação direta com a distribuição binomial. Nesses casos, np é substituído pelo parâmetro λ (letra grega minúscula *lambda*) e a probabilidade de obter x sucessos é calculada pela fórmula

DISTRIBUIÇÃO DE POISSON

$$f(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} \quad \textit{para } x = 0, 1, 2, 3, \ldots$$

onde λ é interpretado como o número esperado, ou médio, de sucessos, como será explicado no último parágrafo da Seção 8.7.

Essa fórmula se aplica a muitas situações em que podemos esperar um número fixo de "sucessos" por unidade de tempo (ou por qualquer outro tipo de unidade), digamos, quando um banco espera receber seis cheques sem cobertura por dia, quando são esperados 1,6 acidentes por dia em um cruzamento perigoso, quando são esperadas oito azeitonas numa pizza média portuguesa, quando esperamos 5,6 imperfeições em uma peça de tecido, ou quando se podem esperar 0,03 reclamações por passageiro em uma companhia aérea, e assim por diante.

Snow666

EXEMPLO 8.14 Dado que um banco recebe em média $\lambda = 6$ cheques sem cobertura por dia, qual é a probabilidade de receber quatro cheques sem cobertura em um dia qualquer?

Solução Substituindo $x = 4$ e $\lambda = 6$ na fórmula precedente da distribuição de Poisson, obtemos

$$f(4) = \frac{6^4 \cdot e^{-6}}{4!} = \frac{1.296(0,002479)}{24} = 0,1339$$

onde o valor de e^{-6} foi obtido da Tabela XII.

EXEMPLO 8.15 Se podemos antecipar $\lambda = 5,6$ imperfeições por peça de um determinado material de cortina, qual é a probabilidade de uma peça conter $x = 3$ imperfeições?

Solução Substituindo $x = 3$ e $\lambda = 5,6$ na fórmula precedente da distribuição de Poisson, obtemos

$$f(3) = \frac{5,6^3 \cdot e^{-5,6}}{3!} = \frac{175,616(0,003698)}{6} \approx 0,1082$$

onde o valor de $e^{-5,6}$ foi obtido da Tabela XII.

EXERCÍCIOS

8.27 Verifique, em cada caso, se os valores de n e p satisfazem a regra empírica dada à página 201 para utilização da aproximação de Poisson da distribuição binomial:

(a) $n = 250$ e $p = \frac{1}{20}$

(b) $n = 400$ e $p = \frac{1}{50}$;

(c) $n = 90$ e $p = \frac{1}{10}$.

8.28 Verifique, em cada caso, se os valores de n e p satisfazem a regra empírica dada à página 201 para utilização da aproximação de Poisson da distribuição binomial:

(a) $n = 3000$ e $p = 0,01$;

(b) $n = 600$ e $p = 0,02$;

(c) $n = 75$ e $p = 0,1$.

8.29 Um administrador de hospital, baseado em sua experiência, sabe que 5% de todos pacientes admitidos devem ser colocados em tratamento intensivo imediatamente. Use esse dado para estimar a probabilidade de que entre 120 pacientes novos admitidos, três tenham que ser postos em tratamento intensivo imediatamente.

8.30 Se 3% das pessoas que visitam um apartamento decorado de um prédio em construção estão seriamente interessadas em comprar um apartamento do prédio, qual é a probabilidade de Poisson de que dentre 180 pessoas que visitam um apartamento decorado de um prédio em construção, cinco estejam seriamente interessadas em comprar um apartamento do prédio?

8.31 Suponha que 5% de todos os motoristas habilitados de uma certa cidade se envolvem em pelo menos um acidente de carro num dado ano. Use a fórmula da distribuição de Poisson para aproximar a probabilidade de que, dentre 150 motoristas habilitados daquela cidade, no máximo dois se envolvam em, pelo menos, um acidente num ano dado?

8.32 Use a Figura 8.4 para

(a) conferir o resultado obtido no Exercício 8.31;

(b) encontrar o erro dessa aproximação de Poisson;

(c) encontrar o erro percentual dessa aproximação de Poisson.

Snow666

8.33 A distribuição hipergeométrica com $n = 120$, $a = 50$ e $b = 3.150$ pode ser aproximada por uma distribuição de Poisson?

8.34 A quantidade de reclamações encaminhadas por dia a uma lavanderia a seco é uma variável aleatória de distribuição de Poisson com $\lambda = 3{,}2$. Encontre as probabilidades de que, num dado dia, sejam encaminhadas à lavanderia

(a) somente duas reclamações;
(b) no máximo duas reclamações.

8.35 O número de quebras mensais do tipo de computador utilizado num escritório é uma variável aleatória de distribuição de Poisson com $\lambda = 1{,}6$. Encontre as probabilidades de que esse tipo de computador vá funcionar durante um mês

(a) sem quebrar;
(b) com apenas uma quebra;
(c) com duas quebras.

8.36 Com referência ao Exercício 8.35, suponha que o escritório tenha quatro daqueles computadores. Qual é a probabilidade de que todos quatro computadores funcionarão durante um mês sem quebrar? Qual hipótese deve ser imposta para poder determinar essa probabilidade?

*8.6 A DISTRIBUIÇÃO MULTINOMIAL

Uma generalização importante da distribuição binomial ocorre quando há mais de dois resultados possíveis em cada prova, as probabilidades dos vários resultados permanecem as mesmas para cada prova e as provas são todas independentes. É o caso, por exemplo, quando jogamos um dado repetidamente, em que cada prova tem seis resultados possíveis; quando estudantes são perguntados se gostam de um novo disco, não gostam, ou são indiferentes; ou quando um inspetor do Ministério de Agricultura classifica a carne como de primeira, especial, ou de segunda.

Se há k resultados possíveis para cada prova e suas probabilidades são $p_1, p_2, \ldots$, e p_k, pode ser mostrado que a probabilidade de x_1 resultados do primeiro tipo, x_2 resultados do segundo tipo,..., e x_k resultados do k-ésimo tipo, em n tentativas, é dado por

DISTRIBUIÇÃO MULTINOMIAL

$$\frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}\cdot p_2^{x_2}\cdots p_k^{x_k}$$

Esta distribuição é denominada **distribuição multinomial**.

EXEMPLO 8.16 A rede de TV aberta de uma grande cidade tem 30% da audiência nas noites de sexta-feira, um canal local tem 20%, a TV a cabo tem 40% e 10% assistem a videocassetes. Qual é a probabilidade de que, entre sete espectadores de televisão selecionados aleatoriamente naquela cidade numa noite de sexta-feira, dois estejam assistindo à TV aberta, um esteja assistindo ao canal local, dois estejam vendo TV a cabo e um esteja assistindo a um videocassete?

Solução Substituindo $n = 7$, $x_1 = 3$, $x_2 = 1$, $x_3 = 2$, $x_4 = 1$, $p_1 = 0{,}30$, $p_2 = 0{,}20$, $p_3 = 0{,}40$ e $p_4 = 0{,}10$ na fórmula da distribuição multinomial, obtemos

$$\frac{7!}{3!\cdot 1!\cdot 2!\cdot 1!}\cdot(0{,}30)^3(0{,}20)^1(0{,}40)^2(0{,}10)^1 = 0{,}036$$

arredondado até a terceira casa.

Snow666

EXERCÍCIOS

***8.37** Um carro sendo testado num posto de inspeção estadual tem uma probabilidade de 0,70 de ser aprovado na primeira tentativa, uma probabilidade de 0,20 de ser aprovado na segunda tentativa, e uma probabilidade de 0,10 de ser aprovado na terceira tentativa. Qual é a probabilidade de que dentre dez carros sendo testados, seis sejam aprovados na primeira tentativa, três sejam aprovados na segunda e um seja aprovado na terceira?

***8.38** De acordo com a teoria mendeliana da hereditariedade, se cruzamos plantas com sementes amarelas redondas com plantas com sementes verdes estriadas, então as probabilidades de obtermos uma planta que produza sementes amarelas redondas, sementes amarelas estriadas, sementes verdes redondas ou sementes verdes estriadas, são, respectivamente, de $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$ e $\frac{1}{16}$. Qual é a probabilidade de que, dentre nove plantas assim obtidas, haja quatro que produzam sementes amarelas redondas, duas que produzam sementes amarelas estriadas, três que produzam sementes verdes redondas e nenhuma que produza sementes verdes estriadas?

***8.39** São de 0,60; 0,20; 0,10 e 0,10 as probabilidades de que o formulário para declaração de rendimento seja preenchido corretamente; que contenha somente erros que beneficiam o contribuinte; que contenha somente erros que beneficiam o fisco; e que contenha ambos tipos de erros. Qual é a probabilidade de que dentre dez tais formulários, selecionados ao acaso para uma auditoria, sete estejam preenchidos corretamente, um contenha somente erros que beneficiam o contribuinte, um que contenha somente erros que beneficiam o fisco e um que contenha ambos tipos de erros?

8.7 A MÉDIA DE UMA DISTRIBUIÇÃO DE PROBABILIDADE

Quando, no Exemplo 7.6, mostramos que a agência de uma companhia aérea num certo aeroporto pode esperar receber 2,75 reclamações sobre extravio de bagagem, chegamos àquele resultado usando a fórmula de uma esperança matemática, a saber, somando os produtos obtidos pela multiplicação de 0, 1, 2, 3, . . . pelas probabilidades correspondentes de a agência receber 0, 1, 2, 3, . . . reclamações sobre extravios de bagagem num dia qualquer dado. Aqui, o número de reclamações é uma variável aleatória, e 2,75 é seu **valor esperado**.

Se aplicarmos o mesmo argumento ao primeiro exemplo dado na Seção 8.2, constatamos que, em conjunto, os dois vendedores podem esperar vender

$$0(0{,}44) + 1(0{,}20) + 2(0{,}18) + 3(0{,}18) = 1{,}10$$

dos caminhões. Nesse caso, a quantidade de caminhões vendidos é uma variável aleatória e 1,10 é seu valor esperado.

Conforme explicamos no Capítulo 7, as esperanças matemáticas devem ser interpretadas como médias, e costumamos referir-nos ao valor esperado de uma variável aleatória como sua **média**, ou como a **média de sua distribuição de probabilidade**. Em geral, se uma variável aleatória assume os valores $x_1, x_2, x_3, \ldots$, ou x_k com as probabilidades $f(x_1), f(x_2), f(x_3), \ldots$, e $f(x_k)$, seu valor esperado é

$$x_1 \cdot f(x_1) + x_2 \cdot f(x_2) + x_3 \cdot f(x_3) + \cdots + x_k \cdot f(x_k)$$

e, na notação sigma, escrevemos

MÉDIA DE UMA DISTRIBUIÇÃO DE PROBABILIDADE

$$\mu = \sum x \cdot f(x)$$

Snow666

Tal como a média de uma população, a média de uma distribuição de probabilidade é denotada pela letra grega minúscula μ (mu). A notação é a mesma, pois, conforme salientamos em relação à distribuição binomial, quando observamos um valor de uma variável aleatória, referimo-nos à sua distribuição como a população da qual estamos extraindo uma amostra. Por exemplo, o histograma da Figura 8.2 à página 192 pode ser encarado como a população da qual estamos extraindo uma amostra quando observamos um valor de uma variável aleatória tendo a distribuição binomial com $n = 6$ e $p = 0{,}30$.

EXEMPLO 8.17 Encontre a média da segunda distribuição de probabilidade da Seção 8.2, a que se refere ao número de pontos obtidos na jogada de um dado equilibrado.

Solução Como as probabilidades de obter 1, 2, 3, 4, 5 ou 6 são todas iguais a $\frac{1}{6}$, obtemos

$$\mu = 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6} = 3\tfrac{1}{2}$$

EXEMPLO 8.18 Com referência ao Exemplo 8.5, encontre o número médio de pessoas, dentre seis que estejam fazendo compras no supermercado, que aproveitarão a promoção especial.

Solução Substituindo $x = 0, 1, 2, 3, 4, 5$ e 6 e as probabilidades da página 192 na fórmula de μ, obtemos

$$\begin{aligned}\mu &= 0(0{,}118) + 1(0{,}303) + 2(0{,}324) + 3(0{,}185\\ &\quad + 4(0{,}060) + 5(0{,}010) + 6(0{,}001)\\ &= 1{,}802\end{aligned}$$

Quando uma variável aleatória pode assumir muitos valores distintos, o cálculo de μ pode tornar-se bastante laborioso. Por exemplo, se quisermos saber quantas pessoas podemos esperar que contribuam para um fundo beneficente, quando se solicita contribuição a 2.000 pessoas e há uma probabilidade de 0,40 de que qualquer uma delas vá dar uma contribuição, poderíamos pensar em calcular as 2.001 probabilidades correspondentes a 0, 1, 2, 3, . . . , 1.999 ou 2.000 dos que dão uma contribuição e então substituir na fórmula de μ. Não pensaríamos muito a respeito dessa opção e, em vez disso, poderíamos argumentar que, no final das contas, 40% das pessoas darão uma contribuição, 40% de 2.000 são 800 e, portanto, podemos esperar que 800 das 2.000 pessoas dêem uma contribuição. Analogamente, se uma moeda equilibrada é jogada 1.000 vezes, poderíamos argumentar que, no final das contas, aparecerá cara em 50% das vezes e, portanto, podemos esperar 1.000(0,50) = 500 caras. Esses dois resultados estão corretos; ambos problemas dizem respeito a variáveis aleatórias com distribuições binomiais e pode ser mostrado que, em geral,

MÉDIA DE UMA DISTRIBUIÇÃO BINOMIAL

$$\boxed{\mu = n \cdot p}$$

Em palavras, a média de uma distribuição binomial é simplesmente o produto do número de provas pela probabilidade de sucesso numa prova individual.

EXEMPLO 8.19 Com referência ao Exemplo 8.5, use a fórmula da média de uma distribuição binomial para encontrar o número médio de pessoas, dentre seis que estão fazendo compras no supermercado, que se pode esperar que aproveitem a promoção especial.

Snow666

Solução Como estamos tratando com a distribuição binomial com $n = 6$ e $p = 0{,}30$, obtemos $\mu = 6(0{,}30) = 1{,}80$. A pequena diferença entre os valores obtidos aqui e no Exemplo 8.18 é devida ao arredondamento das probabilidades usadas no Exemplo 8.18 até a terceira casa decimal.

É importante lembrar que a fórmula $\mu = n \cdot p$ se aplica somente a distribuições binomiais. Para outras distribuições, há outras fórmulas; por exemplo, para a distribuição hipergeométrica, a fórmula da média é

MÉDIA DE UMA DISTRIBUIÇÃO HIPERGEOMÉTRICA

$$\mu = \frac{n \cdot a}{a + b}$$

EXEMPLO 8.20 Dentre doze ônibus escolares, cinco têm freios gastos. Se seis desses doze ônibus escolares forem escolhidos ao acaso para uma inspeção, quantos podemos esperar que tenham freios gastos?

Solução Como estamos fazendo uma amostragem sem reposição, aqui temos uma distribuição hipergeométrica com $a = 5$, $b = 7$ e $n = 6$. Substituindo esses valores na fórmula especial da média de uma distribuição hipergeométrica, obtemos

$$\mu = \frac{6 \cdot 5}{5 + 7} = 2{,}5$$

Isso não deveria constituir uma surpresa: selecionamos metade dos ônibus escolares para a inspeção e, portanto, podemos esperar que metade dos que têm freios gastos tenha sido incluída na amostra.

A média da distribuição de Poisson com parâmetro λ também é $\lambda = \mu$ e isso está de acordo com o que sugerimos anteriormente, a saber, que λ deve ser interpretado como uma média. A dedução de todas essas fórmulas especiais pode ser encontrada em livros de Estatística Matemática.

8.8 O DESVIO-PADRÃO DE UMA DISTRIBUIÇÃO DE PROBABILIDADE

No Capítulo 4, vimos que as medidas de tendência mais usadas são a variância e sua raiz quadrada, o desvio-padrão, que medem a variabilidade através da média dos quadrados dos desvios em relação à média. Para as distribuições de probabilidade, medimos a variabilidade quase da mesma maneira, mas em lugar de tomarmos a média dos quadrados dos desvios da média, determinamos seus valores esperados. Em geral, se uma variável aleatória assume os valores $x_1, x_2, x_3, \ldots$, ou x_k, com as probabilidades $f(x_1), f(x_2), f(x_3), \ldots$, e $f(x_k)$, e se a média dessa distribuição de probabilidade é μ, então os desvios da média são $x_1 - \mu, x_2 - \mu, x_3 - \mu, \ldots$, e $x_k - \mu$, e o valor esperado de seus quadrados é

$$(x_1 - \mu)^2 \cdot f(x_1) + (x_2 - \mu)^2 \cdot f(x_2) + \cdots + (x_k - \mu)^2 \cdot f(x_k)$$

Assim, na notação $\sum$, escrevemos

VARIÂNCIA DE UMA DISTRIBUIÇÃO DE PROBABILIDADE

$$\sigma^2 = \sum (x - \mu)^2 \cdot f(x)$$

Snow666

que designamos como a **variância da variável aleatória** ou a **variância de sua distribuição de probabilidade**. Como na seção precedente, e pela mesma razão, denotamos essa descrição de uma distribuição de probabilidade com o mesmo símbolo do que o da descrição correspondente de uma população. Da mesma forma que a raiz quadrada da variância de uma amostra, a raiz quadrada da variância de uma população é uma importante medida de tendência, que designamos como o **desvio-padrão populacional**.

EXEMPLO 8.21 Com referência ao Exemplo 8.5, encontre o desvio-padrão do número de pessoas, dentre seis que estão fazendo compras no supermercado, que aproveitarão a promoção especial.

Solução No Exemplo 8.19, mostramos que $\mu = 1{,}80$ para essa variável aleatória, de modo que podemos dispor os cálculos como segue:

Número de pessoas	*Probabilidade*	*Desvio da média*	*Quadrado do desvio da média*	$(x-\mu)^2 f(x)$
0	0,118	−1,8	3,24	0,38232
1	0,303	−0,8	0,64	0,19392
2	0.324	0,2	0,04	0,01296
3	0,185	1,2	1,44	0,26640
4	0,060	2,2	4,84	0,29040
5	0,010	3,2	10,24	0,10240
6	0,001	4,2	17,64	0,01764
				$\sigma^2 = 1{,}26604$

Os valores na coluna da direita foram obtidos multiplicando os quadrados dos desvios da média por suas probabilidades, e o total dessa coluna é a variância da distribuição. Assim, o desvio-padrão é

$$\sigma = \sqrt{1{,}26604} = 1{,}13$$

arredondado até a segunda casa decimal.

Neste exemplo, os cálculos foram fáceis, embora pudéssemos tê-los simplificado usando a fórmula reduzida

$$\sigma^2 = \sum x^2 \cdot f(x) - \mu^2$$

Teria sido ainda mais fácil usar a fórmula da variância de uma distribuição binomial, a saber

VARIÂNCIA DA DISTRIBUIÇÃO BINOMIAL

$$\sigma^2 = np(1-p)$$

No Exemplo 8.21, em que tivemos $n = 6$ e $p = 0{,}30$, isso forneceria $\sigma^2 = 6(0{,}30)(0{,}70) = 1{,}26$ e $\sigma = 1{,}12$. A diferença entre os resultados obtidos aqui e anteriormente é devida a arredondamento. Se tivéssemos extraído a raiz quadrada com uma casa decimal a mais, a diferença entre os resultados teria sido de apenas 0,003.

O desvio-padrão de uma distribuição de probabilidade, falando intuitivamente, mede o tamanho esperado das flutuações ao acaso de uma variável aleatória correspondente. Quando σ é pequeno, existe uma alta probabilidade de que resulte um valor próximo da média; quando σ é grande, é mais provável que resulte bem afastado da média. Essa idéia importante é expressa formal-

Snow666

mente pelo teorema de Tchebichev, que foi introduzido na Seção 4.3, em relação a dados numéricos. Para distribuições de probabilidade, o teorema de Tchebichev pode ser enunciado como segue:

TEOREMA DE TCHEBICHEV

A probabilidade de uma variável aleatória tomar um valor a menos de k desvios-padrão de qualquer um dos dois lados da média é pelo menos $1 - \frac{1}{k^2}$.

Assim, a probabilidade de obter um valor a menos de dois desvios-padrão da média (um valor entre $\mu - 2\sigma$ e $\mu + 2\sigma$) é de pelo menos $1 - \frac{1}{2^2} = \frac{3}{4}$, a probabilidade de obter um valor a menos de cinco desvios-padrão da média (um valor entre $\mu - 5\sigma$ e $\mu + 5\sigma$) é de pelo menos $1 - \frac{1}{5^2} = \frac{24}{25}$, e assim por diante. Observe que, na formulação desse teorema, a frase "a menos de k desvios-padrão da média" não inclui os valores extremos $\mu - k\sigma$ e $\mu + k\sigma$.

EXEMPLO 8.22 A quantidade de chamadas telefônicas recebidas por um serviço de atendimento entre as 9 e as 10 horas da manhã é uma variável aleatória cuja distribuição tem a média $\mu = 27{,}5$ e o desvio padrão $\mu = 3{,}2$. O que o teorema de Tchebichev, com $k = 3$, nos diz sobre o número de chamadas telefônicas que o serviço de atendimento pode esperar receber entre as 9 e as 10 horas da manhã?

Solução Como $\mu - 3\sigma = 27{,}5 - 3(3{,}2) = 17{,}9$ e $\mu + 3\sigma = 27{,}5 + 3(3{,}2) = 37{,}1$, podemos garantir com uma probabilidade de pelo menos $1 - \frac{1}{3^2} = \frac{8}{9}$, ou de aproximadamente 0,89, que o serviço de atendimento vá receber entre 17,9 e 37,1 chamadas telefônicas, ou seja, qualquer quantidade entre 18 e 37 chamadas telefônicas.

EXEMPLO 8.23 O que o teorema de Tchebichev, com $k = 5$, pode nos dizer sobre o número de caras e, portanto, sobre a proporção de caras que podemos obter em 400 jogadas de uma moeda equilibrada?

Solução Aqui estamos tratando de uma variável aleatória que tem uma distribuição binomial com $n = 400$ e $p = 0{,}50$, de modo que $\mu = 400(0{,}50) = 200$ e $\sigma = \sqrt{400(0{,}50)(0{,}50)} = 10$. Como $\mu - 5\sigma = 200 - 5 \cdot 10 = 150$ e $\mu + 5\sigma = 200 + 5 \cdot 10 = 250$, podemos garantir com uma probabilidade de pelo menos $1 - \frac{1}{5^2} = \frac{24}{25} = 0{,}96$ que vamos obter entre 150 e 250 caras ou que a proporção de caras vai ficar entre $\frac{150}{400} = 0{,}375$ e $\frac{250}{400} = 0{,}625$.

Para continuar com este exemplo, no Exercício 8.53 pede-se ao leitor mostrar que, para $n =$ 10.000 jogadas de uma moeda equilibrada, há uma probabilidade de pelo menos 0,96 de a proporção de caras estar entre 0,475 e 0,525 e que, para 1.000.000 de jogadas de uma moeda equilibrada, há uma probabilidade de, pelo menos, 0,96 de a proporção de caras estar entre 0,4975 e 0,5025. Tudo isso fornece justificativa para a lei dos grandes números, introduzida no Capítulo 5 em relação à interpretação freqüencial da probabilidade.

Snow666

EXERCÍCIOS

8.40 Suponhamos que as probabilidades de que 1, 2, 3 ou 4 novos medicamentos para reduzir a taxa de colesterol sejam aprovados pelo Ministério da Saúde no próximo ano sejam de 0,40; 0,30; 0,20 e 0,10, respectivamente.

(a) Use a fórmula que define a média de uma distribuição de probabilidade para encontrar a média dessa distribuição.

(b) Use a fórmula que define a variância de uma distribuição de probabilidade para encontrar a variância dessa distribuição.

8.41 Em geral, vale a pena simplificar os cálculos da variância populacional ou do desvio-padrão populacional usando uma fórmula para calcular σ^2. Como a utilizada para a variância amostral, tem a vantagem de não precisar trabalhar com os desvios da média. Use essa fórmula de calcular para refazer a parte (b) do Exercício 8.40.

8.42 Sob condições bastante normais, as probabilidades de que, num dado ano, 0, 1, 2, 3 ou 4 furacões atinjam a ilha caribenha de Martinica são de 0,20; 0,50; 0,10; 0,10 e 0,10, respectivamente. Encontre a média e a variância dessa distribuição de probabilidade.

8.43 Um estudo mostra que 27% de todos pacientes que chegam a uma certa clínica precisam esperar, no mínimo, meia hora para serem atendidos pelo seu médico. Use as probabilidades da Figura 8.6 para calcular μ e σ para o número de pacientes, dentre 18 que chegam a essa clínica, que devem esperar pelo menos meia hora para serem atendidos pelo seu médico.

8.44 Na Seção 8.2, mostramos que as probabilidades de obter 0, 1, 2, 3 ou 4 caras em quatro jogadas de uma moeda equilibrada são $\frac{1}{16}$, $\frac{4}{16}$, $\frac{6}{16}$, $\frac{4}{16}$, e $\frac{1}{16}$ respectivamente. Use a fórmula que define μ e σ^2 para encontrar a média e o desvio-padrão dessa distribuição de probabilidade.

8.45 Refaça o Exercício 8.44 usando as fórmulas especiais da média e do desvio-padrão de uma distribuição binomial.

8.46 Se 80% de certos aparelhos de vídeo funcionarão bem durante os 90 dias de garantia de fábrica, encontre a média e o desvio-padrão do número desses aparelhos de vídeo que funcionarão bem durante os 90 dias de garantia de fábrica usando

(a) a Tabela V, a fórmula que define μ e a fórmula para calcular σ^2;

(b) as fórmulas especiais da média e do desvio-padrão de uma distribuição binomial.

Função Densidade de Probabilidade

Binomial with n = 18 and p = 0.270000

x	P(X = x)
0.00	0.0035
1.00	0.0231
2.00	0.0725
3.00	0.1431
4.00	0.1985
5.00	0.2055
6.00	0.1647
7.00	0.1044
8.00	0.0531
9.00	0.0218
10.00	0.0073
11.00	0.0020
12.00	0.0004
13.00	0.0001
14.00	0.0000

Figura 8.6 Impressão de computador para o Exercício 8.43.

Snow666

8.47 Encontre a média e o desvio-padrão de cada uma das seguintes variáveis aleatórias binomiais:

(a) o número de caras obtidas em 484 jogadas de uma moeda equilibrada;
(b) o número de 3 obtidos em 720 lançamentos de um dado equilibrado;
(c) o número de pessoas que, dentre as 600 convidadas, pode-se esperar que compareçam à inauguração de uma nova agência de um banco, quando a probabilidade de que um convidado qualquer vá comparecer é de 0,30.
(d) o número de peças defeituosas numa amostra de 600 partes fabricadas automaticamente, quando a probabilidade de uma peça qualquer estar defeituosa é de 0,04;
(e) o número de alunos, dentre 800 entrevistados, que não gostam da refeição servida no refeitório da universidade, quando a probabilidade de um deles não gostar da comida é de 0,65.

8.48 Um estudo mostra que 60% toda correspondência de primeira classe entre duas cidades do Rio Grande do Sul é entregue dentro de 48 horas. Encontre a média e a variância do número de tais cartas, dentre oito, que são entregues dentro de 48 horas usando

(a) a Tabela V, a fórmula que define μ e a fórmula para calcular σ^2;
(b) as fórmulas especiais da média e do desvio-padrão de uma distribuição binomial.

8.49 No Exemplo 8.9, que trata da variável aleatória com uma distribuição hipergeométrica com $a = 5$, $b = 11$ e $n = 8$, mostramos que as probabilidades de que vá assumir os valores 3, 4 e 5 são de 0,359, 0,128 e 0,013, respectivamente. Como pode ser verificado facilmente, as probabilidades de 0, 1 ou 2 sucessos são de 0,013, 0,128 e 0,359, respectivamente. Use todas essas probabilidades para calcular a média dessa distribuição hipergeométrica. Também use a fórmula especial da média de uma distribuição hipergeométrica para conferir o resultado obtido.

8.50 Numa cidade do nordeste, o número de dias em que a temperatura passa dos 42 graus é uma variável aleatória, com $\mu = 138$ e $\sigma = 9$. O que o teorema de Tchebichev, com $k = 4$, nos diz sobre o número de dias em que a temperatura passa dos 42 graus naquela cidade num dado ano?

8.51 Se o número de raios gama emitidos por segundo por uma certa substância radiativa é uma variável aleatória com distribuição de Poisson com $\lambda = 2{,}5$, as probabilidades de que a substância vá emitir 0, 1, 2, 3, 4, 5, 6, 7, 8 ou 9 raios gama num dado segundo são, respectivamente, de 0,082; 0,205; 0,256; 0,214; 0,134; 0,067; 0,028; 0,010; 0,003 e 0,001. Calcule a média e use o resultado para verificar a fórmula especial $\mu = \lambda$ para uma variável aleatória com distribuição de Poisson com o parâmetro λ.

8.52 Dentre oito professores de uma faculdade que estão sendo considerados para uma promoção, quatro têm titulo de Doutor e quatro não têm.

(a) Se quatro dos oito professores são escolhidos ao acaso, encontre as probabilidades de que 0, 1, 2, 3 ou todos os 4 terem o título de Doutor.
(b) Use os resultados da parte (a) para determinar a média dessa distribuição de probabilidade.
(c) Use a fórmula especial da média de uma distribuição hipergeométrica para conferir o resultado da parte (b).

8.53 Use o teorema de Tchebichev para mostrar que há uma probabilidade de 0,96, no mínimo, de que

(a) em 10.000 lançamentos de uma moeda equilibrada, a proporção de caras fique entre 0,475 e 0,525;
(b) em 1.000.000 de lançamentos de uma moeda equilibrada, a proporção de caras fique entre 0,4975 e 0,5025.

Snow666

8.9 LISTA DE TERMOS-CHAVE (com indicação das páginas de suas definições)

Amostra de uma população binomial, 195
Desvio-padrão de distribuição de probabilidade, 208
Desvio-padrão populacional, 209
Distribuição binomial, 191
Distribuição de Poisson, 201
Distribuição de probabilidade, 187
Distribuição geométrica, 197
Distribuição hipergeométrica, 198
Distribuição multinomial, 205
Média de uma distribuição de probabilidade, 206
Modelo estatístico, 192
Valor esperado de uma variável aleatória, 206
Variância de uma distribuição de probabilidade, 209
Variância de uma variável aleatória, 209
Variável aleatória, 186
Variável aleatória discreta, 187

8.10 REFERÊNCIAS

Uma grande quantidade de informação sobre as várias distribuições de probabilidade podem ser encontradas em

HASTINGS, N. A. J., and PEACOCK, J. B., *Statistical Distributions*. London: Butterworth & Company (Publishers) Ltd., 1975.

