9 Fundamentos dos Testes de Hipóteses: Testes para Uma Amostra

Objetivos do Aprendizado

Neste capítulo, você aprenderá:

- Os princípios básicos do teste de hipóteses
- Como utilizar testes de hipóteses para testar uma média aritmética ou uma proporção
- Os pressupostos de cada um dos procedimentos de testes de hipóteses, como avaliá-los, e as consequências caso eles sejam seriamente violados
- Como evitar as armadilhas envolvidas nos testes de hipóteses
- Questões éticas envolvidas em testes de hipóteses

UTILIZANDO A ESTATÍSTICA

@ Oxford Cereals, Parte II

Assim como no Capítulo 7, você se encontra novamente no papel de gerente da unidade de operações da Oxford Cereals. Você está encarregado de monitorar a quantidade contida em cada caixa de cereal. As especificações da empresa exigem uma média aritmética de 368 gramas de peso por caixa. É responsabilidade sua ajustar o processo quando a média aritmética do peso abastecido na população de caixas difere de 368 gramas. Como você pode tomar uma decisão no sentido de ajustar ou não o processo em uma situação na qual você não consegue pesar cada uma das caixas individualmente à medida que elas vão sendo abastecidas? Você começa pela seleção e a pesagem de uma amostra aleatória de 25 caixas de cereais. Depois de calcular a média aritmética para a amostra, como você procede?

No Capítulo 7, você aprendeu métodos para determinar se o valor da média aritmética de uma amostra é condizente com a média aritmética conhecida de uma população. Nesse cenário da Oxford Cereals, você quer utilizar a média aritmética de uma amostra para validar uma afirmativa sobre a média aritmética da população, um problema um tanto diferente. Para esse tipo de problema, você utiliza um método inferencial conhecido como **teste de hipóteses**. O teste de hipóteses exige que você faça uma afirmativa de uma maneira não ambígua. Nesse cenário, a afirmativa é de que a média aritmética da população é igual a 368 gramas. Você examina uma estatística da amostra para verificar se ela se coaduna mais com a afirmativa declarada, chamada de *hipótese nula*, ou com a alternativa mutuamente excludente (para o cenário em questão, a hipótese de que a média aritmética da população não é igual a 368 gramas).

Neste capítulo, você aprenderá várias aplicações de testes de hipóteses. Você aprenderá a fazer inferências sobre um determinado parâmetro da população *analisando as diferenças* entre os resultados observados, a estatística da amostra e os resultados que esperaria obter caso alguma hipótese subjacente fosse efetivamente verdadeira. Em relação ao cenário da Oxford Cereals, o teste de hipóteses permitiria inferir um dos seguintes itens:

- A média aritmética do peso de caixas de cereais na amostra é um valor condizente com o que você esperaria caso a média aritmética de toda a população de caixas de cereais fosse igual a 368 gramas.
- A média aritmética da população não é igual a 368 gramas, tendo em vista que a média aritmética da amostra é significativamente diferente de 368 gramas.

9.1 Fundamentos da Metodologia para Testes de Hipóteses

O teste de hipóteses geralmente se inicia com algum tipo de teoria, declaração ou assertiva sobre um determinado parâmetro de uma população. Por exemplo, sua hipótese inicial no exemplo dos cereais é de que o processo está operando adequadamente, de modo tal que a média aritmética do peso abastecido corresponde a 368 gramas, e não é necessária nenhuma ação corretiva.

A Hipótese Nula e a Hipótese Alternativa

A hipótese de que o parâmetro da população é igual à especificação da empresa é chamada de hipótese nula. Uma **hipótese nula** é sempre aquela que corresponde ao *status quo*, e é identificada pelo símbolo H_0. Nesse caso, a hipótese nula é de que o processo de abastecimento está operando adequadamente, e, consequentemente, a média aritmética da quantidade abastecida é igual à especificação apresentada pela Oxford Cereals. Isso é declarado como

$$H_0: \mu = 368$$

Embora estejam disponíveis apenas informações originadas da amostra, a hipótese nula é escrita em termos do parâmetro da população, por você estar centrando seu foco na população de todas as caixas de cereais. Você utiliza a estatística da amostra para realizar inferências sobre o processo de abastecimento como um todo. Uma das inferências pode ser a de que os resultados observados com base nos dados da amostra indicam que a hipótese nula é falsa. Se a hipótese nula for considerada falsa, alguma outra afirmativa deve necessariamente ser verdadeira.

Sempre que uma hipótese nula é enunciada, uma hipótese alternativa é também especificada, e ela deve necessariamente ser verdadeira caso a hipótese nula seja falsa. A **hipótese alternativa, H_1**, corresponde ao oposto da hipótese nula, H_0. Isso é declarado no exemplo dos cereais como

$$H_1: \mu \neq 368$$

A hipótese alternativa representa a conclusão a que se chega ao ser rejeitada a hipótese nula. A hipótese nula é rejeitada quando existem evidências suficientes, a partir dos dados extraídos da amostra, de que a hipótese é falsa. No exemplo dos cereais, caso os pesos das caixas selecionadas para fins de amostra estejam suficientemente acima ou abaixo da média aritmética esperada de 368 gramas especificada pela Oxford Cereals, você rejeita a hipótese nula em favor da hipótese alternativa de que a média aritmética da quantidade abastecida é diferente de 368 gramas. Você interrompe a produção e adota a ação necessária, qualquer que seja, para corrigir o problema. Se a hipótese nula não é rejeitada, você deve continuar a acreditar no *status quo*, de que o processo está operando corretamente e, assim, nenhuma ação corretiva é necessária. Nessa segunda circunstância, você não provou que o processo está operando corretamente. Ao invés disso, você não

conseguiu provar que ele está operando incorretamente, e, desse modo, você continua a acreditar (embora não comprovadamente) na hipótese nula.

Na metodologia do teste de hipóteses, você rejeita a hipótese nula quando as evidências geradas pela amostra sugerem que é muito mais provável que a hipótese alternativa seja verdadeira. No entanto, o fato não conseguir rejeitar a hipótese nula não constitui uma prova de que ela seja verdadeira. Você jamais consegue provar que a hipótese nula está correta em razão de a decisão estar baseada somente em informações relacionadas à amostra, e não à população como um todo.. Por conseguinte, se você não consegue rejeitar a hipótese nula, pode apenas concluir que não existem evidências suficientes para garantir a sua respectiva rejeição. Os pontos-chave apresentados a seguir sintetizam a hipótese nula e a hipótese alternativa:

- A hipótese nula, H_0, representa o *status quo*, ou aquilo que se acredita no momento, em relação a uma determinada situação.
- A hipótese alternativa, H_1, é o oposto da hipótese nula, e representa uma declaração a ser investigada, ou uma inferência específica que você gostaria de provar.
- Se você rejeita a hipótese nula, você tem comprovação estatística de que a hipótese alternativa está correta.
- Se você não rejeita a hipótese nula, você não conseguiu comprovar a hipótese alternativa. O fato de não conseguir comprovar a hipótese alternativa, entretanto, não significa que você tenha comprovado a hipótese nula.
- A hipótese nula, H_0, sempre se refere a um valor especificado para o parâmetro da população (tal como μ), e não a uma estatística da amostra (tal como $\bar{X}$).
- A declaração correspondente à hipótese nula sempre contém um sinal de igualdade com relação ao valor especificado para o parâmetro da população (por exemplo, H_0: $\mu = 368$ gramas).
- A declaração correspondente à hipótese alternativa nunca contém um sinal de igualdade com relação ao valor especificado para o parâmetro da população (por exemplo, H_1: $\mu \neq$ 368 gramas).

EXEMPLO 8.1

A Hipótese Nula e a Hipótese Alternativa

Você é o gerente de uma lanchonete. Você deseja determinar se o tempo de espera para atender a um pedido se modificou, no mês anterior, em relação ao valor anterior de 4,5 minutos, correspondente à média aritmética da população. Declare a hipótese nula e a hipótese alternativa.

SOLUÇÃO A hipótese nula é de que a média aritmética da população não se modificou em relação a seu valor anterior, de 4,5 minutos. Isso é enunciado como

$$H_0: \mu = 4{,}5$$

A hipótese alternativa corresponde ao oposto da hipótese nula. Tendo em vista que a hipótese nula afirma que a média aritmética da população é igual a 4,5 minutos, a hipótese alternativa é de que a média aritmética da população não é igual a 4,5 minutos. Isso é indicado sob a forma

$$H_1: \mu \neq 4{,}5$$

O Valor Crítico da Estatística do Teste

A lógica em que se baseia a metodologia do teste de hipóteses diz respeito a determinar qual é a possibilidade de que a hipótese nula seja verdadeira, ao se considerar as informações coletadas em uma amostra. No cenário da Oxford Cereal Company, a hipótese nula é de que a média aritmética da quantidade de cereal por caixa ao longo de todo o processo de abastecimento é igual a 368 gramas (o parâmetro da população especificado pela empresa). Você seleciona uma amostra de caixas do processo de abastecimento, pesa cada uma delas e calcula a média aritmética correspondente à amostra. Essa estatística é uma estimativa para o parâmetro correspondente (a média aritmética da população, μ). Mesmo se a hipótese nula for verdadeira, a estatística (a média aritmética da amostra, $\bar{X}$) pode vir a ser diferente do valor do parâmetro (a média aritmética da população, μ), devido à variação decorrente do processo de amostragem. No entanto, você espera que a estatística da amostra esteja próxima do parâmetro da população, caso a hipótese nula seja verdadeira. Caso a estatística da amostra esteja próxima do parâmetro da população, você não dispõe de evidências suficientes para rejeitar a hipótese nula. Se, por exemplo, a média aritmética da amostra fosse igual a 367,9, você poderia concluir que a média aritmética da população não se modificou (ou seja, $\mu = 368$), tendo em vista que uma média aritmética da amostra de 367,9 está bastante próxima do valor de 368 gramas, correspondente à hipótese nula. Intuitivamente, você imagina que é provável que você obtenha uma média aritmética de amostra equivalente a 367,9 de uma população cuja média aritmética é igual a 368.

Não obstante, se existe uma grande diferença entre o valor da estatística e o valor da hipótese para o parâmetro da população, você pode concluir que a hipótese nula é falsa. Por exemplo, se a média aritmética da amostra é igual a 320, você poderia concluir que a média aritmética da população não é 368 (ou seja, $\mu \neq 368$), uma vez que uma média aritmética de amostra está bastante distante do valor de 368 estipulado pela hipótese. Nesse tipo de situação, você conclui que é bastante improvável obter uma média aritmética de amostra igual a 320 caso a média aritmética da população seja realmente 368. Portanto, é mais lógico concluir que a média aritmética da população não é igual a 368. Nesse caso, você rejeita a hipótese nula.

No entanto, o processo de tomada de decisão não é sempre tão simples e claro. A determinação sobre o que significa "muito próximo" e o que significa "bem diferente" é arbitrária, sem definições claras. A metodologia para os testes de hipóteses proporciona definições claras para avaliar diferenças. Além disso, ela possibilita que você quantifique o processo de tomada de decisão calculando a probabilidade de obter um determinado resultado para a amostra caso a hipótese nula seja verdadeira. Você calcula essa probabilidade determinando a distribuição de amostragens para a estatística de interesse na amostra (por exemplo, a média aritmética da amostra) e, em seguida, calculando a **estatística do teste** específica, com base no resultado conhecido para a amostra. Uma vez que a distribuição de amostragens para a estatística do teste frequentemente segue uma distribuição estatística bastante conhecida, como a distribuição normal padronizada ou a distribuição *t*, você pode utilizar essas distribuições para ajudar a determinar se a hipótese nula é verdadeira.

Regiões de Rejeição e de Não Rejeição

A distribuição de amostragens da estatística do teste está dividida em duas regiões, uma **região de rejeição** (algumas vezes conhecida como região crítica) e uma **região de não rejeição** (veja a Figura 9.1). Caso a estatística do teste se posicione dentro da região de não rejeição, você não rejeita a hipótese nula. No cenário da Oxford Cereals, você conclui que não existem evidências suficientes de que a média aritmética da população correspondente à quantidade abastecida seja diferente de 368. Caso a estatística do teste se posicione na região de rejeição, você rejeita a hipótese nula. Nesse caso, você conclui que a média aritmética da população não corresponde a 368 gramas.

FIGURA 9.1
Regiões de rejeição e de não rejeição no teste de hipóteses

A região de rejeição consiste nos valores relativos à estatística do teste que são improváveis de ocorrer caso a hipótese nula seja verdadeira. Esses valores estão bem mais propensos a ocorrer caso a hipótese nula seja falsa. Por conseguinte, se um determinado valor da estatística do teste se posiciona nessa *região de rejeição*, você rejeita a hipótese nula, tendo em vista que esse valor é improvável caso a hipótese nula seja verdadeira.

Para tomar uma decisão com relação à hipótese nula, você primeiramente determina o **valor crítico** para a estatística do teste. O valor crítico faz a divisão entre a região de não rejeição e a região de rejeição. A determinação desse valor crítico depende do tamanho da região de rejeição. O tamanho da região de rejeição está diretamente relacionado aos riscos envolvidos na utilização somente de evidências decorrentes de amostras para tomar decisões sobre um parâmetro da população.

Riscos na Tomada de Decisão com o Uso da Metodologia para Testes de Hipóteses

Ao se utilizar uma estatística de amostra com o objetivo de serem tomadas decisões quanto a um determinado parâmetro da população, existe um risco de que você chegue a uma conclusão incorreta. Você pode cometer dois tipos diferentes de erro ao aplicar a metodologia para testes de hipóteses: o erro do Tipo I e o erro do Tipo II.

ERRO DO TIPO I E DO TIPO II

Um **erro do Tipo I** ocorre se você rejeita a hipótese nula, H_0, quando ela é verdadeira e não deve ser rejeitada. A probabilidade de ocorrência de um erro do Tipo I é representada por α.

Um **erro do Tipo II** ocorre se você não rejeita a hipótese nula, H_0, quando ela é falsa e deve ser rejeitada. A probabilidade de ocorrência de um erro do Tipo II é representada por β.

No cenário da Oxford Cereals, você comete um erro do Tipo I caso conclua que a média aritmética da população que corresponde à quantidade abastecida *não* é igual a 368 quando ela *é* igual a 368. Esse erro faz com que você ajuste o processo de abastecimento, não obstante o fato de o processo estar operando adequadamente. Por conseguinte, um erro do Tipo I é um "falso alarme". Você comete um erro do Tipo II se concluir que a média aritmética da população correspondente à quantidade abastecida *é* igual a 368 quando ela *não* é igual a 368. Nesse caso, você permitiria que o processo tivesse continuidade, sem ajustes, ainda que eles se fizessem necessários. Por conseguinte, um erro do Tipo II representa uma "oportunidade perdida".

O Nível de Significância (α) A probabilidade de vir a cometer um erro do Tipo I, representado por α (a letra grega *alfa* minúscula), é identificada como o **nível de significância** do teste estatístico. Tradicionalmente, você controla o erro do Tipo I quando decide sobre o nível de risco, α, que você está disposto a correr ao rejeitar a hipótese nula quando ela é verdadeira. Uma vez que você especifica o nível de significância antes de o teste de hipóteses ser realizado, o risco de cometer um erro do Tipo I, α, está diretamente sob seu controle. Tradicionalmente, você seleciona níveis de 0,01, 0,05 e 0,10. A opção por um determinado nível de risco de cometer um erro do Tipo I depende do custo inerente a cometer um erro do Tipo I. Depois que especifica o valor para α, você pode, então, determinar os valores críticos que dividem as regiões de rejeição e de não rejeição. Você conhece o tamanho da região de rejeição tendo em vista que α representa a probabilidade de rejeição quando a hipótese nula é verdadeira. A partir desse fato, você pode, então, determinar o valor crítico ou os valores críticos que dividem as regiões de rejeição e de não rejeição.

O Coeficiente de Confiança O complemento da probabilidade de um erro do Tipo I, $(1 - \alpha)$, é conhecido como coeficiente de confiança. Ao ser multiplicado por 100%, o coeficiente de confiança resulta no nível de confiança, que foi estudado ao serem construídos intervalos de confiança (veja a Seção 8.1).

O **coeficiente de confiança**, $(1 - \alpha)$, corresponde à probabilidade de que você venha a não rejeitar a hipótese nula, H_0, quando ela é efetivamente verdadeira e não deve ser rejeitada. O **nível de confiança** de um teste de hipóteses é representado por $(1 - \alpha) \times 100\%$.

Em termos da metodologia do teste de hipóteses, o coeficiente de confiança representa a probabilidade de que se conclua que o valor do parâmetro, conforme especificado na hipótese nula, é plausível quando ele é verdadeiro. No cenário da Oxford Cereals, o coeficiente de confiança mede a probabilidade de que se conclua que a média aritmética da população da quantidade abastecida é igual a 368 gramas, quando de fato ela é de 368 gramas.

O Risco β O Risco β é a probabilidade de cometer um erro do Tipo II. Diferentemente de um erro do Tipo I, que você controla pela seleção de α, a probabilidade de cometer um erro do Tipo II depende da diferença entre o valor identificado na hipótese e o verdadeiro valor do parâmetro da população. Uma vez que grandes diferenças são mais fáceis de ser encontradas do que pequenas diferenças, caso a diferença entre o valor identificado na hipótese e o verdadeiro valor para o parâmetro da população seja grande, β será pequeno. Por exemplo, se a média aritmética da população é igual a 330 gramas, existe uma pequena chance (β) de que você venha a concluir que a média aritmética não se modificou em relação a 368. Entretanto, caso a diferença entre o valor identificado na hipótese e o verdadeiro valor para o parâmetro seja pequena, β será grande. Por exemplo, no caso de a média aritmética da população ser verdadeiramente igual a 367 gramas, existe uma chance considerável (β) de que você venha a concluir que a média aritmética permanece ainda igual a 368 gramas.

A Eficácia de um Teste O complemento da probabilidade de um erro do Tipo II, $(1 - \beta)$, é conhecido como a eficácia de um teste estatístico.

A **eficácia de um teste estatístico**, $(1 - \beta)$, é a probabilidade de que você venha a rejeitar a hipótese nula quando ela é falsa e deve, efetivamente, ser rejeitada.

No cenário que trata da Oxford Cereals, a eficácia do teste corresponde à probabilidade de que você venha a concluir, corretamente, que a média aritmética da quantidade abastecida não é igual a 368 gramas quando ela efetivamente não é igual a 368 gramas. Para uma discussão detalhada sobre a eficácia de um teste estatístico, veja o tópico *online* **Seção 9.6**, disponível no site da editora para este livro. (Veja a Seção D.3 do Apêndice deste livro para aprender a acessar os arquivos correspondentes aos tópicos *online*.)

