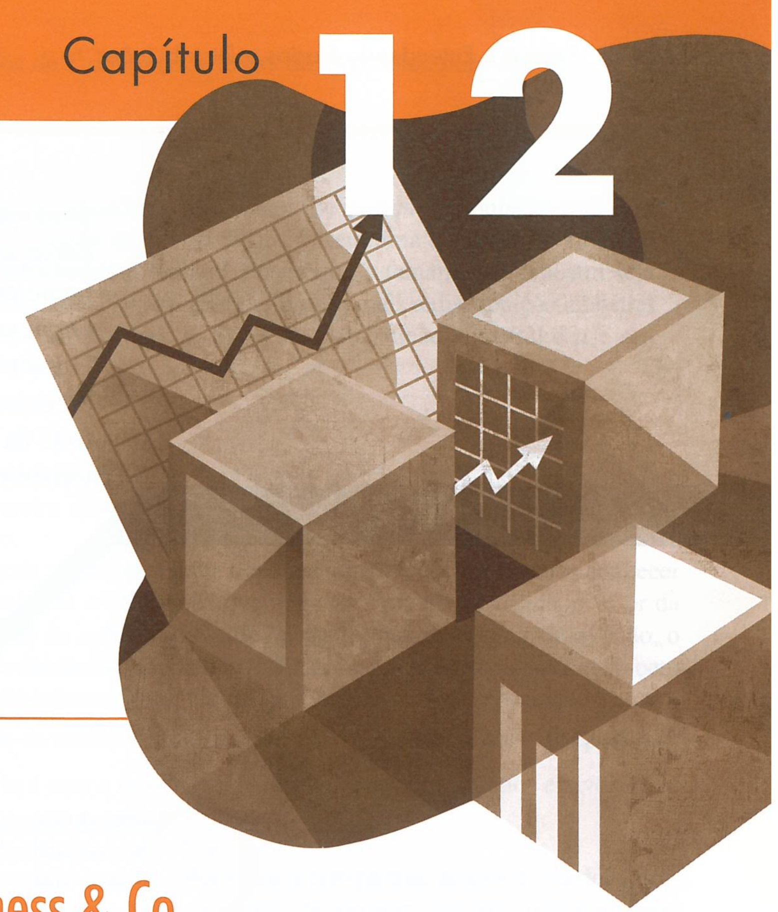
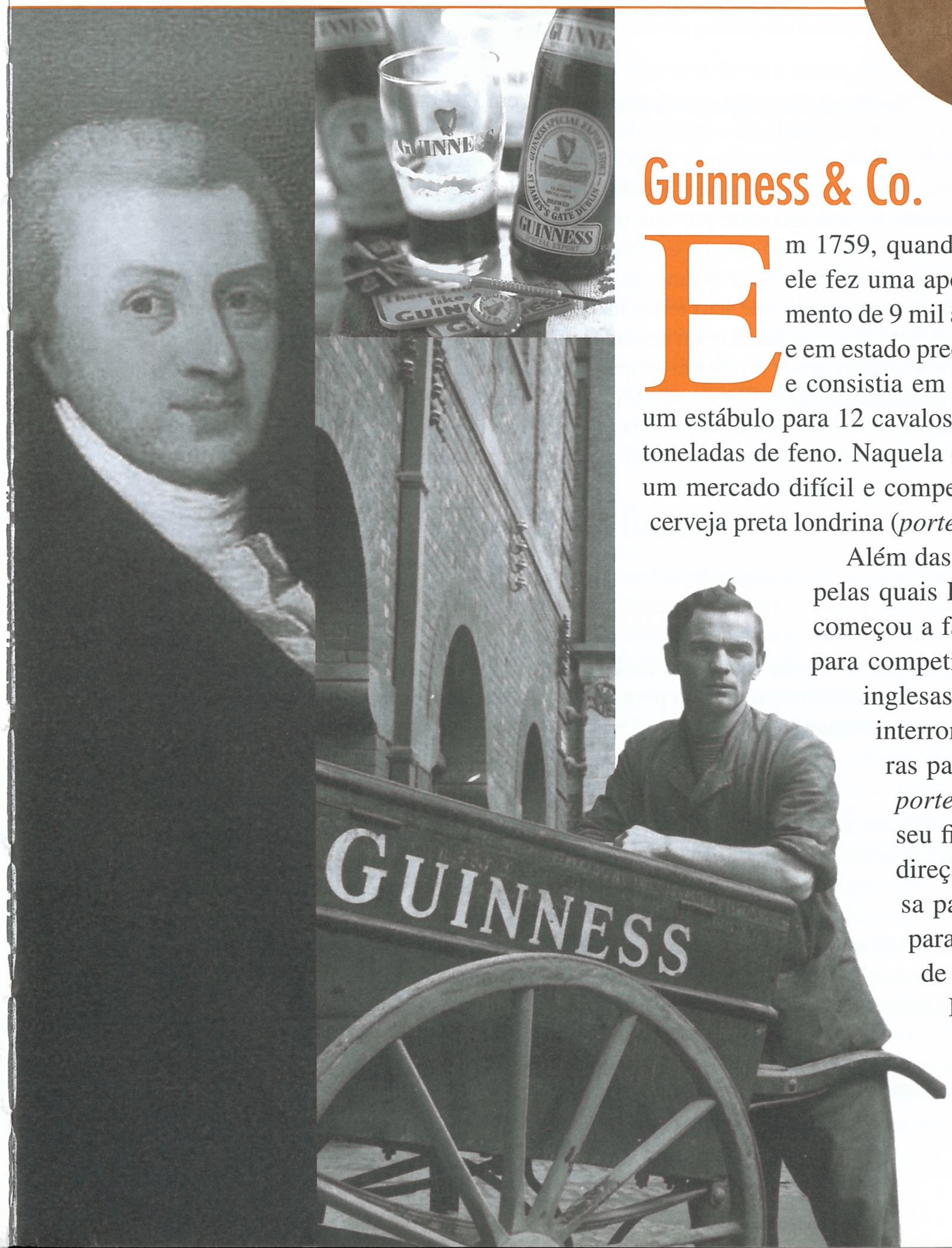
Capítulo 12



Intervalos de Confiança e Testes de Hipóteses para Médias



Guinness & Co.