Tabelas mais detalhadas de probabilidades binomiais podem ser encontradas em

ROMIG, H. G., *50-100 Binomial Tables*. New York: John Wiley & Sons, Inc., 1953.
Tables of the Binomial Probability Distribution, National Bureau of Standards Applied Mathematics Series No. 6. Washington, D.C.: U.S. Government Printing Office, 1950.

e uma tabela detalhada de probabilidades de Poisson é dada em

MOLINA, E. C., *Poisson's Exponential Binomial Limit*. Princeton, N.J.: D. Van Nostrand Company, Inc., 1947.

Com a ampla disponibilidade de programas de computador para probabilidades binomiais e de Poisson, é pouco provável que as tabelas mencionadas acima venham a ser ampliadas ou atualizadas.

Snow666

9

A DISTRIBUIÇÃO NORMAL

9.1 Distribuições Contínuas 215
9.2 A Distribuição Normal 217
*9.3 Verificação da Normalidade 226
9.4 Aplicações da Distribuição Normal 228
9.5 A Aproximação Normal da Distribuição Binomial 231
9.6 Lista de Termos-Chave 236
9.7 Referências 237

Os espaços amostrais contínuos e as **variáveis aleatórias contínuas** ocorrem quando lidamos com grandezas que são medidas numa escala contínua, por exemplo, quando medimos a velocidade do vento ou de um carro, a quantidade de álcool na corrente sangüínea de uma pessoa, o peso líquido de um pacote de lasagna congelada, ou a quantidade de alcatrão num cigarro. Em situações como essas, sempre existe um contínuo de possibilidades, mas na prática não nos resta outra escolha que não seja arredondar as medidas para o inteiro mais próximo ou para algumas casas decimais. Assim, se dissermos que num certo dia a temperatura máxima em Porto Alegre foi de 38 graus, arredondada até o grau mais próximo, isso significa que a temperatura poderia ter sido qualquer uma entre 37,5 e 38,5. Analogamente, se dissermos que um certo avião voava a 1.100 metros, arredondados até os 100 metros mais próximos, isso significa que sua altitude poderia ter sido qualquer uma entre 1.050 e 1.150 metros. Precisamos nos lembrar disso quando perguntamos pelas probabilidades referentes a variáveis contínuas. Em outras palavras, associamos probabilidades a intervalos ou regiões de um espaço amostral, e não a pontos individuais. Por exemplo, podemos querer saber a probabilidade de que, num dado momento, um carro esteja correndo a uma velocidade entre 80 e 85 km/h, e não que esteja correndo a exatamente $27\pi = 84{,}82300167...$ km/h. Analogamente, podemos querer saber se um pacote de alimento congelado pesa pelo menos 195 gramas e não que pese exatamente $\sqrt{38{,}5} = 196{,}21416873...$ gramas.

Neste capítulo, vemos como determinar e manusear probabilidades relativas a variáveis aleatórias contínuas. O lugar dos histogramas é ocupado por curvas contínuas, como na Figura 9.1 e podemos considerá-las como curvas que estão aproximadas por histogramas com intervalos de classe menores e menores. Depois de uma introdução geral às **distribuições contínuas** na Seção 9.1, dedicamos o restante deste capítulo à **distribuição normal**, fundamental para a maior parte

Snow666

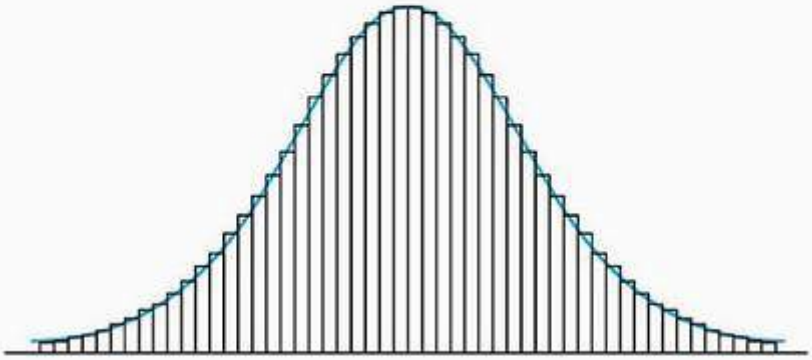

Figura 9.1
Curva de distribuição contínua.

das técnicas "arroz com feijão" da Estatística moderna. Discutem-se várias aplicações da distribuição normal nas Seções 9.4 e 9.5, e segue-se o material opcional da Seção 9.3, em que nos ocupamos de um método para decidir se os dados observados seguem o padrão geral da distribuição normal.

9.1 DISTRIBUIÇÕES CONTÍNUAS

Nos histogramas vistos até agora, as freqüências, percentagens, proporções ou probabilidades eram representadas pelas alturas os retângulos, ou por suas áreas. No caso contínuo, também representamos probabilidades por áreas, não por áreas de retângulos, mas por áreas sob curvas contínuas. Isso é ilustrado pela Figura 9.2, em que o diagrama à esquerda mostra um histograma da distribuição de probabilidade de uma variável aleatória discreta que toma apenas os valores 0, 1, 2, . . . e 10. A probabilidade de que vá tomar o valor 3, por exemplo, é dada pela área do retângulo levemente colorido, e a probabilidade de que vá tomar um valor maior do que ou igual a 8 é dada pela soma das áreas dos três retângulos coloridos mais fortemente. O diagrama à direita na Figura 9.2 refere-se a uma variável aleatória contínua que pode tomar qualquer valor no intervalo de 0 a 10. A probabilidade de que vá tomar um valor no intervalo de 2,5 a 3,5, por exemplo, é dada pela área da região levemente colorida sob a curva, e a probabilidade de que vá tomar um valor maior do que ou igual a 8 é dada pela área colorida mais fortemente sob a curva.

As curvas contínuas, como a mostrada à direita na Figura 9.2, são gráficos de funções denominadas **densidades de probabilidade** ou, informalmente, **distribuições contínuas**. O termo "densidade de probabilidade" tem origem na Física, em que os termos "peso" e "densidade" são empregados praticamente da mesma forma com que usamos os termos "probabilidade" e "densidade de probabilidade" em Estatística. Conforme mostra a Figura 9.3, as densidades de probabilidade se caracterizam pelo fato de que **a área sob a curva entre dois valores quaisquer *a* e *b* dá a probabilidade de uma variável aleatória com essa distribuição contínua tomar um valor no intervalo de *a* a *b*.**

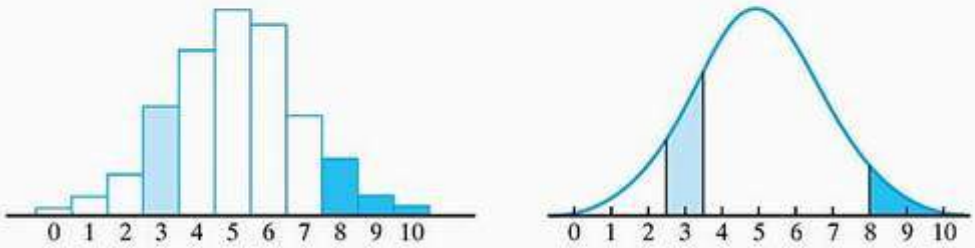

Figura 9.2
Histograma de uma distribuição de probabilidade e gráfico de uma distribuição contínua.

Snow666

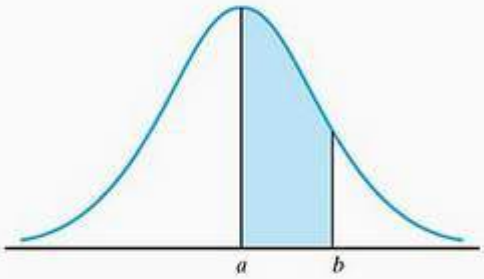

Figura 9.3
Distribuição contínua.

Observe na Figura 9.1 que, quando a e b estão muito próximos, também estão muito próximas à altura da curva e a altura do retângulo cuja base é o intervalo de a a b. Assim, a área sob a curva de a a b é quase igual à do retângulo correspondente.

Decorre que os valores de uma distribuição contínua devem ser não-negativos e que a área total sob a curva, que representa a certeza de que uma variável deve tomar um de seus valores, é sempre igual a 1. Por outro lado, contrastando com as duas regras à página 142, não há exigência de que os valores de uma distribuição contínua sejam menores do que ou iguais a 1.

EXEMPLO 9.1 Verifique que $f(x) = \frac{x}{8}$ pode ser a densidade de probabilidade de uma variável aleatória definida sobre o intervalo de $x = 0$ a $x = 4$.

Solução A primeira condição é verificada, pois $\frac{x}{8}$ é não-negativo (positivo ou nulo) para qualquer valor de x no intervalo de 0 a 4. No que diz respeito à segunda condição, pode-se ver na Figura 9.4 que a área total sob a curva de $x = 0$ a $x = 4$ é a área de um triângulo cuja base é 4 e cuja altura é $\frac{4}{8} = \frac{1}{2}$. A fórmula usual da área de um triângulo dá o valor $\frac{1}{2} \cdot 4 \cdot \frac{1}{2} = 1$ exigido da densidade.

EXEMPLO 9.2 Com referência ao Exemplo 9.1, encontre as probabilidades de uma variável aleatória com a densidade de probabilidade dada vá tomar um valor

(a) menor do que 2;

(b) menor do que ou igual a 2.

Solução **(a)** A probabilidade é dada pela área do triângulo delimitado pela reta tracejada $x = 2$. Sua base é 2, sua altura é $\frac{2}{8} = \frac{1}{4}$, e sua área é $\frac{1}{2} \cdot 2 \cdot \frac{1}{4} = \frac{1}{4}$.

(b) A probabilidade é a mesma que a da parte (a), a saber, $\frac{1}{4}$.

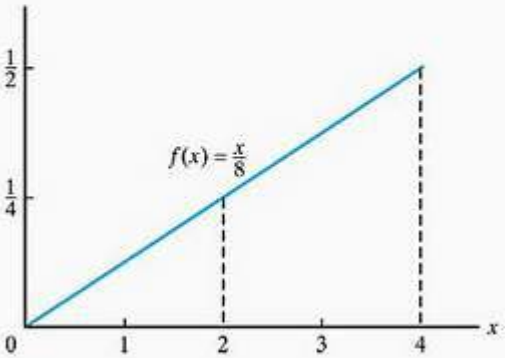

Figura 9.4
Diagrama para os Exemplos 9.1 e 9.2.

Snow666

Este exemplo ilustra o fato importante de que, no caso contínuo, a probabilidade de que uma variável aleatória vá tomar um valor específico é zero. No nosso exemplo, há uma probabilidade zero de que a variável aleatória tome o valor 2, e com isso queremos dizer *exatamente* 2, não incluindo valores vizinhos como 1,9999998 ou 2,0000001.

Uma conseqüência de medição (em vês de contagem) é que devemos atribuir probabilidade zero a qualquer resultado específico. Afirmamos que há uma probabilidade zero de um indivíduo ter um peso de *exatamente* 65,87 kg ou de um cavalo terminar uma corrida em *exatamente* 58,442 segundos. Observe, contudo, que mesmo que cada resultado particular tenha probabilidade zero, o processo ainda assim produz um valor (quer possamos, ou não, medi-lo com um instrumento de alta precisão); assim, eventos com probabilidade zero não só podem, como devem ocorrer, quando estamos lidando com variáveis aleatórias (contínuas) medidas.

As descrições estatísticas de distribuições contínuas são tão importantes como as descrições de distribuições de probabilidade ou como as descrições de distribuições de dados observados, mas a maioria delas, inclusive a média e o desvio-padrão, não podem ser definidas sem usar o Cálculo. Informalmente, entretanto, podemos sempre ver distribuições contínuas como sendo aproximadas por histogramas de distribuições de probabilidade (ver Figura 9.1), cuja média e desvio-padrão sabemos calcular. Então, se escolhermos histogramas com classes cada vez mais estreitas, as médias e os desvios-padrão das distribuições de probabilidade correspondentes se aproximarão cada vez mais da média e do desvio-padrão da distribuição contínua. Na realidade, a média e o desvio-padrão de uma distribuição contínua medem as mesmas propriedades que a média e o desvio-padrão de uma distribuição de probabilidade, a saber, o valor esperado de uma variável aleatória tendo a distribuição dada e a raiz quadrada dos valores esperados dos quadrados dos desvios da média. Mais intuitivamente, a média μ de uma distribuição contínua é uma medida do seu centro, ou meio, e o desvio-padrão σ de uma distribuição contínua é uma medida de sua dispersão, ou espalhamento.

9.2 A DISTRIBUIÇÃO NORMAL

Entre as muitas distribuições contínuas diferentes usadas em Estatística, a mais importante é a **distribuição normal**, cujo estudo remonta a pesquisas desenvolvidas no século XVIII sobre os erros de mensuração. Constatou-se que as discrepâncias entre repetidas medições da mesma grandeza física apresentavam um grau surpreendente de regularidade. A distribuição das discrepâncias podia ser aproximada de perto por uma curva contínua, conhecida como a "curva normal dos erros" e atribuída às leis do acaso. A equação matemática desse tipo de curva é

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

para $-\infty < x < \infty$, onde e é o número irracional 2,71828 . . ., que já encontramos à página 201 no estudo da distribuição de Poisson. Demos essa equação apenas para salientar algumas características essenciais da distribuição normal; ela não será utilizada em nossos cálculos.

O gráfico de uma distribuição normal é uma curva em forma de sino que se prolonga indefinidamente em ambas direções. Embora isso não seja visível num desenho pequeno como o da Figura 9.5, a curva se aproxima cada vez mais do eixo horizontal, sem jamais tocá-lo, por mais afastados que estejamos de um dos dois lados da média. Felizmente, quase nunca é preciso prolongar muito as caudas de uma distribuição normal, porque a área sob a curva a mais de quatro ou cinco desvios-padrão a contar da média é desprezível para a maioria dos fins práticos. (Ver Exercício 9.15.)

Snow666

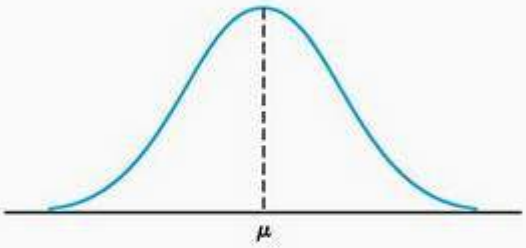

Figura 9.5
Curva da distribuição normal.

Uma característica importante das distribuições normais, visível na equação precedente, é que elas dependem apenas das duas quantidades μ e σ, que são, na realidade, a média e o desvio-padrão. Em outras palavras, há uma, e só uma, distribuição normal com uma dada média μ e um dado desvio-padrão σ. A Figura 9.6 mostra que obteremos curvas diferentes, dependendo dos valores de μ e σ. Em cima, vêem-se duas curvas normais com médias diferentes, mas desvios-padrão iguais; a curva à direita tem uma média maior. No centro, temos duas curvas normais com mesma média, mas desvios-padrão diferentes; a mais baixa e achatada tem um desvio-padrão maior. Embaixo vemos duas curvas normais com médias e desvios-padrão diferentes.

Em todo nosso trabalho com distribuições normais, estaremos interessados apenas nas áreas sob suas curvas, as chamadas **áreas sob a curva normal** que, na prática, se encontram em tabelas como a Tabela I no fim do livro. Como é fisicamente impossível, mas também desnecessário, construir tabelas separadas de áreas sob as curvas normais para todos os pares imagináveis de valores de μ e σ, tabelamos essas áreas apenas para a distribuição normal com $\mu = 0$ e $\sigma = 1$, denominada **distribuição normal padrão**. Podemos, então, obter áreas sob qualquer curva normal através de uma mudança de escala (ver Figura 9.7) que transforma as unidades de medições da escala original, ou escala x, em **unidades padronizadas**, **escores padronizados** ou **escores** z, por meio da fórmula

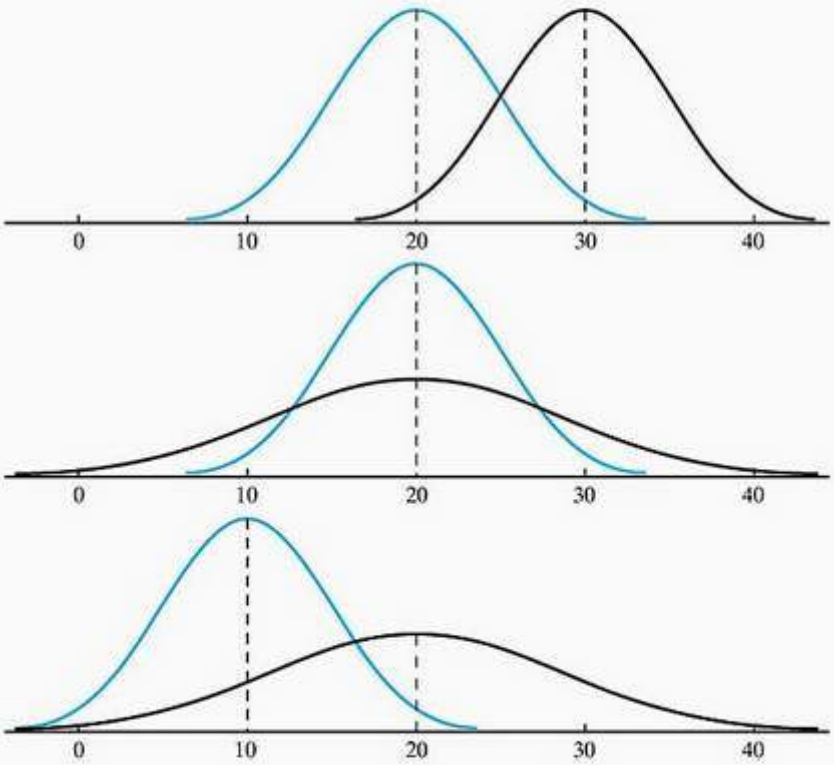

Figura 9.6
Três pares de distribuições normais.

Snow666

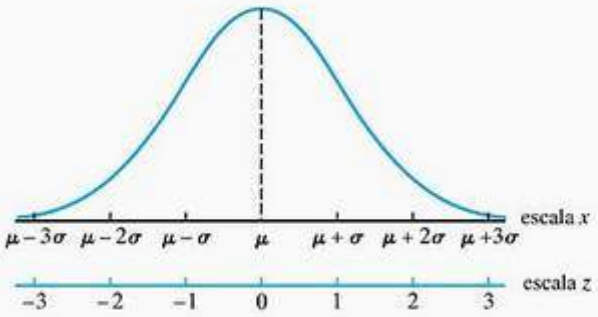

Figura 9.7
Mudança de escala para unidades padronizadas.

UNIDADES PADRONIZADAS

$$z = \frac{x - \mu}{\sigma}$$

Nesta nova escala, a escala z, um valor de z nos diz apenas a quantos desvios-padrão o valor correspondente de x está acima ou abaixo da média de sua distribuição.

Os valores na Tabela I são as áreas sob a curva normal padrão entre a média $z = 0$ e $z = 0{,}00$; 0,01; 0,02; . . .; 3,08 e 3,09 e também $z = 4{,}00$, $z = 5{,}00$ e $z = 6{,}00$. Em outras palavras, os valores da Tabela I são áreas normalizadas como a da região colorida da Figura 9.8.

A Tabela I não tem entradas correspondentes a valores negativos de z, os quais são desnecessários em virtude da simetria de qualquer curva normal em torno de sua média. Isso decorre da equação à página 217, que permanece inalterada quando substituímos $-(x - \mu)$ por $x - \mu$. Especificamente, $f(\mu - a) = f(\mu + a)$, significando que obtemos o mesmo valor para $f(x)$ quando percorremos a distância x para a esquerda ou para a direita de μ.

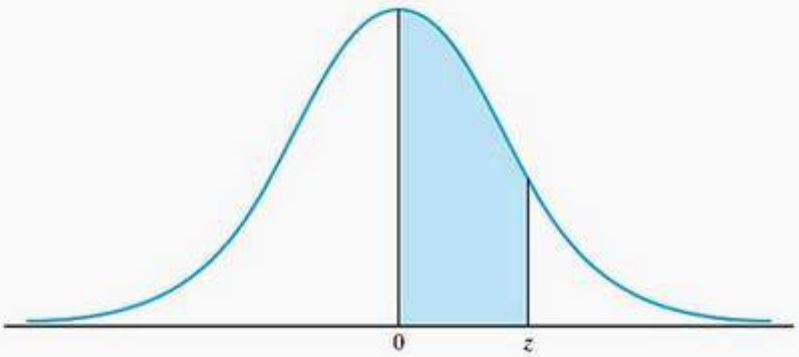

Figura 9.8
Áreas tabeladas da curva normal.

EXEMPLO 9.3 Encontre a área sob a curva normal padrão entre $z = -1{,}20$ e $z = 0$.

Solução Como pode ser visto na Figura 9.9, a área sob a curva entre $z = -1{,}20$ e $z = 0$ é igual à área sob a curva entre $z = 0$ e $z = 1{,}20$. Portanto, procuramos o valor correspondente a $z = 1{,}20$ na Tabela I e obtemos 0,3849.

As questões que envolvem áreas sob distribuições normais surgem em vários contextos, e a capacidade de encontrar rapidamente qualquer área desejada pode ser muito útil. Embora a tabela dê somente áreas entre $z = 0$ e valores positivos selecionados de z, não raro precisamos encontrar áreas à esquerda ou à direita de valores dados de z, positivos ou negativos, ou então áreas entre dois valores dados de z. Isso não é difícil, desde que lembremos exatamente quais áreas estão

Snow666

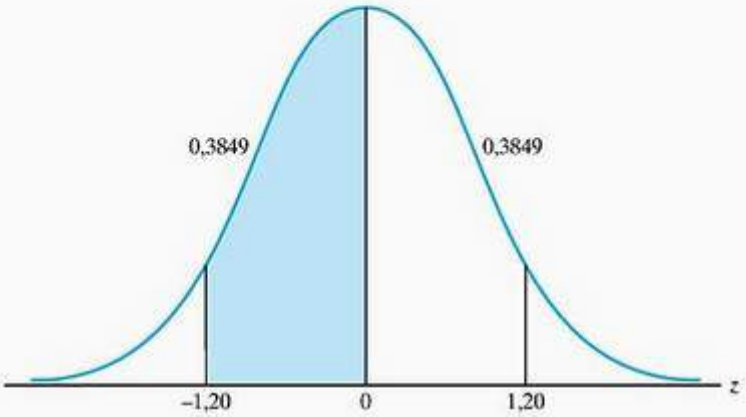

Figura 9.9
Diagrama para o Exemplo 9.3.

representadas pelas entradas na Tabela I, e que não esqueçamos que a distribuição normal padrão é simétrica em torno de $z = 0$, de modo que a área sob a curva à esquerda de $z = 0$ e a área sob a curva à direita de $z = 0$ são ambas iguais a 0,5000.

EXEMPLO 9.4 Encontre a área sob a curva normal padrão

(a) à esquerda de $z = 0{,}94$;

(b) à direita de $z = -0{,}65$;

(c) à direita de $z = 1{,}76$;

(d) à esquerda de $z = -0{,}85$;

(e) entre $z = 0{,}87$ e $z = 1{,}28$;

(f) entre $z = -0{,}34$ e $z = 0{,}62$.

Solução Para cada uma das seis partes, observe o diagrama correspondente na Figura 9.10.

(a) A área à esquerda de $z = 0{,}94$ é igual a 0,5000 mais o valor da Tabela I correspondente a $z = 0{,}94$, ou $0{,}5000 + 0{,}3264 = 0{,}8264$.

(b) A área à direita de $z = -0{,}65$ é 0,5000 mais o valor da Tabela I correspondente a $z = 0{,}65$, ou $0{,}5000 + 0{,}2422 = 0{,}7422$.

(c) A área à direita de $z = 1{,}76$ é 0,5000 menos o valor da Tabela I correspondente a $z = 1{,}76$, ou $0{,}5000 + 0{,}4608 = 0{,}0392$.

(d) A área à esquerda de $z = -0{,}85$ é 0,5000 menos o valor da Tabela I correspondente a $z = 0{,}85$, ou $0{,}5000 - 0{,}3023 = 0{,}1977$.

(e) A área entre $z = 0{,}87$ e $z = 1{,}28$ é a diferença entre os valores da Tabela I correspondentes a $z = 0{,}87$ e $z = 1{,}28$, ou $0{,}3997 - 0{,}3078 = 0{,}0919$.

(f) A área entre $z = -0{,}34$ e $z = 0{,}62$ é a soma dos valores da Tabela I correspondentes a $z = 0{,}34$ e $z = 0{,}62$, ou $0{,}1331 + 0{,}2324 = 0{,}3655$.

Em ambos os exemplos precedentes, lidamos diretamente com a distribuição normal padrão. Consideremos agora um exemplo em que μ e σ não são 0 e 1, o que nos obriga a transformar primeiro para unidades padronizadas.

EXEMPLO 9.5 Se uma variável aleatória tem a distribuição normal com $\mu = 10$ e $\sigma = 5$, qual é a probabilidade de que vá tomar um valor no intervalo de 12 a 15?

Snow666

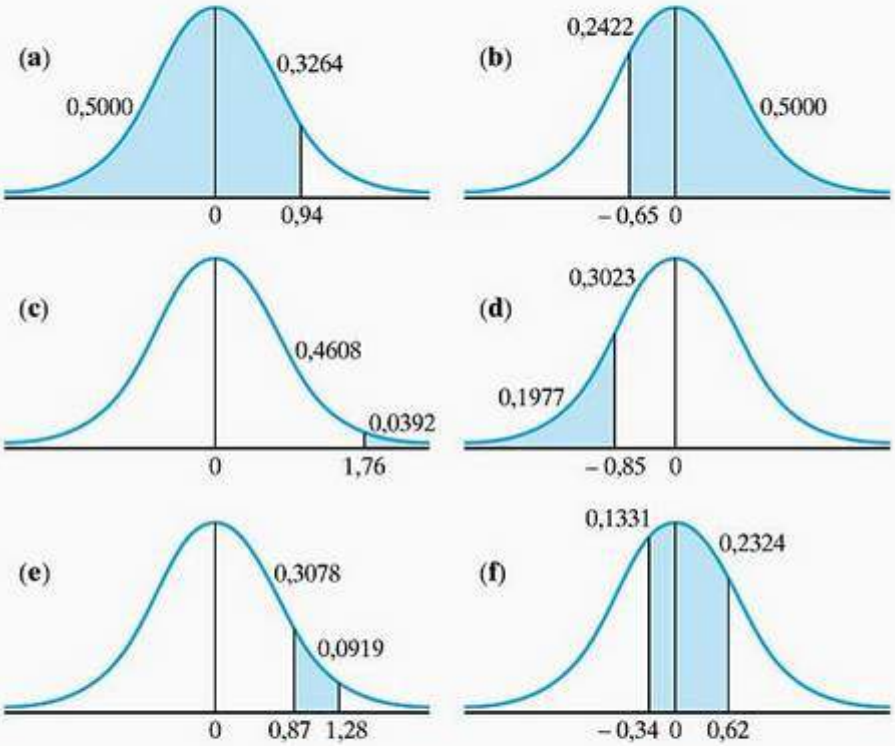

Figura 9.10
Diagrama para o Exemplo 9.4.

Solução A probabilidade é dada pela área da região colorida da Figura 9.11. Transformando $x = 12$ e $x = 15$ em unidades padronizadas, obtemos

$$z = \frac{12 - 10}{5} = 0{,}40 \quad \text{e} \quad z = \frac{15 - 10}{5} = 1{,}00$$

e como os valores correspondentes na Tabela I são 0,1554 e 0,3413, a probabilidade procurada é $0{,}3413 - 0{,}1554 = 0{,}1859$.

EXEMPLO 9.6 Se z_α denota o valor de z para o qual a área sob a curva normal padrão à sua direita é igual a α (letra grega *alfa* minúscula), encontre

(a) $z_{0,01}$; **(b)** $z_{0,05}$.

Solução Para ambas as partes, veja o diagrama da Figura 9.12.

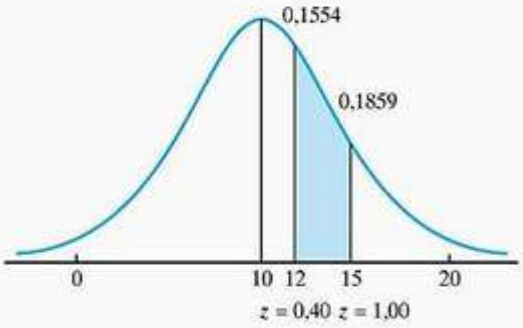

Figura 9.11
Diagrama para o Exemplo 9.5.

Snow666

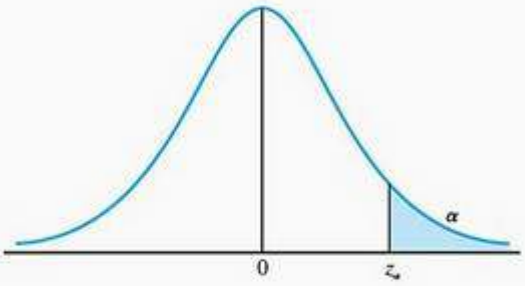

Figura 9.12
Diagrama para o Exemplo 9.6.

(a) Como $z_{0,01}$ corresponde a uma entrada de 0,5000 – 0,0100 = 0,4900 na Tabela I, procuramos o valor mais próximo de 0,4900 e encontramos 0,4901, que corresponde a $z = 2,33$; assim, tomamos $z_{0,01} = 2,33$.

(b) Como $z_{0,05}$ corresponde a uma entrada de 0,5000 – 0,0500 = 0,4500 na Tabela I, procuramos o valor mais próximo de 0,4500 e encontramos 0,4495 e 0,4505, que correspondem a $z = 1,64$ e $z = 1,65$; assim, tomamos $z_{0,05} = 1,645$.

A Tabela I também nos permite verificar a observação à página 90 de que, para distribuições de freqüência com a forma geral da seção transversal de um sino, cerca de 68% dos valores estão a menos de um desvio-padrão da média, cerca de 95% estão a menos de dois desvios-padrão da média, e cerca de 99,7% estão a menos de três desvios-padrão da média. Essas percentagens aplicam-se a distribuições de freqüência com a forma geral de uma distribuição normal, e o leitor pode verificá-las nas três primeiras partes do Exercício 9.15. As outras duas partes daquele exercício mostram que, embora as duas "caudas" se prolonguem indefinidamente em ambas as direções, a área sob a curva normal padrão à direita de $z = 4$ ou $z = 5$, ou à esquerda de $z = -4$ ou $z = -5$, é desprezível.

Embora este capítulo seja dedicado à distribuição normal, e sua importância não pode ser negada, seria um erro grave pensar que a distribuição normal é a única distribuição contínua que interessa ao estado da Estatística. No Capítulo 11 e em capítulos subseqüentes, vamos encontrar outras distribuições contínuas que desempenham papéis importantes em problemas de inferência estatística.