Riscos na Tomada de Decisão: Um Equilíbrio Delicado A Tabela 9.1 ilustra os resultados de duas decisões possíveis (não rejeitar H_0 ou rejeitar H_0), que você pode tomar em qualquer teste de hipóteses. Você pode tomar uma decisão correta ou cometer um de dois tipos de erros.

TABELA 9.1
Testes de Hipóteses e Tomada de Decisão

	Situação Real	
Decisão Estatística	**H_0 Verdadeira**	**H_0 Falsa**
Não rejeitar H_0	Decisão correta Confiança $= (1 - \alpha)$	Erro do Tipo II P(Erro do Tipo II) $= \beta$
Rejeitar H_0	Erro do Tipo I P(Erro do Tipo I) $= \alpha$	Decisão correta Eficácia $= (1 - \beta)$

Um modo de reduzir a probabilidade de cometer um erro do Tipo II é pelo aumento do tamanho da amostra. Amostras com grandes tamanhos geralmente permitem que você detecte diferenças até mesmo pequenas entre os valores apresentados na hipótese e os parâmetros da população. Para um determinado nível de α, aumentar o tamanho da amostra faz com que β diminua e, por conseguinte, faz com que cresça a eficácia do teste para detectar que a hipótese nula, H_0, é falsa. No entanto, existe sempre uma limitação no tocante a seus recursos, e isso afeta a decisão em relação ao tamanho máximo de amostra que você pode adotar. Consequentemente, para um determinado tamanho de amostra, você deve levar em consideração a relação de troca entre ganhos e perdas dos dois possíveis tipos de erro. Uma vez que você pode diretamente controlar o risco do erro do Tipo I, você pode reduzir esse risco selecionando um valor menor para α. Por exemplo, caso as consequências negativas associadas a cometer um erro do Tipo I sejam substanciais, você pode selecionar $\alpha = 0{,}01$, em vez de $\alpha = 0{,}05$. No entanto, quando você diminui α, você faz com que β cresça, o que significa que a redução do risco de um erro do Tipo I resulta em um maior risco de um erro do Tipo II. No entanto, para reduzir β, você poderia selecionar um valor mais alto para α. Por conseguinte, se for importante tentar evitar um erro do Tipo II, você pode selecionar α igual a 0,05 ou 0,10, em vez de 0,01.

No cenário que trata da Oxford Cereals, o risco de ocorrência de um erro do Tipo I envolve concluir que a média aritmética da quantidade abastecida se modificou em relação aos 368 gramas identificados na hipótese, quando, na realidade, a média aritmética não se modificou. O risco de ocorrência de um erro do Tipo II envolve concluir que a média aritmética da quantidade abastecida não se modificou em relação aos 368 gramas identificados na hipótese quando, na realidade, ela se modificou. A escolha de valores razoáveis para α e β depende dos custos inerentes a cada um dos tipos de erro. Por exemplo, caso seja demasiado oneroso modificar o processo de abastecimento de cereais, você desejaria estar bastante confiante de que uma modificação seria

necessária antes de fazer qualquer tipo de mudança. Nesse caso, o risco de que seja cometido um erro do Tipo I é mais importante, e você optaria por um α pequeno. Entretanto, caso você desejasse estar bastante confiante de ser capaz de detectar variações em relação à média aritmética de 368 gramas, o risco de cometer um erro do Tipo II seria mais importante, e você optaria por um nível de α mais elevado.

Agora que você passou a conhecer a metodologia para o teste de hipóteses, tenha sempre em mente que, no cenário Utilizando a Estatística do início deste capítulo, o problema estratégico enfrentado pela Oxford Cereals é determinar se o processo de abastecimento de cereais está operando de maneira adequada (ou seja, se a média aritmética da quantidade abastecida, ao longo de todo o processo de abastecimento, permanece dentro dos 368 gramas especificados, e nenhuma ação corretiva se faz necessária). Para avaliar a exigência de 368 gramas, você seleciona uma amostra aleatória de 25 caixas, pesa cada uma delas, calcula a média aritmética, $\bar{X}$, e, depois, avalia a diferença entre essa estatística da amostra e o parâmetro da população especificado na hipótese, comparando a média aritmética do peso (em gramas) da amostra com a média aritmética esperada para a população, 368 gramas, especificada pela empresa. A hipótese nula e a hipótese alternativa são:

$$H_0: \mu = 368$$
$$H_1: \mu \neq 368$$

Quando o desvio-padrão, σ, é conhecido (o que raramente ocorre), você utiliza o **teste Z para a média aritmética** se a população for distribuída nos moldes da distribuição normal. Se a população não for distribuída e nos moldes de uma distribuição normal, você pode, ainda assim, fazer uso do teste Z, caso o tamanho da amostra seja grande o suficiente para que o Teorema do Limite Central exerça seus efeitos (veja a Seção 7.4). A Equação (9.1) define a estatística do teste Z_{ESTAT} para determinar a diferença entre a média aritmética da amostra, $\bar{X}$, e a média aritmética da população, μ, quando o desvio-padrão, σ, é conhecido.

TESTE Z PARA A MÉDIA ARITMÉTICA (σ CONHECIDO)

$$Z_{ESTAT} = \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \quad \textbf{(9.1)}$$

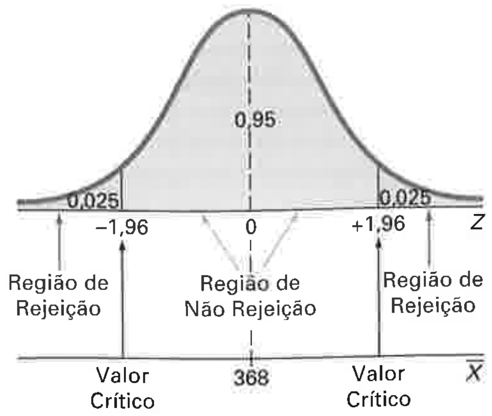
Na Equação (9.1), o numerador mede a diferença entre a média aritmética da amostra observada, $\bar{X}$, e a média aritmética identificada na hipótese, μ. O denominador representa o erro-padrão da média aritmética, de modo tal que Z_{ESTAT} representa a diferença entre $\bar{X}$ e μ, em termos de unidades de erro-padrão.

Testes de Hipóteses Utilizando a Abordagem do Valor Crítico

A abordagem do valor crítico compara o valor calculado da estatística do teste Z_{ESTAT}, extraído da Equação (9.1), aos valores críticos que dividem as regiões de rejeição e a região de não região da distribuição normal. Os valores críticos são expressos sob a forma de valores Z padronizados, que são determinados pelo nível de significância.

Por exemplo, se você utiliza um nível de significância de 0,05, o tamanho da área de rejeição é 0,05. Como a região de rejeição é dividida entre as duas caudas da distribuição, você divide o valor de 0,05 em duas partes iguais de 0,025 cada uma. Para esse **teste bicaudal**, uma região de rejeição equivalente a 0,025 em cada uma das caudas da distribuição normal resulta em uma área acumulada de 0,025 abaixo do valor crítico inferior e em uma área acumulada de 0,975 (1 − 0,025) abaixo do valor crítico superior (que deixa uma área de 0,025 na cauda superior). De acordo com a tabela de distribuição normal padronizada acumulada (Tabela E.2), os valores críticos que dividem as regiões de rejeição e de não rejeição são $-1{,}96$ e $+1{,}96$. A Figura 9.2 ilustra o fato de que, se a média aritmética for efetivamente 368 gramas, conforme declara H_0, os valores para a estatística do teste Z_{ESTAT} terão uma distribuição normal padronizada centralizada em $Z = 0$ (o que corresponde a um valor de $\bar{X}$ correspondente a 368 gramas). Valores de Z_{ESTAT} maiores que $+1{,}96$ ou menores que $-1{,}96$ indicam que $\bar{X}$ é suficientemente diferente de $\mu = 368$ declarado na hipótese, que é improvável que tal valor para $\bar{X}$ viesse a ocorrer caso H_0 fosse verdadeira.

FIGURA 9.2
Testando uma hipótese sobre a média aritmética (σ conhecido) no nível de significância de 0,05



Portanto, a regra de decisão é

$$\text{Rejeitar } H_0 \text{ se } Z_{ESTAT} > +1{,}96$$
$$\text{ou se } Z_{ESTAT} < -1{,}96;$$
$$\text{caso contrário, não rejeitar } H_0.$$

Suponha que a amostra de 25 caixas de cereais indique uma média aritmética de amostra, $\bar{X}$, de 372,5 gramas, e o desvio-padrão da população, σ, seja de 15 gramas. Utilizando a Equação (9.1),

$$Z_{ESTAT} = \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \frac{372{,}5 - 368}{\dfrac{15}{\sqrt{25}}} = +1{,}50$$

Uma vez que $Z_{ESTAT} = +1{,}50$ se posiciona entre $-1{,}96$ e $+1{,}96$, você não rejeita H_0 (veja a Figura 9.3 a seguir). Você continua a acreditar que a média aritmética da quantidade abastecida é de 368 gramas. Para levar em consideração a possibilidade de um erro do Tipo II, você enuncia a conclusão como "existem evidências insuficientes de que a média aritmética da quantidade abastecida seja diferente de 368 gramas".

FIGURA 9.3
Testando uma hipótese sobre a média aritmética do peso do cereal (σ conhecido) no nível de significância de 0,05

A Apresentação 9.1 fornece um resumo da abordagem do valor crítico para o teste de hipóteses. As Etapas 1 a 4 correspondem à tarefa Definir; a etapa 5 combina as tarefas Coletar e Organizar; e a etapa 6 corresponde às tarefas Visualizar e Analisar da metodologia de resolução de problemas discutida inicialmente no Capítulo 2.

APRESENTAÇÃO 9.1 A ABORDAGEM DO VALOR CRÍTICO PARA O TESTE DE HIPÓTESES

1. Declare a hipótese nula, H_0, e a hipótese alternativa, H_1.
2. Escolha o nível de significância, α, e o tamanho da amostra, n. O nível de significância é baseado na importância relativa dos riscos de serem cometidos erros do Tipo I e do Tipo II no problema.
3. Determine a estatística apropriada do teste e a distribuição de amostragens.
4. Determine os valores críticos que dividem as regiões de rejeição e de não rejeição.
5. Colete os dados da amostra, organize os resultados e calcule o valor da estatística do teste.
6. Tome a decisão estatística e expresse a conclusão gerencial. Se a estatística do teste estiver contida na região de não rejeição, você não rejeita a hipótese nula, H_0. Se a estatística de teste estiver contida na região de rejeição, você rejeita a hipótese nula. A conclusão gerencial é expressa no contexto do problema do mundo real.

EXEMPLO 9.2

Aplicando a Abordagem do Valor Crítico ao Teste de Hipóteses na Oxford Cereals

Expresse a abordagem do valor crítico para o teste de hipóteses no caso da Oxford Cereals.

SOLUÇÃO

Etapa 1: Declare a hipótese nula e a hipótese alternativa. A hipótese nula, H_0, é sempre enunciada sob a forma de uma expressão matemática, utilizando parâmetros da população. Ao testar se a média aritmética da quantidade abastecida de cereais é de 368 gramas, a hipótese nula declara que μ é igual a 368. A hipótese alternativa, H_1, é também enunciada sob a forma de uma expressão matemática, utilizando parâmetros da população. Por conseguinte, a hipótese alternativa declara que μ não é igual a 368 gramas.

Etapa 2: Escolha o nível de significância e o tamanho da amostra. Você escolhe o nível de significância, α, de acordo com a importância relativa dos riscos de serem cometidos erros do Tipo I e erros do Tipo II no problema. Quanto menor o valor de α, menor o risco de vir a ser cometido um erro do Tipo I. Neste exemplo, cometer um erro do Tipo I significa que você conclui que a média aritmética da população não corresponde a 368 gramas quando ela é igual a 368 gramas. Consequentemente, você adotará ações corretivas em relação ao processo de abastecimento, não obstante o fato de ele estar operando adequadamente. Nesse caso, $\alpha = 0{,}05$ é selecionado. O tamanho da amostra, n, é igual a 25.

Etapa 3: Selecione a estatística do teste apropriada. Uma vez que σ é conhecido a partir de informações sobre o processo de abastecimento, você utiliza a distribuição normal e a estatística do teste Z_{ESTAT}.

Etapa 4: Determine a região de rejeição. São selecionados valores críticos para a estatística do teste apropriada, de modo a que a região de rejeição contenha uma área total de α quando H_0 for verdadeira e a região de não rejeição contenha uma área total de $1 - \alpha$ quando H_0 for verdadeira. Tendo em vista que $\alpha = 0{,}05$ no exemplo dos cereais, os valores críticos da estatística do teste Z são $-1{,}96$ e $+1{,}96$. A região de rejeição é, portanto, $Z < -1{,}96$ ou $Z_{ESTAT} > +1{,}96$. A região de não rejeição é $-1{,}96 < Z < +1{,}96$.

Etapa 5: Colete os dados da amostra e calcule o valor para a estatística do teste. No exemplo que trata dos cereais, $\bar{X} = 372{,}5$, e o valor da estatística do teste é $Z_{ESTAT} = +1{,}50$.

Etapa 6: Enuncie a decisão estatística e a conclusão gerencial. Em primeiro lugar, determine se a estatística do teste está contida na região de rejeição ou na região de não rejeição. Para o exemplo dos cereais, $Z_{ESTAT} = +1{,}50$ está na região de não rejeição, uma vez que $1{,}96 < Z_{ESTAT} = +1{,}50 \leq +1{,}96$. Como a estatística do teste está contida na região de não rejeição, a decisão estatística corresponde a não rejeitar a hipótese nula, H_0. A conclusão gerencial é de que não existem evidências suficientes para provar que a média aritmética da quantidade abastecida é diferente de 368 gramas. Não é necessária nenhuma ação corretiva quanto ao processo de abastecimento.

EXEMPLO 9.3

Testando e Rejeitando uma Hipótese Nula

Você é o gerente de uma lanchonete, e quer determinar se a média aritmética da população correspondente ao tempo de espera para que um pedido seja atendido se modificou, no mês anterior, em relação a seu valor anterior de 4,5 minutos. Com base em experiências passadas, você consegue pressupor que a população é distribuída nos moldes da distribuição normal, com um desvio-padrão de 1,2 minuto para a população. Você seleciona uma amostra de 25 pedidos durante um período de 1 hora. A média aritmética da amostra é 5,1 minutos. Utilize a abordagem das seis etapas, listada na Apresentação 9.1, para determinar se existem evidências, no nível de significância de 0,05, de que a média aritmética da população do tempo de espera para o atendimento de um pedido se modificou, no mês passado, em relação a seu valor anterior para a média aritmética da população, era de 4,5 minutos.

SOLUÇÃO

Etapa 1: A hipótese nula é de que a média aritmética da população não se modificou em relação ao seu valor anterior, de 4,5 minutos:

$$H_0: \mu = 4{,}5$$

A hipótese alternativa corresponde ao oposto da hipótese nula. Tendo em vista que a hipótese nula é de que a média aritmética da população é de 4,5 minutos, a hipótese alternativa é de que a média aritmética da população não é de 4,5 minutos:

$$H_1: \mu \neq 4{,}5$$

Etapa 2: Você selecionou uma amostra com $n = 25$. O nível de significância é de 0,05 (ou seja, $\alpha = 0{,}05$).

Etapa 3: Uma vez que se pressupõe σ como conhecido, você utiliza a distribuição normal e a estatística do teste Z_{ESTAT}.

Etapa 4: Uma vez que $\alpha = 0{,}05$, os valores críticos da estatística do teste Z_{ESTAT} são $-1{,}96$ e $+1{,}96$. A região de rejeição é $Z_{ESTAT} < -1{,}96$ ou $Z_{ESTAT} > +1{,}96$. A região de não rejeição corresponde a $-1{,}96 \leq Z \leq +1{,}96$.

Etapa 5: Você coleta os dados da amostra e calcula $\bar{X} = 5{,}1$. Utilizando a Equação (9.1), você calcula a estatística do teste:

$$Z_{ESTAT} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{5{,}1 - 4{,}5}{\frac{1{,}2}{\sqrt{25}}} = +2{,}50$$

Etapa 6: Uma vez que $Z_{ESTAT} = +2{,}50 > +1{,}96$, você rejeita a hipótese nula. Você conclui que existem evidências de que a média aritmética da população para o tempo de espera de atendimento de um pedido se modificou em relação a seu valor anterior de 4,5 minutos. A média aritmética do tempo de espera para os clientes é maior agora do que no mês passado. Como gerente, seria desejável que você determinasse um modo de reduzir o tempo de espera para melhorar o serviço.

Testes de Hipóteses Utilizando a Abordagem do Valor-*p*

O uso do valor-*p* para determinar regiões de rejeição e de não rejeição é uma outra abordagem para testes de hipóteses.

VALOR-*p*

O **valor-*p*** é a probabilidade de que seja obtida uma estatística de teste igual ou mais extrema do que o resultado da amostra, considerando-se que a hipótese nula, H_0, é verdadeira. (O valor-*p* é também conhecido como o *nível observado de significância.*)

As regras de decisão para rejeitar H_0 na abordagem do valor-*p* são

- Se o valor-*p* for maior ou igual a α, não rejeitar a hipótese nula.
- Se o valor-*p* for menor do que α, rejeitar a hipótese nula.

Muitas pessoas confundem essas regras, acreditando, equivocadamente, que um valor-*p* alto é uma razão para a rejeição. Você pode evitar essa confusão lembrando-se do seguinte mantra:

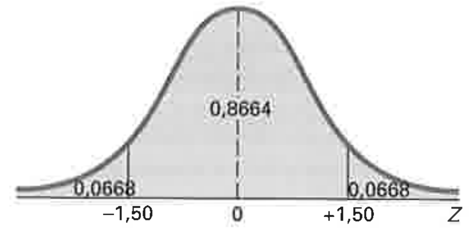
Se o valor-*p* for pequeno, a hipótese nula vai para o dreno.

Para compreender a abordagem do valor-*p*, considere o cenário que trata da Oxford Cereals. Você testou o fato de a média aritmética da quantidade de cereal abastecido ser ou não igual a 368 gramas. A estatística do teste resultou em um valor de Z_{ESTAT} igual a $+1{,}50$, e você não rejeitou a hipótese nula, tendo em vista que $+1{,}50$ era menor que o valor crítico superior, de $+1{,}96$, e maior que o valor crítico inferior, de $-1{,}96$.