Em 1759, quando Arthur Guinness tinha 34 anos, ele fez uma aposta incrível: assinou um arrendamento de 9 mil anos de uma cervejaria abandonada e em estado precário. O terreno cobria quatro acres e consistia em um moinho, duas casas de malte, um estábulo para 12 cavalos e um celeiro que comportava 200 toneladas de feno. Naquela época, a fabricação da cerveja era um mercado difícil e competitivo. Gin, uísque e a tradicional cerveja preta londrina (*porter*) eram as bebidas favoritas.

Além das cervejas claras mais fortes (*ales*), pelas quais Dublin era conhecida, a Guinness começou a fabricar cervejas escuras (*porters*), para competir diretamente com as cervejarias inglesas. Quarenta anos depois, Guinness interrompeu a produção das cervejas claras para se concentrar nas suas *stouts* e *porters*. Quando ele faleceu, em 1803, seu filho Arthur Guinness II assumiu a direção e, alguns anos depois a empresa passou a exportar a Guinness *stout* para outros países da Europa. Por volta de 1830, a Guinness St. James's Gate Brewery havia se tornado a maior cervejaria da Irlanda. Em 1886, a cervejaria Guinness, com uma produção anual de 1,2 milhão de barris, foi a primeira grande cervejaria a ser incorporada no Mercado de Ações de Londres como empresa de capital aberto. Durante os anos de 1890, a empresa começou a empregar cientistas. Um deles, William S. Gosset, foi contratado como químico para testar a qualidade do processo de fermentação. Além de pioneiro do método de controle da qualidade na indústria, o trabalho estatístico de Gosset tornou possível a inferência estatística moderna.[1]

Como químico na cervejaria Guinnes, em Dublin, William S. Gosset estava no comando do controle de qualidade. Seu trabalho era certificar-se de que a *stout* (uma cerveja preta, espessa) que saía da cervejaria tinha qualidade alta o suficiente para satisfazer os padrões dos exigentes consumidores da Guinness. É fácil imaginar por que testar uma grande quantidade de *stout* pode ser indesejável, sem mencionar o perigo para a saúde. Assim, para verificar a qualidade, Gosset geralmente usava uma amostra de somente três ou quatro observações por lote. No entanto, ele notou que seus testes não eram precisos com amostras desse tamanho: quando os lotes rejeitados eram enviados de volta ao laboratório para testes mais intensivos, muitas vezes verificava-se que os resultados do teste estavam errados. Como um estatístico treinado, Gosset sabia que tinha de estar errado *algumas* vezes, mas ele detestava errar com maior frequência do que a teoria previa. O resultado das frustrações de Gosset foi o desenvolvimento de um teste para lidar com pequenas amostras, o principal assunto deste capítulo.

12.1 A distribuição amostral da média

Você aprendeu a criar intervalos de confiança e testar hipóteses sobre proporções. Agora queremos fazer o mesmo para as médias. Para proporções, encontramos o intervalo de confiança como:

$$\hat{p} \pm ME,$$

A ME era igual ao valor crítico, z^*, vezes $EP(\hat{p})$. Nosso intervalo de confiança para as médias será:

$$\bar{y} \pm ME.$$

E nosso ME será um valor crítico vezes $EP(\bar{y})$. Assim, vamos colocar as peças no lugar. O que o Teorema Central do Limite nos ensinou, no Capítulo 9, parece ser o que precisamos.

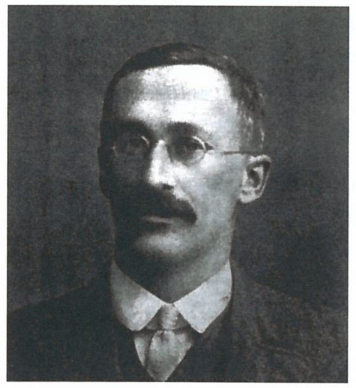
[1] Fonte: Guinness & Co. 2006, www.guinness.co/global/story/history.

Teorema Central do Limite

Quando uma amostra aleatória é selecionada de *qualquer* população com uma média de μ e desvio padrão de σ, sua média, $\bar{y}$, tem uma distribuição amostral cuja *forma* é aproximadamente Normal, desde que o tamanho da amostra seja grande o suficiente. Quanto maior a amostra usada, mais a distribuição amostral da média se aproxima da Normal. A média da distribuição amostral é μ e seu desvio padrão é $DP(\bar{y}) = \frac{\sigma}{\sqrt{n}}$.

Isso nos dá uma distribuição amostral e um desvio padrão para a média. Tudo o que precisamos é uma amostra aleatória de dados quantitativos e o valor real do desvio padrão da população σ.

Mas espere. Isso pode ser um problema. Para calcular $\sigma/\sqrt{n}$, precisamos conhecer σ. Como podemos conhecer σ? Suponha que, para 25 executivos jovens, o valor da média de seus portfólios de ações seja $125672. Isso indicaria o valor de σ? Não, o desvio padrão depende da similaridade do investimento dos executivos, não quão bem eles investiram (a média informa isso). No entanto, precisamos de σ porque é o numerador do desvio padrão da média da amostra: $DP(\bar{y}) = \frac{\sigma}{\sqrt{n}}$. Portanto, o que podemos fazer? A resposta óbvia é usar o desvio padrão da amostra, s, dos dados em vez de σ. O resultado é o erro padrão: $EP(\bar{y}) = \frac{s}{\sqrt{n}}$.