EXERCÍCIOS

9.1 Explique, em cada caso, por que a equação dada não serve como a densidade de probabilidade de uma variável aleatória definida sobre o intervalo de 1 a 4:

(a) $f(x) = 0,25$;

(b) $f(x) = \frac{1}{9}(4x - 7)$.

9.2 Explique, em cada caso, por que a equação dada não serve como a densidade de probabilidade de uma variável aleatória definida sobre o intervalo de 0 a 5:

(a) $f(x) = \frac{1}{10}(x - 4)$;

(b) $f(x) = \frac{1}{50}(x + 1)$.

9.3 A Figura 9.13 mostra o gráfico de uma **distribuição uniforme**, a saber, a de uma variável aleatória que toma um valor constante num intervalo de x dado. Encontre as probabilidades de que essa variável vá tomar

(a) um valor menor do que 6;

(b) o valor 6;

(c) um valor entre 3,5 e 6,8.

Snow666

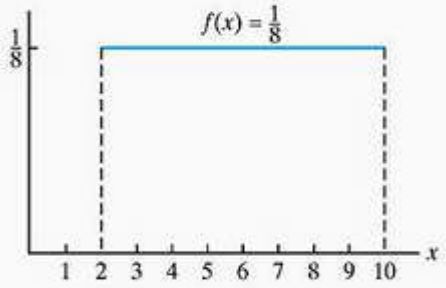

Figura 9.13
Diagrama para o Exercício 9.3.

9.4 Suponha que uma variável aleatória tenha distribuição uniforme (ver Exercício 9.3) no intervalo de 50 a 80. Encontre as probabilidades de que vá tomar um valor

(a) de 60 a 80;
(b) menor do que 50;
(c) maior do que 75;
(d) entre 35 e 45.

9.5 A Figura 9.14 mostra o gráfico da distribuição de uma variável aleatória contínua definida nos valores do intervalo contínuo de – 1 a 1. Encontre as probabilidades de que essa variável aleatória vá tomar um valor

(a) entre $\frac{1}{4}$ e 1;
(b) entre – 0,8 e 0,4;
(c) menor do que 0,65.

9.6 Em cada caso que envolve áreas sob a curva normal padrão, decida se a primeira área é maior, se a segunda área é maior, ou se as duas áreas são iguais:

(a) a área à direita de $z = 1{,}5$ ou a área à direita de $z = 2$;
(b) a área à esquerda de $z = -1{,}5$ ou a área à esquerda de $z = -2$;
(c) a área à direita de $z = 1$ ou a área à esquerda de $z = -1{,}5$;
(d) a área à direita de $z = 2$ ou a área à esquerda de $z = -2$;
(e) a área à direita de $z = -2{,}5$ ou a área à direita de $z = -1{,}5$;
(f) a área à esquerda de $z = 0$ ou a área à direita de $z = -0{,}1$;
(g) a área à direita de $z = 0$ ou a área à esquerda de $z = 0$;
(h) a área à direita de $z = -1{,}4$ ou a área à esquerda de $z = -1{,}4$.

9.7 Em cada caso que envolve áreas sob a curva normal padrão, decida se a primeira área é maior, se a segunda área é maior, ou se as duas áreas são iguais:

(a) a área entre $z = 0$ e $z = 1{,}3$ ou a área entre $z = 0$ e $z = 1$;
(b) a área entre $z = -0{,}2$ e $z = 0{,}2$ ou a área entre $z = -0{,}4$ e $z = 0{,}4$;
(c) a área entre $z = 0{,}15$ e $z = 1{,}5$ ou a área entre $z = 0$ e $z = 3$;
(d) a área entre $z = 0{,}2$ e $z = 0{,}3$ ou a área entre $z = 1{,}2$ e $z = 1{,}3$;

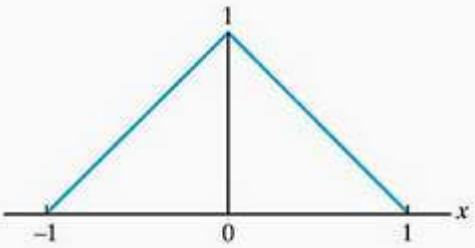

Figura 9.14
Diagrama para o Exercício 9.5.

Snow666

(e) a área entre $z = -1$ e $z = -0,5$ ou a área entre $z = 0,5$ e $z = 1$;
(f) a área à esquerda de $z = -1,5$ ou a área à esquerda de $z = -0,5$;
(g) a área entre $z = -1$ e $z = 1,5$ ou a área entre $z = -1,5$ e $z = 1$.

9.8 Encontre a área sob a curva normal padrão que fica
(a) entre $z = 0$ e $z = 0,87$;
(b) entre $z = -1,66$ e $z = 0$;
(c) à direita de $z = 0,48$;
(d) à direita de $z = -0,27$;
(e) à esquerda de $z = 1,30$;
(f) à esquerda de $z = -0,79$.

9.9 Encontre a área sob a curva normal padrão que fica
(a) entre $z = 0,45$ e $z = 1,23$;
(b) entre $z = -1,15$ e $z = 1,85$;
(c) entre $z = -1,35$ e $z = 0,48$.

9.10 Encontre a área sob a curva normal padrão que fica
(a) entre $z = -0,77$ e $z = 0,77$;
(b) à direita de $z = -1,39$;
(c) à esquerda de $z = 0,27$;
(d) entre $z = 1,69$ e $z = 2,33$.

9.11 Em cada caso que envolve áreas sob a curva normal com $\mu = 115$, decida se a primeira área é maior, se a segunda área é maior, ou se as duas áreas são iguais:
(a) a área à direita de 135 ou a área à esquerda de 85;
(b) a área entre 103 e 127 ou a área entre 115 e 139;
(c) a área à direita de 105 ou a área à esquerda de 125.

9.12 Em cada caso que envolve variáveis aleatórias com distribuições normais, decida se a primeira probabilidade é maior, se a segunda probabilidade é maior, ou se as duas probabilidades são iguais:
(a) a probabilidade de que uma variável aleatória de distribuição normal com $\mu = 50$ e $\sigma = 10$ vá tomar um valor menor do que 60, ou a probabilidade de que uma variável aleatória de distribuição normal com $\mu = 500$ e $\sigma = 100$ vá tomar um valor menor do que 600;
(b) a probabilidade de que uma variável aleatória de distribuição normal com $\mu = 40$ e $\sigma = 5$ vá tomar um valor maior do que 40, ou a probabilidade de que uma variável aleatória de distribuição normal com $\mu = 50$ e $\sigma = 5$ vá tomar um valor maior do que 40;
(c) a probabilidade de que uma variável aleatória de distribuição normal com $\mu = 50$ e $\sigma = 10$ vá tomar um valor menor do que 60, ou a probabilidade de que uma variável aleatória de distribuição normal com $\mu = 50$ e $\sigma = 20$ vá tomar um valor menor do que 60;
(d) a probabilidade de que uma variável aleatória de distribuição normal com $\mu = 100$ e $\sigma = 5$ vá tomar um valor maior do que 110, ou a probabilidade de que uma variável aleatória de distribuição normal com $\mu = 108$ e $\sigma = 5$ vá tomar um valor maior do que 110.

9.13 Encontre z se a área sob a curva normal padrão
(a) entre 0 e z é 0,4788;
(b) à esquerda de z é 0,8365;
(c) entre $-z$ e z é 0,8584;
(d) à esquerda de z é 0,3409.

Snow666

9.14 Encontre z se a área sob a curva normal padrão
(a) entre 0 e z é 0,1480;
(b) à direita de z é 0,7324;
(c) à direita de z é 0,1225;
(d) entre $-z$ e z é 0,9328.

9.15 Encontre a área sob a curva normal padrão entre $-z$ e z, se
(a) $z = 1$;
(b) $z = 2$;
(c) $z = 3$;
(d) $z = 4$;
(e) $z = 5$.

9.16 Com z_α definido como no Exemplo 9.6, verifique que
(a) $z_{0,025} = 1,96$;
(b) $z_{0,005} = 2,575$.

9.17 Se uma variável aleatória tem a distribuição normal com $\mu = 82,0$ e $\sigma = 4,8$, encontre a probabilidade de que ela vá tomar um valor
(a) menor do que 89,2;
(b) maior do que 78,4;
(c) entre 83,2 e 88,0;
(d) entre 73,6 e 90,4.

9.18 Dada uma curva normal com $\mu = 36,3$ e $\sigma = 8,1$, encontre a área sob a curva entre 31,6 e 40,1.

9.19 Uma distribuição normal tem a média $\mu = 62,4$. Encontre seu desvio-padrão se 20% da área sob a curva estão à direita de 79,2.

9.20 Uma variável aleatória tem uma distribuição normal com $\sigma = 10$. Se a probabilidade de que ela vá tomar um valor menor do que 82,5 é 0,8212, qual é a probabilidade de que ela vá tomar um valor maior do que 58,3?

9.21 Uma outra distribuição contínua, denominada **distribuição exponencial**, tem muitas aplicações importantes. Se uma variável aleatória tem uma distribuição exponencial com média μ, a probabilidade de que ela vá tomar um valor entre 0 e um valor não-negativo de x qualquer é $1 - e^{-x/\mu}$ (ver Figura 9.15). Aqui, e é o número irracional que aparece também na fórmula da distribuição normal. Muitas calculadoras têm teclas para o cálculo de expressões de forma $e^{-x/\mu}$, e valores selecionados podem ser obtidos na Tabela XII. Encontre a probabilidade de que uma variável aleatória com distribuição exponencial com média $\mu = 10$ vá tomar um valor

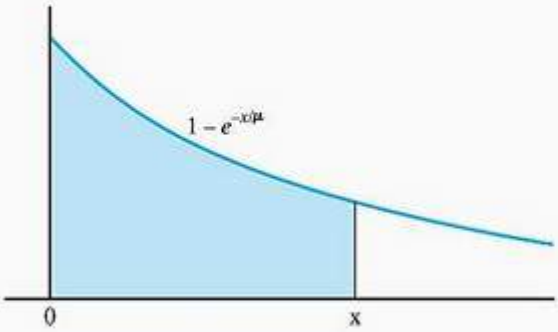

Figura 9.15
Distribuição exponencial.

Snow666

(a) menor do que 4;
(b) entre 5 e 9;
(c) maior do que 16.

9.22 A duração de certo componente eletrônico é uma variável aleatória que tem distribuição exponencial com média $\mu = 2.000$ horas. Use a fórmula do Exercício 9.21 para encontrar as probabilidades de que um tal componente dure

(a) no máximo 2.400 horas;
(b) no mínimo 1.600 horas;
(c) entre 1.800 e 2.200 horas.

9.23 De acordo com uma pesquisa médica, o tempo entre ocorrências sucessivas de uma doença tropical rara é uma variável aleatória que tem distribuição exponencial com média $\mu = 120$ dias. Encontre as probabilidades de que o tempo entre ocorrências sucessivas da doença

(a) exceda 240 dias;
(b) exceda 360 dias;
(c) seja inferior a 60 dias.

9.24 Na região em torno de uma falha geológica, o tempo entre tremores secundários posteriores a um terremoto é uma variável aleatória que tem distribuição exponencial com $\mu = 36$ horas. Encontre as probabilidades de que o tempo entre tremores secundários sucessivos

(a) seja inferior a 18 horas;
(b) exceda 72 horas;
(c) fique entre 36 e 108 horas.

*9.3 VERIFICAÇÃO DA NORMALIDADE

Em muitos dos procedimentos que discutiremos em capítulos subseqüentes, suporemos que nossos dados vêm de populações normais, ou seja, que lidamos com valores de variáveis aleatórias que têm distribuições normais. Por muitos anos, essa hipótese era verificada com o uso de um tipo especial de papel de gráfico, denominado **papel de probabilidade normal** ou **papel de probabilidade aritmética**. Tal papel de gráfico podia ser encontrado na maioria das papelarias, livrarias de universidade ou em lojas especializadas de material de escritório. Hoje em dia, esse papel passou a ser uma curiosidade, sendo difícil se encontrar, pois o trabalho de verificar a normalidade tem sido feito por computadores (com aplicativos estatísticos especiais) e outras ferramentas. Como já indicamos várias vezes, os computadores com aplicativos apropriados ou algum outro recurso tecnológico podem ser muito úteis em Estatística, mas nenhuma dessas ferramentas é essencial para usar este livro. Por isso esta seção está marcada como opcional com o asterisco.

Na realidade, os gráficos mostrados nas Figuras 9.16 e 9.17 compartilham as escalas do papel de probabilidade normal. Os eixos horizontais, usados para os valores da variável aleatória, a mesma em ambos diagramas, têm escalas ordinárias com subdivisões iguais. Por outro lado, a escala nos eixos verticais está marcada de tal modo que o gráfico de qualquer distribuição normal cumulativa fica uma linha reta. Isso é precisamente como julgamos a "normalidade" de qualquer conjunto de dados. *Se o padrão obtido de nossos dados é aproximadamente o de uma linha reta, estamos justificados para concluir que os dados provêm de uma* ***população normal***.

A Figura 9.16 dá um exemplo de um **gráfico de probabilidade normal** obtido com o uso de um computador e o aplicativo MINITAB. A Figura 9.17 faz o mesmo com o uso de uma cal-

Snow666

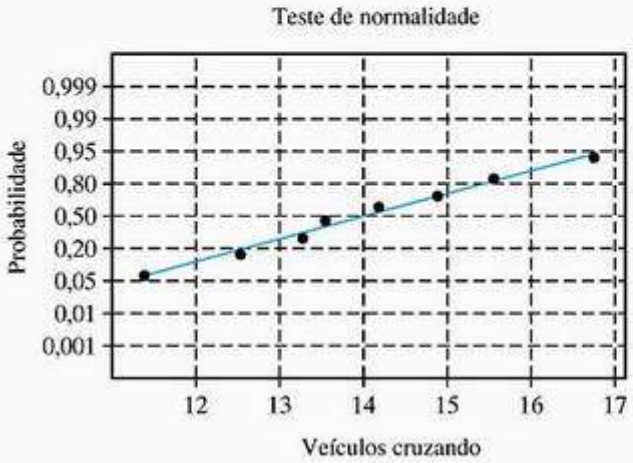

Figura 9.16
Gráfico de probabilidade normal obtido com computador e o aplicativo MINITAB.

culadora gráfica. Qualquer um desses dois gráficos de probabilidade normal poderia servir como procedimento preliminar para estimar, digamos, a durabilidade de uma nova tinta para pintar as linhas divisórias de tráfego das rodovias. Ocorre que o procedimento estatístico utilizado pelo órgão governamental que trata da manutenção das rodovias para julgar a durabilidade de tal tinta só funciona quando seus dados constituem uma amostra de uma população normal. Portanto, esse órgão pintou linhas em rodovias com tráfego pesado em oito locais diferentes e constatou que a pintura se deteriora depois de ter sido cruzada por 14,26; 16,78; 13,65; 11,53; 12,64; 13,37; 15,60 e 14,94 milhões de veículos. Então, digitando esses dados em seu computador e seguindo as instruções de seu aplicativo MINITAB para um gráfico de probabilidade normal, obtiveram o gráfico de probabilidade normal mostrado na Figura 9.16. Como pode ser observado, os oito pontos estão muito próximos de uma linha reta; com certeza suficientemente próximos para concluir que os dados constituem uma amostra de uma população normal. Num nível menos subjetivo, poderíamos continuar com um dos muitos procedimentos para testar a normalidade de um conjunto de dados amostrais. As informações para vários desses procedimentos, dos quais não necessitamos, foram omitidos da janela de MINITAB exibida na Figura 9.16

As calculadoras gráficas geram gráficos de probabilidade de uma maneira um pouco diferente do gráfico gerado por MINITAB da Figura 9.16. O gráfico exibido na Figura 9.17 se refere aos mesmos dados, mas a teoria subjacente é outra. Novamente, a linearidade do gráfico é um sinal de normalidade, ou seja, indica que os dados são valores de uma variável aleatória com uma distribuição normal ou muito próxima da normal.

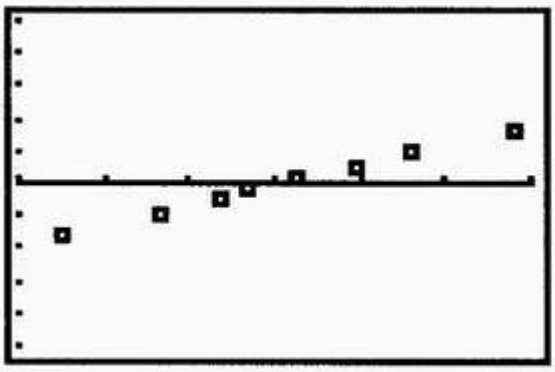

Figura 9.17
Gráfico de probabilidade normal obtido com uma calculadora gráfica TI-83Plus.

Snow666

EXERCÍCIOS

*9.25 Use um aplicativo computacional apropriado ou uma calculadora gráfica para produzir um gráfico de probabilidade normal para os dados de ocupação de quartos de hotel à página 28. Esses dados podem ser interpretados como sendo os valores de uma variável aleatória com uma distribuição normal?

*9.26 Use um aplicativo computacional apropriado ou uma calculadora gráfica para produzir um gráfico de probabilidade normal para os dados do Exercício 2.22 à página 33, referente ao número de cirurgias feitas num hospital durante 80 semanas. Esses dados podem ser interpretados como sendo os valores de uma variável aleatória com uma distribuição normal?

*9.27 Use um aplicativo computacional apropriado ou uma calculadora gráfica para produzir um gráfico de probabilidade normal para os dados do Exercício 2.42 à página 41, referente aos comprimentos das raízes de 120 plantinhas. Esses dados podem ser interpretados como sendo os valores de uma variável aleatória com uma distribuição normal?

*9.28 Use um aplicativo computacional apropriado ou uma calculadora gráfica para produzir um gráfico de probabilidade normal para os dados do Exercício 2.36 à página 40, referente às percentagens de encolhimento de espécimes de cerâmica plástica. Esses dados podem ser interpretados como sendo os valores de uma variável aleatória com uma distribuição normal?

*9.29 Use um aplicativo computacional apropriado ou uma calculadora gráfica para produzir um gráfico de probabilidade normal para os comprimentos das 60 trutas marinhas que constituem o primeiro conjunto de dados da Seção 2.1, à página 26. Esses dados podem ser interpretados como sendo os valores de uma variável aleatória com uma distribuição normal?

9.4 APLICAÇÕES DA DISTRIBUIÇÃO NORMAL

Consideremos, agora, algumas aplicações em que podemos supor, em cada caso, que as variáveis aleatórias sob consideração têm distribuições normais ou podem ser muito bem aproximados por distribuições normais. As aplicações mais diretas das distribuições normais ocorrem quando nos são dados os valores de ambos seus parâmetros, μ e σ. Como já indicamos à página 217 em relação à equação da curva normal, esses dois parâmetros especificam completamente a distribuição normal, ou seja, a área sob a curva normal pode ser determinada convertendo para unidades padronizadas e então utilizando a Tabela I no final do livro. No que segue ilustramos esse tipo de aplicação.

EXEMPLO 9.7 Se a quantidade de radiação cósmica a que uma pessoa está exposta enquanto atravessar o território dos Estados Unidos de avião é uma variável aleatória de distribuição normal com $\mu = 4{,}35$ mrem e $\sigma = 0{,}59$ mrem, encontre as probabilidades de que uma pessoa num tal vôo esteja exposta a

(a) mais de 5,00 mrem de radiação cósmica;

(b) alguma quantidade de 3,00 a 4,00 mrem de radiação cósmica.

Solução **(a)** Essa probabilidade é dada pela área da região colorida sob a curva do diagrama no topo da Figura 9.18, a saber, a área sob a curva à direita de

$$z = \frac{5{,}00 - 4{,}35}{0{,}59} \approx 1{,}10$$

Snow666

Como a entrada na Tabela I correspondente a $z = 1{,}10$ é 0,3643, vemos que a probabilidade de que uma pessoa esteja exposta a mais de 5,00 mrem de radiação cósmica num tal vôo é $0{,}5000 - 0{,}3646 = 0{,}1357$ ou, aproximadamente, 0,14.

(b) Essa probabilidade é dada pela área da região colorida sob a curva do diagrama na base da Figura 9.18, a saber, a área sob a curva à direita de

$$z = \frac{3{,}00 - 4{,}35}{0{,}59} \approx -2{,}29 \quad \text{e} \quad z = \frac{4{,}00 - 4{,}35}{0{,}59} \approx -0{,}59$$

Como as entradas na Tabela I correspondente a $z = 2{,}29$ e $z = 0{,}59$ são, respectivamente, 0,4890 e 0,2224, vemos que a probabilidade de que uma pessoa esteja exposta a alguma quantidade de 3,00 a 4,00 mrem de radiação cósmica num tal vôo é $0{,}4890 - 0{,}2224 = 0{,}2666$ ou, aproximadamente, 0,27.

Embora a distribuição normal seja uma distribuição contínua que é aplicável a variáveis aleatórias contínuas, muitas vezes ela é usada para aproximar distribuições de variáveis aleatórias discretas, que podem tomar apenas um número finito de valores, ou tantos valores quantos são os inteiros positivos. Para tanto, devemos usar a **correção de continuidade** ilustrada no exemplo seguinte. Além disso, esse exemplo é novamente uma aplicação com dados valores (na verdade, estimados) da média e do desvio-padrão.

EXEMPLO 9.8 Num estudo de comportamento agressivo, constatou-se que um conjunto de camundongos brancos do sexo masculino, quando retornaram ao grupo em que viviam após quatro semanas de isolamento, envolveram-se numa média de 18,6 brigas nos primeiros cinco minutos, com desvio-padrão de 3,3 brigas. Se puder ser admitido que a distribuição dessa variável aleatória (o número de brigas em que um tal camundongo se envolve nas condições enunciadas) pode ser muito bem aproximada por uma distribuição normal, qual é a probabilidade de que um tal camundongo vá se envolver em pelo menos 15 brigas nos primeiros cinco minutos?

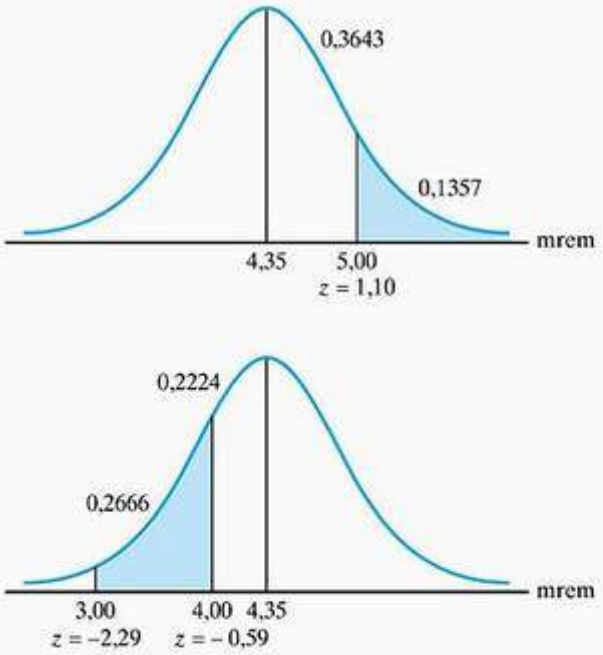

Figura 9.18
Diagrama para o Exemplo 9.7.

Snow666

Solução A resposta é dada pela área da região colorida sob a curva na Figura 9.19, a saber, pela área sob a curva à direita de 14,5, e não de 15. A razão disso é que o número de brigas é um número inteiro. Portanto, se quisermos aproximar a distribuição dessa variável aleatória por uma distribuição normal, devemos "espalhar" seus valores ao longo de uma escala contínua e, para tanto, representamos cada número inteiro k pelo intervalo de $k - \frac{1}{2}$ até $k + \frac{1}{2}$. Por exemplo, 5 é representado pelo intervalo de 4,5 a 5,5, 10 é representado pelo intervalo de 9,5 a 10,5, 20 é representado pelo intervalo de 19,5 a 20,5, e a probabilidade de 15 ou mais é dada pela área sob a curva à direita de 14,5. Conseqüentemente, obtemos

$$z = \frac{14{,}5 - 18{,}6}{3{,}3} \approx -1{,}24$$

e segue pela Tabela I que a área da região colorida sob a curva na Figura 9.19, ou seja, a probabilidade de que um tal camundongo vá se envolver em pelo menos 15 brigas nos primeiros cinco minutos, é de 0,5000 + 0,3925 = 0,8925 ou, aproximadamente, 0,89.

Um pouco mais difíceis são as aplicações em que somente é dado um dos dois parâmetros μ e σ, além de um valor de x e da área sob a curva à sua esquerda ou direita. O que segue ilustra esse tipo de aplicação com dados valores de σ e de x, bem como da área sob a curva à esquerda desse valor de x.

EXEMPLO 9.9 A quantidade efetiva de massa de rejunte que uma máquina deposita em sacos de 6 quilogramas varia de saco para saco, e pode ser considerada uma variável aleatória normal com desvio-padrão de 0,04 quilogramas. Se apenas 2% dos sacos podem conter menos do que 6 quilogramas de massa de rejunte, qual deve ser a quantidade média a ser depositada nesses sacos?

Solução Aqui temos dado $\sigma = 0{,}04$, $x = 6{,}00$, uma área sob a curva normal (a da região colorida na Figura 9.20) e devemos encontrar μ. O valor de z para o qual a entrada na Tabela I está mais próximo de 0,5000 – 0,0200 = 0,4800 é $z = 2{,}05$, que corresponde a 0,4798, de modo que

$$-2{,}05 = \frac{6{,}00 - \mu}{0{,}04}$$

Então, resolvendo em μ, obtemos $6{,}00 - \mu = -2{,}05(0{,}04) = -0{,}082$ e $\mu = 6{,}00 + 0{,}082 = 6{,}082$ ou, aproximadamente, 6,08. A quantidade média a ser depositada deve ser de 6,08 quilogramas.

Todos os exemplos desta seção trataram de variáveis aleatórias com distribuições normais, ou distribuições que podem ser muito bem aproximadas por curvas normais. Quando observamos um valor (ou valores) de uma variável aleatória tendo uma distribuição normal, podemos dizer que estamos **extraindo uma amostra de uma população normal**, o que é consistente com a terminologia introduzida ao final da Seção 8.3.

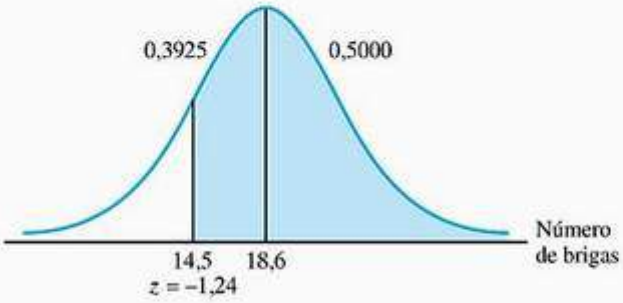

Figura 9.19 Diagrama para o Exemplo 9.8.

Snow666

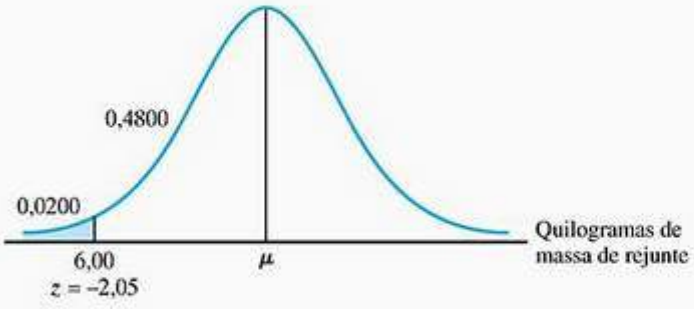

Figura 9.20
Diagrama para o Exemplo 9.9.

9.5 A APROXIMAÇÃO NORMAL DA DISTRIBUIÇÃO BINOMIAL

A distribuição normal fornece uma aproximação bastante boa da distribuição binomial quando n, o número de provas, é grande e p, a probabilidade de sucesso numa prova individual, está próxima de $\frac{1}{2}$. A Figura 9.21 mostra os histogramas das distribuições binomiais com $p = \frac{1}{2}$ e $n = 2, 5, 10$ e 25, e pode ser observado que, com n crescente, essas distribuições tendem para o padrão simétrico em forma de sino da distribuição normal. Na verdade, as distribuições normais com a média $\mu = np$ e o desvio-padrão $\sigma = \sqrt{np(1-p)}$ podem, muitas vezes, ser usadas para aproximar probabilidades binomiais, quando n não é tão grande e p é bastante diferente de $\frac{1}{2}$. Como "não é tão grande" e "é bastante diferente" não são expressões muito precisas, vamos enunciar a seguinte regra empírica:

É considerada prática segura utilizar a aproximação normal da distribuição binomial somente quando np e $n(1-p)$ forem ambos maiores do que 5; simbolicamente, quando

$$np > 5 \text{ e } n(1-p) > 5$$

A Figura 9.21 mostra claramente como o formato da distribuição binomial tende ao formato de uma distribuição normal e chama a atenção para o fato de que estamos aproximando a distribuição de uma variável aleatória discreta (de fato, uma variável aleatória finita) com a distribuição de uma variável que é contínua. Isso foi antecipado na solução do Exemplo 9.8, em que introduzimos a correção de continuidade de espalhar cada número inteiro k ao longo do intervalo contínuo de $k - \frac{1}{2}$ até $k + \frac{1}{2}$.

EXEMPLO 9.10 Use a distribuição normal para aproximar a probabilidade binomial de obter 6 caras e 10 coroas em 16 lançamentos de uma moeda equilibrada, e compare o resultado com o valor correspondente na Tabela V.

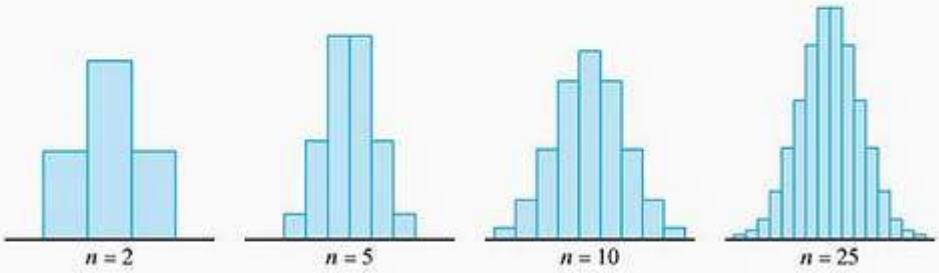

Figura 9.21
Distribuição binomial com $p = \frac{1}{2}$.