Para utilizar a abordagem do valor-*p* em relação ao *teste bicaudal*, você encontra a probabilidade de que seja obtida uma estatística de teste Z_{ESTAT} igual ou *mais extrema* do que 1,50 unidade de desvio-padrão do centro de uma distribuição normal padronizada. Em outras palavras, você

precisa calcular a probabilidade de um valor de Z maior do que $+1{,}50$, juntamente com a probabilidade de que seja obtido um valor de Z menor do que $-1{,}50$. A Tabela E.2 demonstra que a probabilidade de um valor de Z menor do que $-1{,}50$ é de 0,0668. A probabilidade de um valor de Z inferior a $+1{,}50$ corresponde a 0,9332, e a probabilidade de um valor de Z superior a $+1{,}50$ corresponde a $1 - 0{,}9332 = 0{,}0668$. Por conseguinte, o valor-*p* correspondente a esse teste bicaudal é igual a $0{,}0668 + 0{,}0668 = 0{,}1336$ (veja a Figura 9.4). Consequentemente, a probabilidade de um resultado igual ou mais extremo do que aquele observado é de 0,1336. Como 0,1336 é maior do que $\alpha = 0{,}05$, você não rejeita a hipótese nula.

FIGURA 9.4
Encontrando um valor-*p* para um teste bicaudal



Nesse exemplo, a média aritmética da amostra observada é igual a 372,5 gramas, 4,5 gramas acima do valor especificado na hipótese, e o valor-*p* é igual a 0,1336. Assim, se a média aritmética da população é de 368 gramas, existe uma chance de 13,36% de que a média aritmética da amostra difira de 368 gramas em pelo menos 4,5 gramas (isto é, que venha a ser $\geq 372{,}5$ gramas ou $\leq 363{,}5$ gramas). Portanto, ainda que 372,5 estejam acima do valor de 368 especificado na hipótese, um resultado tão extremo ou mais extremo que 372,5 não é altamente improvável quando a média aritmética da população é igual a 368.

A não ser que você esteja lidando com uma estatística de teste que siga o padrão da distribuição normal, você somente será capaz de fazer uma aproximação do valor-*p* com base nas tabelas de distribuição. No entanto, programas como o Microsoft Excel podem calcular o valor-*p* para qualquer teste de hipóteses. Se tiver acesso a um programa que calcule valores-*p*, você pode substituir a abordagem do valor crítico pela abordagem do valor-*p* quando estiver testando hipóteses.

A Figura 9.5 exibe uma solução por meio de planilha para o exemplo que trata do abastecimento de cereais discutido nesta seção. Embora a planilha faça uso da abordagem do valor-*p* na célula A18 para determinar rejeição ou não-rejeição, a planilha também inclui a estatística do teste Z_{ESTAT} e valores críticos

FIGURA 9.5
Planilha do teste Z para a média aritmética (σ conhecido) para o exemplo do abastecimento de cereais

A Figura 9.5 mostra a ***planilha CÁLCULO*** *da* ***pasta de trabalho Z Média Aritmética.*** *Crie essa planilha usando as instruções contidas na Seção GE 9.1.*

	A	B	
1	Teste Z para a Média Aritmética		
2			
3	Dados		
4	Hipótese Nula μ=	368	
5	Nível de Significância	0,05	
6	Desvio-padrão da População	15	
7	Tamanho da Amostra	25	
8	Média Aritmética da Amostra	372,5	
9			
10	Cálculos Intermediários		
11	Erro-padrão da Média Aritmética	3	=B6/RAIZ(B7)
12	Estatística do Teste Z	1,5	=(B8 - B4)/B11
13			
14	Teste Bicaudal		
15	Valor Crítico Inferior	-1,9600	=INV.NORMP(B5/2)
16	Valor Crítico Superior	1,9600	=INV.NORMP(1 - B5/2)
17	Valor *p*	0,1336	=2 * (1 - DIST.NORMP(ABS(B12)))
18	Não rejeitar a hipótese nula		=SE(B17 < B5, "Rejeitar a hipótese nula", "Não rejeitar a hipótese nula")

A Apresentação 9.2 oferece um resumo da abordagem do valor-*p* para o teste de hipóteses.

APRESENTAÇÃO 9.2 A ABORDAGEM DO VALOR-*p* PARA O TESTE DE HIPÓTESES

1. Especifique a hipótese nula, H_0, e a hipótese alternativa, H_1.
2. Escolha o nível de significância, α, e o tamanho da amostra, n. O nível de significância é baseado na importância relativa dos riscos de se cometer um erro do Tipo I e um erro do Tipo II no problema.

3. Determine a estatística do teste e a distribuição de amostragens apropriadas.
4. Colete os dados da amostra, calcule o valor da estatística do teste e calcule o valor-p.
5. Tome a decisão estatística e enuncie a conclusão gerencial. Se o valor-p for maior ou igual a α, você não rejeita a hipótese nula. Se o valor-p for menor do que α, você rejeita a hipótese nula. Lembre-se do mantra: Se o valor-p for pequeno, a hipótese nula vai para o dreno. A conclusão gerencial é enunciada no contexto do problema relacionado ao mundo real.

EXEMPLO 9.4

Testando e Rejeitando uma Hipótese Nula Utilizando a Abordagem do Valor-p

Você é o gerente de uma lanchonete. O problema da empresa é determinar se a média aritmética da população do tempo de espera para atendimento de um pedido se modificou, no mês anterior, em relação a seu valor anterior de 4,5 minutos. Com base na experiência passada, você pode pressupor que o desvio-padrão da população corresponde a 1,2 minuto. Você seleciona uma amostra de 25 pedidos durante um período de 1 hora. A média aritmética da amostra é igual a 5,1 minutos. Utilize a abordagem de cinco etapas para o valor-p, listada na Apresentação 9.2, para determinar se existem evidências de que a média aritmética da população do tempo de espera para o atendimento de um pedido se modificou, no mês passado, em relação a seu valor anterior, de 4,5 minutos, para a média aritmética da população.

SOLUÇÃO

Etapa 1: A hipótese nula é de que a média aritmética da população não se modificou em relação a seu valor anterior de 4,5 minutos:

$$H_0: \mu = 4{,}5$$

A hipótese alternativa é o oposto da hipótese nula. Uma vez que a hipótese nula afirma que a média aritmética da população é de 4,5 minutos, a hipótese alternativa afirma que a média aritmética da população não é de 4,5 minutos:

$$H_1: \mu \neq 4{,}5$$

Etapa 2: Você selecionou uma amostra com $n = 25$ e escolheu um nível de significância de 0,05 (ou seja, $\alpha = 0{,}05$).

Etapa 3: Selecione a estatística de teste apropriada. Uma vez que σ é pressupostamente conhecido, você utiliza a distribuição normal e a estatística de teste Z_{ESTAT}.

Etapa 4: Você coleta os dados e calcula $\bar{X} = 5{,}1$. Utilizando a Equação (9.1), você calcula a estatística do teste da seguinte maneira:

$$Z_{ESTAT} = \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \frac{5{,}1 - 4{,}5}{\dfrac{1{,}2}{\sqrt{25}}} = +2{,}50$$

Para encontrar a probabilidade de obter uma estatística do teste Z_{ESTAT} que seja igual ou mais extrema do que 2,50 unidades de desvio-padrão em relação ao centro da distribuição normal padronizada, você calcula a probabilidade de um valor de Z_{ESTAT} maior do que +2,50, juntamente com a probabilidade de um valor de Z abaixo de −2,50. Com base na Tabela E.2, a probabilidade de um valor de Z_{ESTAT} inferior a −2,50 é 0,0062. A probabilidade de um valor abaixo de +2,50 é 0,9938. Portanto, a probabilidade de um valor acima de +2,50 é igual a $1 - 0{,}9938 = 0{,}0062$. Por conseguinte, o valor-p correspondente a esse teste bicaudal é $0{,}0062 + 0{,}0062 = 0{,}0124$.

Etapa 5: Uma vez que o valor-$p = 0{,}0124 < \alpha = 0{,}05$, você rejeita a hipótese nula. Você conclui que existem evidências de que a média aritmética da população correspondente ao tempo de espera de atendimento de um pedido se modificou em relação a seu valor de 4,5 minutos, correspondente à média aritmética da população. A média aritmética do tempo de espera para os clientes é maior agora do que no mês passado.

Uma Ligação entre a Estimativa do Intervalo de Confiança e o Teste de Hipóteses

Este capítulo e o Capítulo 8 examinam a estimativa para intervalos de confiança e o teste de hipóteses, os dois principais componentes da inferência estatística. Embora compartilhem a mesma fundamentação conceitual, a estimativa do intervalo de confiança e o teste de hipóteses são utilizados para diferentes finalidades. No Capítulo 8, os intervalos de confiança estimaram parâmetros. Neste capítulo, o teste de hipóteses elabora decisões sobre valores especificados de parâmetros da população. Testes de hipóteses são utilizados quando estamos tentando provar que um determinado parâmetro é menor, maior ou não igual a um valor especificado. A interpretação apropriada de um intervalo de confiança, no entanto, pode também indicar se um determinado parâmetro é menor, maior ou não igual a um valor especificado. Por exemplo, nesta seção, você testou se a média aritmética da população correspondente à quantidade abastecida era diferente de 368 gramas utilizando a Equação (9.1):

$$Z_{ESTAT} = \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$$

Em vez de testar a hipótese nula de que $\mu = 368$ gramas, você pode chegar à mesma conclusão construindo uma estimativa para o intervalo de confiança de μ. Se o valor especificado na hipótese para $\mu = 368$ estiver contido dentro do intervalo, você não rejeita a hipótese nula, uma vez que 368 não seria considerado um valor incomum. Entretanto, se o valor especificado na hipótese não se posiciona dentro do intervalo, você rejeita a hipótese nula, uma vez que "$\mu = 368$ gramas" é então considerado um valor incomum. Utilizando a Equação (8.1), na Seção 8.1 do Capítulo 8, e os dados a seguir:

$$n = 25, \bar{X} = 372{,}5 \text{ gramas}, \sigma = 15 \text{ gramas}$$

para um nível de confiança de 95% (ou seja, $\alpha = 0{,}05$),

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$372{,}5 \pm (1{,}96)\frac{15}{\sqrt{25}}$$

$$372{,}5 \pm 5{,}88$$

de modo tal que

$$366{,}62 \leq \mu \leq 378{,}38$$

Uma vez que o intervalo inclui o valor especificado na hipótese que corresponde a 368 gramas, você não rejeita a hipótese nula. Existem evidências insuficientes de que a média aritmética da quantidade abastecida, ao longo de todo o processo, não é igual a 368 gramas. Você chegou a essa mesma decisão ao utilizar o teste de hipóteses bicaudal.

É Realmente Possível Conhecer o Desvio-Padrão da População?

O final da Seção 8.1 abordou a questão de como o aprendizado do método para estimativas de intervalos de confiança que exigiam o conhecimento prévio de σ, o desvio-padrão da população, servia como uma introdução eficaz para o conceito de intervalos de confiança. O texto em questão revelou, então, que seria improvável que você viesse a utilizar o citado procedimento em relação à maior parte das aplicações práticas, por diversas razões.

Por analogia, no que diz respeito à maior parte das aplicações práticas, é improvável que você venha a utilizar o método de teste de hipóteses que exija o conhecimento prévio de σ. Se conhecesse previamente o desvio-padrão da população, você conheceria também a média aritmética da população e não precisaria formular uma hipótese sobre a média aritmética e então testar essa hipótese. Assim, por que estudar um teste de hipóteses para a média aritmética que exija que σ seja conhecido? Utilizar esse tipo de teste torna muito mais fácil explicar os fundamentos de testes de hipóteses. Com um desvio-padrão da população conhecido, você pode utilizar a distribuição normal e calcular valores-p utilizando tabelas da distribuição normal.

Uma vez que é importante que você compreenda o conceito de teste de hipóteses quando estiver lendo o restante deste livro, reexamine esta seção criteriosamente para compreender o conceito subjacente — ainda que você saiba de antemão que jamais terá uma razão prática para utilizar o teste representado pela Equação (9.1).

Problemas para a Seção 9.1

APRENDENDO O BÁSICO

9.1 Se você utilizar um nível de significância de 0,05 em um teste de hipóteses (bicaudal), o que você decide, caso o valor calculado da estatística do teste $Z_{ESTAT} = -0{,}76$?

9.2 Se você utilizar um nível de significância de 0,05 em um teste de hipóteses (bicaudal), o que você decide, caso o valor calculado da estatística do teste $Z_{ESTAT} = +2{,}21$?

9.3 Se você utilizar um nível de significância de 0,10 em um teste de hipóteses (bicaudal), qual seria a sua regra de decisão para rejeitar uma hipótese nula de que a média aritmética da população é igual a 500 se você estivesse utilizando o teste Z?

9.4 Se você está utilizando um nível de significância de 0,01 em um teste de hipóteses (bicaudal), qual seria a sua regra de decisão para rejeitar a hipótese H_0: $\mu = 12{,}5$, se você estivesse utilizando o teste Z?

9.5 Qual seria a sua decisão para o Problema 9.4 se $Z_{ESTAT} = -2{,}61$?

9.6 Qual seria o valor-p se, em um determinado teste de hipóteses bicaudal, $Z_{ESTAT} = +2{,}00$?

9.7 No Problema 9.6, qual seria a sua decisão estatística caso você estivesse testando a hipótese nula no nível de significância de 0,10?

9.8 Qual seria o valor-p se, em um determinado teste de hipóteses bicaudal, $Z_{ESTAT} = -1{,}38$?

APLICANDO OS CONCEITOS

9.9 No sistema judiciário dos EUA, o acusado é presumivelmente inocente até prova em contrário. Considere uma hipótese nula, H_0, de que o acusado seja inocente, e uma hipótese alternativa, H_1, de que o acusado seja culpado. Um júri tem duas decisões possíveis: condenar o acusado (ou seja, rejeitar a hipótese nula) ou não condenar o acusado (ou seja, não rejeitar a hipótese nula). Explique o significado dos riscos de que venha a ser cometido um erro do Tipo I ou um erro do Tipo II neste exemplo.

9.10 Suponha que o acusado no Problema 9.9 fosse presumido culpado até prova em contrário, como ocorre em alguns outros sistemas judiciários. De que maneira a hipótese nula e a hipótese alternativa divergiriam em relação àquelas apresentadas no Problema 9.9? Quais seriam os significados dos riscos de vir a cometer um erro do Tipo I ou um erro do Tipo II, nesse caso?

9.11 A Food and Drug Administration (FDA) é responsável pela aprovação de novos medicamentos nos EUA. Inúmeros grupos de defesa do consumidor acreditam que o processo de aprovação é demasiado fácil e que, consequentemente, é aprovada uma quantidade muito grande de medicamentos que posteriormente se revelam prejudiciais à saúde. Por outro lado, vários lobistas desse setor estão realizando esforços para tornar o processo de aprovação mais flexível, de modo a que as empresas do setor farmacêutico possam ter novos medicamentos aprovados com mais facilidade e maior rapidez (dados extraídos R. Sharpe, "FDA Tries to Find Right Balance on Drug Approvals", *The Wall Street Journal*, 20 de abril de 1999, p. A24). Considere uma hipótese nula de que um novo medicamento, ainda não aprovado, seja prejudicial para a saúde e uma hipótese alternativa de que um novo medicamento, ainda não aprovado, seja seguro.

a. Explique os riscos inerentes a cometer um erro do Tipo I ou um erro do Tipo II.
b. Que tipo de erro os grupos de defesa do consumidor estão tentando evitar? Explique.
c. Que tipo de erro os lobistas do setor estão tentando evitar? Explique.
d. De que maneira seria possível diminuir a chance de ocorrência tanto de erros do Tipo I quanto de erros do Tipo II?

9.12 Como resultado de reclamações relacionadas a atrasos tanto de professores quanto de alunos, o secretário-geral de uma universidade de grande porte deseja fazer um ajuste no horário das aulas para que haja tempo adequado para o deslocamento entre as salas de aula. Ele está pronto para realizar esse estudo. Até o momento, o secretário-geral acreditava que deveria haver 20 minutos de intervalo entre os horários das aulas. Especifique a hipótese nula, H_0, e a hipótese alternativa, H_1.

9.13 Os alunos em sua escola estudam mais, menos ou mais ou menos o mesmo que alunos de outras escolas de administração? A revista *Business Week* relata que nas 50 principais escolas de administração dos EUA os alunos estudavam uma média de 14,6 horas por semana (dados extraídos de "Cracking the Books", Special Report/Online Extra, www.businessweek.com, 19 de março de 2007). Construa um teste de hipóteses para tentar provar que a média aritmética do número de horas estudadas em sua escola é diferente da referência de 14,6 horas por semana apresentada pela *Business Week*.
a. Elabore a hipótese nula e a hipótese alternativa.
b. Qual seria um erro do Tipo I para seu teste?
c. Qual seria um erro do Tipo II para seu teste?

AUTO-AVALIAÇÃO **9.14** O gerente de controle da qualidade de uma fábrica de lâmpadas precisa determinar se a média aritmética da vida útil de uma grande remessa de lâmpadas é igual a 375 horas. O desvio-padrão da população corresponde a 100 horas. Uma amostra aleatória de 64 lâmpadas indica uma média aritmética de vida útil, para a amostra, de 350 horas.
a. No nível de significância de 0,05, existem evidências de que a média aritmética da vida útil seja diferente de 375 horas?
b. Calcule o valor-p e interprete seu significado.
c. Construa a estimativa para o intervalo de confiança de 95% para a média aritmética da população correspondente à vida útil das lâmpadas.
d. Compare os resultados de (a) e (c). A que conclusões você chega?

9.15 O gerente de uma loja de tintas quer determinar se a média aritmética da quantidade de tinta contida em latas de 1 galão, adquiridas de um fabricante nacionalmente conhecido, é, efetivamente, 1 galão. Você sabe, com base nas especificações do fabricante, que o desvio-padrão da quantidade de tinta é de 0,02 galão. Você seleciona uma amostra aleatória de 50 latas, e a média aritmética da quantidade de tinta, por lata de 1 galão, corresponde a 0,995 galão.
a. Existem evidências de que a média aritmética da quantidade seja diferente de 1,0 galão (use $\alpha = 0{,}01$)?
b. Calcule o valor-p e interprete seu significado.
c. Construa a estimativa para o intervalo de confiança de 99% para a média aritmética da população correspondente à quantidade de tinta.
d. Compare os resultados de (a) e (c). A que conclusões você chega?

9.2 O Teste t de Hipóteses para a Média Aritmética (σ Desconhecido)

Em praticamente todas as situações que tratam de testes de hipóteses relacionados à média aritmética da população, μ, não se conhece o desvio-padrão da população, σ. Em vez disso, se utiliza o desvio-padrão da amostra, S. Se se parte do pressuposto de que a população é distribuída, nos moldes de uma distribuição normal, a distribuição de amostragens da média aritmética segue uma distribuição t com $n - 1$ graus de liberdade, e você utiliza o **teste t para a média aritmética**. Se a população não for distribuída, conforme a distribuição normal, você pode, ainda assim, utilizar o teste t caso o tamanho da amostra seja suficientemente grande para que o Teorema do Limite Central possa ser aplicado (veja a Seção 7.4). A Equação (9.2) define a estatística do teste t para determinar a diferença entre a média aritmética da amostra, $\bar{X}$ e a média aritmética da população, μ, quando é utilizado o desvio-padrão da amostra, S.