Para encontrar a distribuição amostral de $\frac{\bar{y}}{s/\sqrt{n}}$, Gosset a simulou à mão. Ele extraiu tiras de papel de amostras pequenas de um chapéu centenas de vezes e calculou as médias e os desvios padrão com uma calculadora à manivela. Hoje, você poderia repetir em segundos no computador esse experimento que levou mais de um ano. O trabalho de Gosset foi tão meticuloso que, além de conseguir a forma aproximadamente certa do novo histograma, ele ainda descobriu a fórmula exata para ela a partir da amostra. A fórmula só foi confirmada matematicamente anos mais tarde por Sir R. A. Fisher.

Há um século, as pessoas apenas colocavam o erro padrão dentro do modelo Normal, assumindo que funcionaria. Para tamanhos de amostra maiores, isso *realmente* funcionava. No entanto, havia problemas com amostras menores. A variação adicional no erro padrão criava um caos com os valores-P e as margens de erro.

William S. Gosset foi o primeiro a investigar esse fenômeno. Ele percebeu que, além de permitir variação extra com as margens de erro maiores e os valores-P, também precisamos de um novo modelo de distribuição amostral. De fato, precisamos de toda uma *família* de modelos, dependendo do tamanho da amostra, n. Esses modelos são unimodais, simétricos e em forma de sino, mas quanto menor for nossa amostra, mais precisaremos expandir as caudas. O trabalho de Gosset revolucionou a estatística, mas muitos que o utilizam sequer conhecem o nome do químico.

O *t* de Gosset

Gosset verificava a qualidade da *stout* executando testes de hipóteses. Ele sabia que se determinasse $\alpha = 0{,}05$, o teste faria alguns erros do Tipo I rejeitando aproximadamente 5% dos lotes bons da *stout*. Entretanto, o laboratório relatou que ele estava rejeitando, na verdade, aproximadamente 15% dos lotes bons. Gosset sabia que algo estava errado e isso o incomodava.

Gosset parou de trabalhar temporariamente para estudar o problema e conseguir um diploma de graduação no campo emergente da estatística. Ele descobriu que, quando usava o erro padrão $\frac{s}{\sqrt{n}}$, a forma do modelo amostral não era mais Normal. Descobriu ainda o que era o novo modelo e o chamou de distribuição t.

A Guinness não forneceu muito apoio ao trabalho de Gosset. De fato, ela tinha uma política contra a publicação de resultados. Gosset precisou convencer a empresa de que ele não estava publicando um segredo industrial e, como parte da permissão para a publicação, teve de usar um pseudônimo. O pseudônimo que ele escolheu foi "Student" (Estudante), e, desde então, o modelo que ele descobriu é conhecido como o ***t* de Student**.

O modelo de Gosset tem sempre a forma de sino, mas os detalhes mudam com o tamanho da amostra. Assim, os modelos t de Student formam uma família de distribuições relacionadas, que dependem de um parâmetro conhecido como **graus de liberdade**. Geralmente, denotamos os graus de liberdade como gl e o modelo como t_{gl}, com o valor numérico dos graus de liberdade como um subscrito.

! ALERTA DE NOTAÇÃO:
Desde Gosset, a letra t tem sido usada em estatística apenas para a sua distribuição.

12.2 Um intervalo de confiança para médias

Para fazer um intervalo de confiança ou testar hipóteses para as médias, precisamos usar o modelo de Gosset. Qual deles? Bem, para médias, verifica-se que o valor correto para os graus de liberdade é gl $= n - 1$.

Modelo da distribuição amostral para médias

Quando certas condições são satisfeitas, a média da amostra padronizada,

$$t = \frac{\bar{y} - \mu}{SE(\bar{y})}$$

segue o modelo t de Student com $n - 1$ graus de liberdade. Encontramos o erro padrão a partir de:

$$EP(\bar{y}) = \frac{s}{\sqrt{n}}$$

Quando Gosset corrigiu o modelo Normal para a incerteza extra, a margem de erro ficou maior, como você já deve ter imaginado. Quando você usa o modelo de Gosset em vez do modelo Normal, seus intervalos de confiança serão um pouco maiores e seus valores-P, mais altos (Figura 12.1). Essa é a correção que você precisa. Ao usar o modelo t, você compensa a variabilidade extra de maneira correta.

Intervalo t de uma amostra

Quando as suposições e condições são satisfeitas, estamos prontos para encontrar o **intervalo de confiança para a média da população,** μ. O intervalo de confiança é:

$$\bar{y} \pm t^*_{n-1} \times SE(\bar{y}),$$

onde o erro padrão da média é:

$$EP(\bar{y}) = \frac{s}{\sqrt{n}}$$

O valor crítico t^*_{n-1} depende do nível de confiança particular, C, que você especifica e do número dos graus de liberdade, $n - 1$, que obtemos a partir do tamanho da amostra.

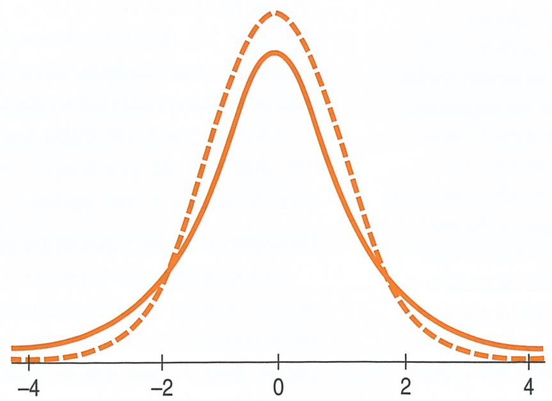


Figura 12.1 O modelo t (curva sólida) com 2 graus de liberdade tem caudas mais grossas do que o modelo Normal Padrão (curva pontilhada). Assim, a Regra 68-95-99,7 não funciona para modelos t com pequenos graus de liberdade.