Snow666

Solução Com $n = 16$ e $p = \frac{1}{2}$, verificamos que $np = 16 \cdot \frac{1}{2} = 8$ e $n(1-p) = 16(1-\frac{1}{2}) = 8$ são ambos maiores do que 5 e, portanto, que a aproximação normal está justificada. A aproximação normal à probabilidade de 6 caras e 10 coroas é dada pela área da região colorida sob a curva na Figura 9.22, com 6 caras representadas pelo intervalo de 5,5 até 6,5. Como $\mu = 16 \cdot \frac{1}{2} = 8$ e $\sigma = \sqrt{16 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 2$, obtemos

$$z = \frac{5{,}5 - 8}{2} = -1{,}25 \quad \text{e} \quad z = \frac{6{,}5 - 8}{2} = -0{,}75$$

em unidades padronizadas para $x = 5{,}5$ e $x = 6{,}5$. As entradas correspondentes a $z = 1{,}25$ e $z = 0{,}75$ na Tabela I são 0,3944 e 0,2734 e, portanto, obtemos 0,3944 – 0,2734 = 0,1210 para a aproximação normal da probabilidade binomial de obter 6 caras e 10 coroas em 16 lançamentos de uma moeda equilibrada. Como o valor correspondente na Tabela V é 0,122, segue que o **erro percentual** é $\frac{0{,}001}{0{,}122} \cdot 100\% = 0{,}82\%$.

A aproximação normal à distribuição binomial é de grande valia quando, sem ela, as probabilidades binomiais individuais teriam de ser calculadas para muitos valores de x. Isso ocorre no exemplo a seguir, em que teríamos que determinar e somar as probabilidades individuais para 9, 10, 11, . . ., 148, 149 e 150 sucessos em 150 provas, a menos que usássemos diretamente algum tipo de aproximação para obter a probabilidades de pelo menos 9 sucessos em 150 provas, digamos, em termos de alguma área de curva normal.

EXEMPLO 9.11 Suponhamos que 5% dos tijolos crus despachados por um fabricante apresentem ligeiros defeitos. Use a aproximação normal da distribuição binomial para aproximar a probabilidade de que dentre 150 tijolos crus despachados pelo fabricante, pelo menos nove apresentem ligeiros defeitos. Também utilize o impresso de computador da Figura 8.4 à página 202 para calcular o erro e o erro percentual desta aproximação.

Solução Como 150(0,05) = 7,5 e 150(1 – 0,05) = 142,5 são ambos maiores do que 5, está satisfeita a regra empírica que autoriza o uso da aproximação normal à distribuição binomial. De acordo com a correção de continuidade explicada à página 230, representamos 9 dos tijolos pelo intervalo de 8,5 a 9,5; assim, devemos determinar a região colorida sob a curva na Figura 9.23. Como $\mu = 150(0{,}05) = 7{,}5$ e $\sigma = \sqrt{150(0{,}05)(0{,}95)} \approx 2{,}67$, obtemos

$$z = \frac{8{,}5 - 7{,}5}{2{,}67} \approx 0{,}37$$

em unidades padronizadas para $x = 8{,}5$. A entrada correspondente na Tabela I é 0,1443 e, portanto, obtemos 0,5000 – 0,1443 = 0,3557 para a aproximação normal da probabilidade de que pelo

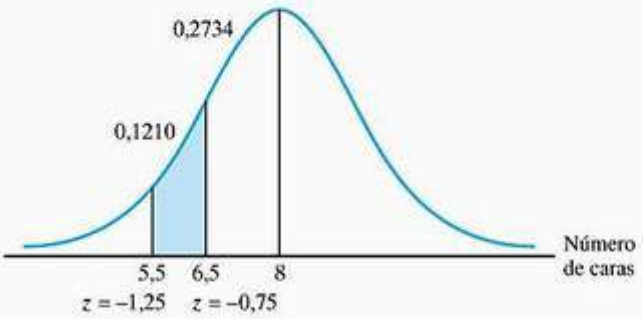

Figura 9.22 Diagrama para o Exemplo 9.10.

Snow666

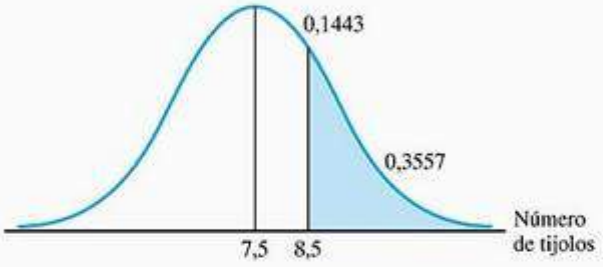

Figura 9.23
Diagrama para o Exemplo 9.11.

menos 9 dos tijolos crus apresentarão ligeiros defeitos. Como o impresso da Figura 8.4 mostra que a probabilidade binomial correspondente é 0,1171 + 0,0869 + 0,0582 + · · · + 0,0001 = 0,3361, verificamos que o erro da nossa aproximação é 0,3557 − 0,3361 = 0,0196 ou, aproximadamente, 0,02. O erro percentual correspondente é $\frac{0,0196}{0,3361}$ 5,8%.

O erro percentual obtido nesse exemplo é um pouco substancial e serve para alertar que satisfazer a regra empírica da página 231 não necessariamente garante que obtenhamos uma boa aproximação. Por exemplo, se p é muito pequeno e se quisermos aproximar a probabilidade de um valor que não está muito próximo da média, a aproximação poderá ser bastante pobre, mesmo se a regra empírica da página 231 tiver sido satisfeita. Conforme o leitor é convidado a verificar no Exercício 9.53, se quiséssemos aproximar a probabilidade de haver apenas um tijolo com um ligeiro defeito dentre os 150 despachados pelo fabricante, o erro percentual teria sido maior do que 100%. Como não estamos esperando que o leitor se transforme instantaneamente num perito do uso da aproximação normal à distribuição binomial, adicionamos que essa aproximação foi apresentada aqui principalmente porque ela será necessária em capítulos subseqüentes para inferências relativas a proporções de grandes amostras.

EXERCÍCIOS

9.30 O tempo necessário para montar uma estante "fácil de montar" de uma fábrica de móveis é uma variável aleatória de distribuição normal com $\mu = 17,40$ minutos e $\sigma = 2,20$ minutos. Qual é a probabilidades de que esse tipo de estante possa ser montada por uma pessoa em

(a) menos do que 19,0 minutos;
(b) em algum tempo entre 12,0 e 15,0 minutos?

9.31 Com referência ao Exercício 9.30, para qual intervalo de tempo a probabilidade de que uma tal montagem possa ser concluída é de 0,20?

9.32 A redução do consumo de oxigênio de uma pessoa durante períodos de meditação transcendental pode ser considerada como uma variável aleatória de distribuição normal com $\mu = 38,6$ cm^3 por minuto e $\sigma = 6,5$ cm^3 por minuto. Encontre as probabilidades de que, durante um período de meditação transcendental, o consumo de oxigênio de uma pessoa sofra uma redução de

(a) pelo menos 33,4 cm^3 por minuto;
(b) no máximo 34,7 cm^3 por minuto.

9.33 As uvas dos cachos de certa cepa criados num vinhedo pesam em média 18,2 gramas com desvio-padrão de 1,2 gramas. Supondo que a distribuição do peso dessas uvas tem aproximadamente o formato de uma distribuição normal, qual é a percentagem das uvas que pesa

(a) pelo menos 16,1 gramas;
(b) mais do que 17,3 gramas;
(c) algum peso entre 16,7 e 18,8 gramas?

Snow666

9.34 Com referência ao Exercício 9.33, encontre o peso acima do qual podemos esperar que estejam 80% das uvas mais pesadas?

9.35 Um montador necessita de molas helicoidais que suportem uma carga de pelo menos 20,0 kgf. Dentre dois fornecedores, o Fornecedor A pode garantir molas que, em média, suportam uma carga de 24,5 kgf com desvio-padrão de 2,1 kgf, enquanto que o Fornecedor B pode garantir molas que, em média, suportam uma carga de 23,3 kgf com desvio-padrão de 1,6 kgf. Supondo que as distribuições dessas cargas possam ser aproximadas com distribuições normais, determine qual dos dois fornecedores pode garantir ao montador a menor percentagem de molas insatisfatórias.

9.36 Os comprimentos de zangões adultos de uma certa variedade têm uma média de 1,96 cm com desvio-padrão de 0,08 cm. Supondo que a distribuição desses comprimentos tenha aproximadamente o formato de uma distribuição normal, encontre a percentagem desses zangões que têm pelo menos 2,0 cm de comprimento.

9.37 Com referência ao Exercício 9.36, encontre o comprimento acima do qual podemos esperar que estejam os 5% zangões mais compridos.

9.38 Em 1945, depois da Segunda Guerra Mundial, todos soldados dos EUA receberam uma pontuação baseada em tempo de serviço, número de condecorações, número de confrontos, etc. Supondo que a distribuição desses pontos possa ser muito bem aproximada por uma distribuição normal com $\mu = 63$ e $\sigma = 20$, quantos soldados de um exército de 8.000.000 seriam dispensados se fossem dispensados todos soldados com mais do que 79,0 pontos? (Os dados são cortesia do Departamento de Matemática da Academia Militar dos EUA.)

9.39 A distribuição dos QI de 4.000 funcionários de uma firma grande tem uma média de 104,5, desvio-padrão de 13,9 e seu formato é aproximadamente o de uma distribuição normal. Sabendo que um certo serviço requer um QI mínimo de 95 e entedia os de QI acima de 110, quantos dos funcionários dessa firma poderiam ser indicados para fazer esse serviço se o único critério fosse o QI? (Use correção de continuidade.)

9.40 O número de reclamações diárias numa loja de departamentos tem uma distribuição aproximadamente normal com $\mu = 25{,}8$. Também, as chances são de 4 para 1 de que num dia qualquer ocorram 18 reclamações.

(a) Qual é o desvio-padrão dessa distribuição normal?

(b) Qual é a probabilidade de que ocorram pelo menos 30 reclamações num dado dia?

9.41 O número anual mundial de terremotos é uma variável aleatória com uma distribuição aproximadamente normal de $\mu = 20{,}8$. Qual é o desvio-padrão dessa distribuição se a probabilidade de que ocorram pelo menos 18 terremotos é de 0,70?

9.42 Sabe-se que o número de ligações externas para fora de um escritório entre as 15 e as 17 horas é uma variável aleatória (ou muito próximo de uma) com uma média de 338. Se a probabilidade de que ocorram mais do que 400 ligações é de 0,06, qual é a probabilidade de que ocorram menos do que 300 ligações?

9.43 Uma variável aleatória tem uma distribuição normal com $\sigma = 4{,}0$. Se a probabilidade de que essa variável tome um valor menor do que 82,6 é de 0,9713, qual é a probabilidade de essa variável vá tomar um valor entre 70,0 e 80,0?

9.44 Um aluno pretende responder um teste de questões do tipo verdadeiro ou falso usando o resultado de lançamentos de uma moeda. Qual é a probabilidade de que o aluno acerte pelo menos 12 de $n = 20$ questões usando

(a) a aproximação normal à distribuição binomial com $n = 20$ e $p = 0{,}50$;

(b) a Tabela V?

Snow666

9.45 Com referência ao Exercício 9.44, encontre a percentagem de erro da aproximação.

9.46 Verifique, em cada caso, se estão satisfeitas as condições para a aproximação normal da distribuição binomial:

(a) $n = 32$ e $p = \frac{1}{7}$;
(b) $n = 75$ e $p = 0{,}10$;
(c) $n = 50$ e $p = 0{,}08$.

9.47 Verifique, em cada caso, se estão satisfeitas as condições para a aproximação normal da distribuição binomial:

(a) $n = 150$ e $p = 0{,}05$;
(b) $n = 60$ e $p = 0{,}92$;
(c) $n = 120$ e $p = \frac{1}{20}$.

9.48 Os registros mostram que 80% dos fregueses de um restaurante pagam com cartão de crédito. Use a aproximação normal da distribuição binomial para encontrar a probabilidade de que pelo menos 170 dentre 200 fregueses do restaurante paguem com cartão de crédito.

9.49 A probabilidade de que uma nuvem borrifada com iodeto de prata deixe de apresentar um crescimento espetacular é de 0,26. Use a aproximação normal da distribuição binomial para encontrar a probabilidade de que entre 30 nuvens borrifadas com iodeto de prata

(a) no máximo 8 deixem de apresentar um crescimento espetacular;
(b) somente 8 deixem de apresentar um crescimento espetacular.

9.50 Com referência ao Exercício 9.49, use a tabela impressa por MINITAB mostrado na Figura 9.24 para encontrar o erro percentual em ambas partes daquele exercício.

9.51 Estudos mostram que 22% de todos pacientes que tomam um certo antibiótico ficam com dor de cabeça. Use a aproximação normal da distribuição binomial para encontrar a probabilidade de que entre 50 pacientes tomando este antibiótico

(a) pelo menos 10 vão ficar com dor de cabeça;
(b) no máximo 15 vão ficar com dor de cabeça.

Função Densidade de Probabilidade

Binomial with n = 30 and p = 0.260000

x	P(X = x)
0.00	0.0001
1.00	0.0013
2.00	0.0064
3.00	0.0210
4.00	0.0499
5.00	0.0911
6.00	0.1334
7.00	0.1606
8.00	0.1623
9.00	0.1394
10.00	0.1028
11.00	0.0657
12.00	0.0365
13.00	0.0178
14.00	0.0076
15.00	0.0028
16.00	0.0009

Figura 9.24 Distribuição binomial com $n = 30$ e $p = 0{,}26$.

Snow666

9.52 Com referência ao Exercício 9.51, use o impresso mostrado na Figura 9.25 para encontrar o erro percentual em ambas partes daquele exercício.

9.53 Use o impresso da Figura 8.4 para mostrar que se utilizarmos a aproximação normal da probabilidade binomial com $x = 1$, $n = 150$ e $p = 0{,}05$, então o erro percentual é de cerca de 117%.

9.54 Use a distribuição normal para aproximar a probabilidade de que pelo menos 55 dentre 90 pessoas que cruzam o Atlântico de avião venham a sentir os efeitos da diferença de fuso horário durante, pelo menos, 24 horas, se é de 0,70 a probabilidade de que uma pessoa qualquer cruzando o Atlântico num avião vá sentir o efeito da diferença de fuso horário durante, pelo menos, 24 horas.

9.55 Use uma calculadora gráfica ou os dados fornecidos por computador da distribuição binomial com $n = 90$ e $p = 0{,}70$ para encontrar o erro percentual da aproximação do Exercício 9.54.

Função Distribuição Cumulativa

Binomial with n = 50 and p = 0.220000

x	P(X <= x)
5.00	0.0233
6.00	0.0555
7.00	0.1126
8.00	0.1991
9.00	0.3130
10.00	0.4448
11.00	0.5799
12.00	0.7037
13.00	0.8058
14.00	0.8819
15.00	0.9335
16.00	0.9653
17.00	0.9832
18.00	0.9925
19.00	0.9969
20.00	0.9988

Figura 9.25 Distribuição binomial com $n = 50$ e $p = 0{,}22$

9.6 LISTA DE TERMOS-CHAVE (com indicação das páginas de suas definições)

Amostragem de uma população normal, 230
Aproximação normal da distribuição binomial, 231
Área sob a curva normal, 218
Correção de continuidade, 229
Densidade de probabilidade, 215
Distribuição contínua, 214,215
Distribuição exponencial, 225
Distribuição normal, 214, 217
Distribuição normal padrão, 218
Distribuição uniforme, 222
Erro percentual, 232
Escores padronizados, 218
Escores *z*, 218
*Gráfico de probabilidade normal, 226
*Papel de probabilidade aritmética, 226
*Papel de probabilidade normal, 226
População normal, 226
Unidades padronizadas, 218
Variável aleatória contínua, 214

Snow666

9.7 REFERÊNCIAS

Informação detalhada sobre várias distribuições contínuas pode ser encontrada em

HASTINGS, N. A. J., and PEACOCK, J. B., *Statistical Distributions*. London: Butterworth & Company (Publishers) Ltd., 1975.

e informações adicionais sobre a aproximação normal da distribuição binomial em

GREEN, J., and ROUND-TURNER, J., "The Error in Approximating Cumulative Binomial and Poisson Probabilities," *Teaching Statistics*, May 1986.

Tabelas mais extensas de áreas sob curvas normais, bem como tabelas de outras distribuições contínuas, podem ser encontradas em

FISHER, R. A., and YATES, F., *Statistical Tables for Biological, Agricultural and Medical Research*. Cambridge: The University Press, 1954.

PEARSON, E. S., and HARTLEY, H. O., *Biometrika Tables for Statisticians*, Vol. 1, New York, John Wiley & Sons, Inc. 1968.

Snow666

10

AMOSTRAGEM E DISTRIBUIÇÕES AMOSTRAIS

10.1 Amostragem Aleatória 239
***10.2** Planejamento de Amostras 245
***10.3** Amostragem Sistemática 245
***10.4** Amostragem Estratificada 245
***10.5** Amostragem por Conglomerado 248
10.6 Distribuições Amostrais 250
10.7 O Erro Padrão da Média 253
10.8 O Teorema do Limite Central 255
10.9 Algumas Considerações Adicionais 257
***10.10** Nota Técnica (Simulação) 260
10.11 Lista de Termos-Chave 262
10.12 Referências 262

A maior parte do trabalho em Estatística tem por objetivo fazer generalizações sensatas, baseadas em amostras, sobre as populações das quais provêm as amostras. Repare na palavra "sensata", pois não é uma questão fácil de responder quando e sob quais circunstâncias as amostras permitem tais generalizações. Por exemplo, se quiséssemos estimar o valor médio dos gastos de uma pessoa durante suas férias, tomaríamos como amostra os valores gastos por passageiros de primeira classe de aviões de longo percurso; tentaríamos estimar o preço médio de produtos hortigranjeiros baseados somente no preço de aspargos frescos; ou tentaríamos prever as temperaturas de janeiro no Recife baseando-nos nas temperaturas de janeiro na Antártica? Obviamente que não, mas exatamente quais pessoas de férias, quais produtos hortigranjeiros, e quais temperaturas de janeiro deveríamos incluir em nossas amostras não é intuitivamente claro nem, de modo algum, evidente.

Na maioria dos métodos estudados no restante deste livro, supomos que as amostras com as quais lidamos são as que se denominam **amostras aleatórias**. Damos tanta ênfase às amostras aleatórias, definidas e estudadas na Seção 10.1 porque elas permitem generalizações válidas, ou lógicas. Como veremos, entretanto, a amostragem aleatória nem sempre é viável, ou sequer desejável, e alguns procedimentos alternativos de amostragem são mencionados nas Seções 10.2 até 10.5.

Na Seção 10.6 introduzimos o conceito relacionado de **distribuição amostral**, que nos diz como as quantidades determinadas de amostras variam de uma para outra amostra. Finalmente, nas Seções 10.7 até 10.9, vemos como tais variações aleatórias podem ser medidas, previstas e talvez até controladas.

Snow666

10.1 AMOSTRAGEM ALEATÓRIA

Na Seção 3.1, fizemos uma distinção entre populações e amostras, afirmando que uma população consiste em todas as observações concebivelmente possíveis (ou hipoteticamente possíveis) de um determinado fenômeno, enquanto uma amostra é simplesmente parte de uma população. No que segue, distinguiremos ainda entre dois tipos de populações, as **populações finitas** e as **populações infinitas**.

Uma população finita consiste em um número finito, ou fixo, de elementos, medidas ou observações. Exemplos de populações finitas são os pesos líquidos das 3.000 latas de tinta de um certo lote de produção, os escores obtidos por todos os candidatos no vestibular de inverno de 2004 de uma certa universidade, as 52 cartas diferentes num baralho comum e o registro diário das temperaturas máximas num posto meteorológico durante os anos de 1995 a 1999.

Por outro lado, uma população é infinita se contém, pelo menos hipoteticamente, uma quantidade infinita de elementos. Esse seria o caso, por exemplo, quando observamos os valores de uma variável aleatória contínua, digamos, quando medimos repetidamente o ponto de ebulição de um composto de silicone. Também é o caso quando observamos os totais obtidos em repetidas jogadas de um par de dados e quando tomamos amostras com reposição de uma população finita. Não há limite para o número de vezes que podemos medir o ponto de ebulição de um composto de silicone, não há limite para o número de vezes que podemos jogar um par de dados, e não há limite para o número de vezes que podemos extrair um elemento de uma população e repô-lo antes de extrair o seguinte.

Para apresentar a idéia de **amostragem aleatória de uma população finita**, vejamos primeiro quantas amostras diferentes de tamanho n podem ser extraídas de uma população finita de tamanho N. Recorrendo à regra à página 117 para encontrar o número de combinações de n objetos tomados r de cada vez, vemos que, com uma troca de letras, a resposta é $\binom{N}{n}$.

EXEMPLO 10.1 Quantas amostras de tamanho n diferentes podem ser extraídas de uma população finita de tamanho N, se

(a) $n = 2$ e $N = 12$;

(b) $n = 3$ e $N = 50$?

Solução **(a)** Há $\binom{12}{2} = \frac{12 \cdot 11}{2} = 66$ amostras diferentes.

(b) Há $\binom{50}{3} = \frac{50 \cdot 49 \cdot 48}{3!} = 19.600$ amostras diferentes.

Com base no resultado de que há $\binom{N}{n}$ amostras de tamanho n diferentes numa população finita de tamanho N, vamos dar a seguinte definição de uma **amostra aleatória** (às vezes também denominada **amostra aleatória simples**) de uma população finita:

Uma amostra de tamanho n extraída de uma população finita de tamanho N é aleatória se é escolhida de tal modo que cada uma das $\binom{N}{n}$ amostras possíveis tem a mesma probabilidade $\frac{1}{\binom{N}{n}}$ de serem escolhida.

Snow666

Por exemplo, se uma população consiste nos $N = 5$ elementos a, b, c, d e e (que podem ser as rendas anuais de cinco pessoas, os pesos de cinco vacas, ou cinco modelos diferentes de aviões), há $\binom{5}{3} = 10$ amostras possíveis de tamanho $n = 3$. Essas amostrar consistem nos elementos abc, abd, abe, acd, ace, ade, bcd, bce, bde e cde. Se escolhermos uma dessas amostras de tal maneira que cada uma tenha a probabilidade $\frac{1}{10}$ de ser escolhida, dizemos que essa amostra é uma amostra aleatória.

Em seguida vem o problema de como as amostras aleatórias são efetivamente extraídas na prática. Numa situação simples como a que acabamos de escrever, poderíamos escrever cada uma das 10 amostras num pedacinho de papel, colocá-los num chapéu, misturá-los totalmente e então tirar um sem olhar. No entanto, obviamente isso seria impraticável numa situação real mais complexa, em que n e N, ou apenas N, são grandes. Por exemplo, para $n = 4$ e $N = 100$, teríamos de escrever e extrair um dentre $\binom{100}{4} = 3.921.225$ pedacinhos de papel.

Felizmente, é possível extrair uma amostra aleatória de uma população finita sem relacionar todas as amostras possíveis, e mencionamos isso apenas para enfatizar que a escolha de uma amostra aleatória deve depender completamente do acaso. Em vez de listar todas as amostras possíveis, podemos escrever cada um dos N elementos da população finita num pedacinho de papel, e extrair n deles, um de cada vez, sem reposição, certificando-nos de que, em cada uma das extrações sucessivas, todos os elementos restantes da população tenham a mesma chance de serem escolhidos. É fácil de mostrar matematicamente que isso também leva à probabilidade $\frac{1}{\binom{N}{n}}$ para cada amostra possível. Por exemplo, para extrair uma amostra aleatória de $n = 12$ de $N = 138$ locais de escavação arqueológica, poderíamos escrever os números 1, 2, 3, . . ., 137 e 138 em 138 pedacinhos de papel, misturá-los completamente (digamos, no chapéu proverbial) e então extrair 12, um de cada vez, sem reposição, sem espiar.

Mesmo um procedimento simples como esse pode ser ainda mais simplificado. Hoje em dia, a maneira mais simples de tomar uma amostra aleatória de uma população finita é basear a seleção em **números aleatórios** gerados por meio de calculadoras estatísticas ou computadores. Por exemplo, a calculadora HP 21S STAT/MATH fornece números aleatórios de quatro dígitos de tal modo que cada número de quatro dígitos entre 0000 e 9999 tem a probabilidade de 0,0001. Para evitar a possibilidade de obter os mesmos números aleatórios em aplicações diferentes, podemos dar a "semente" para a calculadora, digitando um número de quatro dígitos arbitrário com o qual começar a operação.

EXEMPLO 10.2 Com referência aos locais de escavação arqueológica mencionados anteriormente, que supomos terem sido numerados de 001 a 138, use uma calculadora estatística para gerar uma amostra aleatória de $n = 12$ desses locais.

Solução Usando somente os três primeiros dígitos dos números aleatórios de quatro dígitos gerados, omitindo 000 e os maiores do que 138, e também omitindo qualquer número que já tenha sido selecionado, obtemos 041, 021, 079, 084, 016, 108, 029, 003, 100, 046, 136 e 075. Os locais de escavação arqueológica associados com esses números constituem uma amostra aleatória.

Figura 10.1 Amostra gerada com computador para o Exemplo 10.2.

Inteiros Aleatórios 1-138

	C1	C2	C3	C4	C5	C6
1	24	131	69	113	127	5
2	57	52	7	13	11	64

Snow666

Se quiséssemos usar um computador, em vez de uma calculadora estatística, no Exemplo 10.2, um aplicativo como MINITAB poderia ter fornecido um resultado como o mostrado na Figura 10.1. O que mostramos nessa figura consiste nos locais de escavação arqueológica associados com os números 24, 131, 69, 113, 127, 5, 57, 52, 7, 13, 11 e 64. (Os outros números na figura são simplesmente as linhas e colunas da planilha em que o aplicativo colocou os dados.)

Há poucas décadas, as amostras aleatórias eram geradas quase que exclusivamente usando números aleatórios publicados em tabelas. Tais tabelas consistiam em páginas e mais páginas de nada mais do que os dígitos 0, 1, 2, 3, . . ., 8 e 9 arranjados em linhas e colunas. Essas tabelas eram construídas usando dados de censos, expansões decimais de números irracionais, tabelas de logaritmos, resultados de loterias e outras coleções de números. Havia até alguns procedimentos aritméticos, como o em que se começava com um número de três ou quatro dígitos, se elevava o número ao quadrado, e então se tomavam os três ou quatro dígitos do meio para começar a seqüência de dígitos aleatórios. Então, repetidamente se elevava esse número ao quadrado e se repetia o processo. Nos últimos anos esses métodos foram substituídos por métodos computacionais que geravam tabelas de números aleatórios e até esses métodos agora foram substituídos por métodos como o que utilizamos no Exemplo 10.2, em que geramos nossos próprios números aleatórios.

Mesmo assim, como a tecnologia necessária pode não estar sempre à mão, ilustramos aqui como utilizar uma tabela publicada de números aleatórios. Para isso, vamos utilizar a Figura 10.2, que consiste em uma página reproduzida do livro intitulado *Tables of 105,000 Random Decimal Digits*, publicado em 1949 pela Interstate Commerce Comission do Bureau of Transport Economics and Statistics, em Washington, D.C.

EXEMPLO 10.3 Repita o Exemplo 10.2, lendo os números de três dígitos da Figura 10.2, com as mesmas restrições que as dadas na solução do Exemplo 10.2. Use as colunas de número 11, 12 e 13, começando na linha de número 6 e descendo pela página. Se necessário, continue com as colunas 16, 17 e 18, também começando com a linha 6 e descendo pela página.

Solução Novamente omitindo o número 000 e todos números maiores do que 138 e tomando cuidado para não escolher valor algum mais de uma vez, pode ser facilmente verificado que os números aleatórios que obtemos para a amostra são 007, 012, 031, 135, 114, 120, 047, 124, 070, 009, 118 e 094. Assim, a amostra consiste nos locais de escavação arqueológica de números 7, 12, 31, 135, 114, 120, 47, 124, 79, 9, 118 e 94.

Quando os itens estão ordenados segundo listas numeradas, é fácil extrair amostras aleatórias utilizando calculadoras, computadores ou tabelas de números aleatórios. Infelizmente, contudo, há muitas situações em que é impossível proceder da maneira que acabamos de descrever. Por exemplo, se quisermos utilizar uma amostra para estimar a média do diâmetro externo de milhares de mancais embalados num grande engradado, ou se quisermos estimar a altura média das árvores de uma floresta, seria impossível numerar os mancais ou as árvores, escolher números aleatórios e então localizar e medir os mancais ou árvores correspondentes. Nessas e em muitas outras situações análogas, tudo que podemos fazer é proceder de acordo com a definição do dicionário da palavra "aleatório", a saber, "casualmente, sem objetivo ou propósito". Isto é, não devemos selecionar ou rejeitar qualquer elemento de uma população em razão de sua aparência ser típica ou não, nem tampouco devemos favorecer ou desprezar qualquer parcela de uma população em razão de ser acessível ou não, e assim por diante. Com algumas reservas, tais amostras são comumente consideradas como se fossem, na realidade, amostras aleatórias.

Até aqui abordamos a amostragem aleatória apenas em relação a populações finitas. Para populações infinitas, dizemos que

Uma amostra de tamanho n de uma população infinita é aleatória se consiste em valores de variáveis aleatórias independentes que têm a mesma distribuição.

Snow666

94620	27963	96478	21559	19246	88097	44926
60947	60775	73181	43264	56895	04232	59604
27499	53523	63110	57106	20865	91683	80688
01603	23156	89223	43429	95353	44662	59433
00815	01552	06392	31437	70385	45863	75971
83844	90942	74857	52419	68723	47830	63010
06626	10042	93629	37609	57215	08409	81906
56760	63348	24949	11859	29793	37457	59377
64416	29934	00755	09418	14230	62887	92683
63569	17906	38076	32135	19096	96970	75917
22693	35089	72994	04252	23791	60249	83010
43413	59744	01275	71326	91382	45114	20245
09224	78530	50566	49965	04851	18280	14039
67625	34683	03142	74733	63558	09665	22610
86874	12549	98699	54952	91579	26023	81076
54548	49505	62515	63903	13193	33905	66936
73236	66167	49728	03581	40699	10396	81827
15220	66319	13543	14071	59148	95154	72852
16151	08029	36954	03891	38313	34016	18671
43635	84249	88984	80993	55431	90793	62603
30193	42776	85611	57635	51362	79907	77364
37430	45246	11400	20986	43996	73122	88474
88312	93047	12088	86937	70794	01041	74867
98995	58159	04700	90443	13168	31553	67891
51734	20849	70198	67906	00880	82899	66065
88698	41755	56216	66852	17748	04963	54859
51865	09836	73966	65711	41699	11732	17173
40300	08852	27528	84648	79589	95295	72895
02760	28625	70476	76410	32988	10194	94917
78450	26245	91763	73117	33047	03577	62599
50252	56911	62693	73817	98693	18728	94741
07929	66728	47761	81472	44806	15592	71357
09030	39605	87507	85446	51257	89555	75520
56670	88445	85799	76200	21795	38894	58070
48140	13583	94911	13318	64741	64336	95103
36764	86132	12463	28385	94242	32063	45233
14351	71381	28133	68269	65145	28152	39087
81276	00835	63835	87174	42446	08882	27067
55524	86088	00069	59254	24654	77371	26409
78852	65889	32719	13758	23937	90740	16866
11861	69032	51915	23510	32050	52052	24004
67699	01009	07050	73324	06732	27510	33761
50064	39500	17450	18030	63124	48061	59412
93126	17700	94400	76075	08317	27324	72723
01657	92602	41043	05686	15650	29970	95877
13800	76690	75133	60456	28491	03845	11507
98135	42870	48578	29036	69876	86563	61729
08313	99293	00990	13595	77457	79969	11339
90974	83965	62732	85161	54330	22406	86253
33273	61993	88407	69399	17301	70975	99129

Figura 10.2
Exemplo de página com dígitos aleatórios.