TESTE t DE HIPÓTESES PARA A MÉDIA ARITMÉTICA (σ DESCONHECIDO)

$$t_{ESTAT} = \frac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}} \tag{9.2}$$

em que a estatística do teste t_{ESTAT} segue uma distribuição t que tem $n - 1$ graus de liberdade.

Para ilustrar a utilização do teste t para a média aritmética, retorne ao cenário que trata da Saxon Home Improvement, no início do Capítulo 8. O objetivo da empresa é determinar se a média aritmética do montante por fatura de vendas permanece inalterada em relação aos \$120 dos últimos cinco anos. Como um dos contabilistas da empresa, você precisa informar se esse valor se modifica. Em outras palavras, o teste de hipóteses é utilizado com o objetivo de tentar provar que a média aritmética do montante por fatura de vendas está aumentando ou diminuindo.

A Abordagem do Valor Crítico

Para realizar esse teste de hipóteses bicaudal, você utiliza o método das seis etapas, listadas na Apresentação 9.1, na Seção 9.1.

Etapa 1: Você define as seguintes hipóteses:

$$H_0: \mu = \$120$$
$$H_1: \mu \neq \$120$$

A hipótese alternativa contém a declaração que você está tentando provar. Se a hipótese nula for rejeitada, existe então comprovação estatística de que a média aritmética da população correspondente ao montante por fatura de vendas não é mais de \$120. Se a conclusão estatística for "não rejeitar H_0", você irá então concluir que não existem evidências suficientes para provar que a média aritmética do montante difere da média aritmética de longo prazo, que corresponde a \$120.

Etapa 2: Você coleta os dados de uma amostra com $n = 12$ faturas de vendas. Você decide utilizar $\alpha = 0{,}05$.

Etapa 3: Uma vez que α é desconhecido, você utiliza a distribuição t e a estatística do teste t_{ESTAT}. Você deve necessariamente pressupor que a população correspondente a faturas de vendas seja distribuída nos moldes da distribuição normal, uma vez que um tamanho de amostra igual a 12 é demasiado pequeno para que o Teorema do Limite Central possa ser aplicado. Esse pressuposto é examinado mais adiante nesta seção.

Etapa 4: Para um determinado tamanho de amostra, n, a estatística do teste t_{ESTAT} segue uma distribuição t com $n - 1$ graus de liberdade. Os valores críticos da distribuição t com $12 - 1 = 11$ graus de liberdade são encontrados na Tabela E.3, conforme ilustrado na Tabela 9.2 e na Figura 9.6. A hipótese alternativa, H_1: $\mu \neq \$120$, é bicaudal. A área na região de rejeição da cauda esquerda (inferior) da distribuição t é igual a 0,025, e a área na região de rejeição da cauda direita (superior) da distribuição t é também igual a 0,025.

Tomando como base a tabela t, em conformidade com a Tabela E.3, da qual uma parte está ilustrada na Tabela 9.2, os valores críticos correspondem a ±2,2010. A regra de decisão é

$$\text{Rejeitar } H_0 \text{ se } t_{ESTAT} < -t_{\alpha/2} = -2{,}2010$$
$$\text{ou se } t_{ESTAT} > t_{\alpha/2} = +2{,}2010;$$
$$\text{caso contrário, não rejeitar } H_0.$$

TABELA 9.2
Determinando o Valor Crítico a partir da Tabela t para uma Área de 0,025 em Cada Cauda, com 11 Graus de Liberdade

	Probabilidades Acumuladas					
	0,75	0,90	0,95	0,975	0,99	0,995
	Áreas da Cauda Superior					
Graus de Liberdade	0,25	0,10	0,05	0,025	0,01	0,005
1	1,0000	3,0777	6,3138	12,7062	31,8207	63,6574
2	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8409
4	0,7407	1,5332	2,1318	2,7764	3,7469	4,6041
5	0,7267	1,4759	2,0150	2,5706	3,3649	4,0322
6	0,7176	1,4398	1,9432	2,4469	3,1427	3,7074
7	0,7111	1,4149	1,8946	2,3646	2,9980	3,4995
8	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554
9	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498
10	0,6998	1,3722	1,8125	2,2281	2,7638	3,1693
11	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058

Fonte: *Extraída da Tabela E.3.*

FIGURA 9.6
Testando uma hipótese sobre a média aritmética (σ desconhecido) no nível de significância de 0,05 com 11 graus de liberdade

Etapa 5: Você organiza e armazena os dados a partir de uma amostra aleatória de 12 faturas de vendas em **Faturas** :

108,98	152,22	111,45	110,59	127,46	107,26
93,32	91,97	111,56	75,71	128,58	135,11

Utilizando as Equações (3.1) e (3.7), nas Seções 3.1 e 3.2 do Capítulo 3,

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \$112{,}85 \text{ e } S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} = \$20{,}80$$

Com base na Equação (9.2),

$$t_{ESTAT} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{112{,}85 - 120}{\frac{20{,}80}{\sqrt{12}}} = -1{,}1908$$

A Figura 9.7 mostra uma solução, por meio de uma planilha, para esse teste de hipóteses.

FIGURA 9.7
Planilha para o teste t de faturas de vendas

A Figura 9.7 exibe a planilha CÁLCULO da pasta de trabalho T Média Aritmética. Crie essa planilha utilizando as instruções na Seção GE9.2.

	A	B	
1	Teste t para a Hipótese da Média Aritmética		
2			
3	Dados		
4	Hipótese Nula	120	
5	Nível de Significância	0,05	
6	Tamanho da Amostra	12	
7	Média Aritmética da Amostra	112,85	
8	Desvio-padrão da Amostra	20,8	
9			
10	Cálculos Intermediários		
11	Erro-padrão da Média Aritmética	6,0044	=B8/RAIZ(B6)
12	Graus de Liberdade	11	=B6 - 1
13	Estatística do Teste t	-1,1908	=(B7 - B4)/B11
14			
15	Teste Bicaudal		
16	Valor Crítico Inferior	-2,2010	=-INVT(B5, B12)
17	Valor Crítico Superior	2,2010	=INVT(B5, B12)
18	Valor p	0,2588	=DISTT(ABS(B13), B12, 2)
19	Não rejeitar a hipótese nula		=SE(B18 < B5, "Rejeitar a hipótese nula", "Não rejeitar a hipótese nula")

Etapa 6: Para analisar os resultados, uma vez que $-2{,}2010 < t_{ESTAT} = -1{,}1908 < 2{,}2010$, você não rejeita H_0. Você não tem evidências suficientes para concluir que a média aritmética do montante por fatura de vendas difere de \$120. A auditoria sugere que a média aritmética do montante por fatura não se modificou.

A Abordagem do Valor-p

Etapa 1-3. Essas etapas são as mesmas utilizadas na abordagem do valor crítico.

Etapa 4. Com base na planilha do Microsoft Excel da Figura 9.10, $t_{ESTAT} = -1{,}19$ e o valor-p = 0,2588.

Etapa 5. Os resultados da planilha apresentada na Figura 9.10 mostram que o valor-p para esse teste bicaudal é 0,2588. Uma vez que o valor-p de 0,2588 é maior do que $\alpha = 0{,}05$, você não rejeita H_0. Os dados não fornecem evidências suficientes para concluir que a média aritmética do valor por fatura de vendas difere de \$120. A auditoria sugere que a média aritmética do montante por fatura de vendas não se modificou. O valor-p indica que, caso a hipótese nula seja verdadeira, a probabilidade de que uma amostra de 12 faturas possa vir a ter uma média aritmética mensal que difira em \$7,15 ou mais em relação aos \$120 declarados é de 0,2588. Em outras palavras, se a média aritmética do montante por fatura de vendas for verdadeiramente \$120, existe então uma chance de 25,88% de que seja observada uma média aritmética de amostra abaixo de \$112,85 ou acima de \$127,15.

No exemplo que acabamos de apresentar, seria incorreto afirmar que existe uma chance de 25,88% de que a hipótese nula seja verdadeira. Lembre-se de que o valor-p corresponde a uma probabilidade condicional, calculada com base no *pressuposto* de que a hipótese nula seja verdadeira. De modo geral, é apropriado afirmar o seguinte:

> Caso a hipótese nula seja verdadeira, existe uma chance de (valor-p) × 100% de se observar uma estatística de teste pelo menos tão contraditória em relação à hipótese nula quanto o resultado da amostra.

Verificando o Pressuposto da Normalidade

Você utiliza o teste t de uma amostra quando o desvio-padrão da população, σ, não é conhecido e é estimado pelo uso do desvio-padrão da amostra, S. Para utilizar o teste t, você pressupõe que os dados representam uma amostra aleatória extraída de uma população que seja distribuída nos moldes da distribuição normal. Na prática, contanto que o tamanho da amostra não seja demasiadamente pequeno e a população não seja demasiadamente assimétrica, a distribuição t fornece uma boa aproximação para a distribuição de amostragens da média aritmética quando σ é desconhecido.

Existem várias maneiras de avaliar o pressuposto da normalidade, necessário para que seja utilizado o teste t. Você pode observar o quão estreitamente as estatísticas da amostra estão relacionadas com as propriedades teóricas da distribuição normal. Você pode, também, construir um histograma, uma disposição ramo e folha, um *box-plot* ou um gráfico da probabilidade normal para visualizar a distribuição dos valores relacionados às faturas de vendas. Para detalhes sobre a avaliação da normalidade, veja a Seção 6.3 no Capítulo 6.

Estatísticas descritivas, um *box-plot* e um gráfico da probabilidade normal para os dados sobre faturas de vendas são ilustrados nas Figuras 9.8 a 9.10.

FIGURA 9.8
Planilha para as estatísticas descritivas dos dados relativos às faturas de vendas

Calcule estatísticas descritivas utilizando as instruções apresentadas na Seção GE3.2.

	A	B
1	*Montante das Faturas*	
2		
3	Média Aritmética	112,8508
4	Erro-padrão	6,0039
5	Mediana	111,02
6	Moda	#N/A
7	Desvio-padrão	20,7980
8	Variância da Amostra	432,5565
9	Curtose	0,1727
10	Assimetria	0,1336
11	Amplitude	76,51
12	Mínimo	75,71
13	Máximo	152,22
14	Soma	1354,21
15	Contagem	12

FIGURA 9.9
Box-plot para os dados sobre faturas de vendas

Crie box-plots utilizando as instruções apresentadas na Seção GE3.3.

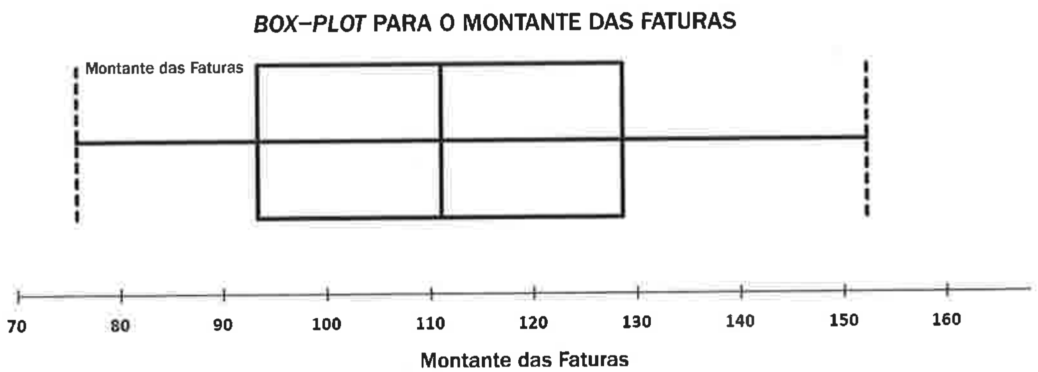
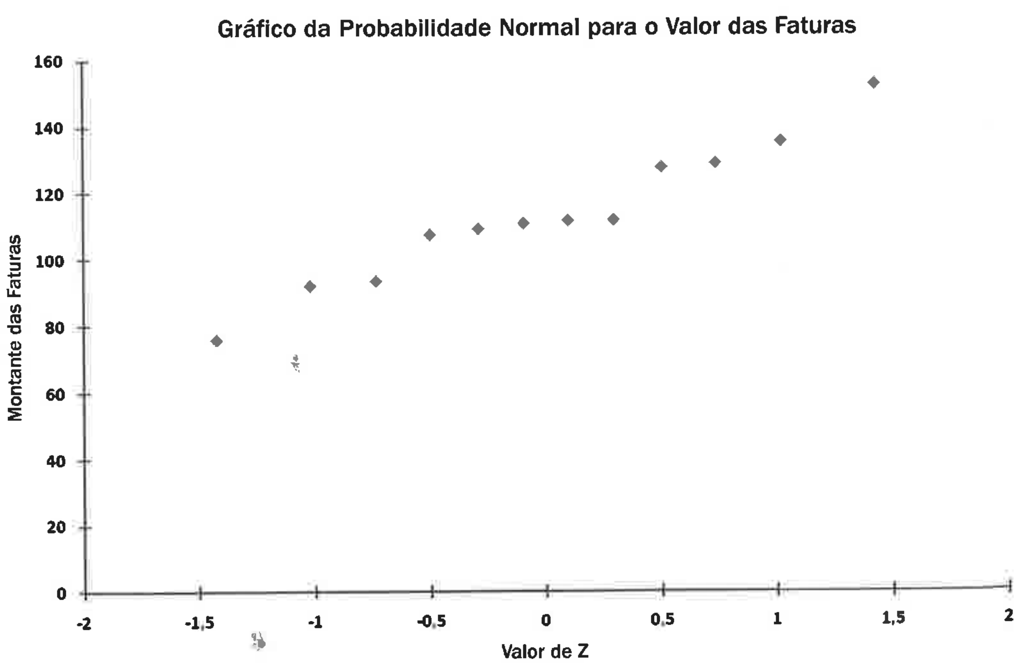


FIGURA 9.10
Gráfico da probabilidade normal para os dados sobre faturas de vendas

Crie gráficos da probabilidade normal utilizando as instruções apresentadas na Seção GE6.3.



A média aritmética se encontra bastante próxima da mediana, e os pontos no gráfico da probabilidade normal aparentam estar crescendo aproximadamente no formato de uma linha reta. O *box-plot* aparenta ser aproximadamente simétrico. Por conseguinte, você pode pressupor que a população de faturas de vendas é distribuída aproximadamente nos moldes de uma distribuição normal. O pressuposto da normalidade é válido, e, por conseguinte, os resultados obtidos pelo auditor são válidos.

O teste t é um teste **robusto**. Ele não perde a sua eficácia caso o formato da população se afaste um pouco de uma distribuição normal, particularmente quando o tamanho da amostra é grande o suficiente para possibilitar que a estatística do teste t seja influenciada pelo Teorema do Limite Central (veja a Seção 7.4). No entanto, você pode tirar conclusões equivocadas e perder a eficácia estatística caso utilize incorretamente o teste t. Se o tamanho da amostra, n, for pequeno (ou seja, menor do que 30) e você não consiga facilmente adotar o pressuposto de que a população subjacente seja pelo menos aproximadamente distribuída conforme a distribuição normal, outros procedimentos de testes *não paramétricos* serão mais apropriados (veja as referências 1 e 2).

Problemas para a Seção 9.2

APRENDENDO O BÁSICO

9.16 Se, em uma amostra de tamanho $n = 16$, selecionada de uma população normal, $\bar{X} = 56$ e $S = 12$, qual é o valor da estatística de t_{ESTAT} se você estiver testando a hipótese nula, H_0: $\mu = 50$?

9.17 No Problema 9.16, quantos graus de liberdade existem no teste t?

9.18 Nos Problemas 9.16 e 9.17, quais são os valores críticos de t se o nível de significância, α, for 0,05 e a hipótese alternativa, H_1, for $\mu \neq 50$?

9.19 Nos Problemas 9.16, 9.17 e 9.18, qual é a sua decisão estatística se a hipótese alternativa, H_1, for $\mu \neq 50$?

9.20 Se, em uma amostra de tamanho $n = 16$, selecionada de uma população assimétrica à esquerda, $\bar{X} = 65$ e $S = 21$, você utilizaria o teste t para testar a hipótese nula H_0: $\mu = 60$? Discuta.

9.21 Se, em uma amostra de tamanho $n = 160$, selecionada de uma população assimétrica à esquerda, $\bar{X} = 65$ e $S = 21$, você utilizaria o teste t para testar a hipótese nula H_0: $\mu = 60$? Discuta.

APLICANDO OS CONCEITOS

AUTO-AVALIAÇÃO **9.22** Você é o gerente de uma franquia de uma lanchonete. No mês passado, a média aritmética do tempo de espera no guichê de atendimento a automóveis das lojas em sua região geográfica, medido desde a hora em que o cliente faz um pedido até o momento em que recebe o seu lanche, foi de 3,7 minutos. Você seleciona uma amostra aleatória de 64 pedidos. A média aritmética da amostra para o tempo de espera é de 3,57 minutos, com um desvio-padrão de 0,8 minuto.

a. No nível de significância de 0,05, existem evidências de que a média aritmética da população relativa ao tempo de espera seja diferente de 3,7 minutos?

b. Uma vez que o tamanho da amostra é 64, você precisa se preocupar com o formato da distribuição da população ao conduzir o teste t em (a)? Explique.

9.23 Um fabricante de chocolates utiliza máquinas para embalar as barras à medida que elas percorrem uma linha de produção. Embora as embalagens apresentem no rótulo o peso de 8 onças, a empresa quer que as embalagens contenham uma média aritmética de 8,17 onças, de modo tal que efetivamente nenhuma das embalagens contenha menos de 8 onças. Uma amostra de 50 embalagens é selecionada periodicamente, e o processo de embalagem é interrompido caso existam evidências de que a média aritmética da quantidade embalada é diferente de 8,17 onças. Suponha que em uma determinada amostra com 50 embalagens a média aritmética da quantidade contida seja de 8,159 onças, com um desvio-padrão de amostra igual a 0,051 onça.

a. Existem evidências de que a média aritmética da população seja diferente de 8,17 onças? (Utilize um nível de significância de 0,05.)

b. Determine o valor-p e interprete o seu significado.