Os modelos t de Student são unimodais, simétricos e com forma de sino, como o modelo Normal. No entanto, modelos t com pequenos graus de liberdade têm um pico mais estreito do que o modelo Normal e têm caudas mais grossas (isso é o que torna a margem de erro maior.) À medida que os graus de liberdade aumentam, os modelos t se parecem cada vez mais com o modelo Normal Padrão. Na verdade, o modelo t com grau de liberdade infinito é exatamente igual ao Normal Padrão.[2] Isso é ótimo se você tiver um número infinito de valores dos dados. Infelizmente, não é algo prático. Felizmente, acima de algumas centenas de graus de liberdade, é muito difícil notar a diferença. É claro, numa situação rara em que *conhecemos* σ, seria ridículo não usar essa informação. Se não tivermos de estimar σ, podemos usar o modelo Normal. Tipicamente, o conhecimento do valor de σ pode estar relacionado à (muita) experiência ou a um modelo teórico. Normalmente, entretanto, estimamos σ por s a partir dos dados e usamos, então, o modelo t.

z ou t?
Se você conhece σ, use z (isso é raro!). Sempre que você usar s para estimar σ, use t.

12.3 Suposições e condições

Gosset descobriu o modelo t por simulação. Anos mais tarde, quando Sir Ronald Fisher mostrou matematicamente que Gosset estava certo, ele precisou fazer algumas suposições para que a prova funcionasse. Essas são as suposições que precisamos usar nos modelos t de Student.

Não *queremos* parar
Verificamos as condições a fim de fazer uma análise significativa dos nossos dados. As condições servem como critério de *desqualificação* – continuamos, a não ser que exista um problema sério. Se encontrarmos problemas menores, os observamos e manifestamos cautela sobre os nossos resultados. Se a amostra não for AAS, mas acreditamos que ela seja representativa da população, limitamos nossas conclusões de forma correspondente. Se houver valores atípicos, em vez de parar, executamos nossa análise com e sem eles. Se a amostra parece bimodal, tentamos analisar os subgrupos separadamente. Somente quando existe um problema maior – como uma amostra pequena fortemente assimétrica ou uma amostra obviamente não representativa – não podemos proceder de forma alguma.

Suposição de independência

Suposição de independência: Os valores dos dados devem ser independentes. Não há como verificar a independência dos dados olhando para a amostra, mas devemos pensar se essa suposição é razoável.

Condição de aleatoriedade: Os dados resultam de uma amostra aleatória ou de um experimento padronizado adequado. Dados amostrados aleatoriamente – especialmente dados de uma Amostra Aleatória Simples (AAS) – são ideais.

Quando uma amostra é coletada sem substituição, tecnicamente temos de confirmar que não amostramos uma fração grande da população, o que poderia colocar em risco a independência das nossas seleções.

Condição dos 10%: O tamanho da amostra não deve exceder mais de 10% da população. Na prática, não mencionamos a Condição dos 10% quando estimamos as médias. Por quê? Quando fizemos inferências sobre as proporções, essa condição era crucial, já que normalmente tínhamos amostras grandes. Porém, nossas amostras, em geral, são menores para as médias, assim, esse problema surge somente se estivermos amostrando uma população pequena (e, nesse caso, existe uma fórmula de correção que poderemos utilizar).

Suposição de normalidade

Os modelos t de Student não funcionarão para dados extremamente assimétricos. Quão assimétrico é muito assimétrico? Formalmente, assumimos que os dados são de uma população que segue o modelo Normal. De maneira prática, não há como ter certeza de que isso é verdade.

Quase certamente *não* é verdadeiro. Os modelos são idealizados; os dados reais são, bem, reais. A boa notícia, entretanto, é que, mesmo para pequenas amostras, é suficiente verificar uma condição.

Condição de normalidade. Os dados vêm de uma distribuição unimodal e simétrica. Trata-se de uma condição muito mais prática, que podemos verificar com um histograma.[3] Para amostras menores, pode ser difícil ver uma forma de distribuição

[2] Formalmente, no limite, à medida que o número de graus de liberdade vai ao infinito.

[3] Ou podemos verificar uma representação mais adequada, chamada de diagrama de probabilidade normal, discutida no Capítulo 16.

ALERTA DE NOTAÇÃO:

Quando determinamos valores críticos de um modelo Normal, os representamos por z^*. Quando usamos um modelo t de Student, os valores críticos são representados por t^*.

no histograma. Infelizmente, a condição é mais importante quanto mais difícil for de verificar.[4]

Para uma amostra pequena ($n < 15$ mais ou menos), os dados devem seguir um modelo Normal à risca. É claro que, com tão poucos dados, é difícil verificar isso. Porém, se você encontrar valores atípicos ou forte assimetria, não use esses métodos.

Para tamanhos da amostra moderados (n entre 15 e 40, mais ou menos), os métodos t funcionarão bem para dados unimodais e razoavelmente assimétricos. Faça um histograma para verificar.

Quando o tamanho da amostra for maior que 40 ou 50, os métodos t podem ser usados com segurança, a não ser que os dados sejam extremamente assimétricos. De qualquer modo, faça um histograma. Se você encontrar valores atípicos nos dados e eles não forem erros fáceis de consertar, é recomendável executar a análise duas vezes, uma com e outra sem os valores atípicos, mesmo para amostras grandes. Os valores atípicos podem conter informações adicionais sobre os dados, assim, merecem uma atenção especial. Se você encontrar múltiplas modas, pode ter diferentes grupos que devem ser analisados e entendidos separadamente.