Como salientamos em relação às distribuições binomial e normal, é a essa "mesma" distribuição que nos referimos como a população da qual estamos extraindo uma amostra. Além disso, por "independente" queremos dizer que a probabilidade relativa a qualquer uma das variáveis aleatórias é a mesma, independentemente de quais valores podem ter sido observados para as outras variáveis aleatórias.

Snow666

Por exemplo, se obtemos 2, 5, 1, 3, 6, 4, 4, 5, 2, 4, 1 e 2 em doze jogadas de um dado, esses números constituem uma amostra aleatória se são valores de variáveis aleatórias independentes com a mesma distribuição de probabilidade

$$f(x) = \frac{1}{6} \quad \text{para} \quad x = 1, 2, 3, 4, 5 \text{ ou } 6$$

Como outro exemplo de amostra aleatória de uma população infinita, suponhamos que oito estudantes tenham obtido as seguintes medições do ponto de ebulição de um composto de silício: 136, 153, 170, 148, 157, 152, 143 e 150 graus Celsius. De acordo com a definição, esses valores constituem uma amostra aleatória se forem valores de variáveis aleatórias independentes com a mesma distribuição, digamos, a distribuição normal com $\mu = 152$ e $\sigma = 10$. Para julgar se esse é realmente o caso, teríamos de averiguar, entre outras coisas, se as técnicas de medida adotadas pelos oito estudantes são igualmente precisas (de modo que σ seja o mesmo para cada uma das variáveis aleatórias) e se não há colaboração (que poderia tornar dependentes as variáveis aleatórias). Na prática, não é tarefa fácil julgar se um conjunto de dados pode ser encarado como uma amostra aleatória e voltaremos a esse assunto com maior detalhe no Capítulo 18.

EXERCÍCIOS

10.1 Quantas amostras diferentes de tamanho $n = 2$ podem ser extraídas de uma população finita de tamanho

(a) $N = 6$; (c) $N = 32$;
(b) $N = 20$; (d) $N = 75$?

10.2 Quantas amostras diferentes de tamanho $n = 3$ podem ser extraídas de uma população finita de tamanho

(a) $N = 8$; (c) $N = 30$;
(b) $N = 26$; (d) $N = 40$?

10.3 Qual é a probabilidade de cada amostra possível se extraímos uma amostra aleatória de tamanho $n = 4$ de uma população finita de tamanho

(a) $N = 12$;
(b) $N = 20$?

10.4 Qual é a probabilidade de cada amostra possível se extraímos uma amostra aleatória de tamanho $n = 6$ de uma população finita de tamanho

(a) $N = 10$;
(b) $N = 15$?

10.5 Liste as $\binom{6}{3} = 20$ amostras possíveis e tamanho $n = 3$ que podem ser extraídas de uma população finita cujos elementos são denotados por u, v, w, x, y e z.

10.6 Com referência ao Exercício 10.5, qual é a probabilidade de que uma amostra aleatória vá incluir os elementos denotados por u?

10.7 Com referência ao Exercício 10.5, qual é a probabilidade de que uma amostra aleatória vá incluir os elementos denotados por u e v?

10.8 Uma livraria de campus universitário mantém em estoque sete livros diferentes de história da arte, oferecendo um desconto de 10% na compra de quaisquer três deles. De quantas escolhas diferentes dispõe um cliente?

10.9 Uma pessoa planejando uma viagem aos EUA tem amigos em Los Angeles, San Francisco, Chicago, Nova York e Boston. Se ela escolhe três dessas cidades ao acaso, quais são as probabilidades

Snow666

(a) de cada uma das seleções possíveis;
(b) de que a seleção vá incluir San Francisco;
(c) de que a seleção vá incluir San Francisco e Chicago?

10.10 Uma organização quer incluir 6 dos 50 estados norte-americanos numa pesquisa de mercado. Se os estados fossem numerados 01, 02, 03,..., 49 e 50 em ordem alfabética e se a organização usasse as colunas 6 e 7 e descesse a página começando na linha 3 da tabela na Figura 10.2, quais estados seriam incluídos na pesquisa? Uma lista que associa números aos elementos de uma população para fins de obtenção de uma amostragem é denominada uma **estrutura amostral**. Para este exercício, tal lista pode ser obtida da lista de prefixos nacionais de um catálogo telefônico.

10.11 Uma hematologista deseja conferir novamente uma amostra de $n = 10$ dos 653 espécimes de sangue analisados em seu laboratório num determinado mês. Em seus registros, esses espécimes de sangue estão numerados 3250, 3251, 3252, . . ., 3901 e 3902. Quais espécimes ela selecionaria se usasse as colunas 21, 22 e 23 e descesse a página começando na linha 16 da tabela na Figura 10.2? (Como todos números começam com 3, esse dígito pode ser omitido na seleção da amostra.)

10.12 Use uma calculadora estatística, uma calculadora gráfica ou um computador para refazer o Exercício 10.11.

10.13 Ao longo de duas semanas, a seção de calçados de uma loja de departamentos registrou trezentas vendas, com respectivas notas fiscais numeradas de 251 a 550. Se um auditor quisesse conferir $n = 12$ dessas notas fiscais selecionadas ao acaso, quais ele conferiria se escolhesse as colunas 18, 19 e 20 e descesse pelas colunas começando na linha do alto da tabela na Figura 10.2?

10.14 Use uma calculadora estatística, uma calculadora gráfica ou um computador para refazer o Exercício 10.13.

10.15 Um assessor da prefeitura deseja reavaliar uma amostra aleatória de 50 das 7.964 residências unifamiliares do município e ele pede sua ajuda na seleção da amostra. Primeiro, você cria uma estrutura amostral atribuindo a essas residências os números 0001, 0002, 0003, . . ., 7963 e 7964, e então usa as quatro primeiras colunas da tabela na Figura 10.2, descendo pela página a partir da linha do alto, continuando com as próximas quatro colunas, também descendo pela página a partir da linha do alto. Em termos da numeração, quais residências seriam selecionadas dessa maneira?

10.16 Use um computador para refazer o Exercício 10.15.

10.17 Na página 240, dissemos que se pode extrair uma amostra aleatória de uma população finita sem listar todas as amostras possíveis; em vez disso, simplesmente numeramos (ou etiquetamos) os N elementos da população finita e então extraímos n deles, um de cada vez, sem reposição, certificando-nos de que, em cada uma das extrações sucessivas, todos os elementos restantes da população tenham a mesma probabilidade de serem escolhidos. Confira isso para o exemplo à página 240, em que tratamos de amostras aleatórias de tamanho $n = 3$ da população finita que consiste nos elementos a, b, c, d e e, mostrando que a probabilidade de que uma amostra particular seja escolhida por esse método (digamos, a amostra bce) é, como antes, $\frac{1}{10}$.

10.18 Com o mesmo tipo de argumento que o do Exercício 10.17, verifique que cada amostra aleatória possível de tamanho $n = 3$, extraída, um elemento de cada vez, de uma população finita de tamanho $N = 100$, tem a probabilidade $1 \Big/ \binom{100}{3} = \frac{1}{161.700}$.

Snow666

10.19 Com o mesmo tipo de argumento que o do Exercício 10.17, verifique que cada amostra aleatória possível de tamanho n, extraída, um elemento de cada vez, de uma população finita de tamanho N, tem a probabilidade $1 \Big/ \binom{N}{n}$.

*10.2 PLANEJAMENTO DE AMOSTRAS

O único tipo de amostras que estudamos até agora são as amostras aleatórias, e sequer consideramos a possibilidade de que, sob certas circunstâncias, possa haver amostras melhores (isto é, mais fáceis de serem obtidas, mais baratas ou mais informativas) do que as amostras aleatórias e tampouco entramos em detalhes sobre o que deve ser feito quando a amostragem aleatória for impossível. Na verdade, há muitas outras maneiras de extrair uma amostra de uma população, e existe uma extensa literatura dedicada ao assunto de planejamento de procedimentos de amostragem.

Em Estatística, um **planejamento de amostra** é um plano definido, completamente determinado antes da coleta de quaisquer dados, de obter uma amostra de uma dada população. Assim, o plano para extrair uma amostra aleatória simples de tamanho 12 das 247 farmácias de uma cidade, utilizando uma tabela de números aleatórios de uma maneira predeterminada, constitui um planejamento de amostra. Nas três seções seguintes, discutiremos alguns planejamentos de amostras que se aplicam especialmente a operações de larga escala, levantamentos e semelhantes, e costumamos denominar esse material **amostragem por levantamento**. A seção foi marcada como opcional não por que não é importante. Todo o assunto de amostragem por levantamento raramente é coberto em disciplinas introdutórias gerais de Estatística, simplesmente por que não há tempo suficiente no tipo de disciplina para a qual foi projetado este livro.

*10.3 AMOSTRAGEM SISTEMÁTICA

Em alguns casos, a maneira mais prática de se extrair uma amostra consiste em selecionar, digamos, cada 20º nome de uma lista, cada 12ª casa de um lado de uma rua, cada 50º peça que sai de uma linha de montagem, e assim por diante. Isso é o que se denomina **amostragem sistemática**, e um elemento de aleatoriedade pode ser introduzido nesse tipo de amostragem, utilizando-se números aleatórios para escolher a unidade com que se começa. Embora uma amostra sistemática possa não ser uma amostra aleatória de acordo com a definição, muitas vezes é razoável tratar as amostras sistemáticas como se fossem amostras aleatórias; na verdade, em alguns casos, as amostras sistemáticas de fato apresentam uma melhora em relação às amostras aleatórias simples, pois as amostras se dispersam mais uniformemente sobre toda a população.

O perigo real da amostragem sistemática reside na possível presença de periodicidades ocultas. Por exemplo, se inspecionássemos cada 40ª peça produzida por determinada máquina, os resultado seriam enganosos se, em virtude de um falha regularmente recorrente, toda 10ª peça produzida pela máquina apresentasse defeito. Também, a amostragem sistemática poderia produzir resultados tendenciosos, se entrevistássemos os moradores de cada 12ª casa ao longo de certa rua e acontecesse que toda 12ª casa ao longo dessa rua fosse justamente uma casa de esquina de dois lotes.

*10.4 AMOSTRAGEM ESTRATIFICADA

Se dispusermos de dados sobre a composição de uma população, e isto é relevante para nossa pesquisa, poderemos eventualmente melhorar uma amostragem aleatória mediante **estratificação**. Trata-se de um processo que consiste em estratificar (ou dividir) a população em um certo núme-

Snow666

ro de subpopulações que não se sobrepõem, chamadas **estratos** ou **camadas**, e então extrair uma amostra de cada estrato. Se os elementos selecionados em cada estrato constituírem amostras aleatórias simples, então o processo global – estratificação seguida de amostragem aleatória – é denominado **amostragem aleatória estratificada (simples)**.

Suponha, por exemplo, que queiramos estimar o peso médio de quatro pessoas com base em uma amostra de tamanho 2, e que os pesos (desconhecidos) das quatro pessoas sejam 115, 135, 185 e 205 libras (lembre que uma libra equivale a um pouco menos do que meio quilograma). Assim, o peso médio que queremos estimar é

$$\mu = \frac{115 + 135 + 185 + 205}{4} = 160 \text{ libras}$$

Se extrairmos uma amostra aleatória de tamanho 2 dessa população, as $\binom{4}{2} = 6$ amostras possíveis são 115 e 135, 115 e 185, 115 e 205, 135 e 185, 135 e 205 e 185 e 205, e as médias correspondentes são 125, 150, 160, 160, 170 e 195. Observe que, como cada uma dessas amostras tem probabilidade $\frac{1}{6}$, as probabilidades são $\frac{1}{3}$, $\frac{1}{3}$ e $\frac{1}{3}$ de que nosso erro (a diferença entre a média amostral e $\mu = 160$) seja 0, 10 ou 35.

Suponhamos, agora, que saibamos que duas dessas pessoas são homens e duas são mulheres, e suponha que os pesos (desconhecidos) dos homens sejam 185 e 205 libras, enquanto que os pesos (desconhecidos) das mulheres sejam 115 e 135 libras. Estratificando nossa amostra (por sexo) e escolhendo aleatoriamente um dos dois homens e uma das duas mulheres, vemos que há apenas quatro amostras estratificadas, 115 e 185, 115 e 205, 135 e 185 e 135 e 205. As médias dessas amostras são 150, 160, 160 e 170, e agora as probabilidades são de $\frac{1}{2}$ e $\frac{1}{2}$ de que nosso erro seja 0 ou 10. É claro que a estratificação melhorou enormemente nossa chance de obter uma boa (precisa) estimativa do peso médio das quatro pessoas. Veja, entretanto, o Exercício 10.24.

Essencialmente, o objetivo de estratificar é formar os estratos de tal modo que haja alguma relação entre o fato de estar em determinado estrato e a resposta procurada no estudo estatístico, e que, dentro de cada estrato, haja tanta homogeneidade (uniformidade) quanto possível. Em nosso exemplo existe essa conexão entre sexo e peso, e há muito menos variabilidade de peso dentro de cada um dos dois grupos do que dentro da população como um todo.

No exemplo precedente, utilizamos uma **alocação proporcional**, o que significa que os tamanhos das amostras dos diferentes estratos são proporcionais aos tamanhos dos estratos. Em geral, se dividirmos uma população de tamanho N em k estratos de tamanhos $N_1, N_2, \ldots,$ e N_k e extrairmos uma amostra de tamanho n_1 do primeiro estrato, uma amostra de tamanho n_2 do segundo estrato, . . ., e uma amostra de tamanho n_K do k-ésimo estrato, dizemos que a alocação é proporcional se

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_k}{N_k}$$

ou se essas razões são tão aproximadamente iguais quanto possível. No exemplo relativo aos pesos, tínhamos $N_1 = 2$, $N_2 = 2$, $n_1 = 1$ e $n_2 = 1$, de forma que

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{1}{2}$$

e a alocação foi, de fato, proporcional.

Como pode ser facilmente verificado, uma alocação é proporcional se

TAMANHOS DE AMOSTRA PARA ALOCAÇÃO PROPORCIONAL

$$n_i = \frac{N_i}{N} \cdot n \quad \textit{para } i = 1, 2, \ldots, \textit{e } k$$

Snow666

onde $n = n_1 + n_2 + \cdots + n_k$ é o tamanho total da amostra. Quando necessário, utilizamos os inteiros mais próximos dos valores dados por essa fórmula.

EXEMPLO 10.4 Deve-se extrair uma amostra estratificada de tamanho $n = 60$ de uma população de tamanho $N = 4.000$, que consista em três estratos de tamanhos $N_1 = 2.000$, $N_2 = 1.200$ e $N_3 = 800$. Para que a alocação seja proporcional, qual deve ser o tamanho da amostra a ser extraída de cada estrato?

Solução Substituindo na fórmula, obtemos

$$n_1 = \frac{2.000}{4.000} \cdot 60 = 30 \qquad n_2 = \frac{1.200}{4.000} \cdot 60 = 18$$

e

$$n_3 = \frac{800}{4.000} \cdot 60 = 12$$

Este exemplo ilustra a alocação proporcional, mas é preciso acrescentar que existem outras maneiras de alocar porções de uma amostra a estratos diferentes. Uma delas, denominada **alocação ótima**, é descrita no Exercício 10.29. Ela leva em conta não só o tamanho dos estratos, como na alocação proporcional, mas também a variabilidade (de qualquer que seja a característica de interesse) dentro dos estratos.

Além disso, estratificar não está limitado a uma única variável de classificação, ou característica e, muitas vezes, as populações são estratificadas de acordo com várias características. Por exemplo, numa pesquisa no âmbito de todo sistema escolar, projetada para determinar a atitude dos estudantes em relação a, digamos, um novo plano de pagamentos, um sistema estadual de 17 faculdades poderia estratificar sua amostra não só em relação a essas faculdades, mas também em relação aos conceitos dos alunos por turma, ao seu sexo e ao seu curso. Assim, parte da amostra seria alocada às mulheres calouras na faculdade A cursando Engenharia, outra parte aos alunos homens do segundo ano da faculdade L fazendo curso de Letras, e assim por diante. Até certo ponto, uma estratificação como essa, denominada **estratificação cruzada**, aumenta a precisão (confiabilidade) das estimativas e outras generalizações, e é utilizada amplamente, particularmente em amostragem de opinião e pesquisa de mercado.

Na amostragem estratificada, o custo da extração de amostras aleatórias dos estratos individuais é, muitas vezes, tão elevado que os entrevistadores simplesmente recebem quotas a serem preenchidas dos diferentes estratos, com poucas restrições (se houver alguma) sobre como devem ser preenchidas. Por exemplo, na determinação da atitude dos eleitores em relação a uma ampliação da assistência média a pessoas idosas, um entrevistador que trabalha em determinada área pode ser instruído a entrevistar 6 homens com menos de 30 anos de idade que trabalhem por conta própria e que tenham casa própria, 10 mulheres assalariadas na faixa etária dos 45 aos 60 anos que vivem em apartamentos, 3 homens aposentados de mais de 60 anos que vivem casas alugadas, e assim por diante, deixando a critério do entrevistador a efetiva escolha dos indivíduos. Isto é o que se denomina **amostragem por quotas**, e é um processo conveniente, relativamente barato e, por vezes, necessário, mas da maneira pela qual é muitas vezes posto em prática, as amostras resultantes não têm as características essenciais de amostras aleatórias. Na ausência de qualquer controle sobre suas escolhas, os entrevistadores tendem naturalmente a selecionar indivíduos mais imediatamente disponíveis, como pessoas que trabalham no mesmo edifício, fazem compras no mesmo supermercado, ou talvez residam na mesma grande área. As amostras por quotas são, portanto, essencialmente **amostras de julgamento**, e as inferências baseadas em tais amostras geralmente não se prestam a qualquer tipo de avaliação estatística formal.

Snow666

*10.5 AMOSTRAGEM POR CONGLOMERADO

Para ilustrar um outro tipo importante de amostragem, suponhamos que uma grande fundação queira estudar a variação dos padrões de despesas de famílias na área de Salvador. Ao tentar completar um cronograma para entrevistar 1.200 famílias, a fundação constata que a amostragem aleatória simples é praticamente impossível, pois não há disponibilidade de listas adequadas e o custo de contatar as famílias espalhadas numa área grande (com possíveis retornos devido a não encontrar ninguém em casa) é muito elevado. Uma forma de extrair uma amostra nessa situação consiste em dividir a área total de interesse em diversas áreas menores, sem sobreposição, digamos, em quarteirões. Selecionam-se, então, aleatoriamente alguns desses quarteirões, com a amostra final consistindo em todas as famílias (ou uma amostra delas) residentes nesses quarteirões.

Nesse tipo de amostragem, denominado **amostragem por conglomerado**, a população total é dividida em várias subdivisões relativamente pequenas, e algumas dessas subdivisões, ou conglomerados, são selecionadas aleatoriamente para integrarem a amostra global. Se os conglomerados forem subdivisões geográficas, como no exemplo precedente, esse tipo de amostragem é também denominado **amostragem por área**. Para dar um outro exemplo de amostragem por conglomerado, suponhamos que o pró-reitor de graduação de uma universidade deseje saber o que os alunos que freqüentam diretórios acadêmicos pensam sobre um novo regulamento. Ele pode obter uma amostra por conglomerado entrevistando alguns ou todos os alunos de vários diretórios acadêmicos escolhidos aleatoriamente.

Embora as estimativas baseadas em amostras por conglomerado em geral não sejam tão confiáveis quanto as baseadas em amostras aleatórias simples de mesmo tamanho (ver Exercício 10.28), muitas vezes elas são mais confiáveis por custo unitário. Voltando à pesquisa das despesas familiares na área de Salvador, é fácil ver que é possível extrair, pelo mesmo custo, uma amostra por conglomerado de tamanho correspondente a várias amostras aleatórias simples. É muito mais barato visitar e entrevistar famílias que vivem próximas em conglomerados do que famílias selecionadas aleatoriamente numa grande área.

Na prática, num mesmo estudo podemos perfeitamente usar vários dos métodos de amostragem estudados. Por exemplo, se os estatísticos do governo quiserem estudar a atitude dos professores do ensino fundamental em relação a certos programas federais, podem começar estratificando o país por estados ou alguma outra subdivisão geográfica. Para extrair uma amostra de cada estrato, eles podem, então, usar a amostragem por conglomerado, subdividindo cada estrato em várias partes geográficas menores (digamos, distritos escolares) e, finalmente, utilizar a amostragem aleatória simples ou sistemática para selecionar uma amostra de professores do ensino fundamental dentro de cada conglomerado.

EXERCÍCIOS

***10.20** A seguir estão as percentagens de pessoas com 25 anos de idade ou mais com alguma formação superior, mas sem título, nos 50 estados norte-americanos, listados em ordem alfabética por estado, como revela o censo dos EUA de 1990:

16,8	27,6	25,4	16,6	22,6	24,0	15,9	16,9	19,4	17,0
20,1	24,2	19,4	16,6	17,0	21,9	15,2	17,2	16,1	18,6
15,8	20,4	19,0	16,9	18,4	22,1	21,1	25,8	18,0	15,5
20,9	15,7	16,8	20,5	17,0	21,3	25,0	12,9	15,0	15,8
18,8	16,9	21,1	27,9	14,7	18,5	25,0	13,2	16,7	24,2

Relacione as dez amostras sistemáticas possíveis de tamanho $n = 5$ que podem ser extraídas dessa lista, partindo de um dos primeiros dez números da primeira linha e então tomando cada décimo número na lista.

Snow666

***10.21** Com referência ao Exercício 10.20, liste as cinco amostras sistemáticas possíveis de tamanho $n = 10$ que podem ser extraídas dessa lista começando com um dos primeiros cinco números da primeira linha e então tomando cada quinto número da lista.

***10.22** A seguir estão os números relativos ao volume mensal de correio aéreo (em milhões de toneladas) transportada em rotas domésticas dos EUA durante um período de quatro anos:

67	62	75	67	70	68	64	70	66	73	73	97
76	73	80	78	78	72	75	75	73	83	76	108
84	78	86	85	81	78	78	75	78	86	76	111
79	77	87	84	82	77	79	77	80	84	78	117

Relacione as seis amostras sistemáticas possíveis de tamanho $n = 8$ que podem ser extraídas dessa lista começando com um dos seis primeiros números da primeira linha e então tomando cada sexto número.

***10.23** Se uma das seis amostras sistemáticas do Exercício 10.23 for escolhida aleatoriamente para estimar o volume médio mensal de correio, explique por que existe um sério risco de se obter um resultado muito enganador.

***10.24** Para generalizar o exemplo à página 246, suponha que queiramos estimar o peso médio de seis pessoas, cujos pesos (desconhecidos) são 115, 125, 135, 185, 195 e 205 libras.

(a) Relacione todas as amostras aleatórias possíveis de tamanho 2 que podem ser extraídas dessa população, calcule suas médias e determine a probabilidade de que a média de uma tal amostra vá diferir por mais de 5 de 160, o peso médio real das seis pessoas.

(b) Suponha que os três primeiros pesos sejam de mulheres e que os outros três sejam de homens. Relacione todas as amostras estratificadas possíveis de tamanho 2 que podem ser extraídas escolhendo aleatoriamente uma das três mulheres e um dos três homens, calcule suas médias e determine a probabilidade de que a média de tal amostra vá diferir por mais de 5 de 160, o peso médio real das seis pessoas.

(c) Suponha que três das pessoas, aquelas com pesos 125, 135 e 185 libras, tenham menos de 25 anos de idade, enquanto que as restantes tenham mais do que 25 anos de idade. Relacione todas as amostras estratificadas possíveis de tamanho 2 que podem ser extraídas escolhendo aleatoriamente uma das três pessoas mais jovens e uma das três mais velhas, calcule suas médias e determine a probabilidade de que a média de tal amostra vá diferir por mais de 5 de 160, o peso médio real das seis pessoas.

(d) Compare os resultados das partes (a), (b) e (c).

***10.25** Baseando-se em seus volumes de vendas, 9 de 12 revendedores de automóveis zero quilômetro de uma cidade estão classificados como sendo pequenos, e as outros três como sendo grandes. Quantas amostras estratificadas diferentes de quatro desses revendedores de automóveis zero quilômetro podemos escolher, se

(a) a metade da amostra deve ser alocada a cada um dos estratos;

(b) a alocação deve ser proporcional?

***10.26** Dentre 36 pessoas escolhidas para as secretarias de um município, 18 são advogados, 12 são executivos de firmas e 6 são professores. Quantas amostras estratificadas diferentes de seis dessas pessoas podemos escolher, se

(a) um terço da amostra deve ser atribuído a cada um dos estratos;

(b) a alocação deve ser proporcional?

***10.27** Uma amostra de tamanho $n = 40$ deve ser extraída de uma população de tamanho $N = 1.000$, que consiste em quatro estratos de tamanhos $N_1 = 250$, $N_2 = 600$, $N_3 = 100$ e $N_4 = 50$. Se a alocação deve ser proporcional, qual deve ser o tamanho da amostra a ser extraída de cada um dos quatro estratos?

Snow666

***10.28** Com referência à parte (b) do Exercício 10.24, relacione todas as amostras por conglomerado possíveis de tamanho 2 que podem ser extraídas escolhendo aleatoriamente duas das três mulheres ou dois dos três homens, calcule suas médias e determine a probabilidade de que a média de tal amostra vá diferir por mais de 5 de 160, o peso médio real das seis pessoas. Compare essa probabilidade com aquelas obtidas nas partes (a) e (b) do Exercício 10.24. O que isso mostra quanto aos méritos relativos da amostragem aleatória simples, da amostragem estratificada e da amostragem por conglomerado na situação dada?

***10.29** Na amostragem estratificada com alocação proporcional, a importância das diferenças entre os tamanhos dos estratos é levada em conta fazendo os estratos maiores contribuírem com um número relativamente maior de elementos para a amostra. No entanto, os estratos não diferem somente em tamanho, mas também em variabilidade e, portanto, pareceria razoável tomar amostras maiores de extratos mais variáveis, e amostras menores de estratos menos variáveis. Denotando por $\sigma_1, \sigma_2, \ldots$, e σ_k os desvios-padrão dos k estratos, podemos levar em conta tanto as diferenças nos tamanhos dos estratos quanto as diferenças na variabilidade dos mesmos, exigindo que

$$\frac{n_1}{N_1\sigma_1} = \frac{n_2}{N_2\sigma_2} = \cdots = \frac{n_k}{N_k\sigma_k}$$

Os tamanhos das amostras para esse tipo de alocação, denominada **alocação ótima**, são dados pela fórmula

$$n_i = \frac{n \cdot N_i\sigma_i}{N_1\sigma_1 + N_2\sigma_2 + \cdots + N_k\sigma_k}$$

para $i = 1, 2, \ldots$, e k, onde, se necessário, arredondamos para o inteiro mais próximo.

(a) Deve-se extrair uma amostra de tamanho $n = 100$ de uma população que consiste em dois estratos para os quais $N_1 = 10.000$, $N_2 = 30.000$, $N_3 = 45$ e $\sigma_2 = 60$. Para obter uma alocação ótima, qual é o tamanho da amostra que deve ser extraída de cada um dos dois estratos?

(b) Deve-se extrair uma amostra de tamanho $n = 84$ de uma população que consiste em três estratos para os quais $N_1 = 5.000$, $N_2 = 2.000$, $N_3 = 3.000$, $\sigma_1 = 15$, $\sigma_2 = 18$ e $\sigma_3 = 5$. Para obter uma alocação ótima, qual é o tamanho da amostra que deve ser extraída de cada um dos três estratos?

10.6 DISTRIBUIÇÕES AMOSTRAIS

A média amostral, a mediana amostral e o desvio-padrão amostral são exemplos de variáveis aleatórias cujos valores variam de amostra a amostra. Suas distribuições, que refletem tais variações aleatórias, desempenham papel fundamental na inferência estatística, e são denominadas **distribuições amostrais**. Neste capítulo, enfocamos principalmente a média amostral e sua distribuição amostral mas, em alguns dos exercícios às páginas 261-262, e em capítulos posteriores, vamos considerar também as distribuições amostrais de outras estatísticas.

Para dar um exemplo de uma distribuição amostral, vamos construir a da média de uma amostra aleatória de tamanho $n = 2$ extraída, sem reposição, de uma população finita de tamanho $N = 5$, cujos elementos são os números 3, 5, 7, 9 e 11. A média dessa população é

$$\mu = \frac{3+5+7+9+11}{5} = 7$$

Snow666

e seu desvio-padrão é

$$\sigma = \sqrt{\frac{(3-7)^2 + (5-7)^2 + (7-7)^2 + (9-7)^2 + (11-7)^2}{5}} = \sqrt{8}$$

Se tomamos, agora, uma amostra aleatória de tamanho $n = 2$ dessa população, existem $\binom{5}{2} = 10$ possibilidades 3 e 5, 3 e 7, 3 e 9, 3 e 11, 5 e 7, 5 e 9, 5 e 11, 7 e 9, 7 e 11 e 9 e 11. Suas médias são 4, 5, 6, 7, 6, 7, 8, 8, 9 e 10 e, como a amostragem é aleatória, cada um desses dez valores tem a probabilidade $\frac{1}{10}$. Assim, chegamos à seguinte distribuição amostral da média:

$\bar{x}$	*Probabilidade*
4	$\frac{1}{10}$
5	$\frac{1}{10}$
6	$\frac{2}{10}$
7	$\frac{2}{10}$
8	$\frac{2}{10}$
9	$\frac{1}{10}$
10	$\frac{1}{10}$

A Figura 10.3 mostra um histograma dessa distribuição.