9.24 Em um ano recente, a Federal Communications Commission (Comissão Federal de Comunicações) relatou que a média aritmética do tempo de espera por consertos dos clientes da Verizon era de 36,5 horas. Em um esforço para melhorar esse serviço, suponha que tenha sido desenvolvido um novo processo de serviços de conserto. Esse novo processo, quando utilizado para uma amostra de 100 consertos, resultou em uma média aritmética de amostra correspondente a 34,5 horas e um desvio-padrão de amostra de 11,7 horas.

a. Existem evidências de que a média aritmética da população do número de horas seja diferente de 36,5 horas? (Utilize um nível de significância de 0,05.)

b. Determine o valor-p e interprete o seu significado.

9.25 Em um ano recente, a Federal Communications Commission (Comissão Federal de Comunicações) relatou que a média aritmética do tempo de espera por consertos dos clientes da AT&T era de 25,3 horas. Em um esforço para melhorar esse serviço, suponha que tenha sido desenvolvido um novo processo de serviços de conserto. Esse novo processo, quando utilizado para uma amostra de 100 consertos, resultou em uma média aritmética de amostra de 22,3 horas e um desvio-padrão de amostra de 8,3 horas.

a. Existem evidências de que a média aritmética da população do número de horas seja diferente de 25,3 horas? (Utilize um nível de significância de 0,05.)

b. Determine o valor-p e interprete o seu significado.

9.26 Os dados no arquivo **PreçosCinema** contêm preços (em dólares) para dois ingressos, incluídos a taxa de serviços de emissão de ingressos pela Internet, uma embalagem grande de pipocas e dois refrigerantes médios, em uma amostra de seis cadeias de cinema.

36,15 31,00 35,05 40,25 33,75 43,00

Fonte: *Dados extraídos de K. Kelly, "The Multiplex Under Siege",* The Wall Street Journal, *24-25 de dezembro de 2005, pp. P1, P5.*

a. No nível de significância de 0,05, existem evidências de que a média aritmética do preço correspondente a dois ingressos, incluídos a taxa de serviços de emissão de ingressos pela Internet, uma embalagem grande de pipocas e dois refrigerantes médios seja diferente de $35?
b. Determine o valor-p e interprete o seu significado.
c. Qual pressuposto você deve adotar sobre a distribuição da população para conduzir o teste t em (a) e (b)?
d. Uma vez que o tamanho da amostra é 6, você precisa se preocupar com o formato da distribuição da população ao conduzir o teste t em (a)? Explique.

9.27 No estado de Nova York, bancos de investimentos podem vender um tipo de seguro de vida chamado de *savings bank life insurance* (SBLI). O processo de aprovação consiste na subscrição, que inclui uma revisão da proposta, uma seção para confirmação do histórico médico, possíveis pedidos de informações ou exames médicos complementares e o estágio de compilação da apólice, durante o qual as páginas da apólice são geradas e enviadas ao banco para entrega. A capacidade de emitir as apólices aprovadas em tempo hábil para o cliente é crítica para a eficácia desse serviço. Durante o período de um mês, foi selecionada uma amostra aleatória composta por 27 apólices aprovadas, e foram registrados os tempos totais de processamento, em dias (conforme armazenado no arquivo **Seguro**):

73	19	16	64	28	28	31	90	60	56	31	56	22	18
45	48	17	17	17	91	92	63	50	51	69	16	17	

a. No passado, a média aritmética do tempo de processamento era de 45 dias. No nível de significância de 0,05, existem evidências de que a média aritmética do tempo de processamento tenha se modificado em relação a 45 dias?
b. Qual pressuposto em relação à distribuição da população é necessário para conduzir o teste t em (a)?
c. Construa um *box-plot* e um gráfico da probabilidade normal para avaliar o pressuposto adotado em (b).
d. Você acha que o pressuposto necessário para conduzir o teste t em (a) é válido? Explique.

9.28 Os dados a seguir (veja o arquivo **Refrigerante**) representam a quantidade de refrigerante abastecido em uma amostra de 50 garrafas consecutivas de 2 litros. Os resultados, apresentados horizontalmente na ordem em que as garrafas são abastecidas, foram:

2,109	2,086	2,066	2,075	2,065	2,057	2,052	2,044	2,036	2,038
2,031	2,029	2,025	2,029	2,023	2,020	2,015	2,014	2,013	2,014
2,012	2,012	2,012	2,010	2,005	2,003	1,999	1,996	1,997	1,992
1,994	1,986	1,984	1,981	1,973	1,975	1,971	1,969	1,966	1,967
1,963	1,957	1,951	1,951	1,947	1,941	1,941	1,938	1,908	1,894

a. No nível de significância de 0,05, existem evidências de que a média aritmética da quantidade de refrigerante abastecida seja diferente de 2,0 litros?
b. Determine o valor-p em (a) e interprete o seu significado.
c. Em (a), você pressupôs que a distribuição correspondente à quantidade de refrigerante abastecida era distribuída nos moldes de uma distribuição normal. Avalie esse pressuposto por meio da construção de um *box-plot* ou um gráfico da probabilidade normal.
d. Você acha que o pressuposto necessário para conduzir o teste t em (a) é válido? Explique.
e. Examine os valores relativos às 50 garrafas, em sua respectiva ordem sequencial, conforme apresentado no problema. Existe um padrão em relação aos resultados? Em caso afirmativo, que impacto esse padrão exerce sobre a validade dos resultados em (a)?

9.29 Um dos principais indicadores da qualidade de serviços oferecidos por qualquer organização é a velocidade com que ela responde às reclamações dos clientes. Uma grande loja de departamentos, de propriedade familiar, que vende mobiliário e cobertura para pisos, incluindo carpetes, passou por uma grande expansão ao longo dos últimos anos. Em particular, o departamento de coberturas para pisos expandiu de 2 especialistas em instalação para um supervisor de instalações, um responsável pela medição e 15 especialistas em instalação. A loja tinha como objetivo estratégico melhorar sua resposta às reclamações. A variável de interesse foi definida como o número de dias entre o momento em que a reclamação era feita e o momento em que era solucionado o problema. Foram coletados dados de 50 reclamações feitas ao longo do ano passado. Os dados foram armazenados em **Mobiliário** e são os seguintes:

54	5	35	137	31	27	152	2	123	81	74	27
11	19	126	110	110	29	61	35	94	31	26	5
12	4	165	32	29	28	29	26	25	1	14	13
13	10	5	27	4	52	30	22	36	26	20	23
33	68										

a. O supervisor de instalações alega que a média aritmética do número de dias entre o recebimento de uma reclamação e a resolução do problema relacionado à essa reclamação é de 20 dias. No nível de significância de 0,05, existem evidências de que a declaração não seja verdadeira (ou seja, de que a média aritmética do número de dias é diferente de 20)?
b. Qual pressuposto sobre a distribuição da população é necessário para construir o teste t em (a)?
c. Construa um *box-plot* e um gráfico da distribuição normal para avaliar o pressuposto adotado em (b).
d. Você acredita que o pressuposto necessário para construir o teste t em (a) é válido? Explique.

9.30 Uma indústria produz suportes de aço para equipamentos elétricos. O principal componente do suporte é uma placa de aço em baixo-relevo que é produzida a partir de uma bobina de aço de calibre 14. Ela é produzida por uma força progressiva de uma prensa de 250 toneladas, com uma operação de limpeza que coloca duas fôrmas de 90 graus no aço plano, para fabricar o baixo-relevo. A distância de um lado da fôrma até o outro é crítica em razão da necessidade de o suporte ser à prova d'água quando utilizado em ambientes externos. A empresa exige que a largura do baixo-relevo esteja entre 8,31 polegadas e 8,61 polegadas. O arquivo de dados **BaixoRelevo** contém as larguras dos baixos-relevos, em polegadas, para uma amostra de tamanho $n = 49$.

8,312	8,343	8,317	8,383	8,348	8,410	8,351	8,373	8,481	8,422
8,476	8,382	8,484	8,403	8,414	8,419	8,385	8,465	8,498	8,447
8,436	8,413	8,489	8,414	8,481	8,415	8,479	8,429	8,458	8,462
8,460	8,444	8,429	8,460	8,412	8,420	8,410	8,405	8,323	8,420
8,396	8,447	8,405	8,439	8,411	8,427	8,420	8,498	8,409	

a. No nível de significância de 0,05, existem evidências de que a média aritmética da largura dos baixos-relevos seja diferente de 8,46 polegadas?

b. Qual pressuposto sobre a distribuição da população é necessário para construir o teste t em (a)?
c. Avalie o pressuposto adotado em (b).
d. Você acha que o pressuposto necessário para construir o teste t em (a) é válido? Explique.

9.31 Uma operação de uma usina é cortar fragmentos de aço em pedaços que são utilizados na estrutura de assentos dianteiros em automóveis. O aço é cortado com uma serra de diamante e requer que os pedaços resultantes estejam contidos no intervalo entre $\pm 0{,}005$ polegada em relação ao comprimento especificado pela montadora de automóveis. O arquivo **Aço** contém uma amostra de 100 peças de aço. A medição apresentada corresponde à diferença, em polegadas, entre o comprimento efetivo do pedaço de aço, medido por um dispositivo de mensuração a *laser*, e o comprimento especificado para o pedaço de aço. Por exemplo, um valor de $-0{,}002$ representa um pedaço de aço que é 0,002 polegada mais curto do que o comprimento especificado.
a. No nível de significância de 0,05, existem evidências de que a média aritmética da diferença não seja igual a 0,0 polegada?
b. Construa uma estimativa para o intervalo de confiança de 95% para a média aritmética da população. Interprete esse intervalo.
c. Compare as conclusões tiradas em (a) e (b).
d. Uma vez que $n = 100$, você precisa se preocupar com o pressuposto da normalidade necessário para o teste t e o intervalo de t?

9.32 No Problema 3.67, Capítulo 3, você conheceu a operação de abastecimento de saquinhos de chá. Uma importante característica de interesse relacionada à qualidade desse processo é o peso do chá nos saquinhos individuais. O arquivo **SaquinhosChá** contém uma disposição ordenada do peso, em gramas, de uma amostra de 50 saquinhos de chá, produzidos durante um turno de oito horas.
a. Existem evidências de que a média aritmética da quantidade de chá por saquinho seja diferente de 5,5 gramas (utilize $\alpha = 0{,}01$)?
b. Construa uma estimativa para o intervalo de confiança de 99% para a média aritmética da população correspondente à quantidade de chá por saquinho. Interprete esse intervalo.
c. Compare as conclusões tiradas em (a) e (b).

9.33 Embora muitas pessoas acreditem que podem colocar uma refeição na mesa em um curto período de tempo, um artigo recente relatou que elas acabam gastando cerca de 40 minutos fazendo isso (dados extraídos de N. Hellmich, "Americans Go for the Quick Fix for Dinner", *USA Today*, 14 de fevereiro de 2006). Suponha que outro estudo seja conduzido para testar a validade dessa afirmativa. É selecionada uma amostra de 25 pessoas, e é registrada a extensão do tempo necessário para preparar e cozinhar o jantar (em minutos), com os seguintes resultados (no arquivo **Jantar**).

44,0	51,9	49,7	40,0	55,5	33,0	43,4	41,3	45,2	40,7	41,1	49,1	30,9
45,2	55,3	52,1	55,1	38,8	43,1	39,2	58,6	49,8	43,2	47,9	46,6	

a. Existem evidências de que a média aritmética da população correspondente ao tempo necessário para preparar e cozinhar o jantar seja diferente de 40 minutos? Utilize a abordagem do valor-p e um nível de significância de 0,05.
b. Que pressuposto sobre a distribuição da população é necessário para conduzir o teste t em (a)?
c. Faça uma lista dos vários modos pelos quais você poderia avaliar o pressuposto observado em (b).
d. Avalie o pressuposto observado em (b) e determine se o teste t em (a) é válido.

9.3 Testes Unicaudais

Na Seção 9.1, foi utilizada a metodologia do teste de hipóteses para examinar a questão de a média aritmética da população correspondente à quantidade de cereal abastecido ser ou não igual a 368 gramas. A hipótese alternativa (H_1: $\mu \neq 368$) inclui duas possibilidades: a média aritmética é inferior a 368 gramas ou a média aritmética é superior a 368 gramas. Por essa razão, a região de rejeição é dividida entre as duas caudas da distribuição de amostragens para a média aritmética. Na Seção 9.2, foi utilizado um teste bicaudal para determinar se a média aritmética do valor por fatura se modificou em relação a $120.

Em contraposição a esses dois exemplos, muitas situações requerem uma hipótese alternativa que se concentre em uma *determinada direção*. Por exemplo, a média aritmética da população é *menor do que* um valor especificado. Uma situação desse tipo envolve um problema de uma empresa com relação ao tempo de atendimento em um guichê em uma lanchonete do tipo *drive-thru*. A velocidade com que os clientes são atendidos é de importância crucial para o sucesso do serviço (veja **www.qsrmagazine.com/reports/drive-thru_time_study**). Em um estudo recente, o McDonald's tinha uma média aritmética de 158,77 segundos para o tempo de atendimento, que correspondia ao quinto melhor no setor. Suponha que o McDonald's tenha dado início a um esforço de melhoria de qualidade de modo a reduzir o tempo de atendimento e tenha aperfeiçoado o processo de atendimento a automóveis nos guichês. Dados extraídos do novo processo serão coletados em um teste em uma amostra de 25 lojas. Como o McDonald's teria interesse em instituir o novo processo em todas as suas lojas somente se ele resultasse em uma *redução* no tempo de atendimento nos guichês, toda a região de rejeição está localizada na cauda inferior da distribuição.

A Abordagem do Valor Crítico

Você deseja determinar se um novo processo de atendimento a automóveis em guichês de uma lanchonete tipo *drive-thru* tem uma média aritmética inferior a 158,77 segundos. Para realizar esse teste de hipóteses unicaudal, você utiliza o método de seis etapas ilustrado na Apresentação 9.1, Seção 9.1.

Etapa 1: Você define a hipótese nula e a hipótese alternativa:

$$H_0: \mu \geq 158{,}77$$

$$H_1: \mu < 158{,}77$$

A hipótese alternativa contém a afirmativa que você está tentando provar. Caso a conclusão do teste seja "rejeitar H_0", existe comprovação estatística de que a média aritmética do tempo de atendimento nos guichês destinados a automóveis (*drive-thru*) é menor do que o tempo de atendimento nos guichês destinados a automóveis no processo antigo. Isso constituiria razão para modificar o processo desse tipo de atendimento em toda a população de lojas. Caso a conclusão do teste seja "não rejeitar H_0", não existem então evidências suficientes para provar que a média aritmética do tempo de atendimento nos guichês destinados a automóveis no novo processo é significativamente menor do que no processo antigo. Caso isso ocorra, não haveria razão suficiente para instituir o novo processo de atendimento nos guichês destinados a automóveis na população de lojas.

Etapa 2: Você coleta os dados selecionando uma amostra de $n = 25$ lojas. Você decide utilizar $\alpha = 0{,}05$.

Etapa 3: Uma vez que σ é desconhecido, você utiliza a distribuição t e a estatística do teste t_{ESTAT}. Você deve pressupor que o tempo de atendimento é distribuído nos padrões de uma distribuição normal.

Etapa 4: A região de rejeição está inteiramente contida na cauda inferior da distribuição de amostragens da média aritmética, tendo em vista que você deseja rejeitar H_0 somente quando a média aritmética da amostra for significativamente inferior a 158,77 segundos. Quando toda a região de rejeição está contida em uma das caudas da distribuição de amostragens da estatística do teste, o teste é chamado de **teste unicaudal** ou **teste direcional**. Se a hipótese alternativa inclui o sinal que corresponde a *menor que*, o valor crítico de t é negativo. Conforme ilustrado na Tabela 9.3 e na Figura 9.11, tendo em vista que toda a região de rejeição está contida na cauda inferior da distribuição t e contém uma área de 0,05, em razão da simetria da distribuição t, o valor crítico da estatística do teste t com $25 - 1 = 24$ graus de liberdade é de $-1{,}7109$.

A regra de decisão é

$$\text{Rejeitar } H_0 \text{ se } t_{ESTAT} < -1{,}7109;$$
$$\text{caso contrário, não rejeitar } H_0.$$

TABELA 9.3
Determinando o Valor Crítico, a partir da Tabela t, para uma Área de 0,025 na Cauda Inferior, com 24 Graus de Liberdade

	Probabilidades Acumuladas					
	0,75	0,90	$\boxed{0{,}95}$	0,975	0,99	0,995
	Áreas da Cauda Superior					
Graus de Liberdade	0,25	0,10	$\boxed{0{,}05}$	0,025	0,01	0,005
1	1,0000	3,0777	6,3138	12,7062	31,8207	63,6574
2	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8409
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
23	0,6853	1,3195	1,7139	2,0687	2,4999	2,8073
$\boxed{24}$	0,6848	1,3178	$\boxed{1{,}7109}$	2,0639	2,4922	2,7969
25	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874

Fonte: *Extraída da Tabela E.3.*

FIGURA 9.11
Teste de hipóteses unicaudal para a média aritmética (σ desconhecido) no nível de significância de 0,05

Etapa 5: Com base na amostra de 25 lojas que selecionou, você descobre que a média aritmética do tempo de atendimento nos guichês destinados a automóveis é de 147,5 segundos e que o desvio-padrão da amostra é de 20 segundos. Utilizando $n = 25$, $\bar{X} = 147{,}5$, $S = 20$ e a Equação (9.2),

$$t_{ESTAT} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{147{,}5 - 158{,}77}{\frac{20}{\sqrt{25}}} = -2{,}82$$

Etapa 6: Uma vez que $t_{ESTAT} = -2{,}82 < -1{,}7109$, você rejeita a hipótese nula (veja a Figura 9.11). Você conclui que a média aritmética para o tempo de atendimento nos guichês destinados a automóveis é de menos do que 158,77 segundos. Existem razões suficientes para modificar o processo de atendimento nos guichês destinados a automóveis para toda a população de lojas.

A Abordagem do Valor-*p*

Utilize as cinco etapas listadas na Apresentação 9.2 para realizar o teste t para o estudo sobre o tempo de atendimento nos guichês destinados a automóveis utilizando agora a abordagem do valor-*p*.

Etapas 1-3. Essas etapas são as mesmas utilizadas na abordagem do valor crítico.

Etapa 4. $t_{ESTAT} = -2{,}82$ (veja a etapa 5 da abordagem do valor crítico). Uma vez que a hipótese alternativa indica uma região de rejeição inteiramente contida na cauda *inferior* da distribuição de amostragens, para calcular o valor-*p* você precisa encontrar a probabilidade de que a estatística do teste t_{ESTAT} será menor do que −2,82. Com base na Figura 9.12, o valor-*p* é 0,0048.