Se os dados são extremamente assimétricos (como os dados do CEO, no Capítulo 6 – veja a Figura 12.2), a média pode não ser o representante mais adequado. No entanto, em problemas de negócios, muitas vezes estamos preocupados com custos e lucros. Quando os nossos dados consistem em uma coleção de ocorrências cujo *total* é uma consequência dos negócios – como quando somamos os lucros (ou perdas) de muitas transações ou os custos de muitos materiais – então, a média é apenas o total dividido por n. Esse é um valor com uma consequência para os negócios. Felizmente, nesse exemplo, o Teorema Central do Limite vem ao nosso socorro. Mesmo quando só podemos amostrar a partir de uma distribuição muito assimétrica, a distribuição amostral da média da nossa amostra será próxima da Normal; portanto, poderemos usar os métodos do t de Student sem muita preocupação, desde que o tamanho da amostra seja *grande o suficiente*.

Quão grande é grande o suficiente? Veja um histograma dos salários do CEO ($000).

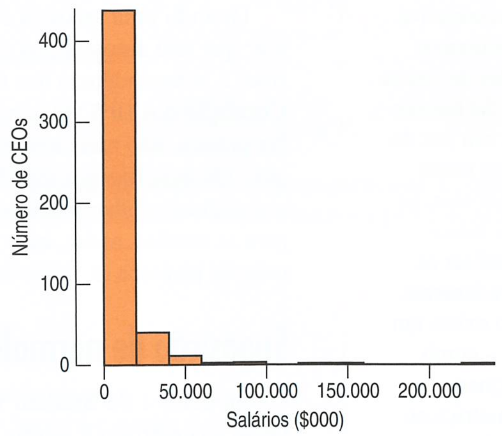


Figura 12.2 É difícil imaginar uma distribuição mais assimétrica do que esses salários anuais dos CEOs da *Fortune 500*.

Embora essa distribuição seja muito assimétrica, o Teorema Central do Limite irá tornar a distribuição amostral das médias das amostras dessa distribuição cada vez mais Normal, à medida que o tamanho da amostra aumenta. Eis um histograma e um diagrama da probabilidade Normal das médias das amostras de tamanho de 100 CEOs.

[4] Existem testes formais de Normalidade, mas eles não ajudam. Quando temos uma amostra pequena – quando realmente nos importamos com a verificação da Normalidade – esses testes têm pouco poder. Assim, não faz muito sentido usá-los para decidir se faremos um teste t.

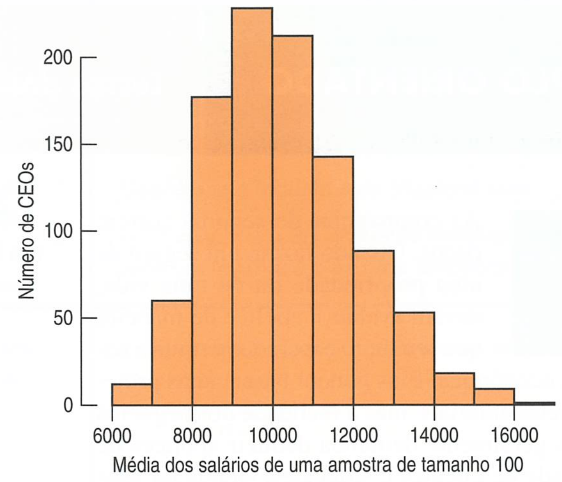
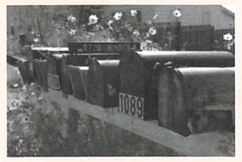


Figura 12.3 Mesmo amostras tão pequenas como 100 do conjunto de dados dos CEOs produzem médias cuja distribuição amostral é quase normal. Amostras maiores terão distribuições amostrais ainda mais próximas da Normal.

Muitas vezes, em aplicações modernas de negócios, temos amostras de muitas centenas ou milhares. Devemos, ainda, prestar atenção aos valores típicos e múltiplas modas e ter certeza de que as observações são independentes. Mas se a média for de interesse, o Teorema Central do Limite funciona muito bem, assegurando que a distribuição amostral da média estará próxima à Normal para amostras desse tamanho.

TESTE RÁPIDO



A cada 10 anos, os Estados Unidos fazem um censo que busca contar cada residente. O censo também coleta informações sobre uma variedade de questões econômicas e sociais. Todos os tipos de negócios usam os dados do censo para planejar vendas e estratégias de *marketing* e entender as variáveis demográficas subjacentes das áreas que eles servem.

Existem dois formulários para o censo: o "formulário resumido", respondido pela maioria das pessoas, e o "formulário completo", enviado aproximadamente para um em cada seis ou sete domicílios escolhidos aleatoriamente. De acordo com o Serviço de Recenseamento (factfinder.census.gov), "... cada estimativa baseada nas respostas do formulário completo tem um intervalo de confiança associado".

1. Por que o Serviço de Recenseamento precisa de um intervalo de confiança para a informação do formulário completo, mas não para as questões que aparecem em ambos os formulários, completo e resumido?

2. Por que o Serviço de Recenseamento baseia esses intervalos de confiança nos modelos t?

O Serviço de Recenseamento prossegue afirmando: "Esses intervalos de confiança são maiores ... para áreas geográficas com populações menores e para características que ocorrem com menor frequência na área que está sendo examinada (como a proporção de pessoas pobres num bairro de classe média)".

3. Por que é assim? Por exemplo, por que um intervalo de confiança para a média do gasto mensal familiar em habitação deve ser maior para uma área pouco povoada de fazendas do meio-oeste do que para uma área densamente povoada de um centro urbano? Como a fórmula de um intervalo t de uma amostra indica que isso irá acontecer?

Para lidar com esse problema, o Serviço de Recenseamento registra dados do formulário completo para "... áreas geográficas para as quais aproximadamente duzentos ou mais formulários completos foram completados – o que é grande o suficiente para produzir estimativas de boa qualidade. Se áreas ponderadas menores forem usadas, os intervalos de confiança em torno das estimativas seriam significativamente maiores, fornecendo resultados com menor utilidade".