Um exame dessa distribuição amostral revela algumas informações apropriadas sobre as variações ao acaso da média de uma amostra aleatória de tamanho $n = 2$ de uma dada população finita. Por exemplo, vemos que a probabilidade é de $\frac{6}{10}$ de que uma média amostral vá diferir por 1 ou menos da média populacional $\mu = 7$, e que a probabilidade é de $\frac{8}{10}$ de que uma média amostral vá diferir por 2 ou menos da média populacional $\mu = 7$. O primeiro caso corresponde a $\bar{x} = 6$, 7 ou 8, e o segundo caso corresponde a $\bar{x} = 5$, 6, 7, 8 ou 9. Assim, se não conhecêssemos a média da população dada e se quiséssemos estimá-la com uma amostra aleatória de tamanho $n = 2$, isso nos daria uma idéia do tamanho potencial de nosso erro.

Podemos obter informação adicional útil sobre essa distribuição amostral da média calculando sua média e seu desvio-padrão, denotados, respectivamente, por $\mu_{\bar{x}}$ e $\sigma_{\bar{x}}$. Aqui, o índice $\bar{x}$ é usado para distinguir entre os parâmetros da distribuição amostral e os da população original. Usando novamente as definições de média e de desvio-padrão de uma distribuição de probabilidade, obtemos

$$\begin{aligned}\mu_{\bar{x}} &= 4 \cdot \frac{1}{10} + 5 \cdot \frac{1}{10} + 6 \cdot \frac{2}{10} + 7 \cdot \frac{2}{10} + 8 \cdot \frac{2}{10} + 9 \cdot \frac{1}{10} + 10 \cdot \frac{1}{10} \\ &= 7\end{aligned}$$

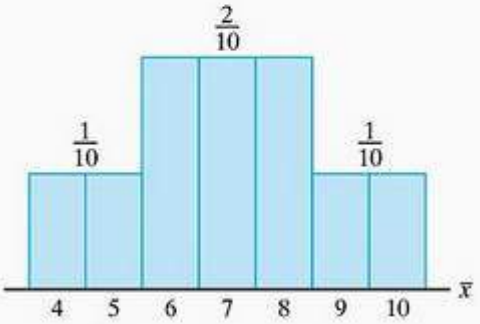

Figura 10.3
Distribuição amostral da média.

Snow666

e

$$\begin{aligned}\sigma_{\bar{x}}^2 &= (4-7)^2 \cdot \frac{1}{10} + (5-7)^2 \cdot \frac{1}{10} + (6-7)^2 \cdot \frac{2}{10} + (7-7)^2 \cdot \frac{2}{10} \\ &\quad + (8-7)^2 \cdot \frac{2}{10} + (9-7)^2 \cdot \frac{1}{10} + (10-7)^2 \cdot \frac{1}{10} \\ &= 3\end{aligned}$$

de forma que $\sigma_{\bar{x}} = \sqrt{3}$. Observe que, pelo menos para este exemplo,

> **$\mu_{\bar{x}}$, a média da distribuição amostral de $\bar{x}$, é igual a μ, a média da população;**
>
> **$\sigma_{\bar{x}}$, o desvio-padrão da distribuição amostral de $\bar{x}$, é menor do que σ, o desvio-padrão populacional.**

Essas relações são de importância fundamental, e voltaremos a elas na Seção 10.7.

Para ilustrar o conceito de distribuição amostral, tomamos uma amostra muito pequena, de tamanho $n = 2$, de uma população finita muito pequena, de tamanho $N = 5$, mas seria difícil reproduzir esse método para construir uma distribuição amostral da média de uma amostra aleatória grande de uma população grande. Por exemplo, para $n = 10$ e $N = 100$, teríamos de relacionar mais do que 17 trilhões de amostras.

Para termos alguma idéia da distribuição amostral da média de uma amostra um pouco maior de uma população finita grande, lançaremos mão de uma **simulação por computador**. Em outras palavras, deixaremos para um computador a tarefa de extrair repetidas amostras de uma dada população, determinar suas médias e descrever de várias maneiras a distribuição dessas médias. Isso nos dará alguma idéia sobre a forma global e sobre algumas das características essenciais da autêntica distribuição amostral da média para amostras aleatórias extraídas do tamanho dado da população dada.

Sem um computador, podemos imaginar a simulação como segue: em primeiro lugar, escrevermos os números de 1 a 1.000 em 1.000 pedacinhos de papel (ou fichas de pôquer, pequenas bolinhas, ou o que quer que se preste a extrair amostras aleatórias). Então extraímos, sem reposição, uma amostra aleatória de tamanho $n = 15$ dessa população e registramos seus valores. Recolocamos a amostra antes de extrair a próxima e repetimos esse processo até que tenhamos obtido 100 amostras aleatórias.

A população com que estamos tratando aqui é denominada **distribuição inteira** (também dita **distribuição discreta uniforme**), em que cada inteiro de 1 até N tem a probabilidade $\frac{1}{N}$. Usando as fórmulas para a soma e a soma dos quadrados dos inteiros de 1 a N, é fácil mostrar que a média e o desvio-padrão dessa distribuição são $\mu = \frac{N+1}{2}$ e $\sigma = \sqrt{\frac{N^2-1}{12}}$. Para $N = 1.000$, seus valores são $\mu = 500{,}5$ e $\sigma = 288{,}67$, arredondados até a segunda casa decimal.

Na verdade, usando um pacote computacional adequado, como MINITAB 13, neste caso, obtemos o impresso mostrado na Figura 10.4. O computador seguiu as instruções de gerar 100 amostras aleatórias de tamanho $n = 15$, cada uma sem reposição, mas repondo cada amostra entes de extrair a próxima.

Como pode ser visto na Figura 10.4, a distribuição de 100 médias amostrais é bastante simétrica e em forma de sino. Na verdade, o padrão global segue de perto o de uma curva normal. O computador também nos diz (ver Figura 10.5) que a média das 100 médias amostrais é 502,40 e que seu desvio-padrão é 75,64. De acordo com a relação indicada anteriormente, a média das 100 médias amostrais é muito parecida com a média populacional, e seu desvio-padrão é (muito) menos do que o da população.

Snow666

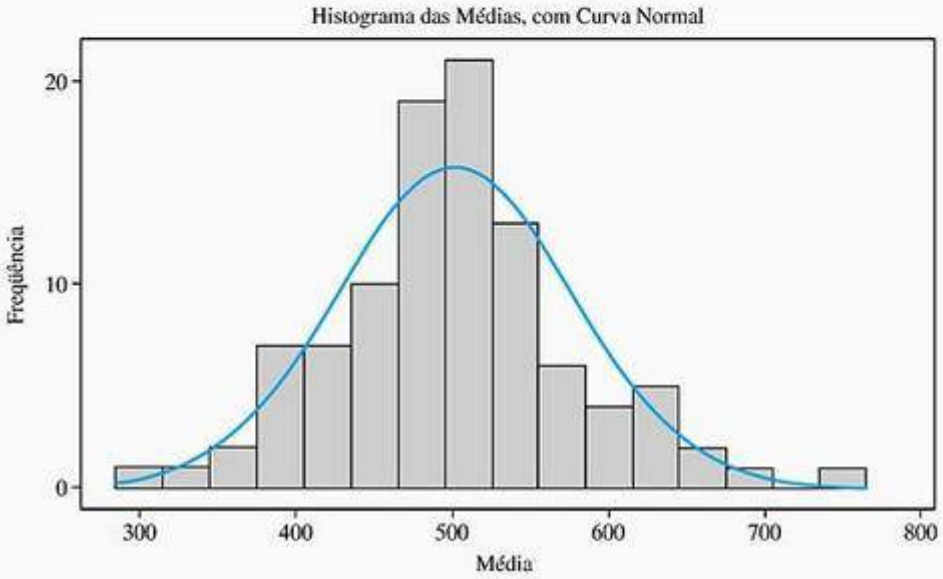

Figura 10.4
Simulação por computador de uma distribuição amostral da média.

10.7 O ERRO-PADRÃO DA MÉDIA

Na maioria das situações práticas, não podemos proceder como nas duas exemplificações da Seção 10.6. Ou seja, não podemos enumerar todas as amostras possíveis ou simular uma distribuição amostral a fim de julgar quão próxima uma média amostral está da média da população da qual provém a amostra. Felizmente, entretanto, em geral podemos obter a informação da qual necessitamos de teoremas, que expressam fatos essenciais sobre distribuições amostrais da média. Um desses teoremas é abordado nesta seção e o outro será desenvolvido na Seção 10.8.

O primeiro desses dois teoremas expressa formalmente o que descobrimos em ambos os exemplos da seção precedente, ou seja, que a média da distribuição amostral de $\bar{x}$ é igual à média da população da qual provém a amostra, e o desvio-padrão da distribuição amostral de $\bar{x}$ é menor do que o desvio-padrão daquela população. Podemos reformulá-lo como segue: para amostras aleatórias de tamanho n extraídas de uma população com a média μ e o desvio-padrão σ, a distribuição amostral de $\bar{x}$ tem média

MÉDIA DA DISTRIBUIÇÃO AMOSTRAL DE $\bar{x}$

$$\mu_{\bar{x}} = \mu$$

Estatística Descritiva: Médias

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Means	100	502.40	497.47	501.15	75.64	7.56

Variable	Minimum	Maximum	Q1	Q3
Means	307.93	750.93	461.68	540.75

Figura 10.5
Descrição da distribuição amostral.

Snow666

e o desvio-padrão

ERRO-PADRÃO DA MÉDIA

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad ou \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

dependendo de a população ser infinita ou finita de tamanho N.

Costumamos referir-nos a $\sigma_{\bar{x}}$ como o **erro-padrão da média**, sendo que a expressão "padrão" é empregada no sentido de uma média, como em "desvio-padrão". Seu papel é fundamental na Estatística, pois mede a extensão da flutuação ou variação esperada das médias amostrais, em função do acaso. Se $\sigma_{\bar{x}}$ é pequeno, há boas chances de a média de uma amostra estar próxima da média da população; se $\sigma_{\bar{x}}$ é grande, é mais provável que obtenhamos uma média amostral consideravelmente diferente da média da população.

Podemos ver o que determina o tamanho de $\sigma_{\bar{x}}$ a partir das duas fórmulas precedentes. Ambas as fórmulas (para populações infinitas e finitas) mostram que $\sigma_{\bar{x}}$ cresce quando a variabilidade da população aumenta, e decresce quando o tamanho da população aumenta. Na realidade, é diretamente proporcional a σ e inversamente proporcional à raiz quadrada de n. (Para populações finitas, decresce ainda mais rapidamente por virtude do n que aparece em $\sqrt{\frac{N-n}{N-1}}$.)

EXEMPLO 10.5 Quando extraímos uma amostra aleatória de uma população infinita, o que acontece com o erro-padrão da média e, portanto, com o erro que poderíamos esperar quando utilizamos a média da amostra para estimar a média da população, se o tamanho da amostra é

(a) aumentado de 50 para 200;

(b) diminuído de 360 para 40?

Solução **(a)** A razão dos dois erros-padrão é

$$\frac{\frac{\sigma}{\sqrt{200}}}{\frac{\sigma}{\sqrt{50}}} = \frac{\sqrt{50}}{\sqrt{200}} = \sqrt{\frac{50}{200}} = \sqrt{\frac{1}{4}} = \frac{1}{2}$$

e quando n é quadruplicado, o erro-padrão da média é reduzido, mas apenas dividido por 2.

(b) A razão dos dois erros padrão é

$$\frac{\frac{\sigma}{\sqrt{40}}}{\frac{\sigma}{\sqrt{360}}} = \frac{\sqrt{360}}{\sqrt{40}} = \sqrt{9} = 3$$

e quando n é dividido por 9, o erro padrão da média é aumentado, mas apenas multiplicado por 3.

O fator $\sqrt{\frac{N-n}{N-1}}$ na segunda fórmula para $\sigma_{\bar{x}}$ é denominado **fator de correção para população finita**, porque, sem ele, as duas fórmulas para $\sigma_{\bar{x}}$ (para populações infinitas e finitas) são iguais. Na prática, omitimos esse fator, a menos que a amostra constitua pelo menos 5% da população, pois, de outra forma, tal fator está tão próximo de 1 que quase não influi no valor de $\sigma_{\bar{x}}$.

EXEMPLO 10.6 Encontre o valor do fator de correção para população finita para $N = 10.000$ e $n = 100$.

Snow666

Solução Substituindo $N = 10.000$ e $n = 100$ na fórmula do fator de correção para população finita, obtemos

$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{10.000-100}{10.000-1}} = 0,995$$

e esse valor está tão próximo de 1 que o fator de correção pode ser omitido para fins práticos.

Já que enunciamos as fórmulas para o erro-padrão da média sem prova, verifiquemos a que se refere a populações finitas usando os resultados das duas exemplificações da Seção 10.6.

EXEMPLO 10.7 Com referência ao exemplificado à página 250, em que tínhamos $N = 5$, $n = 2$ e $\sigma = \sqrt{8}$, verifique que a segunda das duas fórmulas para $\sigma_{\bar{x}}$ fornece $\sigma = \sqrt{3}$, ou seja, o valor que obtivemos à página 252.

Solução Substituindo $N = 5$, $n = 2$ e $\sigma = \sqrt{8}$, na segunda das duas fórmulas para $\sigma_{\bar{x}}$, obtemos

$$\sigma_{\bar{x}} = \frac{\sqrt{8}}{\sqrt{2}} \cdot \frac{\sqrt{5-2}}{5-1} = \frac{\sqrt{8}}{\sqrt{2}} \cdot \frac{\sqrt{3}}{4} = \sqrt{\frac{8}{2} \cdot \frac{3}{4}} = \sqrt{3}$$

EXEMPLO 10.8 Com referência à simulação por computador da Figura 10.4, onde tínhamos $N = 1.000$, $n = 15$ e $\sigma = 288,67$, que valor poderíamos ter esperado para o desvio-padrão das 100 médias amostrais?

Solução Substituindo $N = 1.000$, $n = 15$ e $\sigma = 288,67$ na segunda das duas fórmulas para $\sigma_{\bar{x}}$, obtemos

$$\sigma_{\bar{x}} = \frac{288,67}{\sqrt{15}} \cdot \sqrt{\frac{1.000-15}{1.000-1}} = 74,01$$

e isso está bem próximo de 75,64, que é o valor realmente obtido na simulação por computador da Figura 10.5.

10.8 O TEOREMA DO LIMITE CENTRAL

Quando utilizamos uma média amostral para estimar a média de uma população, as incertezas sobre o erro potencial podem ser expressas de várias maneiras. Se soubéssemos a distribuição amostral exata da média, o que, é claro, nunca sabemos, poderíamos proceder como na primeira exemplificação da Seção 10.6 e calcular as probabilidades associadas com erros de diversos tamanhos. Uma outra coisa que raramente, ou nunca, fazemos é usar o teorema de Tchebichev e afirmar, com uma probabilidade de pelo menos $1 - \frac{1}{k^2}$, que a média de uma amostra aleatória diferirá da média da população correspondente por menos do que $k \cdot \sigma_{\bar{x}}$.

EXEMPLO 10.9 Usando o teorema de Tchebichev com $k = 2$, o que podemos dizer sobre o tamanho potencial de nosso erro, se utilizarmos uma amostra aleatória de tamanho $n = 64$ para estimar a média de uma população infinita com $\sigma = 20$?

Solução Substituindo $n = 64$ e $\sigma = 20$ na fórmula apropriada do erro-padrão da média, obtemos

$$\sigma_{\bar{x}} = \frac{20}{\sqrt{64}} = 2,5$$

Snow666

e segue que podemos afirmar, com uma probabilidade de pelo menos $1-\frac{1}{2^2}=0{,}75$, que o erro será inferior a $k \cdot \sigma_{\bar{x}} = 2(2{,}5) = 5$.

A importância desse exemplo é que ele mostra *como* podemos fazer afirmações probabilísticas exatas sobre o erro potencial quando estimamos a média de uma população. O problema com o uso do teorema de Tchebichev é que "pelo menos 0,75" não nos diz o suficiente quando na realidade aquela probabilidade pode ser, digamos, 0,998, ou mesmo 0,999. Enquanto o teorema de Tchebichev estabelece uma conexão lógica entre os erros e as probabilidades que podem ser cometidos, existe um outro teorema matemático que, em muitas circunstâncias, nos permite fazer afirmações probabilísticas muito mais fortes sobre tais erros potenciais.

Esse teorema, que é o segundo dos dois teoremas mencionados à página 253, é denominado o **teorema do limite central** e, informalmente, afirma que para grandes amostras, a distribuição amostral da média pode ser muito bem aproximada por uma distribuição normal. Lembrando que na Seção 10.7 vimos que

$$\mu_{\bar{x}} = \mu \quad \text{e} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

para amostras aleatórias de populações infinitas, podemos, então, dizer formalmente que

TEOREMA DO LIMITE CENTRAL

Se $\bar{x}$ é a média de uma amostra aleatória de tamanho n de uma população infinita com a média μ e o desvio-padrão σ e se n é grande, então

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

tem aproximadamente a distribuição normal padrão.

Esse teorema é de importância fundamental em Estatística, pois justifica a aplicação de métodos da curva normal a uma ampla gama de problemas; esse teorema se aplica a populações infinitas e também a populações finitas quando n, embora grande, constitui uma porção pequena da população. Não podemos dizer precisamente quão grande n deve ser para que possamos aplicar o teorema do limite central mas, a menos que a distribuição da população tenha um formato muito raro, em geral, $n = 30$ é considerado suficientemente grande. Quando a própria população da qual estamos extraindo amostras tem aproximadamente o formato de uma curva normal, a distribuição amostral da média pode ser aproximada muito bem com uma distribuição normal, independentemente do tamanho de n.

O teorema do limite central também pode ser usado para populações finitas, mas é bastante complicado dar uma descrição precisa das situações em que isso é possível. A utilização adequada mais comum é o caso em que n é grande enquanto $\frac{n}{N}$ é pequeno.

Vejamos, agora, qual probabilidade tomará o lugar de "pelo menos 0,75" se utilizarmos o teorema do limite central, em vez do teorema de Tchebichev, no Exemplo 10.9.

EXEMPLO 10.10 Usando o teorema do limite central, qual é a probabilidade de o erro ser inferior a 5 quando usamos a média de uma amostra aleatória de tamanho $n = 64$ para estimar a média de uma população infinita com $\sigma = 20$?

Solução A probabilidade é dada pela área da região colorida sob a curva na Figura 10.6, ou seja, pela área sob a curva normal padrão entre

$$z = \frac{-5}{20/\sqrt{64}} = -2 \quad \text{e} \quad z = \frac{5}{20/\sqrt{64}} = 2$$

Snow666

Como a entrada na Tabela I correspondente a $z = 2{,}00$ é 0,4772, a probabilidade procurada é $0{,}4772 + 0{,}4772 = 0{,}9544$. Assim, a afirmação de que a probabilidade é de "pelo menos 0,75" é substituída pela afirmação muito mais forte de que a probabilidade é da ordem de 0,95.

10.9 ALGUMAS CONSIDERAÇÕES ADICIONAIS

Nas Seções 10.6 até 10.8, nosso objetivo principal foi introduzir o conceito de uma distribuição amostral, e a que escolhemos como ilustração foi a distribuição amostral da média. Devemos deixar claro, entretanto, que, em lugar da média, poderíamos ter utilizado a mediana, o desvio-padrão, ou alguma outra estatística, e estudado suas flutuações aleatórias. No que diz respeito à teoria envolvida, isso teria exigido uma fórmula diferente para o desvio-padrão e uma teoria análoga, ainda que diferente, ao teorema do limite central.

Por exemplo, para grandes amostras de populações contínuas, o **erro-padrão da mediana** é aproximadamente

$$\sigma_{\tilde{x}} = 1{,}25 \cdot \frac{\sigma}{\sqrt{n}}$$

onde n é o tamanho da amostra e σ é o desvio-padrão populacional. Note que a comparação das fórmulas

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{e} \quad \sigma_{\tilde{x}} = 1{,}25 \cdot \frac{\sigma}{\sqrt{n}}$$

reflete o fato de que a média é geralmente mais confiável do que a mediana (isto é, ela tende a nos expor a erros menores) quando estimamos a média de uma população simétrica. Para populações simétricas, as médias das distribuições amostrais de $\bar{x}$ e $\tilde{x}$ são ambas iguais à média populacional μ

EXEMPLO 10.11 Quão grande precisamos tomar uma amostra aleatória para que sua média seja uma estimativa tão confiável da média de uma população contínua simétrica quanto a mediana de uma amostra aleatória de tamanho $n = 200$?

Solução Igualando as duas fórmulas do erro padrão e substituindo $n = 200$ na fórmula do erro padrão da mediana, obtemos

$$\frac{\sigma}{\sqrt{n}} = 1{,}25 \cdot \frac{\sigma}{\sqrt{200}}$$

que, resolvida em relação a n, dá $n = 128$. Assim, para o objetivo proposto, a média de uma amostra aleatória de tamanho $n = 128$ é tão "boa" quanto a mediana de uma amostra aleatória de tamanho $n = 200$.

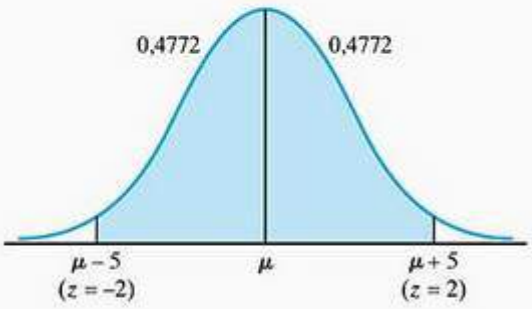

Figura 10.6 Diagrama para o Exemplo 10.10.

Snow666

Além disso, um ponto que vale a pena enfatizar é que as exemplificações da Seção 10.6 foram utilizadas como recurso didático, projetadas para transmitir a idéia de uma distribuição amostral, mas elas não refletem o que fazemos na prática. Na prática, raramente podemos relacionar todas as amostras possíveis e é comum basearmos uma inferência numa única amostra, e não em 100 amostras. No Capítulo 11 e em capítulos subseqüentes, abordaremos em maior profundidade o problema de traduzir a teoria sobre distribuições amostrais em métodos de avaliar os méritos e as deficiências de processos estatísticos.

Um outro ponto que vale a pena enfatizar se refere a $\sqrt{n}$ que aparece no denominador das fórmulas do erro-padrão da média. É claro que, quando n se torna cada vez maior, nossas generalizações estarão sujeitas a erros cada vez menores, mas a $\sqrt{n}$ nas fórmulas do erro-padrão da média nos diz que o ganho em confiabilidade não é proporcional ao aumento no tamanho da amostra. Como vimos, quadruplicar o tamanho da amostra apenas duplica a confiabilidade de uma média amostral como uma estimativa da média de uma população. Na realidade, para quadruplicar a confiabilidade, deveríamos multiplicar por 16 o tamanho da amostra. Usando a linguagem dos economistas, essa relação entre confiabilidade e tamanho da amostra indica que há diminuição de lucros com um aumento do tamanho da amostra. Raramente compensa trabalhar com amostras extremamente grandes.

EXERCÍCIOS

10.30 Suponha que na exemplificação à página 245, em que extraímos, sem reposição, amostras aleatórias de tamanho $n = 2$ da população finita cujos elementos são os números 3, 5, 7, 9 e 11, a extração tivesse sido feita com reposição.

(a) Relacione as 25 amostras ordenadas que podem ser extraídas com reposição da população dada e calcule suas médias. (Aqui, "ordenada" significa que 3 e 7 é uma amostra diferente de 7 e 3.)

(b) Supondo que a amostragem é aleatória, a saber, que cada uma das amostras ordenadas da parte (a) tenha a probabilidade $\frac{1}{25}$, construa a distribuição amostral da média para amostras aleatórias de tamanho $n = 2$ extraídas, com reposição, da população dada.

10.31 Com referência ao Exercício 10.30, encontre as probabilidades de que a média de uma amostra aleatória de tamanho $n = 2$ extraída, com reposição, da população dada, seja diferente de $\mu = 7$ por

(a) 1 ou menos;

(b) no máximo 2.

10.32 Calcule o desvio-padrão da distribuição amostral obtida na parte (b) do Exercício 10.30 e verifique o resultado substituindo $n = 2$ e $\sigma = \sqrt{8}$ na fórmula do erro-padrão à página 254.

10.33 Uma população finita consiste nos $N = 8$ elementos 12, 12, 12, 12, 12, 14, 20 e 42. Como pode ser facilmente verificado, a média dessa população é $\mu = 17$ e seu desvio-padrão é $\sigma = \sqrt{96} = 4\sqrt{6}$.

(a) Relacione as $\binom{8}{2}$ amostras possíveis de tamanho $n = 2$ que podem ser extraídas, sem reposição, dessa população finita. [*Sugestão*: em $\binom{5}{2} = 10$ dessas amostras, ambos os valores são iguais a 12.]

(b) Calcule a média de cada uma das amostras obtidas na parte (a).

Snow666

(c) Atribuindo a cada uma das amostras obtidas na parte (a) a probabilidade $\frac{1}{28}$, construa a distribuição amostral da média para amostras aleatórias de tamanho $n = 2$ extraídas, sem reposição, da população dada.

(d) Encontre a média e o desvio-padrão da distribuição amostral da média obtida na parte (c).

10.34 Use a fórmula do erro padrão à página 254 para conferir o resultado na parte (d) do Exercício 10.33.

10.35 Quando extraímos amostras de uma população infinita, o que acontece com o erro-padrão da média quando o tamanho da amostra é

(a) aumentado de 30 para 120;

(b) diminuído de 245 para 5?

10.36 Quando extraímos amostras de uma população infinita, o que acontece com o erro-padrão da média quando o tamanho da amostra é

(a) diminuído de 1.000 para 10;

(b) aumentado de 80 para 500?

10.37 Qual é o valor do fator de correção para população finita quando

(a) $N = 200$ e $n = 10$;

(b) $N = 300$ e $n = 25$;

(c) $N = 5.000$ e $n = 100$?

10.38 Mostre que se a média de uma amostra aleatória de tamanho n é usada para estimar a média de uma população infinita com desvio-padrão σ e n é grande, então há uma chance meio a meio de que a magnitude do erro seja menor do que

$$0{,}6745 \cdot \frac{\sigma}{\sqrt{n}}$$

É costume designar essa quantidade como o **erro provável da média**; hoje em dia, é usada principalmente em aplicações militares.

(a) Se uma amostra aleatória de tamanho $n = 64$ é extraída de uma população infinita com $\sigma = 24{,}8$, qual é o erro provável da média?

(b) Se uma amostra aleatória de tamanho $n = 144$ é extraída de uma população muito grande (consistindo, digamos, nas multas pagas por várias infrações de tráfego num certo estado em 1999) com $\sigma = 219{,}12$ unidades monetárias, qual é o erro provável da média? Explique seu significado.

10.39 Na exemplificação à página 246, comparamos amostras estratificadas de uma população de quatro pesos com amostras aleatórias comuns de mesmo tamanho.

(a) Atribuindo a cada uma das amostras aleatórias da página 246 a probabilidade $\frac{1}{6}$, mostre que a média e o desvio-padrão dessa distribuição amostral da média são $\mu_{\bar{x}} = 160$ e $\sigma_{\bar{x}} = 21{,}0$.

(b) Atribuindo a cada uma das amostras aleatórias da página 246 a probabilidade $\frac{1}{4}$, mostre que a média e o desvio-padrão dessa distribuição amostral da média são $\mu_{\bar{x}} = 160$ e $\sigma_{\bar{x}} = 7{,}1$.

10.40 A média de uma amostra aleatória de tamanho $n = 36$ é usada para estimar a média de uma população infinita com desvio-padrão $\sigma = 9$. O que pode ser dito sobre a probabilidade de o erro dessa estimativa ser menor do que 4,5, se aplicarmos

(a) o teorema de Tchebichev;

(b) o teorema do limite central?

Snow666

10.41 A média de uma amostra aleatória de tamanho $n = 25$ é usada para estimar o intervalo de tempo médio durante o qual uma pessoa com mais de 65 anos de idade consegue se concentrar. Sabendo que o desvio-padrão dessa população é $\sigma = 2{,}4$ minutos, o que pode ser afirmado sobre a probabilidade de o erro da estimativa ser menor do que 1,2 minutos, se aplicarmos

(a) o teorema de Tchebichev;
(b) o teorema do limite central?

10.42 A média de uma amostra aleatória de tamanho $n = 100$ será utilizada para estimar a produção média diária de leite num rebanho muito grande de vacas leiteiras. Sabendo que o desvio-padrão dessa população é $\sigma = 3{,}6$ litros, o que pode ser afirmado sobre a probabilidade de o erro da estimativa ser

(a) maior do que 0,72 litros;
(b) menor do que 0,45 litros?

10.43 Se as medições da gravidade específica de uma lata de metal podem ser consideradas como uma amostra aleatória de uma população normal com desvio-padrão $\sigma = 0{,}025$ unidades de medição, qual é a probabilidade de a média de uma amostra aleatória de tamanho $n = 16$ estar errada em, no máximo, 0,01 unidades de medição?

10.44 Verifique que a média de uma amostra aleatória de tamanho $n = 256$ é uma estimativa tão confiável da média de uma população contínua simétrica quanto a mediana de uma amostra aleatória de tamanho $n = 400$.

10.45 Quão grande deve ser uma amostra aleatória para que sua mediana seja uma estimativa tão confiável da média de uma população contínua simétrica quanto a média de uma amostra aleatória de tamanho $n = 144$?

*10.10 NOTA TÉCNICA (SIMULAÇÃO)

A simulação fornece uma das maneiras mais efetivas de ilustrar e, conseqüentemente, de ensinar, alguns dos conceitos básicos da Estatística. Ela serve para demonstrar a validade da teoria quando as provas matemáticas rigorosas estão além dos pré-requisitos deste livro.

Também, como veremos em capítulos que seguem, a avaliação e a interpretação de técnicas estatísticas freqüentemente exigirão que imaginemos o que aconteceria se os experimentos fossem repetidos um grande número de vezes. Como, na maioria das vezes, tais repetições não são nem práticas nem viáveis, podemos recorrer, em vez disso, a simulações, preferivelmente com o uso de computadores. A simulação tem importância, também, no desenvolvimento da teoria estatística, pois há situações em que a simulação é mais fácil do que uma análise matemática detalhada.

As simulações de amostras aleatórias também podem ser feitas usando tabelas de números aleatórios, mas para o uso nos exercícios a seguir, apresentamos na Figura 10.7 quarenta amostras aleatórias simuladas por computador, cada uma delas consistindo em $n = 5$ valores de uma variável aleatória com distribuição de Poisson com $\lambda = 16$, e portanto $\mu = 16$ e $\sigma = 4$. (O leitor pode imaginar esses números como dados referentes ao número de chamadas de emergência que um serviço de ambulâncias recebe numa tarde, o número de ligações que uma operadora recebe durante um intervalo de dez minutos, ou o número de panfletos de propaganda que um motorista recebe num certo sinal de trânsito.)