FIGURA 9.12
Planilha do teste t para o estudo sobre o tempo de atendimento nos guichês destinados a automóveis

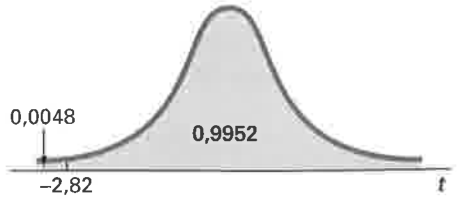
A Figura 9.12 exibe a planilha CÁLCULO_INFERIOR da pasta de trabalho T Média Aritmética. Crie essa planilha utilizando as instruções apresentadas na Seção GE9.3.

	A	B	
1	Teste *t* para a Hipótese da Média Aritmética		
2			
3	Dados		
4	Hipótese Nula	158,77	
5	Nível de Significância	0,05	
6	Tamanho da Amostra	25	
7	Média Aritmética da Amostra	147,5	
8	Desvio-padrão da Amostra	20	
9			
10	Cálculos Intermediários		
11	Erro-padrão da Média Aritmética	4,0000	=B8/RAIZ(B6)
12	Graus de Liberdade	24	=B6 - 1
13	Estatística do Teste *t*	-2,8175	=(B7 - B4)/B11
14			
15	Teste Bicaudal		
16	Valor Crítico Inferior	-1,7109	=-INVT(2 * B5, B12)
17	Valor *p*	0,0048	=SE(B13 < 0, E22, E23)
18	Rejeitar a hipótese nula		=SE(B17 < B5, "Rejeitar a hipótese nula", "Não rejeitar a hipótese nula")

Não ilustrado
Célula E22: =DISTT(ABS(B13), B12, 1)
Célula E23: =1 - E22

Etapa 5. O valor-p de 0,0048 é menor que a $= 0{,}05$ (veja a Figura 9.13). Você rejeita H_0 e conclui que a média aritmética do tempo de atendimento nos guichês destinados a automóveis é menor do que 158,77 segundos. Existem evidências suficientes para modificar o processo de atendimento nos guichês destinados a automóveis para a população inteira de lojas.

FIGURA 9.13
Determinando o valor-p para um teste bicaudal



EXEMPLO 9.5
Um Teste Unicaudal para a Média Aritmética

Uma empresa que fabrica barras de chocolate está particularmente preocupada com o fato de a média aritmética do peso de uma barra de chocolate não ser superior a 6,03 onças. É selecionada uma amostra com 50 barras de chocolate; a média aritmética da amostra é 6,034 onças, e o desvio-padrão da amostra é de 0,02 onça. Utilizando o nível de significância $\alpha = 0{,}01$, existem evidências de que a média aritmética da população correspondente ao peso das barras de chocolate seja maior do que 6,03 onças?

SOLUÇÃO Utilizando a abordagem do valor crítico,

Etapa 1: Inicialmente, você define as suas hipóteses:

$$H_0: \mu \leq 6{,}03$$
$$H_1: \mu > 6{,}03$$

Etapa 2: Você coleta os dados de uma amostra com tamanho $n = 50$. Você decide utilizar $\alpha = 0{,}01$.

Etapa 3: Uma vez que σ é desconhecido, você utiliza a distribuição t e a estatística do teste t_{ESTAT}.

Etapa 4: A região de rejeição está inteiramente contida na cauda superior da distribuição de amostragens para a média aritmética, uma vez que você deseja rejeitar H_0 exclusivamente quando a média aritmética da amostra for significativamente maior do que 6,03 onças. Tendo em vista que toda a região de rejeição está contida na cauda superior da distribuição t, e contém uma área correspondente a 0,01, o valor crítico para a distribuição t, com $50 - 1 = 49$ graus de liberdade é 2,4049 (veja a Tabela E.3).

A regra de decisão é

Rejeitar H_0 se $t_{ESTAT} > 2{,}4049$;
caso contrário, não rejeitar H_0.

Etapa 5: Tendo como base a sua amostra de 50 barras de chocolate, você descobre que a média aritmética do peso na amostra é igual a 6,034 onças e o desvio-padrão da amostra é de 0,02 onças. Utilizando $n = 50$, $\bar{X} = 6{,}034$, $S = 0{,}02$ e a Equação (9.2) no início da Seção 9.2,

$$t_{ESTAT} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{6{,}034 - 6{,}03}{\frac{0{,}02}{\sqrt{50}}} = 1{,}414$$

Etapa 6: Tendo em vista que $t_{ESTAT} = 1{,}414 < 2{,}4049$, ou utilizando o Microsoft Excel, o valor-p é igual a $0{,}0818 > 0{,}01$, você não rejeita a hipótese nula. Não existem evidências suficientes para concluir que a média aritmética do peso da população é maior do que 6,03 onças.

Para realizar um teste de hipóteses unicaudal, você deve formular apropriadamente H_0 e H_1. Um resumo da hipótese nula e da hipótese alternativa para testes unicaudais é apresentado a seguir:

- A hipótese nula, H_0, representa o *status quo* ou aquilo que se acredita no momento sobre uma determinada situação.

- A hipótese alternativa, H_1, é o oposto da hipótese nula, e representa uma declaração a ser investigada ou uma inferência específica que você gostaria de provar.
- Se você rejeita a hipótese nula, você tem comprovação estatística de que a hipótese alternativa está correta.
- Se você não rejeita a hipótese nula, você não conseguiu então comprovar a hipótese alternativa. O fato de não conseguir comprovar a hipótese alternativa, no entanto, não significa que você tenha comprovado a hipótese nula.
- A hipótese nula sempre se refere a um valor especificado para o *parâmetro da população* (tal como μ), e não a uma *estatística da amostra* (tal como $\bar{X}$).
- A declaração da hipótese nula *sempre* contém um sinal de igualdade com relação ao valor especificado para o parâmetro (por exemplo, H_0: $\mu \geq 158{,}77$).
- A declaração da hipótese alternativa *jamais* contém um sinal de igualdade com relação ao valor especificado para o parâmetro (por exemplo, H_1: $\mu < 158{,}77$).

Problemas para a Seção 9.3

APRENDENDO O BÁSICO

9.34 Em um teste de hipóteses unicaudal no qual você rejeita H_0 somente na cauda *superior*, qual é o valor-p se $Z_{ESTAT} = +2{,}00$?

9.35 No Problema 9.34, qual será a sua decisão estatística se você testar a hipótese nula no nível de significância de 0,05?

9.36 Em um teste de hipóteses unicaudal no qual você rejeita H_0 somente na cauda *inferior*, qual é o valor-p se $Z_{ESTAT} = -1{,}38$?

9.37 No Problema 9.36, qual será sua decisão estatística se você testar a hipótese nula no nível de significância de 0,01?

9.38 Em um teste de hipóteses unicaudal no qual você rejeita H_0 somente na cauda *inferior*, qual é o valor-p se $Z_{ESTAT} = +1{,}38$?

9.39 No Problema 9.38, qual será sua decisão estatística se você testar a hipótese nula no nível de significância de 0,01?

9.40 Em um teste de hipóteses unicaudal no qual você rejeita H_0 somente na cauda *superior*, qual é o valor da estatística do teste t, com 10 graus de liberdade, no nível de significância de 0,01?

9.41 No Problema 9.40, qual seria a sua decisão estatística se $t_{ESTAT} = +2{,}39$?

9.42 Em um teste de hipóteses unicaudal no qual você rejeita H_0 somente na cauda *inferior*, qual é o valor da estatística do teste t_{ESTAT} com 20 graus de liberdade, no nível de significância de 0,01?

9.43 No Problema 9.42, qual seria a sua decisão estatística se $t_{ESTAT} = -1{,}15$?

APLICANDO OS CONCEITOS

9.44 Em um ano recente, a Federal Communications Commission (Comissão Federal de Comunicações) relatou que a média aritmética do tempo de espera por consertos dos clientes da Verizon era de 36,5 horas. Em um esforço para melhorar esse serviço, suponha que tenha sido desenvolvido um novo processo de serviços de conserto. Esse novo processo, quando utilizado para uma amostra de 100 consertos, resultou em uma média aritmética de amostra de 34,5 horas e um desvio-padrão de amostra de 11,7 horas.

a. Existem evidências de que a média aritmética da população do número de horas seja menor do que 36,5 horas? (Utilize um nível de significância de 0,05.)
b. Determine o valor-p e interprete o seu significado.
c. Compare os resultados em (a) e (b) com os resultados do Problema 9.24 (a) e (b) nos Problemas para a Seção 9.2.

9.45 Em um ano recente, a Federal Communications Commission (Comissão Federal de Comunicações) relatou que a média aritmética do tempo de espera por consertos dos clientes da AT&T era de 25,3 horas. Em um esforço para melhorar esse serviço, suponha que tenha sido desenvolvido um novo processo de serviços de conserto. Esse novo processo, quando utilizado para uma amostra de 100 consertos, resultou em uma média aritmética de amostra de 22,3 horas e um desvio-padrão de amostra de 8,3 horas.

a. Existem evidências de que a média aritmética da população do número de horas seja menor do que 25,3 horas? (Utilize um nível de significância de 0,05.)
b. Determine o valor-p e interprete o seu significado.
c. Compare os resultados em (a) e (b) com os resultados do Problema 9.25 (a) e (b) nos Problemas para a Seção 9.2.

AUTO-AVALIAÇÃO **9.46** A Glen Valley Steel Company fabrica barras de aço. Se o processo de produção estiver operando adequadamente, serão produzidas barras de aço normalmente distribuídas, com uma média aritmética de comprimento equivalente a *pelo menos* 2,8 pés. Barras de aço mais longas podem ser utilizadas ou alteradas, e as barras mais curtas devem ser descartadas. Você seleciona uma amostra de 25 barras; a média aritmética do comprimento é de 2,73 pés e o desvio-padrão da amostra é de 0,20 pé.

a. Caso você teste a hipótese nula no nível de significância de 0,05, que decisão você tomaria, utilizando a abordagem do valor crítico para o teste de hipóteses?
b. Caso você teste a hipótese nula no nível de significância de 0,05, que decisão você tomaria, utilizando a abordagem do valor-p para o teste de hipóteses?
c. Interprete o significado do valor-p neste problema.
d. Compare suas conclusões em (a) e (b).

9.47 Você é o gerente de um restaurante que entrega *pizzas* em dormitórios de faculdades. Você acabou de modificar o seu processo de entregas, em um esforço para reduzir a média aritmética do tempo entre o pedido e a entrega, em relação aos atuais 25 minutos. Uma amostra de 36 pedidos utilizando o novo processo de entrega resulta em uma média aritmética de amostra de 22,4 minutos e um desvio-padrão de amostra de 6 minutos.

a. Utilizando a abordagem de seis etapas para o valor crítico, no nível de significância de 0,05, existem evidências de que a média aritmética da população correspondente ao tempo de entrega tenha sido reduzida para menos do que a média aritmética anterior da população, de 25 minutos?

b. No nível de significância de 0,05, utilize a abordagem de cinco etapas para o valor-*p*.

c. Interprete o significado do valor-*p* em (b).

d. Compare suas conclusões em (a) e (b).

9.48 As crianças nos Estados Unidos respondem diretamente por $36 bilhões em vendas a cada ano. Quando se considera a influência indireta dessas crianças sobre decisões relacionadas a produtos, desde equipamentos de som até férias, o gasto econômico total influenciado por crianças nos Estados Unidos chega a $290 bilhões. Estima-se que por volta dos 10 anos de idade uma criança faz uma média de mais de cinco visitas por semana a estabelecimentos comerciais (dados extraídos de M. E. Goldberg, G. J. Gorn, L. A. Peracchio, and G. Bamossy, "Understanding Materialism Among Youth", *Journal of Consumer Psychology*, 2003, 13(3), pp. 278-288). Suponha que você queira provar que as crianças em sua cidade fazem, em média, mais de cinco visitas por semana a estabelecimentos comerciais. Seja μ a média aritmética da população correspondente ao número de vezes que as crianças em sua cidade fazem visitas a estabelecimentos comerciais.

a. Especifique a hipótese nula e a hipótese alternativa.

b. Explique o significado do erro do Tipo I e do erro do Tipo II no contexto desse cenário.

c. Suponha que você conduza um estudo semelhante na cidade em que vive. Você extrai uma amostra de 100 crianças e descobre que a média aritmética correspondente ao número de visitas a estabelecimentos comerciais é de 5,47 e o desvio-padrão do número de visitas a estabelecimentos comerciais é de 1,6. No nível de significância de 0,01, existem evidências de que a média aritmética da população correspondente ao número de visitas a estabelecimentos comerciais seja maior do que 5 por semana?

d. Interprete o significado do valor-*p* em (c).

9.49 A média aritmética relativa ao tempo de espera para pagar as compras em um supermercado tem sido de 10,73 minutos. Recentemente, em um esforço para reduzir o tempo de espera, o supermercado realizou experiências com um sistema de uma única fila de espera com vários caixas para pagamento. Foi selecionada uma amostra de 100 consumidores, e a média aritmética relativa ao tempo de espera para pagar as compras foi de 9,52 minutos, com um desvio-padrão de amostra de 5,8 minutos.

a. No nível de significância de 0,05, utilizando a abordagem do valor crítico para o teste de hipóteses, existem evidências de que a média aritmética da população correspondente ao tempo de espera para pagar as compras seja menor do que 10,73 minutos?

b. No nível de significância de 0,05, utilizando a abordagem do valor-*p* para o teste de hipóteses, existem evidências de que a média aritmética da população correspondente ao tempo de espera para um pedido ser atendido seja menor do que 10,73 minutos?

c. Interprete o significado do valor-*p* neste problema.

d. Compare suas conclusões em (a) e (b).

9.4 O Teste *Z* de Hipóteses para a Proporção

Em algumas situações, você deseja testar uma hipótese relacionada à proporção de eventos de interesse na população, π, em vez de testar a média aritmética da população. Para começar, você seleciona uma amostra aleatória e calcula a **proporção da amostra**, $p = X/n$. Você compara então o valor dessa estatística com o valor do parâmetro especificado na hipótese, π, para decidir se deve ou não rejeitar a hipótese nula. Se o número de eventos de interesse (X) e o número de eventos que não são de interesse ($n - X$) forem, cada um deles, pelo menos iguais a cinco, a distribuição de amostragens de uma proporção segue, aproximadamente, uma distribuição normal. Você utiliza o **teste *Z* para a proporção**, fornecido na Equação (9.3), para realizar o teste de hipóteses para a diferença entre a proporção da amostra, p, e a proporção da população especificada na hipótese, π.

TESTE *Z* PARA A PROPORÇÃO

$$Z_{ESTAT} = \frac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}} \tag{9.3}$$

em que

$$p = \text{Proporção da amostra} = \frac{X}{n} = \frac{\text{Número de eventos de interesse na amostra}}{\text{Tamanho da amostra}}$$

π = Proporção especificada na hipótese para os eventos de interesse na população

A estatística do teste Z_{ESTAT} segue aproximadamente uma distribuição normal padronizada quando X e $(n - X)$ são, cada um deles, pelo menos iguais a 5.

De maneira alternativa, ao multiplicar o numerador e o denominador por n, você pode escrever a estatística do teste Z_{ESTAT} em termos do número de eventos de interesse, X, conforme demonstrado na Equação (9.4).

TESTE *Z* PARA A PROPORÇÃO EM TERMOS DO NÚMERO DE EVENTOS DE INTERESSE

$$Z_{ESTAT} = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \tag{9.4}$$

A Abordagem do Valor Crítico

Para ilustrar o teste *Z* para uma proporção, considere um estudo que buscasse determinar se o serviço de atendimento a clientes é melhor ou pior em portais de comércio virtual (*e-commerce*) do que em lojas físicas (dados extraídos de "Consumers Happier with E-Commerce", *USA Today*, 13 de março de 2007, p. 1B). De 1.100 respondentes, 561 afirmaram que o serviço de atendimento a clientes era melhor em portais de comércio virtual do que em lojas físicas. Para essa pesquisa, a hipótese nula e a hipótese alternativa são formuladas do seguinte modo:

H_0: $\pi = 0{,}50$ (ou seja, metade de todos os consumidores acredita que o serviço de atendimento a clientes é melhor em portais de comércio virtual do que em lojas físicas)

H_1: $\pi \neq 0{,}50$ (ou seja, menos da metade ou mais da metade de todos os consumidores acredita que o serviço de atendimento a clientes é melhor em portais de comércio virtual do que em lojas físicas)

Uma vez que você está interessado em determinar se a proporção da população de consumidores que acredita que o serviço de atendimento a clientes é melhor em portais de comércio virtual do que em lojas físicas é menor ou maior do que 0,50, você utiliza um teste bicaudal. Caso você selecione o nível de significância $\alpha = 0{,}05$, a região de rejeição e a região de não rejeição são construídas na forma da Figura 9.14, e a regra de decisão é

Rejeitar H_0 se $Z_{ESTAT} < -1{,}96$ ou se $Z_{ESTAT} > +1{,}96$;

caso contrário, não rejeitar H_0.

FIGURA 9.14
Teste de hipóteses bicaudal para a proporção no nível de significância de 0,05

Tendo em vista que 561 dos 1.100 respondentes afirmaram que o serviço de atendimento a clientes é melhor em portais de comércio virtual do que em lojas físicas,

$$p = \frac{561}{1.100} = 0{,}51$$

Utilizando a Equação (9.3),

$$Z_{ESTAT} = \frac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}} = \frac{0{,}51 - 0{,}50}{\sqrt{\dfrac{0{,}50(1-0{,}50)}{1.100}}} = \frac{0{,}01}{0{,}0151} = 0{,}6633$$

ou utilizando a Equação (9.4),

$$Z_{ESTAT} = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} = \frac{561 - (1.100)(0{,}50)}{\sqrt{1.100(0{,}50)(0{,}50)}} = \frac{11}{16{,}5831} = 0{,}6633$$

Como $-1{,}96 < Z_{ESTAT} = 0{,}6633 < 1{,}96$, você não rejeita H_0. Não existem evidências suficientes para comprovar que a proporção da população correspondente a todos os consumidores que acreditam que o serviço de atendimento a clientes é melhor em portais de comércio virtual do que em lojas físicas não é igual a 0,50. A Figura 9.15 apresenta em uma planilha os resultados relativos a esses dados.