4. Suponha que o Serviço de Recenseamento decidiu incluir nos relatos áreas das quais somente 50 formulários completos foram preenchidos. Que efeito isso teria num intervalo de confiança de 95% para, digamos, o custo médio da habitação? Especificamente, quais dos valores usados na fórmula para a margem de erro mudariam? Quais valores mudariam muito e quais valores mudariam apenas um pouco? Aproximadamente quão maior seria o intervalo de confiança com base em 50 formulários do que um com base em 200 formulários?

EXEMPLO ORIENTADO — *Lucros dos seguros*

As companhias de seguros correm riscos. Quando fazem um seguro de uma propriedade ou de uma vida, devem avaliar a apólice de maneira que seu lucro esperado permita a sobrevivência econômica. Elas podem basear suas projeções em tabelas atuariais, mas a realidade dos negócios dos seguros geralmente demanda deduzir apólices de uma variedade de clientes e situações. Gerenciar esse risco é ainda mais difícil, porque, até o vencimento da apólice, a companhia não saberá se teve lucro, independentemente do prêmio cobrado.

Uma gerente quer verificar o desempenho do seu representante de vendas, por isso selecionou 30 apólices vencidas vendidas pelo representante de vendas e calculou o lucro líquido (prêmio do seguro menos a indenização paga) para cada uma delas.

A gerente quer que você, como consultor, construa um intervalo de confiança de 95% para o lucro médio das apólices vendidas pelo representante de vendas.

Lucro (em $) das 30 apólices

222,80	463,35	2089,40
1756,23	-66,20	2692,75
1100,85	57,90	2495,70
3340,66	833,95	2172,70
1006,50	1390,70	3249,65
445,50	2447,50	-397,70
3255,60	1847,50	-397,31
3701,85	865,40	186,25
-803,35	1415,65	590,85
3865,90	2756,34	578,95

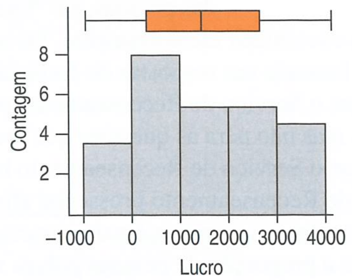
PLANEJAR

Especificação: Declare o que queremos saber. Identifique as variáveis e o seu contexto.

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

Faça uma figura. Verifique a forma da distribuição e procure assimetrias, modas múltiplas e valores atípicos.

Eis um diagrama de caixa e bigode e um histograma para esses valores



A amostra parece ser unimodal e simétrica com valores do lucro entre –$1000 e $4000 e sem valores atípicos.

Modelo: Pense sobre as suposições e verifique as condições

✓ **Suposição da independência**
Essa é uma amostra aleatória, portanto, as observações deveriam ser independentes.

✓ **Condição de aleatoriedade**
Essa amostra foi selecionada aleatoriamente das apólices vencidas vendidas pelo representante de vendas da companhia.

✓ **Condição de normalidade**
A distribuição dos lucros é unimodal e bem simétrica, sem assimetria forte.

Declare o modelo da distribuição amostral para a estatística sendo utilizada.

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

FAZER

Mecânica: Calcule as estatísticas básicas e construa o intervalo de confiança.

Usando um *software*, obtemos as seguintes estatísticas básicas:

$n = 30$

$\bar{y} = \$1438{,}90$

$s = \$1329{,}60$

Lembre que o erro padrão da média é igual ao desvio padrão dividido pela raiz quadrada de n.

O erro padrão da média é:

$$SE(\bar{y}) = \frac{s}{\sqrt{n}} = \frac{1329{,}60}{\sqrt{30}} = \$242{,}75$$

O valor crítico de que precisamos para fazer um intervalo de confiança de 95% vem da tabela t de Student, um programa de computador ou uma calculadora. Temos $30 - 1 = 29$ graus de liberdade.
O nível de confiança selecionado indica que queremos que 95% da probabilidade seja capturada no meio da distribuição, assim, excluímos 2,5% em cada cauda, para um total de 5%. Os graus de liberdade e os 2,5% da probabilidade da cauda é o que precisamos saber para encontrar o valor crítico. Aqui ele é 2,045.

Existem $30 - 1 = 29$ graus de liberdade. A gerente especificou um nível de confiança de 95%, assim, o valor crítico (da tabela T) é 2,045.

A margem de erro é:

$$\begin{aligned} ME &= 2{,}045 \times SE(\bar{y}) \\ &= 2{,}045 \times 242{,}75 \\ &= \$496{,}42 \end{aligned}$$

O intervalo de confiança de 95% para a média do lucro é:

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

$= (\$942{,}48; \$1935{,}32)$.

RELATAR

Conclusão: Interprete o intervalo de confiança no contexto apropriado.

Memorando:

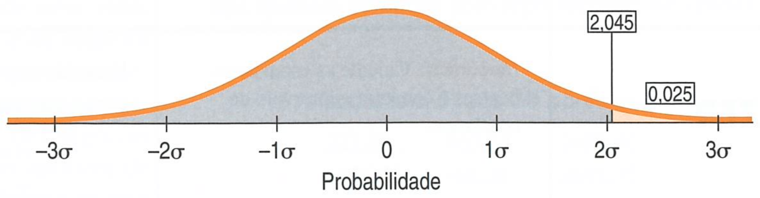
Re: Lucro das apólices

A partir da nossa análise das apólices selecionadas, estamos 95% confiantes de que a média verdadeira do lucro das apólices vendidas por esse representante de vendas está contida num intervalo de $942,48 a $1935,32.