Snow666

EXERCÍCIOS

***10.46** Cada linha na Figura 10.7 constitui uma amostra aleatória de tamanho $n = 5$ de uma população de Poisson com $\lambda = 16$, e portanto com $\mu = 16$ e $\sigma = 4$.

(a) Calcule as médias das 40 amostras na Figura 10.7.

(b) Calcule a média e o desvio-padrão das 40 médias obtidas na parte (a) e compare os resultados com os correspondentes valores esperados a partir da teoria da Seção 10.7.

***10.47** Determine as medianas das 40 amostras mostradas na Figura 10.7, calcule seu desvio-padrão e compare o resultado com o correspondente valor esperado a partir da teoria da Seção 10.9.

Amostras Simuladas de Poisson com n = 5

```
MTB > Random 40 c1-c5;
SUBC>  Poisson 16.
MTB > Print c1-c5.
```

Row					
1	9	15	6	19	11
2	16	15	19	14	14
3	14	20	11	22	19
4	14	20	17	22	14
5	21	11	13	18	14
6	13	11	13	15	12
7	14	12	14	17	10
8	17	13	25	17	20
9	21	16	16	18	21
10	15	12	16	11	14
11	21	12	19	14	14
12	20	22	16	19	17
13	25	15	8	16	21
14	15	19	18	12	18
15	17	23	20	11	13
16	18	16	16	21	22
17	20	19	21	17	9
18	19	17	11	14	19
19	12	18	16	10	14
20	11	14	11	12	26
21	17	16	11	11	9
22	15	16	16	19	18
23	16	12	18	16	15
24	20	19	23	14	14
25	19	18	16	24	13
26	13	18	14	17	25
27	16	17	18	14	22
28	15	17	11	15	13
29	23	12	13	13	16
30	16	28	11	14	11
31	14	15	18	7	16
32	19	17	11	16	13
33	13	14	16	12	17
34	25	14	8	15	16
35	12	17	12	12	13
36	12	16	17	15	25
37	20	14	14	16	17
38	13	19	19	16	17
39	12	14	11	19	14
40	14	9	17	24	19

Figura 10.7 Simulação por computador de dados de Poisson.

Snow666

***10.48** No Exercício 11.35 apresentaremos uma maneira de estimar o desvio-padrão populacional em termos da amplitude (maior valor menos menor valor). Com esse propósito, a amplitude é dividida por uma constante, que depende do tamanho da amostra; por exemplo, dividimos por 2,33 para $n = 5$. Uma outra maneira de dizer isso é que para amostras de tamanho $n = 5$, a média da distribuição amostral da amplitude é $2{,}33\sigma$. Para verificar isso, determine as amplitudes das 40 amostras mostradas na Figura 10.7 e então calcule sua média. Isso é uma estimativa da média da distribuição amostral e, como o tamanho da amostra é $n = 5$, dividindo-a por 2,33 fornece uma estimativa para σ, que se sabe ser igual a 4. Encontre o erro percentual dessa estimativa.

***10.49** Nas páginas 87/88 explicamos que a divisão por $n - 1$ nas fórmulas do desvio-padrão amostral e da variância amostral serve para tornar s^2 um estimador não tendencioso para σ^2, ou seja, para fazer a média da distribuição amostral de s^2 igual a σ^2. Para verificar isso, encontre a média das variâncias das 40 amostras na Figura 10.7, que são 26,0; 4,3; 20,7; 12,8; 16,3; 2,2; 6,8; 19,8; 6,3; 4,3; 14,5; 5,7; 41,5; 8,3; 24,2; 7,8; 23,2; 12,0; 10,0; 40,7; 12,2; 2,7; 4,8; 15,5; 16,5; 22,3; 8,8; 5,2; 20,3; 49,5; 17,5; 10,2; 4,3; 37,3; 4,7; 23,5; 6,2; 6,2; 9,5 e 31,3. Como σ^2 é sabidamente igual a 16, calcule o erro percentual dessa estimativa.

10.11 LISTA DE TERMOS-CHAVE (com indicação das páginas de suas definições)

*Alocação ótima, 247, 250
Alocação proporcional, 246
Amostra aleatória, 238, 239
Amostra aleatória simples, 239
Amostragem aleatória (simples) estratificada, 246
Amostragem aleatória de uma população finita, 239
*Amostragem estratificada, 245
*Amostragem por área, 248
*Amostragem por conglomerado, 248
*Amostragem por cotas, 247
*Amostragem por julgamento, 247
Amostragem por levantamento, 245
*Amostragem sistemática, 245
Distribuição amostral, 238
Distribuição discreta uniforme, 252
Distribuição inteira, 252
**Erro provável da média*, 259
Erro-padrão da média, 254
Erro-padrão da mediana, 257
*Estratificação, 245
*Estratificação cruzada, 247
*Estratos, 246
Estrutura amostral, 244
Fator de correção para população finita, 254
Números aleatórios, 240
*Planejamento de amostra, 245
População finita, 239
População infinita, 239
Simulação por computador, 252
Teorema do limite central, 256

10.12 REFERÊNCIAS

Dentre as muitas tabelas publicadas de números aleatórios, uma das mais largamente utilizadas é

RAND CORPORATION, *A Million Random Digits with 100,000 Normal Deviates*. New York: Macmillan Publishing Co., Inc., third printing, 1996.

Também existem calculadoras programadas para gerar números aleatórios e é razoavelmente fácil programar um computador para que permita ao usuário gerar seus próprios números aleatórios. Um dos muitos artigos a respeito desse assunto é o seguinte

KIMERLING, C., "Generate Your Own Random Numbers", *Mathematics Teacher*, February 1984.

Snow666

Material interessante sobre o início do desenvolvimento de tabelas de números aleatórios pode ser encontrado em

BENNETT, J. D., *Randomness*. Cambridge, Mass.: Harvard University Press, 1998.

Conseqüências das várias fórmulas do erro-padrão e formulações (e provas) mais gerais do teorema do limite central podem ser encontradas na maioria dos livros didáticos de Estatística matemática. Tudo que é tipo de informação sobre amostragem é dada em

COCHRAN, W. G., *Sampling Techniques*, 3rd ed. New York: John Wiley & Sons, Inc., 1977.

SCHAFFER, R. L., MENDENHALL, W., and OTT, L., *Elementary Survey Sampling*, 4th ed. Boston: PWS-Kent Publishing Co., 1990.

SLONIN, M. J., *Sampling in a Nutshell*. New York: Simon and Schuster, 1973.

WILLIAMS, W. H., *A Sampler on Sampling*. New York: John Wiley & Sons, Inc., 1978.

Snow666

EXERCÍCIOS DE REVISÃO PARA OS CAPÍTULOS 8, 9 E 10

R.91 Dentre 18 trabalhadores num piquete de grevistas, dez são homens e oito são mulheres. Se uma equipe de TV selecionar aleatoriamente quatro deles para serem filmados, qual é a probabilidade de que vá incluir

(a) somente homens;
(b) dois homens e duas mulheres?

R.92 Encontre a área sob a curva normal padrão que está

(a) à esquerda de $z = 1{,}65$;
(b) à esquerda de $z = -0{,}44$;
(c) entre $z = 1{,}15$ e $z = 1{,}85$;
(d) entre $z = -0{,}66$ e $z = 0{,}66$.

R.93 Encontre z_α sabendo que $\alpha = 0{,}2709$.

R.94 Um agente alfandegário quer inspecionar 12 de 875 volumes listados no manifesto da carga de um navio. Quais volumes (por números) o agente alfandegário inspecionará se ele usar as colunas 28, 29 e 30 e descer pela página a partir da sexta linha da tabela na Figura 10.2?

R.95 Verifique, em cada caso, se está satisfeita a condição para a aproximação binomial da distribuição hipergeométrica:

(a) $a = 40$, $b = 160$ e $n = 8$;
(b) $a = 100$, $b = 60$ e $n = 10$;
(c) $a = 68$, $b = 82$ e $n = 12$;

R.96 Se a quantidade de tempo que um turista permanece no interior de uma catedral é uma variável aleatória de distribuição normal com $\mu = 23{,}4$ minutos e $\sigma = 6{,}8$ minutos, encontre a probabilidade de que o turista vá permanecer

(a) pelo menos 15 minutos;
(b) algum tempo entre 20 e 30 minutos.

R.97 Qual é a probabilidade de cada amostra possível se são extraídas amostras aleatórias de tamanho $n = 6$ de uma população finita de tamanho $N = 45$?

***R.98** A probabilidade de que um certo jogador de basquete vá converter um lance livre qualquer em cesta é de 0,36. Qual é a probabilidade de que o primeiro lance livre que ele vá converter em cesta seja

(a) seu segundo lance livre;
(b) seu quinto lance livre?

(*Sugestão*: Use a fórmula da distribuição geométrica.)

R.99 Encontre a média da distribuição binomial com $n = 8$ e $p = 0{,}40$, usando

(a) a Tabela V e a fórmula que define μ;
(b) a fórmula especial para a média de uma distribuição binomial.

R.100 Use a distribuição normal para aproximar a distribuição binomial de que pelo menos 25 dentre 60 picadas de abelha causem algum desconforto, se a probabilidade é de 0,48 de que uma picada de abelha qualquer vá causar algum desconforto.

R.101 Verifique, em cada caso, se estão satisfeitas as condições para a aproximação de Poisson da distribuição binomial:

(a) $n = 180$ e $p = \frac{1}{9}$;

Snow666

(b) $n = 480$ e $p = \frac{1}{60}$;

(c) $n = 575$ e $p = \frac{1}{100}$.

R.102 Sabe-se que 6% de todos os ratos são portadores de determinada doença. Se examinarmos uma amostra aleatória de 120 ratos, estaremos satisfazendo a condição exigida para utilizar a aproximação de Poisson à distribuição binomial? Caso estejamos, use a distribuição de Poisson para aproximar a probabilidade de que somente 5 dos ratos sejam portadores da doença.

R.103 Uma variável aleatória tem uma distribuição normal com $\sigma = 4{,}0$. Se há uma probabilidade de 0,9713 de que essa variável assuma um valor menor do que 82,6, qual é a probabilidade de que ela vá assumir um valor entre 70,0 e 80,0?

R.104 Transforme as 40 amostras na Figura 10.7 em 20 amostras de tamanho $n = 10$ combinando as amostras 1 e 2, as amostras 3 e 4, . . ., e as amostras 39 e 40. Calcule a média de cada uma dessas 20 amostras e determine sua média e seu desvio-padrão. Compare os resultados com os valores que poderíamos esperar obter de acordo com o teorema à página 253/254.

R.105 Um pequeno navio de cruzeiro tem cabines externas de luxo, cabines externas normais e cabines internas, e as probabilidades de que um agente de viagens vá receber pedidos de reservas para a primeira, a segunda e a terceira dessas categorias são de 0,30, 0,60 e 0,10. Se um agente de viagens receber pedidos de nove reservas, qual é a probabilidade de que quatro deles sejam para as cabinas externas de luxo, quatro para as cabines externas normais e um para uma cabine interna?

R.106 Num certo bairro, o tempo de que uma ambulância leva para atender um chamado pode ser considerado uma variável aleatória de distribuição normal com $\mu = 5{,}8$ minutos e desvio padrão $\sigma = 1{,}2$ minutos. Qual é a probabilidade de que uma ambulância vá levar no máximo 8,0 minutos para atender uma chamada?

R.107 Um certo zoológico tem uma coleção grande de tamanduás, com cinco machos e quatro fêmeas. Se um veterinário escolher ao acaso três deles para examinar, quais são as probabilidades de que

(a) nenhum deles seja fêmea;

(b) dois deles sejam fêmeas?

R.108 Alega-se que

(a) se o tamanho da amostra é aumentado em 44%, então o erro-padrão da média é reduzido em 20%;

(b) se o erro-padrão da média deve ser reduzido em 20%, então o tamanho da amostra deve ser aumentado em 56,25%.

Qual dessas duas afirmações é verdadeira e qual é falsa?

R.109 A Figura R.4 mostra a densidade de probabilidade de uma variável aleatória contínua definida no intervalo de 0 a 3.

(a) Verifique que a área total sob a curva é igual a 1.

(b) Encontre a probabilidade de que a variável aleatória vá tomar um valor maior do que 1,5.

***R.110** Suponha que queiramos encontrar a probabilidade de que uma variável aleatória de distribuição hipergeométrica com $n = 14$, $a = 180$ e $b = 120$ vá tomar o valor $x = 5$.

(a) Verifique que a distribuição binomial com $n = 14$ e $p = \frac{180}{180+120} = \frac{3}{5}$ pode ser usada para aproximar essa distribuição hipergeométrica.

(b) Verifique que a distribuição normal com $\mu = np = 14 \cdot \frac{3}{5} = 8{,}4$ e $\sigma = \sqrt{np(1-p)} = \sqrt{14(0{,}6)(0{,}4)} \approx 1{,}83$ pode ser usada para aproximar a distribuição binomial da parte (a).

Snow666

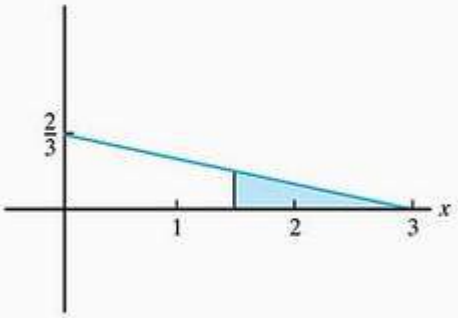

Figura R.4
Diagrama para o Exercício R.109.

(c) Use a distribuição normal com $\mu = 8{,}4$ e $\sigma = 1{,}83$ para aproximar a probabilidade hipergeométrica com $x = 5$, $n = 14$, $a = 180$ e $b = 120$.

R.111 O tempo que um eletricista leva para consertar um ventilador de teto pode ser tratado como uma variável aleatória de distribuição normal com $\mu = 24{,}55$ minutos e $\sigma = 3{,}16$ minutos. Encontre a probabilidade de que um eletricista leve algum tempo entre 20,00 e 30,00 minutos para consertar um ventilador de teto.

R.112 Quantas amostras diferentes de tamanho

(a) $n = 3$ podem ser extraídas dentre $N = 14$ revistas diferentes da sala de espera de um consultório médico?

(b) $n = 5$ podem ser extraídas para um comprador em potencial dentre $N = 24$ casas à venda num certo bairro da periferia?

R.113 Um grupo de 300 pessoas sorteadas para um júri inclui apenas 30 pessoas com idade inferior a 25 anos. Para um caso específico relacionado a narcóticos, o júri de 12 realmente selecionado do grupo não incluiu pessoa alguma com idade inferior a 25 anos. O jovem advogado de defesa queixou-se de que este júri de 12 não é representativo. Ele argumentou que a probabilidade de que um dos jurados tenha idade inferior a 25 anos deve ser *muitas vezes* a probabilidade de que nenhum deles tenha idade inferior a 25 anos.

(a) Encontre a razão dessas duas probabilidades, usando a distribuição hipergeométrica.

(b) Encontre a razão dessas duas probabilidades, usando a distribuição binomial como aproximação.

R.114 Durante uma noite de sexta-feira, 50 motos de tele-entrega de uma rede de farmácias percorreram

23,2	26,7	21,5	23,8	19,1	22,3	27,4	22,4	20,6	23,5
16,5	22,2	21,9	14,4	25,6	23,0	25,4	21,2	16,8	28,4
20,5	21,5	22,6	19,8	20,5	21,7	16,3	18,9	24,0	21,3
22,2	24,8	17,5	18,0	21,4	22,5	20,6	17,7	15,9	22,5
26,7	21,3	24,5	19,3	25,4	20,0	16,5	21,1	23,8	20,5

quilômetros. Use um computador ou uma calculadora gráfica para conferir se esses dados podem ser considerados como os valores de uma variável aleatória de distribuição normal.

R.115 Utilizando a parte superior do impresso na Figura R.5, encontre as probabilidades de que uma variável aleatória de distribuição de Poisson com $\lambda = 1{,}6$ vá tomar

(a) um valor menor do que 3;

(b) o valor 3, 4 ou 5;

(c) um valor maior do que 4.

R.116 Repita o Exercício R.115 usando a parte inferior do impresso na Figura R.5, a saber, as probabilidades cumulativas.

Snow666

R.117 Use a parte superior da Figura R.5 para calcular a média da distribuição de Poisson com $\lambda = 1{,}6$ verificando, assim, a fórmula $\mu = \lambda$.

R.118 Use a parte superior da Figura R.5 e a fórmula de cálculo à página 199 para calcular a variância da distribuição de Poisson com $\lambda = 1{,}6$ verificando, assim, a fórmula $\sigma^2 = \lambda$.

R.119 Qual é o fator de correção para população finita se

(a) $N = 120$ e $n = 30$;
(b) $N = 400$ e $n = 50$?

R.120 Determine, em cada caso, se as seguintes funções podem ser distribuições de probabilidade (definidas, em cada caso, para os valores dados de x) e explique suas respostas:

(a) $f(x) = \frac{1}{8}$ para $x = 0, 1, 2, 3, 4, 5, 6$ e 7;

(b) $f(x) = \frac{x+1}{16}$ para $x = 1, 2, 3$ e 4;

(c) $f(x) = \frac{(x-1)(x-2)}{20}$ para $x = 2, 3, 4$ e 5.

R.121 O número de flores num cacto raro é uma variável aleatória de distribuição de Poisson com $\lambda = 2{,}3$. Qual é a probabilidade de que tal planta vá ter

(a) nenhuma flor;
(b) pelo menos duas flores?

R.122 As probabilidades de que 0, 1, 2, 3, 4 ou 5 dos pacientes de um médico peguem a gripe durante a semana de férias de fim de ano do médico são, respectivamente, de 0,22; 0,34; 0,24; 0,13; 0,06 e 0,01.

(a) Encontre a média dessa distribuição de probabilidade.

Função Densidade de Probabilidade

```
Poisson with mu = 1.60000

        x   P( X = x )
     0.00       0.2019
     1.00       0.3230
     2.00       0.2584
     3.00       0.1378
     4.00       0.0551
     5.00       0.0176
     6.00       0.0047
     7.00       0.0011
     8.00       0.0002
     9.00       0.0000
```

Função Distribuição Cumulativa

```
Poisson with mu = 1.60000

        x   P( X <= x )
     0.00       0.2019
     1.00       0.5249
     2.00       0.7834
     3.00       0.9212
     4.00       0.9763
     5.00       0.9940
     6.00       0.9987
     7.00       0.9997
     8.00       1.0000
     9.00       1.0000
```

Figura R.5 Distribuição de Poisson com $\lambda = 1{,}6$.

Snow666

(b) Use a fórmula de cálculo para determinar a variância dessa distribuição de probabilidade.

***R.123** Dentre 80 pessoas entrevistadas para ocuparem certos empregos numa repartição pública, 40 são casadas, 20 são solteiras, 10 divorciadas e 10 viúvas. De quantas maneiras podemos extrair dessas pessoas entrevistadas uma amostra estratificada de 10 por cento se:

(a) um quarto da amostra deve ser alocado a cada grupo;
(b) a alocação deve ser proporcional?

R.124 Use a aproximação normal da distribuição binomial para aproximar as probabilidades de uma variável aleatória de distribuição binomial com $n = 18$ e $p = 0{,}27$ tomar um valor

(a) menor do que 6;
(b) entre 4 e 8;
(c) maior do que 6.

R.125 Use a parte superior da Figura R.6 para determinar as probabilidades de que uma variável aleatória de distribuição binomial com $n = 12$ e $p = 0{,}46$ vá tomar

(a) o valor 6, 7 ou 8;
(b) um valor maior do que ou igual a 9.

Função Densidade de Probabilidade

Binomial with n = 12 and p = 0.460000

x	P (X = x)
0.00	0.0006
1.00	0.0063
2.00	0.0294
3.00	0.0836
4.00	0.1602
5.00	0.2184
6.00	0.2171
7.00	0.1585
8.00	0.0844
9.00	0.0319
10.00	0.0082
11.00	0.0013
12.00	0.0001

Função Distribuição Cumulativa

Binomial with n = 12 and p = 0.460000

x	P (X <= x)
0.00	0.0006
1.00	0.0069
2.00	0.0363
3.00	0.1199
4.00	0.2802
5.00	0.4986
6.00	0.7157
7.00	0.8742
8.00	0.9585
9.00	0.9905
10.00	0.9986
11.00	0.9999
12.00	1.0000

Figura R.6
Distribuição binomial com $n = 12$ e $p = 0{,}46$.

Snow666

R.126 Use a parte inferior da Figura R.6 para repetir o Exercício R.125.

R.127 Verifique, em cada caso, se estão satisfeitas as condições para a aproximação normal da distribuição binomial:

(a) $n = 55$ e $p = \frac{1}{5}$;

(b) $n = 105$ e $p = \frac{1}{35}$;

(c) $n = 210$ e $p = \frac{1}{30}$;

(d) $n = 40$ e $p = 0{,}95$.

R.128 Qual é a probabilidade de cada amostra possível se são extraídas amostras aleatórias de tamanho $n = 3$ de uma população finita de tamanho $N = 70$?

R.129 Dentre 12 moedas do império, sete são autênticas, três são falsificações banhadas em ouro e as duas restantes são falsificações em bronze puro. Se um comprador leigo escolher três dessas moedas, qual é a probabilidade de que ele pegue uma de cada tipo?

R.130 Encontre a média da distribuição binomial com $n = 9$ e $p = 0{,}40$, usando

(a) a Tabela V e a fórmula que define μ;

(b) a fórmula especial para a média de uma distribuição binomial.

Snow666

11

PROBLEMAS DE ESTIMATIVA

11.1 Estimativa de Médias 271

11.2 Estimativa de Médias (σ Desconhecido) 275

11.3 Estimativa de Desvios-Padrão 281

11.4 Estimativa de Proporções 286

11.5 Lista de Termos-Chave 292

11.6 Referências 292

Os problemas de inferência estatística tradicionalmente tem sido classificados como problemas de **estimativa**, em que tentamos determinar (dentro de limites razoáveis) os valores de parâmetros populacionais; ou como **testes de hipóteses**, em que aceitamos ou rejeitamos afirmações sobre os parâmetros ou formas de populações; ou ainda problemas de **previsão**, em que prevemos valores futuros de variáveis aleatórias. Em cada caso, as inferências são baseadas em dados amostrais, embora em alguns métodos, os assim denominados **métodos bayesianos**, não tratados neste livro, as inferências também são baseadas em informação colateral e/ou julgamentos subjetivos. Neste capítulo, enfocamos os problemas de estimativa. Os testes de hipóteses são tratados em capítulos subseqüentes, e os problemas de previsão são estudados nos Capítulos 16 e 17.

Os problemas de estimativas são fáceis de exemplificar porque surgem praticamente em qualquer lugar, na ciência, na administração e na vida cotidiana. Em ciência, um psicólogo pode querer determinar o tempo médio de reação de um adulto a um estímulo visual; em administração, um sindicalista pode querer saber quanta variabilidade existe no tempo que os operários filiados ao sindicato levam para chegar ao trabalho; e na vida cotidiana, podemos querer descobrir qual é a percentagem de todos acidentes envolvendo um único carro causados por cansaço do motorista. Nesses três exemplos, o problema do psicólogo diz respeito a uma média populacional, o problema do sindicalista diz respeito a uma medida de variação (talvez um desvio-padrão), e nosso problema de vida cotidiana aborda uma percentagem. Conceitualmente, todos esses problemas são tratados da mesma maneira, mas há diferenças nos métodos específicos que são utilizados. Os métodos de estimar médias populacionais ocupam as Seções 11.1 e 11.2; os relacionados a medições da variabilidade são tratados na Seção 11.3 e os relacionados a estimativas de percentagens (também de proporções e probabilidades) são discutidos na Seção 11.4.

Snow666

11.1 ESTIMATIVA DE MÉDIAS

Para ilustrar alguns dos problemas com que nos deparamos quando estimamos a média de uma população a partir de dados amostrais, vamos recorrer a um estudo em que planejadores industriais procuram determinar o tempo médio que um adulto leva para montar um brinquedo que é "fácil de montar". Com uma amostra aleatória, eles obtêm os dados seguintes (em minutos) para 36 pessoas que montaram o brinquedo:

17	13	18	19	17	21	29	22	16	28
21	15	26	23	24	20	8	17	17	21
32	18	25	22	16	10	20	22	19	14
30	22	12	24	28	11				

A média dessa amostra é $\bar{x} = 19{,}9$ minutos e, na ausência de qualquer outra informação, podemos tomar esse número como uma estimativa de μ, o "verdadeiro" tempo médio que um adulto leva para montar aquele brinquedo.

Esse tipo de estimativa é denominado **estimativa pontual**, pois consiste em um único número, ou um único ponto na escala dos números reais. Embora se trate da forma mais comum de expressar estimativas, ela deixa margem para não poucas questões. O número em si não nos diz em quanta informação se baseia a estimativa, nem tampouco nos informa sobre o tamanho possível do erro. E, naturalmente, devemos esperar um erro. Isso deveria ter ficado claro de nossa discussão da distribuição amostral da média no Capítulo 10, em que vimos que as flutuações aleatórias da média (e, conseqüentemente, sua confiabilidade como estimativa de μ) dependem de duas coisas: do tamanho da amostra e do tamanho do desvio-padrão populacional σ. Assim, poderíamos suplementar a estimativa, $\bar{x} = 19{,}9$ minutos, com a informação de que se trata da média de uma amostra aleatória de tamanho $n = 36$, cujo desvio-padrão é $s = 5{,}73$ minutos. Embora isso não nos dê o valor efetivo de σ, o desvio-padrão amostral pode servir como uma estimativa desse parâmetro.

Muitas vezes, os relatórios científicos apresentam as médias amostrais dessa maneira, juntamente com os valores de n e s, mas isso não fornece um quadro coerente aos leitores do relatório, a menos que eles tenham algum treinamento formal em Estatística. Para levar isso em conta, utilizamos a teoria das Seções 10.7 e 10.8 e a definição no Exemplo 9.6, segundo as quais z_α é tal que a área à sua direita sob a curva normal padrão é α e, portanto, a área sob a curva normal padrão entre $-z_{\alpha/2}$ e $z_{\alpha/2}$ é igual a $1 - \alpha$. Levando em conta que, para amostras aleatórias grandes de populações infinitas, a distribuição amostral da média é aproximadamente uma distribuição normal com $\mu_{\bar{x}} = \mu$ e $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, vemos, pela Figura 11.1, que é de $1 - \alpha$ a probabilidade de que a média de uma amostra aleatória grande de uma população infinita vá diferir da média populacional por no máximo $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$. Em outras palavras,

Quando utilizamos $\bar{x}$ como uma estimativa de μ, a probabilidade é $1 - \alpha$ de que essa estimativa vá diferir para um ou para o outro lado por, no máximo

ERRO MÁXIMO DE ESTIMATIVA

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Esse resultado é aplicável se n for grande, $n \geq 30$, e a população for infinita (ou suficientemente grande para não ser preciso aplicar o fator de correção para população finita). Os dois valores mais comumente usados, embora não necessariamente, para $1 - \alpha$ são 0,95 e 0,99, e os valores correspondentes de $\alpha/2$ são 0,025 e 0,005. Conforme foi solicitado ao leitor mostrar no Exercício 9.16, temos $z_{0{,}025}$ e $z_{0{,}005} = 2{,}575$.

Snow666

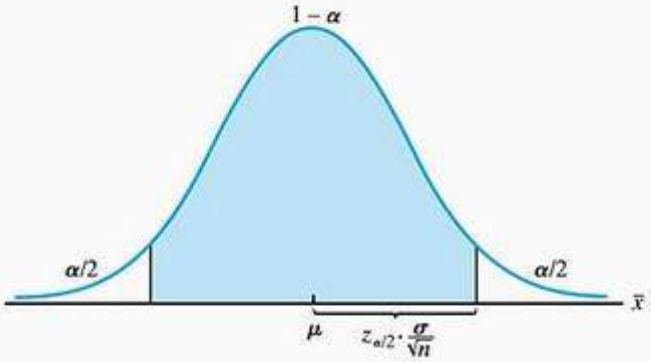

Figura 11.1
Distribuição amostral da média.

EXEMPLO 11.1 Um grupo de técnicos em eficiência pretende utilizar a média de uma amostra aleatória de tamanho $n = 150$ para estimar a aptidão mecânica média (avaliada por certo teste padronizado) dos operários da linha de montagem de uma grande indústria. Se, com base na experiência, os técnicos podem supor que $\sigma = 6{,}2$ para tais dados, o que podem eles afirmar, com uma probabilidade de 0,99, sobre o erro máximo de sua estimativa?

Solução Substituindo $n = 150$, $\sigma = 6{,}2$ e $z_{0{,}005} = 2{,}575$ na fórmula de E, temos

$$E = 2{,}575 \cdot \frac{6{,}2}{\sqrt{150}} \approx 1{,}30$$

Assim, os técnicos em eficiência podem afirmar, com 0,99 de probabilidade, que seu erro será de, no máximo, 1,30.

Suponhamos, agora, que os técnicos coletem efetivamente os dados necessários para que obtenham $\bar{x} = 69{,}5$. Podem eles ainda afirmar, com 0,99 de probabilidade, que o erro de sua estimativa é, no máximo, 1,30? Afinal de contas, $\bar{x} = 69{,}5$ difere de μ, a média populacional, por no máximo 1,30, ou não, e os técnicos não têm como saber se é um ou o outro. Na verdade, eles podem fazer essa afirmação, mas deve ficar entendido que a probabilidade de 0,99 se aplica ao *método* utilizado (obtenção dos dados amostrais e utilização da fórmula de E), e não diretamente ao problema particular que está sendo tratado.

Para fazer essa distinção, tornou-se praxe usar aqui a palavra **confiança**, em vez de "probabilidade".

Em geral, fazemos afirmações probabilísticas sobre valores futuros de variáveis aleatórias (digamos, o erro potencial de uma estimativa) e afirmações de confiança, assim que os dados tenham sido obtidos.

Conseqüentemente, em nosso exemplo, diríamos que os técnicos em eficiência podem ter uma confiança de 99% em que o erro de sua estimativa, $\bar{x} = 69{,}5$, seja de 1,30, no máximo.

O uso da fórmula de E para o erro máximo envolve uma complicação. Precisamos saber o valor do desvio-padrão populacional σ e isso só ocorre raramente. Assim, podemos substituí-lo por um palpite plausível e, sendo conservativos, podemos ser levados a sobrestimar o erro. Alternativamente, podemos substituir σ com uma estimativa, comumente o desvio-padrão amostral s. Em geral, isso é considerado razoável, desde que o tamanho da amostra seja suficientemente grande, e com isso queremos dizer, novamente, $n \geq 30$.