FIGURA 9.15
Planilha para o teste Z correspondente à pesquisa sobre o serviço de atendimento a clientes ser melhor em portais de comércio virtual do que em lojas físicas

*A Figura 9.15 exibe a **planilha CÁLCULO** da **pasta de trabalho Z Proporção**. Crie essa planilha utilizando as instruções na Seção GE9.4.*

	A	B	
1	Teste Z das Hipóteses para a Proporção		
2			
3	Dados		
4	Hipótese Nula p =	0,5	
5	Nível de Significância	0,05	
6	Número de Itens de Interesse	561	
7	Tamanho da Amostra	1100	
8			
9	Cálculos Intermediários		
10	Proporção da Amostra	0,5100	=B6/B7
11	Erro-padrão	0,0151	=RAIZ(B4*(1 - B4)/B7)
12	Estatística do Teste Z	0,6633	=(B10 - B4)/B11
13			
14	Teste Bicaudal		
15	Valor Crítico Inferior	-1,9600	=INV.NORMP(B5/2)
16	Valor Crítico Superior	1,9600	=INV.NORMP(1 - B5/2)
17	Valor p	0,5071	=2 * (1 - DIST.NORMP(ABS(B12)))
18	Não rejeitar a hipótese nula		=SE(B17 < B5, "Rejeitar a hipótese nula", "Não rejeitar a hipótese nula")

A Abordagem do Valor-*p*

Como uma alternativa à abordagem do valor crítico, você pode calcular o valor-p. Para esse teste bicaudal no qual a região de rejeição está localizada na cauda inferior e na cauda superior, você precisa encontrar a área abaixo de um valor Z correspondente a $-0{,}6633$ e acima de um valor Z correspondente a $+0{,}6633$. A Figura 9.15 apresenta um valor-p correspondente a 0,5071. Como esse valor é maior do que o nível de significância selecionado ($\alpha = 0{,}05$), você não rejeita a hipótese nula.

EXEMPLO 9.6

Testando uma Hipótese para uma Proporção

Uma cadeia de lanchonetes desenvolveu um novo processo para garantir que os pedidos feitos em guichês para atendimento de automóveis (*drive-thru*) sejam preenchidos corretamente. O problema da empresa é definido como determinar se o novo processo pode fazer com que cresça a percentagem de pedidos processados corretamente. O processo anterior fazia com que fossem preenchidos pedidos corretamente em 85% dos casos. Dados são coletados a partir de uma amostra de 100 pedidos com o uso do novo processo. Os resultados indicam que 94 pedidos foram preenchidos corretamente. No nível de significância de 0,01, você pode concluir que o novo processo fez crescer a proporção de pedidos preenchidos corretamente?

SOLUÇÃO A hipótese nula e a hipótese alternativa são

H_0: $\pi \leq 0{,}85$ (ou seja, a proporção de pedidos preenchidos corretamente com o uso do novo processo é menor ou igual a 0,85)

H_1: $\pi > 0{,}85$ (ou seja, a proporção de pedidos preenchidos corretamente com o uso do novo processo é maior que 0,85)

Utilizando a Equação (9.3),

$$p = \frac{X}{n} = \frac{94}{100} = 0{,}94$$

$$Z_{ESTAT} = \frac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}} = \frac{0{,}94 - 0{,}85}{\sqrt{\dfrac{0{,}85(1-0{,}85)}{100}}} = \frac{0{,}09}{0{,}0357} = 2{,}52$$

O valor-p para $Z_{ESTAT} > 2{,}52$ é 0,0059.

Utilizando a abordagem do valor crítico, você rejeita H_0 caso $Z_{ESTAT} > 2{,}33$. Utilizando a abordagem do valor-p, você rejeita H_0 caso o valor-$p < 0{,}01$. Uma vez que $Z_{ESTAT} = 2{,}52 > 2{,}33$ ou o valor-$p = 0{,}0059 < 0{,}01$, você rejeita H_0. Você tem evidências de que o novo processo fez a proporção de pedidos preenchidos corretamente crescer para mais de 0,85.

Problemas para a Seção 9.4

APRENDENDO O BÁSICO

9.50 Se, em uma amostra aleatória de 400 itens, 88 forem defeituosos, qual é a proporção da amostra para os itens defeituosos?

9.51 No Problema 9.50, se a hipótese nula é de que 20% dos itens na população são defeituosos, qual é o valor de Z_{ESTAT}?

9.52 Nos Problemas 9.50 e 9.51, suponha que você esteja testando a hipótese nula, H_0: $\pi = 0{,}20$ contra a hipótese alternativa bicaudal, H_1: $\pi \neq 0{,}20$, e escolha o nível de significância $\alpha = 0{,}05$. Qual seria a sua decisão estatística?

APLICANDO OS CONCEITOS

9.53 O Departamento Norte-Americano de Educação relata que 46% dos alunos de faculdade que estudam em horário integral estão empregados ao mesmo tempo em que cursam a faculdade (dados extraídos de "The Condition of Education 2009", *National Center for Education Statistics*, **nces.ed.gov**). Um estudo recente realizado junto a 60 alunos de horário integral na Universidade de Miami descobriu que 29 deles estavam empregados.

a. Utilize a abordagem de cinco etapas do valor-p para o teste de hipóteses e um nível de significância de 0,05 para determinar se a proporção de alunos que estão cursando em horário integral na Universidade de Miami e que estão empregados é maior do que a média nacional de 0,46.
b. Suponha que o estudo descobriu que 36 dos 60 alunos que estão cursando em horário integral estavam empregados e repita (a). As conclusões são as mesmas?

9.54 Revistas virtuais fazem com que seja fácil para o leitor direcionar-se para um portal do anunciante dentro da Web diretamente de um anúncio na revista digital. Uma pesquisa recente indicou que 56% dos leitores de revistas virtuais clicaram em algum anúncio e se direcionaram diretamente para o portal do anunciante na Web. A pesquisa se baseou em uma amostra com tamanho $n = 6.403$ (dados extraídos de "Metrics", *EContent*, janeiro/fevereiro de 2007, p. 20).

a. Utilize a abordagem de cinco etapas do valor-p para tentar determinar se existem evidências de que mais da metade dos leitores de revistas virtuais já se direcionou ao portal de um anunciante na Web.
b. Suponha que o tamanho da amostra fosse de apenas $n = 100$, e, do mesmo modo que antes, 56% dos leitores de revistas virtuais indicassem que já clicaram em algum anúncio de modo a se direcionar diretamente para o portal do anunciante na Web. Utilize a abordagem de cinco etapas do valor-p para tentar determinar se existem evidências de que mais da metade dos leitores de revistas virtuais já se direcionou ao portal de um anunciante na Web. (Utilize o nível de significância de 0,05.)
c. Discuta os efeitos que o tamanho da amostra exerce sobre o teste de hipóteses.
d. Quais você acredita que são as suas chances de rejeitar qualquer hipótese nula no que diz respeito à proporção da população, caso venha a ser utilizado um tamanho de amostra $n = 20$?

9.55 Uma das questões com que as organizações se deparam é o crescimento da diversidade em toda a organização. Uma das maneiras de avaliar o sucesso de uma organização em relação ao crescimento da diversidade é pela comparação da percentagem de trabalhadores na organização em uma determinada posição com um perfil específico com a percentagem em uma posição específica com aquele perfil específico na força de trabalho geral. Recentemente, um grande centro médico acadêmico determinou que 9 dos 17 empregados em uma determinada posição eram do sexo feminino, enquanto 55% dos empregados nessa posição na força de trabalho geral eram do sexo feminino. No nível de significância de 0,05, existem evidências de que a proporção de pessoas do sexo feminino ocupando essa posição nesse centro médico seja diferente daquilo que se esperaria na força de trabalho geral?

AUTO-AVALIAÇÃO **9.56** De 1.000 respondentes entre 24 e 35 anos de idade, 65% relataram que prefeririam "procurar um emprego em um lugar onde gostaria de viver" em vez de "procurar o melhor emprego que possa encontrar, o lugar onde vivo é secundário" (dados extraídos de L. Belkin, "What Do Young Jobseekers Want? (Something Other Than a Job)", *The New York Times*, 6 de setembro de 2007, p. G2). No nível de significância de 0,05, existem evidências de que a proporção de todos os jovens que estão na busca de empregos entre 24 e 35 anos de idade e que prefeririam "procurar um emprego em um lugar onde eu gostaria de viver" em vez de "procurar o melhor emprego que eu possa encontrar, o lugar onde vivo é secundário" seja diferente de 60%?

9.57 Uma das questões mais importantes que os varejistas na Internet enfrentam é a capacidade de reduzir a proporção de consumidores que cancelam suas transações depois de terem selecionado seus produtos. Estima-se que cerca de metade dos consumidores potenciais cancela suas transações depois de terem selecionado seus produtos (B. Tedeschi, "E-Commerce, a Cure for Abandoned Shopping Carts: A Web Checkout System That Eliminates the Need for Multiple Screens", *The New York Times*, 14 de fevereiro de 2005, p. C3). Suponha que uma empresa tenha modificado sua página na Web de modo a que os consumidores possam utilizar um processo de saída com uma única página, em vez de inúmeras páginas. Uma amostra de 500 consumidores que haviam selecionado seus produtos foi apresentada a esse novo sistema de saída. Desses 500 consumidores, 210 cancelaram suas transações depois de terem selecionado seus produtos.

a. No nível de significância de 0,01, existem evidências de que a proporção da população de consumidores que selecionaram produtos e cancelaram suas transações tenha sido inferior a 0,50 com o uso do novo sistema?
b. Suponha que tenha sido apresentado o novo sistema de saída a uma amostra de 100 consumidores (em vez de $n = 500$ consumidores) e que 42 deles tenham cancelado suas transações depois de terem selecionado suas compras. No nível de significância de 0,01, existem evidências de que a proporção da população de consumidores que selecionaram seus produtos e que depois cancelaram suas transações tenha sido menor do que 0,50 com o novo sistema?
c. Compare os resultados de (a) e (b) e discuta o efeito que o tamanho da amostra exerce sobre o resultado e, em geral, no teste de hipóteses.

9.58 Um estudo recente realizado pela Pew Internet e American Life Project (**pewinternet.org**) descobriu que os norte-americanos têm um comportamento complexo e ambivalente em relação à tecnologia (dados extraídos de M. Himowitz, "How to Tell What Kind of Tech User You Are", *Newsday*, 27 de maio de 2007, p. F6). O estudo relatou que 8% dos respondentes eram "Onívoros" que são amantes eletrônicos, mensageiros de texto ou jogadores virtuais (frequentemente com seus próprios *blogs* ou páginas na Web), *video makers* e com *parts* no YouTube. Você acredita que a percentagem de alunos em sua escola que são Onívoros é maior do que 8%, e você planeja conduzir um estudo que prove que isso é verdadeiro.

a. Expresse a hipótese nula e a hipótese alternativa.

Você seleciona uma amostra de 200 alunos em sua escola e descobre que 30 deles podem ser classificados como Onívoros.

b. Utilize a abordagem das seis etapas do valor crítico para o teste de hipóteses ou a abordagem de cinco etapas do valor-*p* de modo a determinar, no nível de significância de 0,05, se existem evidências de que a percentagem de Onívoros em sua escola é maior do que 8%.

9.5 Armadilhas Potenciais dos Testes de Hipóteses e Questões Éticas

Até este ponto, você estudou os conceitos fundamentais dos testes de hipóteses. Você utilizou testes de hipóteses para analisar diferenças entre estatísticas de amostras e parâmetros da população formulados em hipóteses para tomar decisões estratégicas em relação às características da população subjacente. Você aprendeu também a avaliar os riscos envolvidos na tomada dessas decisões.

Ao planejar a realização de um teste de hipóteses baseado em uma pesquisa, estudo investigatório ou experimento projetado, você precisa efetuar diversas perguntas para garantir que você está utilizando a metodologia apropriada. Você precisa trazer à tona e responder a diversas perguntas, como as apresentadas a seguir, no estágio de planejamento:

1. Qual é o objetivo da pesquisa, do estudo ou do experimento? Você é capaz de traduzir o objetivo em uma hipótese nula e uma hipótese alternativa?
2. O teste de hipóteses é um teste bicaudal ou um teste unicaudal?
3. Você consegue selecionar uma amostra aleatória a partir da população subjacente de interesse?
4. Que tipos de dados você coletará a partir da amostra? As variáveis são numéricas ou categóricas?
5. Sob qual nível de significância você deve conduzir o teste de hipóteses?
6. O tamanho de amostra pretendido é grande o suficiente para atingir a eficácia desejada para o teste no nível de significância escolhido?
7. Que procedimento de teste estatístico você deve utilizar, e por quê?
8. A que tipo de conclusões e interpretações você consegue chegar a partir dos resultados do teste de hipóteses?

Deixar de considerar essas questões logo nos primeiros estágios do processo de planejamento pode acarretar resultados com viés ou incompletos. O planejamento apropriado pode ajudar a assegurar que o estudo estatístico proporcione as informações objetivas necessárias para a tomada de decisões estratégicas bem fundamentadas.

Significado Estatístico *Versus* Significado Prático Você precisa fazer a distinção entre a existência de um resultado estatisticamente significativo e o seu significado prático no contexto de um determinado campo de aplicação. Algumas vezes, em decorrência de um tamanho de amostra demasiado grande, você pode obter um resultado estatisticamente significativo mas de pouco significado prático. Por exemplo, suponha que antes de uma campanha nacional de marketing que se concentre em uma série de comerciais de televisão com alto custo você ache que a proporção de pessoas que reconhecem a sua marca seja de 0,30. Ao final da campanha, uma pesquisa realizada junto a 20.000 pessoas indica que 6.168 reconheceram a sua marca. Um teste unicaudal tentando provar que a proporção é, agora, maior do que 0,30 resulta em um valor-*p* igual a 0,0047, e a conclusão estatística correta é que a proporção de consumidores que reconhecem a sua marca é, agora, maior. A campanha foi bem-sucedida? O resultado do teste de hipóteses indica um crescimento estatisticamente significativo em relação à conscientização sobre a marca, mas esse crescimento é importante em termos práticos? A proporção da população é agora estimada em 6.168/20.000 = 0,3084, ou 30,84%. Esse crescimento é menos de 1% a mais do que o valor de 30% especificado na hipótese. O grande volume de despesas associado à campanha de marketing produz um resultado com um crescimento significativo na conscientização sobre a marca? Em razão do impacto mínimo em relação ao mundo real que um crescimento de menos de 1% exerce sobre a estratégia de marketing em termos gerais e do gigantesco volume de despesas associado

à campanha de marketing, você deve concluir que a campanha não foi bem-sucedida. Por outro lado, se a campanha tivesse feito crescer em 20% a conscientização da marca, você poderia concluir que a campanha foi bem-sucedida.

Relatando as Descobertas Ao conduzir uma pesquisa, você deve documentar tanto os bons resultados quanto os maus resultados. Você não deve relatar simplesmente aqueles resultados do teste de hipóteses que demonstram ter significado estatístico e omitir aqueles para os quais existem evidências insuficientes nas descobertas. Em situações nas quais existem evidências insuficientes para rejeitar H_0, você deve deixar claro que isso não prova que a hipótese nula seja verdadeira. O que o resultado efetivamente indica é que, com o tamanho de amostra utilizado, não existem informações suficientes para *reprovar* a hipótese nula.

Questões Éticas Você precisa também fazer a distinção entre uma metodologia de pesquisas precária e um comportamento fora dos padrões da ética. Considerações éticas surgem quando o processo relativo ao teste de hipóteses é manipulado. Algumas das áreas nas quais podem surgir questões éticas incluem o uso de sujeitos humanos em experimentos, o método de coleta de dados, o tipo de teste (unicaudal ou bicaudal), a escolha do nível de significância, a limpeza e o descarte dos dados, bem como deixar de divulgar descobertas pertinentes.

9.6 *Tópico Online:* A EFICÁCIA DE UM TESTE

A Seção 9.1 define o erro do Tipo I e o erro do Tipo II, bem como a eficácia de um teste. Para examinar com mais profundidade a eficácia de um teste, leia o arquivo com o tópico *online* **Seção 9.6**, que está disponível no site da LTC Editora (veja a Seção D.8 do Apêndice para aprender a acessar os arquivos correspondentes aos tópicos *online*).

UTILIZANDO A ESTATÍSTICA

@ Oxford Cereals, Parte II Revisitada

Como gerente de operações da unidade de produção da Oxford Cereals, você era responsável pelo processo de abastecimento de cereais. Era de sua responsabilidade ajustar o processo quando a média aritmética do peso abastecido na população de caixas se desviava das especificações da empresa de 368 gramas. Uma vez que a pesagem de todas as caixas de cereais demandaria uma quantidade demasiada de tempo e não seria prática, você precisava selecionar e pesar uma amostra de caixas e conduzir um teste de hipóteses.

Você determinou que a hipótese nula deveria ser de que a média aritmética da população da quantidade abastecida correspondia a 368 gramas. Se os pesos das caixas selecionadas na amostra estivessem suficientemente acima ou abaixo da média aritmética esperada de 368 gramas, especificada pela Oxford Cereals, você rejeitaria a hipótese nula em favor da hipótese alternativa de que a média aritmética do abastecimento era diferente de 368 gramas. Caso isso acontecesse, você interromperia o processo de produção e adotaria as medidas necessárias para corrigir o problema. Se a hipótese nula não foi rejeitada, você continua a acreditar no *status quo*, de que o processo está operando corretamente, e, portanto, você não adota nenhuma medida corretiva.

Antes de prosseguir, você considerava os riscos envolvidos nos testes de hipóteses. Se você rejeitou a verdadeira hipótese nula, então cometeria um erro do Tipo I e concluiria que a média aritmética da população da quantidade abastecida não correspondia a 368 gramas quando ela efetivamente correspondia a 368 gramas. Esse erro resultaria em ajustar o processo de abastecimento, ainda que o processo estivesse funcionando adequadamente. Caso você não rejeitasse a falsa hipótese nula, cometeria então um erro do Tipo II, e concluiria que a média aritmética da população correspondente à quantidade abastecida era de 368, quando, na realidade, ela não era 368. Nesse caso, você permitiria que o processo continuasse sem ajustes, não obstante o fato de o processo não estar operando da maneira apropriada.

Depois de coletar uma amostra aleatória de 25 caixas de cereal, você utilizou a abordagem de seis etapas para o valor crítico para testar hipóteses. Uma vez que a estatística do teste se posicionou na região de não rejeição, você não rejeitou a hipótese nula. Você concluiu que não existiam evidências suficientes para comprovar que a média aritmética correspondente à quantidade abastecida era diferente de 368 gramas. Nenhuma ação corretiva era necessária em relação ao processo de abastecimento.

RESUMO

Este capítulo apresentou os fundamentos de testes de hipóteses. Você aprendeu a realizar testes Z e testes em relação à média aritmética da população e à proporção da população. O capítulo desenvolveu tanto a abordagem do valor crítico quanto a abordagem do valor-p para testes de hipóteses.

Ao decidir qual teste utilizar, você deve fazer a seguinte pergunta: O teste envolve uma variável numérica ou uma variável categórica? Se o teste envolve uma variável categórica, utilize o teste Z para a proporção. Se o teste envolve uma variável numérica, utilize o teste t para a média aritmética. A Tabela 9.4 fornece uma lista de testes de hipóteses cobertos neste capítulo.