Quando construímos intervalos de confiança dessa forma, esperamos que 95% deles cubram a média verdadeira e 5% não encontrem o valor verdadeiro. Isso é o que significa "95% de confiança".

Advertência: as perdas nos seguros são sujeitas a valores atípicos. Uma perda muito grande poderia influenciar substancialmente o lucro médio. Entretanto, não houve tais casos nesse conjunto de dados.

O valor crítico do Exemplo Orientado foi encontrado na Tabela t de Student no Apêndice C. Para encontrar o valor crítico, localize a linha da tabela correspondente aos graus de liberdade e a coluna correspondente à probabilidade que você deseja. Como um intervalo de confiança de 95% deixa 2,5% dos valores em cada lado, procuramos por 0,025 no topo da coluna ou procuramos por 95% de confiança diretamente na última linha da tabela. O valor na tabela na intersecção é o valor crítico de que precisamos. No Exemplo Orientado, o número de graus de liberdade era 30 – 1 = 29, assim, localizamos o valor de 2,045.



	0,25	0,2	0,15	0,1	0,05	0,025	0,02
24	,6848	,8569	1,059	1,318	1,711	2,064	2,172
25	,6844	,8562	1,058	1,316	1,708	2,060	2,167
26	,6840	,8557	1,058	1,315	1,706	2,056	2,162
27	,6837	,8551	1,057	1,314	1,703	2,052	2,158
28	,6834	,8546	1,056	1,313	1,701	2,048	2,154
29	**,6830**	**,8542**	**1,055**	**1,311**	**1,699**	**2,045**	**2,150**
30	,6828	,8538	1,055	1,310	1,697	2,042	2,147
31	,6825	,8534	1,054	1,309	1,696	2,040	2,144
32	,6822	,8530	1,054	1,309	1,694	2,037	2,141

Figura 12.4. Usando a Tabela t para consultar o valor crítico t^* para um nível de 95% de confiança com 29 graus de liberdade.

Então, o que devemos dizer?

Como 95% das amostras aleatórias produzem um intervalo que captura a média verdadeira, você deveria dizer: "Estou 95% confiante de que o intervalo entre \$942,48 e \$1935,32 contém a média dos lucros de todas as apólices vendidas por esse representante de vendas". Também é correto afirmar algo menos formal: "Estou 95% confiante de que a média dos lucros para todas as apólices vendidas por esse representante de vendas está entre \$942,48 e \$1935,32". Lembre-se: sua incerteza é sobre o intervalo, não sobre a média verdadeira. O intervalo varia aleatoriamente. A média verdadeira do lucro não é nem variável nem aleatória – apenas desconhecida.

12.4 Precauções na interpretação dos intervalos de confiança

Os intervalos de confiança para as médias oferecem interpretações novas e erroneamente tentadoras. Eis algumas maneiras de evitar erros:

- **Não diga:** "*95% de todas as apólices* vendidas por esse representante de vendas têm lucros entre \$942,48 e \$1935,32". O intervalo de confiança diz respeito à *média,* não a mensurações de apólices individuais.
- **Não diga:** "Estamos 95% confiantes de que *uma apólice aleatoriamente selecionada* terá um lucro líquido entre \$942,48 e \$1.935,32". Essa interpretação falsa também é sobre apólices individuais, em vez da *média* das apólices. Estamos 95% confiantes de que a *média* do lucro de todas as apólices (similares) vendidas por esse representante de vendas está entre \$942,48 e \$1935,32.
- **Não diga:** "A média do lucro é de \$ 1438,90 *95% das* vezes". Isso diz respeito a médias, mas ainda assim errado. A afirmação implica que a média verdadeira varia, quando na verdade é o intervalo de confiança que seria diferente se tivéssemos uma amostra diferente.
- Finalmente, **não diga:** "95% *de todas as amostras* terão uma média de lucro entre \$942,48 e \$1935,32". Essa afirmação sugere que *esse* intervalo de alguma forma fixa um padrão para intervalos alternados. Na verdade, esse intervalo não é mais (nem menos) provável de estar correto do que qualquer outro. Você poderia dizer que 95% de todas as amostras possíveis poderiam produzir intervalos que contenham a média verdadeira do lucro. (O problema é que, por nunca sabermos o que é a média verdadeira do lucro, não somos capazes de saber se nossa amostra estava dentro dos 95%.)

12.5 Teste *t* de uma amostra

A gerente tem uma preocupação mais específica. A política da companhia afirma que, se o lucro médio do representante de vendas está abaixo de \$1500, ele tem dado muito desconto e terá de ajustar sua estratégia de preços. Existe evidência a partir dessa amostra de que o lucro médio é realmente menor que \$1500? A pergunta pede por um teste de hipótese chamado de **teste t de uma amostra para a média**.

Você já sabe o suficiente para construir esse teste. Esse teste estatístico é parecido com os outros que já vimos. Sempre comparamos a diferença entre a estatística observada e um valor hipotético ao erro padrão. Para médias que se parecem com $\frac{\bar{y} - \mu_0}{EP(\bar{y})}$, já sabemos que o modelo de probabilidade adequado é o t de Student com n – 1 graus de liberdade.

Teste *t* de uma amostra para a média

As condições para o teste t de uma amostra para a média são as mesmas do intervalo t de uma amostra. Testamos a hipótese H_0: $\mu = \mu_0$ usando a estatística

$$t_{n-1} = \frac{\bar{y} - \mu_0}{EP(\bar{y})}$$

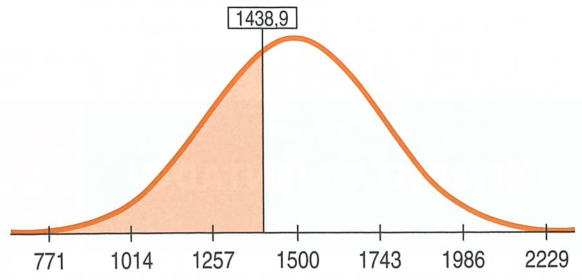
onde o erro padrão de $\bar{y}$ é: $EP(\bar{y}) = \frac{s}{\sqrt{n}}$

Quando as condições forem satisfeitas e a hipótese nula for verdadeira, essa estatística segue um modelo t de Student com n – 1 graus de liberdade. Usamos esse modelo para obter um valor-P.