Snow666

EXEMPLO 11.2 Com referência ao exemplificado na página 271, o que podemos afirmar, com 95% de confiança, sobre o erro máximo, se tomarmos $\bar{x} = 19{,}9$ minutos como uma estimativa do tempo médio que um adulto leva para montar o dado tipo de brinquedo?

Solução Substituindo $n = 36$, $s = 5{,}73$ no lugar de σ, e $z_{0{,}025} = 1{,}96$ na fórmula de E, vemos que é possível afirmar, com 95% de confiança, que o erro é, no máximo,

$$E = 1{,}96 \cdot \frac{5{,}73}{\sqrt{36}} \approx 1{,}87 \text{ minuto}$$

Naturalmente, o erro é de, no máximo, 1,87 minuto, ou não, e não sabemos qual é o caso, mas se tivéssemos que apostar, uma aposta equilibrada de que o erro é, no máximo, de 1,87 minuto, seria de 95 contra 5 (ou 19 contra 1).

Pode-se aplicar a fórmula de E também para determinar o tamanho da amostra que é necessário para atingir um grau de precisão desejado. Suponha que queiramos utilizar a média de uma amostra aleatória grande para estimar a média de uma população, e que pretendamos ser capazes de afirmar, com $1 - \alpha$ de probabilidade, que o erro dessa estimativa não excederá uma quantidade pré-determinada E. Como anteriormente, escrevemos $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ e, resolvendo essa equação em n, obtemos

TAMANHO DA AMOSTRA PARA ESTIMAR μ

$$n = \left[\frac{z_{\alpha/2} \cdot \sigma}{E}\right]^2$$

EXEMPLO 11.3 A diretora de uma faculdade pretende utilizar a média de uma amostra aleatória para estimar o tempo médio que os alunos gastam para ir de uma sala de aula a outra, e ela quer ser capaz de afirmar com 0,95 de probabilidade que seu erro será de no máximo 0,30 minuto. Se, por estudos anteriores, ela sabe que é razoável supor $\sigma = 1{,}50$ minuto, qual é o tamanho da amostra de que ela necessita?

Solução Substituindo $E = 0{,}30$, $\sigma = 1{,}50$ e $z_{0{,}025} = 1{,}96$ na fórmula de n, obtemos

$$n = \left(\frac{1{,}96 \cdot 1{,}50}{0{,}30}\right)^2 \approx 96{,}04$$

que arredondamos até o inteiro mais próximo, 97. Assim, é requerida uma amostra aleatória de tamanho $n = 97$ para a estimativa. (Note que o tratamento teria sido o mesmo se tivéssemos dito "ela quer ser capaz de afirmar com 95% de confiança que seu erro *é*, no máximo, de 0,30 minuto", em vez de "ela quer ser capaz de afirmar com 0,95 de probabilidade que seu erro *será* de no máximo 0,30 minuto". Tudo depende de quando a afirmativa é feita, antes ou depois de ela coletar os dados.)

Como pode ser visto na fórmula para n e também no Exemplo 11.3, esse método tem o mesmo inconveniente que a fórmula para E, ou seja, que precisamos saber (pelo menos aproximadamente) o valor do desvio-padrão populacional, σ. Por esse motivo, às vezes começamos com uma amostra relativamente pequena e então usamos seu desvio-padrão para ver se serão necessários mais dados.

Vamos, agora, introduzir uma maneira diferente de apresentar uma média amostral juntamente com uma avaliação do erro que poderíamos cometer se a utilizássemos para estimar a média da população da qual provém a amostra. Tal como na página 271, lançamos mão do fato de que, pa-

Snow666

ra amostras aleatórias grandes de populações infinitas, a distribuição amostral da média é aproximadamente uma distribuição normal com a média $\mu_{\bar{x}} = \mu$ e desvio padrão $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, de forma que

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

é um valor de uma variável aleatória que tem aproximadamente a distribuição normal padrão. Como há uma probabilidade de $1 - \alpha$ de que uma variável aleatória com essa distribuição vá assumir um valor entre $-z_{\alpha/2}$ e $z_{\alpha/2}$, a saber, que

$$-z_{\alpha/2} < z < z_{\alpha/2}$$

podemos substituir nessa desigualdade a expressão precedente de z e obter

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

Agora, se multiplicarmos cada termo por $\sigma/\sqrt{n}$,, subtrairmos $\bar{x}$ de cada termo e, finalmente, multiplicarmos cada termo por -1, obteremos

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

[Como multiplicamos por -1, tivemos que reverter os sinais das desigualdades, como é sempre o caso quando multiplicamos as expressões em ambos lados de uma desigualdade por um número negativo. Por exemplo, enquanto 4 é maior do que (está à direita de) 3, -4 é menor do que (está à esquerda de) -3.] O resultado que acabamos de obter também pode ser escrito como

INTERVALO DE CONFIANÇA PARA μ

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

e podemos afirmar, com probabilidade $1 - \alpha$, que ela será satisfeita por qualquer amostra dada. Em outras palavras, podemos afirmar com $(1 - \alpha)100\%$ de confiança que o intervalo de $\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ a $\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$, determinado a partir de uma amostra aleatória grande, contém a média populacional que estamos procurando estimar. Quando σ é desconhecido e n é, no mínimo, 30, substituímos σ pelo desvio-padrão amostral s.

Um intervalo como esse é denominado **intervalo de confiança**, suas extremidades são os **limites de confiança**, e $1 - \alpha$ ou $(1 - \alpha)100\%$ é denominado o **grau de confiança**. Como anteriormente, os valores mais usados para o grau de confiança são 0,95 e 0,99 (ou 95% e 99%), e os valores correspondentes de $z_{\alpha/2}$ são 1,96 e 2,575. Ao contrário do que ocorre com as estimativas pontuais, as estimativas dadas sob forma de intervalo de confiança são chamadas **estimativas intervalares**. Elas têm a vantagem de não exigir maiores elaborações sobre sua confiabilidade. Esta é garantida, indiretamente, por sua amplitude e pelo grau de confiança.

Como n deve ser grande para justificar a aproximação normal da distribuição amostral da média, denominamos um intervalo calculado por meio da fórmula precedente como um **intervalo de confiança de grandes amostras** para μ. Também é denominado **intervalo z**, por ser baseado na estatística z que tem a distribuição normal padrão.

EXEMPLO 11.4 Com referência ao Exemplo 11.1, no qual tínhamos $n = 150$ e $\sigma = 6{,}2$, use a informação adicional de que os técnicos em eficiência obtiveram a média amostral $\bar{x} = 69{,}5$ para calcular um intervalo de 95% de confiança para a aptidão mecânica média dos operários da linha de montagem da dada indústria.

Snow666

Solução Substituindo $n = 150$, $\sigma = 6{,}2$, $\bar{x} = 69{,}5$ e $z_{0{,}025} = 1{,}96$ na fórmula do intervalo de confiança, obtemos

$$69{,}5 - 1{,}96 \cdot \frac{6{,}2}{\sqrt{150}} < \mu < 69{,}5 + 1{,}96 \cdot \frac{6{,}2}{\sqrt{150}}$$

$$68{,}5 < \mu < 70{,}5$$

onde os limites de confiança foram arredondados até a primeira casa decimal. É claro que a afirmação de que o intervalo de 68,5 a 70,5 contém o verdadeiro escore médio de aptidão mecânica dos operários da linha de montagem da dada indústria é ou verdadeira ou falsa e não se sabe se é verdadeira ou falsa, mas podemos ter 95% de confiança em que seja verdadeira. Por quê? Porque o método que usamos funciona 95% das vezes. Dito de outra forma, o intervalo pode conter, ou não, μ, mas, se tivéssemos que apostar, uma aposta equilibrada de que o intervalo contém, seria de 95 contra 5 (ou 19 contra 1).

Se quiséssemos construir um intervalo de 99% de confiança no Exemplo 11.4, deveríamos tomar 2,575 no lugar de 1,96 para $z_{\alpha/2}$, e obteríamos $68{,}2 < \mu < 70{,}8$. O intervalo de 99% de confiança é mais amplo do que o intervalo de 95% de confiança, indo de 68,2 a 70,8 em vez de ir de 68,5 a 70,5, assim ilustrando o fato importante de que

Quando aumentamos o grau de confiança, o intervalo de confiança se torna mais amplo, dando-nos, assim, menos informação sobre a grandeza que estamos procurando estimar.

Na verdade, podemos dizer que, "quanto mais seguros quisermos ficar, de menos coisas podemos estar seguros".

11.2 ESTIMATIVA DE MÉDIAS (σ DESCONHECIDO)

Na Seção 11.1, admitimos que as amostras fossem suficientemente grandes, $n \geq 30$, para podermos aproximar a distribuição amostral da média por uma distribuição normal e, quando necessário, substituir σ por s. Para estabelecer métodos correspondentes que se apliquem em geral quando σ é desconhecido, precisamos supor que as populações das quais estamos extraindo amostras tenham aproximadamente a forma de distribuições normais. Podemos, então, basear nossos métodos na **estatística *t*** dada por

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

que é um valor de uma variável aleatória com a **distribuição *t***. Mais especificamente, essa distribuição é denominada a **distribuição *t* de Student**, ou **distribuição de Student**, porque foi estabelecida originalmente por um estatístico, W. S. Gosset, que publicava seus trabalhos sob o pseudônimo de "Student" ("estudante", em inglês). Conforme mostra a Figura 11.2, a forma dessa distribuição contínua é muito semelhante à da distribuição normal padrão, tendo, como esta, a forma de um sino e sendo simétrica em relação à média zero. A forma exata da distribuição t depende de um parâmetro denominado **número de graus de liberdade**, ou simplesmente, **graus de liberdade**, o qual, para os métodos desta seção, é igual a $n - 1$, o tamanho da amostra menos 1.

Para a distribuição normal padrão, definimos $z_{\alpha/2}$ de tal forma que a área sob a curva à sua direita seja igual a $\alpha/2$ e, portanto, a área sob a curva entre $-z_{\alpha/2}$ e $z_{\alpha/2}$ seja igual a $1 - \alpha$. Como se vê na Figura 11.3, os valores correspondentes para a distribuição t são $-t_{\alpha/2}$ e $t_{\alpha/2}$. Como esses valores dependem de $n - 1$, o número de graus de liberdade, eles devem ser obtidos numa tabela es-

Snow666

Figura 11.2
Distribuição normal padrão e distribuição t.

pecial, como a Tabela II no final do livro, ou talvez, num computador. A Tabela II contém, entre outros, os valores de $t_{0,025}$ e $t_{0,005}$ de 1 a 30 graus de liberdade e alguns valores selecionados de graus de liberdade maiores. Como pode ser visto, $t_{0,025}$ e $t_{0,005}$ aproximam os valores correspondentes da distribuição normal padrão quando o número de graus de liberdade se torna grande.

Procedendo como nas páginas 273/274, podemos afirmar com probabilidade $1 - \alpha$ de que uma variável aleatória com distribuição t vá assumir um valor entre $-t_{\alpha/2}$ e $t_{\alpha/2}$, isto é, que

$$-t_{\alpha/2} < t < t_{\alpha/2}$$

Então, substituindo isso na desigualdade da expressão de t à página 275, obtemos

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

e os mesmos passos que usamos nas páginas 274 e 275 fornecem o seguinte intervalo de confiança para μ:

INTERVALO DE CONFIANÇA PARA μ (σ DESCONHECIDO)

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

O grau de confiança é $1 - \alpha$ e a única diferença entre esse intervalo de confiança e o intervalo z com s substituído por σ é que $t_{\alpha/2}$ toma o lugar de $z_{\alpha/2}$. Esse intervalo de confiança para μ costuma ser denominado **intervalo t** e, como a maioria das tabelas da distribuição t dão os valores de $t_{\alpha/2}$ somente para um número pequeno de graus de liberdade, também é conhecido como um **in-**

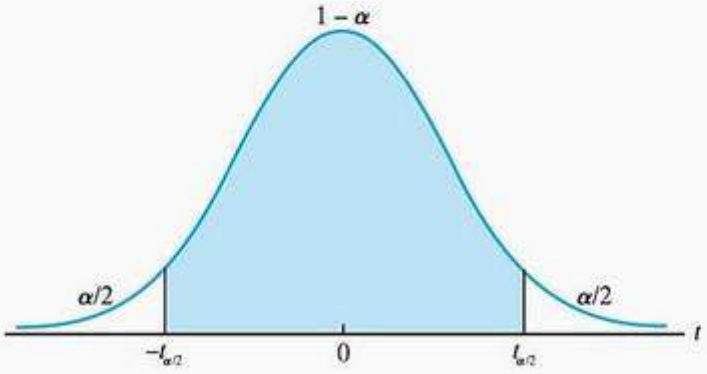

Figura 11.3
Distribuição t.

Snow666

tervalo de confiança para pequenas amostras de μ. Hoje em dia, os computadores e outros recursos tecnológicos fornecem os valores de $t_{\alpha/2}$ para centenas de graus de liberdade, de modo que essa distinção não é mais relevante.

Deve ser lembrado, entretanto, que para o intervalo t existe a hipótese adicional de que a amostra tenha sido extraída de uma população normal, ou pelo menos de uma população que tenha aproximadamente o formato de uma distribuição normal. Isso é importante e será discutido mais nos dois exemplos que seguem.

No Exemplo 11.5 receberemos somente os valores de n, $\bar{x}$ e s, mas não os dados originais, de modo que realmente não há como conferir a "normalidade" da população na amostra. Tudo que podemos fazer nesse caso é esperar pelo melhor. No Exemplo 11.6 receberemos os próprios dados, de modo que, agora, poderemos formar um gráfico de probabilidade normal (ver Seção 9.3) para julgar se é razoável interpretar os dados como uma amostra de uma população normal. Isso requererá o uso de tecnologias apropriadas (aplicativos de computador ou calculadora gráfica) mas há procedimentos alternativos que costumam ser ensinados somente em disciplinas mais avançadas de Estatística.

EXEMPLO 11.5 Durante a execução de determinada tarefa sob condições simuladas de imponderabilidade, a média da taxa de batimentos cardíacos de 12 astronautas aumentou em 27,33 batimentos por minuto, com desvio-padrão de 4,28 batimentos por minuto. Construa um intervalo de 99% de confiança para o verdadeiro aumento médio da taxa de batimentos cardíacos dos astronautas no desempenho daquela tarefa (nas mesmas condições enunciadas).

Solução Como já observamos anteriormente, sem os próprios dados não há como julgar a normalidade da população estudada. Mesmo assim, deixando claro que o resultado está sujeito à validação da suposição, procedemos como segue. Substituindo $n = 12$, $\bar{x} = 27{,}33$, $s = 4{,}28$ e $t_{0{,}005} = 3{,}106$ (a entrada na Tabela II para $12 - 1 = 11$ graus de liberdade) na fórmula do intervalo t, obtemos

$$27{,}33 - 3{,}106 \cdot \frac{4{,}28}{\sqrt{12}} < \mu < 27{,}33 + 3{,}106 \cdot \frac{4{,}28}{\sqrt{12}}$$

e, portanto,

$$23{,}49 < \mu < 31{,}17$$

batimentos por minuto.

No Exemplo 11.6 a seguir vamos nos referir à exemplificação na Seção 9.3, em que já verificamos que os dados podem ser tratados como uma amostra aleatória de uma população normal. Naquela ilustração, que tratava da durabilidade de uma tinta utilizada pelo órgão governamental que trata da manutenção das rodovias para pintar as linhas divisórias de tráfego.

EXEMPLO 11.6 À página 227, mostramos que nos oito locais a pintura deteriorou-se depois de ter sido cruzada por 14,26; 16,78; 13,65; 11,53; 12,64; 13,37; 15,60 e 14,94 milhões de veículos, e que esses são os dados para os quais os gráficos de probabilidade normal das Figuras 9.16 e 9.17 mostraram que podem ser tratados como uma amostra aleatória de uma população normal. Todos esses números estão arredondados até a segunda casa decimal, bem como sua média $\bar{x} = 14{,}10$ e seu desvio-padrão $s = 1{,}67$ milhões de veículos. Calcule um intervalo de 95% de confiança para a média da população considerada.

Solução Substituindo $\bar{x} = 14{,}10$, $s = 1{,}67$, $n = 8$ e $t_{0{,}025} = 2{,}365$ para $8 - 1 = 7$ graus de liberdade na fórmula do intervalo t, obtemos

Snow666

$$14{,}10 - 2{,}365 \cdot \frac{1{,}67}{\sqrt{8}} < \mu < 14{,}10 + 2{,}365 \cdot \frac{1{,}67}{\sqrt{8}}$$

$$12{,}70 < \mu < 15{,}50$$

milhões de veículos cruzando.

Esse resultado é o intervalo de 95% de confiança desejado para o valor médio do tráfego (número de veículos que cruzam) que a tinta pode suportar antes de se deteriorar. Novamente, não podemos afirmar com certeza se o intervalo de 12,70 milhões a 15,50 milhões contém o verdadeiro número médio de veículos que a tinta pode suportar antes de se deteriorar, mas 95 contra 5 seria uma aposta equilibrada de que vai se deteriorar. Essas chances se baseiam no fato de que o método utilizado, o de extrair uma amostra aleatória de uma população normal e aplicar a fórmula anteriormente dada, funciona em 95% das vezes.

O método usado anteriormente para determinar o erro máximo que corremos com $(1 - \alpha)100\%$ de confiança, quando usamos uma média amostral para estimar a média de uma população, é facilmente adaptado a problemas em que σ é desconhecido, desde que a população amostrada tenha aproximadamente a forma de uma distribuição normal. Tudo que precisamos fazer é substituir s no lugar de σ e $t_{\alpha/2}$ no lugar de $z_{\alpha/2}$ na fórmula do erro máximo à página 271.

EXEMPLO 11.7 Com referência ao Exemplo 11,5, suponha que $\bar{x} = 27{,}33$ esteja sendo usada como uma estimativa do verdadeiro aumento médio da taxa de batimentos cardíacos de astronautas no desempenho da tarefa dada. O que pode ser dito sobre o erro máximo do aumento médio com 99% de confiança?

Solução Substituindo $s = 4{,}28$, $n = 12$ e $t_{0{,}005} = 3{,}106$ (a entrada na Tabela II para $12 - 1 = 11$ graus de liberdade) na fórmula modificada de E, obtemos

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 3{,}106 \cdot \frac{4{,}28}{\sqrt{12}} \approx 3{,}84$$

Assim, se usarmos $\bar{x} = 27{,}33$ batimentos por minuto como estimativa do verdadeiro aumento médio da taxa de batimentos cardíacos dos astronautas no desempenho da tarefa dada (sob as condições indicadas), podemos afirmar, com 99% de confiança, que nosso erro é, no máximo, de 3,84 batimentos por minuto.

EXERCÍCIOS

11.1 Um estudo feito por um oficial de uma divisão de veículos blindados mostrou que numa amostra aleatória de $n = 40$ dias, a divisão possuía, em média, 1.126 veículos em condições operacionais. Sabendo que $\sigma = 135$ para tais dados, o que esse oficial pode afirmar com 95% de confiança sobre o erro máximo se ele utilizar $\bar{x} = 1.126$ como uma estimativa do verdadeiro número médio diário de veículos que essa divisão de veículos blindados possui em condições operacionais?

11.2 Com referência ao Exercício 11.1, construa um intervalo de 99% de confiança para o verdadeiro número médio diário de veículos que essa divisão de veículos blindados possui em condições operacionais.

11.3 Um estudo sobre o crescimento anual de certo tipo de orquídeas mostrou que (sob condições controladas) uma amostra aleatória de $n = 40$ dessas orquídeas cresceu uma média de 32,5 mm por ano. Sabendo que $\sigma = 3{,}2$ mm para tais dados, o que podemos concluir com 99% de confiança sobre o erro máximo se $\bar{x} = 35$ mm for usado como uma estimativa do

Snow666

verdadeiro crescimento anual médio desse tipo de orquídea (sob as condições controladas dadas)?

11.4 Com referência ao Exercício 11.3, construa um intervalo de 98% de confiança para o verdadeiro crescimento anual médio desse tipo de orquídea.

11.5 Num estudo de custos de seguro contra colisão de automóveis, uma amostra de $n = 35$ consertos de colisões frontais contra um muro a uma velocidade específica teve um custo médio de 1.438 unidades monetárias. Sabendo que $\sigma = 269$ unidades monetárias para esses dados, o que pode ser dito com 98% de confiança sobre o erro máximo se $\bar{x} = 1.438$ unidades monetárias for usado como uma estimativa do custo médio de tais consertos? Também construa um intervalo de 90% de confiança para o custo médio de tais consertos.

11.6 O que acontece com o erro-padrão da média quando o tamanho da amostra é reduzido de 288 para 32?

11.7 Uma amostra aleatória de tamanho 64 deve ser extraída de uma população que é suficientemente grande para ser tratada como sendo infinita. Sabendo que a média e o desvio-padrão dessa população são $\mu = 23{,}5$ e $\sigma = 3{,}3$, encontre a probabilidade de que a média dessa amostra vá cair entre 23,0 e 24,0.

11.8 Antes de apresentar sua proposta, um empreiteiro quer ter 95% de confiança em que seu erro máximo não é maior do que 6 minutos, ao tomar a média de uma amostra aleatória para estimar o tempo médio necessário para o endurecimento de certo tipo de tijolo cru. Qual deve ser o tamanho da amostra que ele necessita se ele puder admitir que $\sigma = 22$ minutos é o tempo que tais tijolos levam para endurecer?

11.9 Deseja-se estimar o número médio de horas de vôo contínuo que um certo modelo de avião leva até precisar de consertos. Se pudermos supor que $\sigma = 138$ horas para esses dados, qual deve ser o tamanho da amostra que precisamos para poder afirmar com uma probabilidade de 0,99 que o erro da média amostral é no máximo de 40 horas?

11.10 Antes de comprar um grande lote de carne de porco moída, um salsicheiro quer "ter certeza" de que seu erro não é maior do que 2,5 gramas ao usar a média de uma amostra aleatória para estimar o real conteúdo de gordura por quilo. Se o desvio-padrão do conteúdo de gordura é sabidamente de 7,7 gramas por quilo, quantas amostras de um quilo ele necessita se ele interpreta "ter certeza" por ter 95% de confiança.

11.11 Refaça o Exercício 11.10 trocando "ter certeza" por ter 99% de confiança.

11.12 Quando uma amostra compreende uma porção razoável de uma população finita, devemos usar a fórmula do erro-padrão à página 254 e, portanto, para determinar o erro máximo devemos usar a fórmula

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Tomando uma amostra aleatória de $n = 200$ dentre $N = 800$ contas fraudulentas, uma perita-contadora investigando uma companhia de energia descobriu que as quantias devidas tinham uma média de 48,15 unidades monetárias com desvio-padrão de 6,19 unidades monetárias. Usando $s = 6{,}19$ unidades monetárias como estimativa de σ, o que ela pode afirmar com 95% de confiança sobre o erro máximo se utilizar 48,15 como uma estimativa da quantidade devida por todas as 800 contas fraudulentas?

11.13 Um computador é programado para dar valores de uma variável aleatória com uma distribuição normal, da qual só o programador conhece a média e o desvio-padrão. Pede-se a

Snow666

cada um de 30 alunos que use o computador para simular uma amostra aleatória de tamanho $n = 5$ e utilizá-la para construir em seguida um intervalo de 90% de confiança para μ. Os resultados obtidos pelos alunos são os seguintes:

$6{,}30 < \mu < 8{,}26$,	$6{,}50 < \mu < 7{,}72$,	$6{,}93 < \mu < 8{,}01$,
$6{,}60 < \mu < 8{,}00$,	$6{,}51 < \mu < 7{,}51$,	$6{,}82 < \mu < 8{,}66$,
$7{,}02 < \mu < 8{,}11$,	$6{,}94 < \mu < 7{,}64$,	$6{,}24 < \mu < 7{,}26$,
$6{,}87 < \mu < 8{,}17$,	$6{,}77 < \mu < 8{,}13$,	$6{,}14 < \mu < 6{,}82$,
$6{,}83 < \mu < 7{,}93$,	$6{,}66 < \mu < 8{,}10$,	$6{,}73 < \mu < 7{,}49$,
$6{,}41 < \mu < 7{,}67$,	$6{,}76 < \mu < 7{,}57$,	$6{,}97 < \mu < 7{,}47$,
$6{,}01 < \mu < 7{,}43$,	$7{,}15 < \mu < 7{,}89$,	$6{,}87 < \mu < 7{,}81$,
$7{,}35 < \mu < 7{,}99$,	$6{,}60 < \mu < 8{,}16$,	$6{,}47 < \mu < 7{,}81$,
$7{,}01 < \mu < 8{,}33$,	$6{,}97 < \mu < 7{,}55$,	$6{,}56 < \mu < 7{,}48$,
$7{,}13 < \mu < 8{,}03$,	$7{,}39 < \mu < 8{,}01$,	$5{,}98 < \mu < 7{,}68$.

(a) Quantos desses 30 intervalos de confiança poderíamos esperar que contivessem a verdadeira média da população amostrada?

(b) Considerando que o computador foi programado de forma que $\mu = 7{,}30$, quantos dos intervalos de confiança realmente contêm a média da população amostrada? Discuta o resultado.

11.14 Com referência ao Exemplo 11.4, em que usamos $n = 150$, $\bar{x} = 69{,}5$ e $\sigma = 6{,}2$, suponha que tivéssemos pedido um intervalo de 97% de confiança.

(a) Encontre o valor de $z_{0{,}015}$ pela Tabela I.

(b) Use o valor de $z_{0{,}015}$ obtido na parte (a) para calcular um intervalo z de 97% para a aptidão mecânica média dos operários da linha de montagem da dada indústria.

11.15 Num estudo da poluição do ar numa certa área do centro, um técnico da Agência de Proteção do Meio Ambiente obteve uma média de 2,34 microgramas de matéria orgânica solúvel em benzeno suspensa no ar, com desvio-padrão de 0,48 microgramas para uma amostra de tamanho $n = 9$. Supondo que a população amostrada é normal,

(a) construa um intervalo de 95% de confiança para a média dessa população;

(b) o que o técnico pode assegurar com 99% de confiança sobre o erro máximo se $\bar{x} =$ 2,34 microgramas por metro cúbico for usado como uma estimativa da média dessa população?

11.16 Um serviço de testes de produtos de consumo quer estudar o nível de ruído de um novo aspirador de pó. Medindo o nível de ruído de $n = 12$ desses aparelhos, obtém os seguintes dados, em decibéis: 74,0; 78,6; 76,8; 75,5; 73,8; 75,6; 77,3; 75,8; 73,9; 70,2; 81,0 e 73,9.

(a) Use um gráfico de probabilidade normal para verificar se é razoável tratar esses dados como uma amostra de uma população normal.

(b) Construa um intervalo de 95% de confiança para o nível de ruído médio de tais aspiradores de pó.

11.17 Uma amostra aleatória de $n = 9$ pedaços de cabo de manilha (para uso marítimo) tem uma resistência à ruptura média de 41,250 kgf e um desvio-padrão de 1,527 kgf. Supondo que é razoável tratar esses dados como uma amostra de uma população normal, o que pode ser afirmado com 95% de confiança sobre o erro máximo se $\bar{x} = 41{,}250$ kgf for usado como uma estimativa da resistência à ruptura média desses cabos?

11.18 Com referência ao Exercício 11.17, construa um intervalo de 98% de confiança para a resistência à ruptura média dos cabos dados.

11.19 Use a Tabela II para encontrar

(a) $t_{0{,}050}$ para 13 graus de liberdade;

Snow666

(b) $t_{0,025}$ para 18 graus de liberdade;
(c) $t_{0,010}$ para 22 graus de liberdade;
(d) $t_{0,005}$ para 15 graus de liberdade.

11.20 Dez mancais fabricados por determinado processo têm um diâmetro médio de 0,406 cm com desvio-padrão de 0,003 cm. Construa um intervalo de 99% de confiança para o diâmetro médio de mancais fabricados por esse processo. Suponha que é normal a população da qual foi extraída a amostra.

11.21 As medidas da eficiência térmica de $n = 15$ máquinas a diesel fabricadas por uma empresa renomada são as seguintes: 30,7; 35,0; 34,9; 33,6; 28,7; 32,1; 29,0; 31,4; 31,7; 31,8; 33,6; 29,7; 33,4; 28,2 e 31,6.

(a) Use um gráfico de probabilidade normal para verificar se é razoável tratar esses dados como uma amostra de uma população normal.
(b) Construa um intervalo de 99% de confiança para a eficiência térmica média de tais máquinas a diesel.

11.22 Cinco recipientes de um solvente comercial selecionados aleatoriamente de um lote de produção grande, pesaram 19,5; 19,3; 20,0; 19,0 e 19,7 quilogramas. Supondo que esses dados podem ser vistos como uma amostra de uma população normal, construa um intervalo t de 99% o peso médio dos recipientes de solvente do lote de produção.

11.23 Na digitação de um texto, um autor cometeu 10, 11, 14, 8, 12, 17, 9,12, 15 e 12 erros numa amostra aleatória de dez páginas. Supondo que é razoável aproximar a população da qual foi extraída essa amostra por uma distribuição normal, construa um intervalo de 98% de confiança para o número médio de tais erros que o autor comete por página.

11.24 Com referência ao Exercício 11.23, mude o grau de confiança para 0,93 e use um aplicativo de computador ou uma calculadora gráfica para refazer o exercício.

11.25 Em sete estações climáticas na Serra do Mar, o nível de precipitação durante uma tempestade de verão foi medido em 1,2; 1,4; 1,8; 2,0; 1,5; 1,2 e 1,4 cm. Supondo que é razoável tratar esses dados como uma amostra aleatória de uma população normal, construa um intervalo de 95% de confiança para o nível de precipitação médio na Serra do Mar durante aquela tempestade de verão.

11.26 Use um aplicativo de computador ou uma calculadora gráfica para refazer o Exercício 11.25 com o grau de confiança alterado para 0,98.

11.27 Durante uma visita rotineira a um presídio, uma dentista constata que 12 detentos, escolhidos aleatoriamente, necessitam de 2, 3, 6, 1, 4, 2, 4, 5, 0, 3, 5 e 1 obturações. Se ela supuser que esses dados constituem uma amostra de uma população que pode ser muito bem aproximada por uma distribuição normal, o que ela pode afirmar com 99% de confiança sobre o erro máximo cometido usando a média dessa amostra como uma estimativa do número médio de obturações de que os detentos desse presídio grande necessitam.

11.3 ESTIMATIVA DE DESVIOS-PADRÃO

Até aqui neste capítulo, estivemos aprendendo como julgar o erro máximo na estimativa da média de uma população e como construir um intervalo de confiança para a média populacional. Essas técnicas são importantes, pois seguidamente fazemos inferências sobre médias, mas ainda mais importante é o fato de que os conceitos nos quais elas se baseiam podem ser traduzidos para outros parâmetros populacionais.