TABELA 9.4
Resumo dos Tópicos no Capítulo 9

	Tipo de Dados	
Tipo de Análise	**Numéricos**	**Categóricos**
Teste de hipóteses com relação a um único parâmetro	Teste t de hipóteses para a média aritmética (Seção 9.2)	Teste Z de hipóteses para a proporção (Seção 9.4)

EQUAÇÕES-CHAVE

Teste Z de Hipóteses para a Média Aritmética (σ Conhecido)

$$Z_{ESTAT} = \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \tag{9.1}$$

Teste t de Hipóteses para a Média Aritmética (σ Desconhecido)

$$t_{ESTAT} = \frac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}} \tag{9.2}$$

Teste Z para a Proporção

$$Z_{ESTAT} = \frac{p - \pi}{\sqrt{\dfrac{\pi(1 - \pi)}{n}}} \tag{9.3}$$

Teste Z para a Proporção em Termos do Número de Eventos de Interesse

$$Z_{ESTAT} = \frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}} \tag{9.4}$$

TERMOS-CHAVE

coeficiente de confiança
eficácia de um teste estatístico
erro do Tipo I
erro do Tipo II
estatística do teste
hipótese alternativa (H_1)
hipótese nula (H_0)
nível de confiança
nível de significância (α)
proporção da amostra
região de não rejeição
região de rejeição
risco β
robusto
teste bicaudal
teste de hipóteses
teste direcional
teste t para a média aritmética
teste unicaudal
teste Z para a média aritmética
teste Z para a proporção
valor crítico
valor-p

PROBLEMAS DE REVISÃO DO CAPÍTULO 9

AVALIANDO O SEU ENTENDIMENTO

9.59 Qual é a diferença entre uma hipótese nula, H_0, e uma hipótese alternativa, H_1?

9.60 Qual é a diferença entre um erro do Tipo I e um erro do Tipo II?

9.61 Qual o significado da eficácia de um teste?

9.62 Qual é a diferença entre um teste unicaudal e um teste bicaudal?

9.63 O que significa um valor-p?

9.64 Como uma estimativa de intervalo de confiança para a média aritmética da população pode fornecer conclusões para o correspondente teste de hipóteses bicaudal para a média aritmética da população?

9.65 O que representa a abordagem das seis etapas para o valor crítico para testes de hipóteses?

9.66 O que representa a abordagem das cinco etapas para o valor-p para testes de hipóteses?

APLICANDO OS CONCEITOS

9.67 Um artigo na *Marketing News* (T. T. Semon, "Consider a Statistical Insignificance Test", *Marketing News*, 1º de fevereiro de 1999) argumenta que o nível de significância utilizado na comparação entre dois produtos geralmente é demasiadamente baixo — ou seja, algumas vezes você deveria utilizar um valor de α maior do que 0,05. Especificamente, o artigo refaz um teste sobre a proporção de potenciais consumidores com uma preferência pelo produto 1 em relação ao produto 2. A hipótese nula era de que a proporção da população de potenciais consumidores que preferiam o produto 1 correspondia a 0,50, e a hipótese alternativa era de que a proporção não seria igual a 0,50. O valor-p para o teste era 0,22. O artigo sugere que, em alguns casos, isso deve constituir evidência suficiente para rejeitar a hipótese nula.

a. Especifique, em termos estatísticos, a hipótese nula e a hipótese alternativa para esse exemplo.
b. Explique os riscos associados ao erro do Tipo I e ao erro do Tipo II no caso presente.
c. Quais seriam as consequências caso você rejeitasse a hipótese nula para um valor-p de 0,22?
d. Por que você imagina que o artigo sugeriu aumentar o valor de α?
e. O que você faria nessa situação?
f. Qual seria sua resposta para (e) caso o *valor-p* fosse igual a 0,12? E se ele fosse igual a 0,06?

9.68 A rede La Quinta Motor Inns desenvolveu um modelo informatizado para ajudar a prever a rentabilidade de terrenos que estão sendo avaliados como localizações para novos hotéis. Se o modelo informatizado prevê uma rentabilidade alta, a rede La Quinta compra o terreno proposto e constrói um hotel. Se o modelo informatizado prevê uma rentabilidade pequena ou moderada, a rede La Quinta opta por não prosseguir com a negociação em relação àquele terreno (dados extraídos de S. E. Kimes e J. A. Fitzsimmons, "Selecting Profitable Hotel Sites at La Quinta Motor Inns", *Interfaces*, Vol. 20, março-abril de 1990, pp. 12-20.) Esse processo de tomada de decisão pode ser expresso em uma estrutura de teste de hipóteses. A hipótese nula é de que o terreno não é um local rentável. A hipótese alternativa é de que o terreno é um local rentável.

a. Explique os riscos associados a cometer um erro do Tipo I nesse caso.
b. Explique os riscos associados a cometer um erro do Tipo II nesse caso.
c. Qual entre os tipos de erro você imagina que os executivos da rede La Quinta Motor Inns desejariam evitar? Explique.
d. De que modo alterações no critério de rejeição afetam as probabilidades de serem cometidos erros do Tipo I e do Tipo II?

9.69 A Webcredible, uma empresa de consultoria com sede no Reino Unido, especializada em portais da Web, redes informatizadas internas (*intranets*), dispositivos móveis e aplicações, conduziu uma pesquisa junto a 1.132 usuários de aparelhos de telefone móveis entre fevereiro e abril de 2009. A pesquisa descobriu que 52% dos usuários de aparelhos de telefone móveis estão atualmente utilizando a Internet móvel (dados extraídos de "Email and Social Networking Most Popular Mobile Internet Activities", **www.webcredible.co.uk**, 13 de maio de 2009). Os autores do artigo deduzem que a pesquisa prova que mais da metade de todos os usuários de aparelhos de telefone móveis está atualmente utilizando a Internet móvel.

a. Utilize a abordagem de cinco etapas do valor-p para o teste de hipóteses e um nível de significância de 0,05 para tentar provar que mais da metade de todos os usuários de aparelhos de telefone móveis está atualmente utilizando a Internet móvel.
b. Com base na sua resposta para (a), a afirmativa deduzida pelos autores seria válida?
c. Suponha que a pesquisa tivesse descoberto que 53% dos usuários de aparelhos de telefone móveis estão atualmente utilizando a Internet móvel. Repita os itens (a) e (b).
d. Compare os resultados de (a) e (c).

9.70 O proprietário de um posto de gasolina deseja estudar os hábitos relacionados à compra de gasolina pelos motoristas em seu posto. Ele seleciona uma amostra aleatória de 60 motoristas, durante uma determinada semana, com os seguintes resultados:

- A quantidade adquirida era $\bar{X} = 11{,}3$ galões, $S = 3{,}1$ galões.
- Onze motoristas adquiriram gasolina do tipo *premium* aditivada.

a. No nível de significância de 0,05, existem evidências de que a média aritmética da população relacionada à aquisição de gasolina seja diferente de 10 galões?
b. Determine o valor-p em (a).
c. No nível de significância de 0,05, existem evidências de que menos de 20% de todos os motoristas adquirem gasolina do tipo *premium* aditivada?
d. Qual seria sua resposta em (a) se a média aritmética da amostra correspondesse a 10,3 galões?
e. Qual seria sua resposta em (c) se 7 motoristas tivessem adquirido gasolina do tipo *premium* aditivada?

9.71 É atribuída a uma auditora de uma agência governamental a tarefa de avaliar os reembolsos pagos aos médicos pelo Medicare por consultas médicas realizadas em consultório. A auditoria foi conduzida em uma amostra de 75 dos reembolsos, com os seguintes resultados:

- Em 12 das consultas em consultório, foi reembolsada uma quantia incorreta.
- O volume de reembolso foi $\bar{X} = \$93,70$, $S = \$34,55$.

a. Em um nível de significância de 0,05, existem evidências de que a média aritmética do volume de reembolso seja menor do que \$100?
b. No nível de significância de 0,05, existem evidências de que a proporção de reembolsos incorretos na população seja maior do que 0,10?
c. Discuta os pressupostos subjacentes para o teste utilizado em (a).
d. Qual seria a sua resposta em (a) se a média aritmética da amostra fosse igual a \$90?
e. Qual seria a sua resposta em (b) se 15 consultas em consultórios tivessem tido reembolsos incorretos?

9.72 Uma agência bancária localizada em um bairro comercial de uma cidade desenvolveu um processo de aperfeiçoamento para atendimento aos clientes durante o horário de pico do almoço, do meio-dia às 13 horas. O tempo de espera (definido como o intervalo de tempo entre o momento em que o cliente entra na fila até o seu atendimento no caixa) de todos os clientes atendidos naquele horário foi registrado ao longo de uma semana. Foi selecionada uma amostra aleatória de 15 clientes, e os resultados (armazenados em **Banco1**) são os seguintes:

4,21	5,55	3,02	5,13	4,77	2,34	3,54	3,20
4,50	6,10	0,38	5,12	6,46	6,19	3,79	

a. No nível de significância de 0,05, existem evidências de que a média aritmética do tempo de espera seja inferior a 5 minutos?
b. Que pressuposto em relação à distribuição da população precisa ser adotado para realizar o teste t em (a)?
c. Construa um *box-plot* ou um gráfico da distribuição normal para avaliar o pressuposto adotado em (b).
d. Você acredita que o pressuposto necessário para conduzir o teste t em (a) é válido? Explique.
e. Tão logo uma cliente entra na agência durante o horário do almoço, ela pergunta ao gerente da agência quanto tempo deve esperar para ser atendida. O gerente responde: "Quase certamente não mais de 5 minutos." Com base nos resultados de (a), avalie essa afirmativa.

9.73 Uma empresa de produção industrial fabrica isoladores elétricos. Se os isoladores quebram durante o uso, existe a possibilidade de que venha a ocorrer um curto-circuito. Para testar a resistência dos isoladores, é realizado um teste de destruição para determinar a quantidade de força necessária para quebrar os isoladores. A força é medida observando-se a quantidade de libras de força que pode ser aplicada ao isolador, antes que ele quebre. Os dados a seguir (armazenados no arquivo **Força**) são de 30 isoladores submetidos a esse teste.

1.870	1.728	1.656	1.610	1.634	1.784	1.522	1.696	1.592	1.662
1.866	1.764	1.734	1.662	1.734	1.774	1.550	1.756	1.762	1.866
1.820	1.744	1.788	1.688	1.810	1.752	1.680	1.810	1.652	1.736

a. Em um nível de significância de 0,05, existem evidências de que a média aritmética da população correspondente à força seja maior do que 1.500 libras?
b. Que pressuposto em relação à distribuição da população precisa ser adotado para realizar o teste t em (a)?
c. Construa um histograma, um *box-plot* ou um gráfico da distribuição normal para avaliar o pressuposto adotado em (b).
d. Você acha que o pressuposto necessário para conduzir o teste t em (a) é válido? Explique.

9.74 Uma importante característica da qualidade utilizada pelo fabricante de placas de asfalto das marcas Boston e Vermont corresponde à taxa de umidade que as placas contêm quando estão embaladas. Os clientes podem achar que adquiriram um produto de baixa qualidade se encontrarem umidade e placas ainda úmidas dentro da embalagem. Em alguns casos, a umidade excessiva pode causar a queda de grãos fixados à placa. Esses grãos servem para dar coloração e textura às placas, resultando em problemas de aparência. Uma placa é pesada e depois seca. A placa é novamente pesada, e, tomando-se por base a quantidade de umidade extraída do produto, são calculadas as libras de umidade por 100 pés quadrados. A empresa gostaria de demonstrar que a média aritmética do teor de umidade é menor do que 0,35 libra por 100 pés quadrados. O arquivo de dados **Umidade** inclui 36 medições (em libras por 100 pés quadrados) para as placas Boston e 31 medições para as placas Vermont.

a. No que se refere às placas Boston, existem evidências de que, em um nível de significância equivalente a 0,05, a média aritmética do teor de umidade seja menor do que 0,35 libra por 100 pés quadrados?
b. Interprete o significado do valor-p em (a).
c. No que se refere às placas Vermont, existem evidências de que, em um nível de significância de 0,05, a média aritmética do teor de umidade seja menor do que 0,35 libra por 100 pés quadrados?
d. Interprete o significado do valor-p em (c).
e. Que pressuposto em relação à distribuição da população precisa ser adotado para realizar os testes t em (a) e (c)?
f. Construa um histograma, um *box-plot* ou um gráfico da distribuição normal para avaliar o pressuposto adotado em (a) e (c).
g. Você acha que o pressuposto necessário para conduzir o teste t em (a) e (c) é válido? Explique.

9.75 Estudos conduzidos pelo fabricante das placas de asfalto Boston e Vermont demonstraram que o peso das placas é um fator importante na percepção da qualidade pelos consumidores. Mais do que isso, o peso corresponde à quantidade de matéria-prima que está sendo utilizada e é, por conseguinte, muito importante para a empresa, do ponto de vista de custos. O último estágio da linha de produção embala as placas antes de as embalagens serem colocadas em paletes de madeira. Uma vez preenchido um palete (um palete, para a maioria das marcas, contém 16 pés quadrados de placas), ele é pesado, e a medição é registrada. O arquivo de dados **Palete** contém o peso (em libras) de uma amostra com 368 paletes de placas Boston e 330 paletes de placas Vermont.

a. Para as placas Boston, existem evidências de que a média aritmética da população correspondente ao peso seja diferente de 3.150 libras?
b. Interprete o significado do valor-p em (a).
c. Para as placas Vermont, existem evidências de que a média aritmética da população correspondente ao peso seja diferente de 3.700 libras?
d. Interprete o significado do valor-p em (c).

e. Nos itens de (a) a (d), você precisa se preocupar com o pressuposto da normalidade? Explique.

9.76 O fabricante das placas de asfalto Boston e Vermont dá a seus clientes uma garantia de 20 anos para a maior parte de seus produtos. Para determinar se uma placa durará todo o período da garantia, são conduzidos, na área de produção, testes de aceleração de vida útil. O teste de aceleração de vida útil expõe a placa, em um laboratório, às condições de desgaste às quais ela estaria sujeita ao longo de uma vida útil de utilização normal, por meio de um experimento que leva apenas alguns poucos minutos para ser realizado. Nesse teste, uma placa é repetidamente esfregada com uma escova, por um curto período de tempo, e os grãos da placa removidos por meio da escovação são então pesados (em gramas). Espera-se que as placas que sofrem a perda de poucas quantidades de grãos durem mais, com a utilização normal, do que as placas que sofrem a perda de grandes quantidades de grãos. O arquivo de dados **Grão** contém uma amostra com 170 medições realizadas nas placas Boston e 140 medições realizadas nas placas Vermont.

a. No que se refere às placas Boston, existem evidências de que a média aritmética da perda de grãos seja diferente de 0,50 grama?
b. Interprete o significado do valor-p em (a).
c. No que se refere às placas Vermont, existem evidências de que a média aritmética da perda de grãos seja diferente de 0,50 grama?
d. Interprete o significado do valor-p em (c).
e. Nos itens de (a) a (d), você precisa se preocupar com o pressuposto da normalidade? Explique.

EXERCÍCIO DE REDAÇÃO DE RELATÓRIOS

9.77 Com referência aos resultados dos Problemas 9.74 a 9.76, que tratam das placas Boston e Vermont, redija um relatório que avalie o teor de umidade, o peso e a perda de grãos nos dois tipos de placas.

ADMINISTRANDO O *SPRINGVILLE HERALD*

Dando continuidade ao monitoramento da intensidade de tinta na impressão do jornal, inicialmente descrito no caso "Administrando o *Springville Herald*" do Capítulo 6, o departamento de produção do jornal quer garantir que a média aritmética da intensidade de tinta na impressão de todos os jornais seja igual a pelo menos 0,97, em uma escala padronizada na qual a meta é 1,0. Foi selecionada uma amostra aleatória com 50 jornais e mensurada a intensidade de tinta em um ponto em cada um dos 50 jornais (e armazenada no arquivo **SH9**).

0,854	1,023	1,005	1,030	1,219	0,977	1,044	0,778	1,122	1,114
1,091	1,086	1,141	0,931	0,723	0,934	1,060	1,047	0,800	0,889
1,012	0,695	0,869	0,734	1,131	0,993	0,762	0,814	1,108	0,805
1,223	1,024	0,884	0,799	0,870	0,898	0,621	0,818	1,113	1,286
1,052	0,678	1,162	0,808	1,012	0,859	0,951	1,112	1,003	0,972

Calcule as estatísticas da amostra e determine se existem evidências de que a média aritmética da população correspondente à intensidade de tinta é menor do que 0,97. Redija para a administração do jornal um memorando que sintetize as suas conclusões.

CASO DE INTERNET

Aplique seus conhecimentos sobre testes e hipóteses neste Caso de Internet, que dá continuidade à controvérsia sobre o abastecimento e empacotamento de cereais apresentado no Caso de Internet do Capítulo 7.

Em resposta às declarações negativas realizadas por grupos de defesa do consumidor como o CCACC — Consumers Concerned About Cereal Cheaters (Consumidores Preocupados com Fraudadores de Cereais), no Caso de Internet do Capítulo 7, a Oxford Cereals conduziu, recentemente, um experimento relativo ao empacotamento de cereais. A empresa declara que os resultados dos experimentos refutam as alegações do CCACC de que a Oxford Cereals vem fraudando os consumidores por colocar nas embalagens de cereais uma quantidade inferior ao peso especificado no rótulo.

Utilizando o site da LTC Editora para este livro, abra a página correspondente ao Caso de Internet para o Capítulo 9, ou abra diretamente o arquivo **OC_FullUp.htm** caso já tenha baixado para seu computador de uso pessoal os arquivos contendo os Casos de Internet, para examinar o artigo divulgado pela Oxford Cereals e os documentos de apoio que descrevem o experimento. Depois disso, responda às seguintes perguntas:

1. Os resultados para o experimento são válidos? Por que sim ou por que não? Se você estivesse conduzindo esse experimento, você modificaria alguma coisa?
2. Os resultados respaldam a declaração de que a Oxford Cereals não está enganando os consumidores?
3. A declaração da alta direção da Oxford Cereals de que muitas caixas de cereal contêm *mais* de 368 gramas é surpreendente? Ela é verdadeira?
4. Poderia existir alguma circunstância em que os resultados do experimento realizado pela Oxford Cereals *e* os resultados do CCACC estejam ambos corretos? Explique.

REFERÊNCIAS

1. Bradley, J. V., *Distribution-Free Statistical Tests* (Upper Saddle River, NJ: Prentice Hall, 1968).
2. Daniel, W., *Applied Nonparametric Statistics*, 2nd ed. (Boston: Houghton Mifflin, 1990).
3. *Microsoft Excel 2007* (Redmond, WA: Microsoft Corp., 2007).