EXEMPLO ORIENTADO *Lucros dos seguros revisitado*

Vamos aplicar o teste t de uma amostra para as 30 apólices vencidas amostradas pela gerente. Dessas 30 apólices, a gerência quer saber se existe evidência de que o lucro médio das apólices vendidas pelo representante de vendas é menor que \$1500.

PLANEJAR	**Especificação** Declare o que queremos saber. Deixe claro qual é a população e qual é o parâmetro. Identifique as variáveis e o seu contexto.	*Queremos testar se o lucro médio das vendas das apólices do representante é menor que \$1500. Temos uma amostra aleatória de 30 apólices vencidas para julgar essa hipótese.*
	Hipótese Damos o benefício da dúvida para o representante de vendas. A hipótese nula é que a média verdadeira do lucro é igual a \$1500. Por estarmos interessados em saber se o lucro é menor, a alternativa é unilateral.	H_0: μ = \$1500 H_A: μ < \$1500

	Faça um gráfico. Verifique se a distribuição tem assimetria, múltiplas modas ou valores atípicos	Verificamos o histograma desses dados no Exemplo Orientado anterior e vimos que ele tinha uma distribuição unimodal e simétrica.
	Modelo Verifique as condições.	Verificamos as Condições de Aleatoriedade e Normalidade no Exemplo Orientado anterior.
	Declare o modelo da distribuição amostral. Escolha o método.	As condições foram satisfeitas, então usaremos o modelo t de Student com $n - 1 = 29$ graus de liberdade e um teste t de uma amostra para a média.
FAZER	**Mecânica** Calcule as estatísticas da amostra. Certifique-se de incluir as unidades quando você escrever o que sabe sobre os dados.	Usando um *software*, obtemos as seguintes estatísticas básicas: $n = 30$ Média = \$1438,90 Desv. Padrão = \$1329,60
	O cálculo da estatística-t é apenas um valor padronizado. Subtraímos a média hipotética e dividimos o resultado pelo erro padrão. Assumimos que o modelo nulo é verdadeiro para encontrar o valor-p. Faça uma figura do modelo t, centrado em μ_0. Visto que esse é um teste unilateral à esquerda, marque a região correspondente do lucro médio observado. O valor-P é a probabilidade de observar a média da amostra tão pequena quanto \$1438,90 (ou menor) se a média verdadeira fosse \$1500, como a hipótese nula declara. Podemos encontrar esse valor-p a partir de uma tabela, usando uma calculadora ou um programa de computador.	$t = \dfrac{1438{,}90 - 1500}{1329{,}60/\sqrt{30}} = -0{,}2517$ (A média observada está menos abaixo de um erro padrão abaixo do valor suposto.)  Valor-P = $P(t_{29} < -0{,}2517) = 0{,}4015$ (ou, a partir da tabela, $0{,}1 < P < 0{,}5$)
RELATAR	**Conclusão** Relacione o valor-P à sua decisão sobre H_0 e declare sua conclusão no contexto.	**Memorando:** **Re: Desempenho das vendas** O lucro médio dos 30 contratos amostrados fechados pelo representante de vendas em questão ficou abaixo do nosso padrão de \$1500, mas não existe evidência suficiente, a partir dessa amostra de apólices, para indicar que a média verdadeira está abaixo de \$1500. Se a média fosse \$1500, esperaríamos que um tamanho da amostra de 30 tivesse uma média tão baixa quanto a observada aproximadamente em 40,15% das vezes.

Observe que, da forma como essa hipótese foi estabelecida, o lucro médio do representante de vendas deveria ser bem mais baixo que \$1500 para rejeitar a hipótese nula. Como a hipótese nula afirmava que a média era de \$1500 e a hipótese alternativa afirmava que era menos, essa forma de determinar as hipóteses garantiu o benefício da dúvida ao representante de vendas. Nada há de errado com isso, mas sempre tenha certeza de que as hipóteses estejam declaradas de modo a auxiliá-lo a tomar uma decisão de negócios correta.

Encontrando os valores *t* manualmente

O modelo t de Student é diferente para cada valor do grau de liberdade. Poderíamos ter imprimido uma tabela como a Tabela Z (do Apêndice C) para cada valor do grau de liberdade, o que gastaria muitas páginas e provavelmente não seria um *best-seller*. Uma forma de resumir o livro é nos limitarmos a intervalos de confiança de 80, 90, 95 e 99%. Portanto, os livros de estatística geralmente têm uma tabela dos valores críticos do modelo t para um conjunto selecionado de níveis de confiança. Esse livro também tem: veja a Tabela T no Apêndice C. (Você também pode buscar tabelas na Internet.)

As tabelas t apresentam tantos graus de liberdade quantos couberem na página e são mais fáceis de usar do que as Tabelas da Normal (Figura 12.5). É claro, para graus de liberdade *suficientes*, o modelo t fica cada vez mais próximo ao Normal, assim, a tabela fornece uma linha final com os valores críticos do modelo Normal e a rotula de "∞ gl."

Probabilidade bilateral / Probabilidade unilateral		0,20 / 0,10	0,10 / 0,05	0,05 / 0,025
Tabela T	gl			
Valores de t_α	1	3,078	6,314	12,706
	2	1,886	2,920	4,303
	3	1,638	2,353	3,182
	4	1,533	2,132	2,776
	5	1,476	2,015	2,571
	6	1,440	1,943	2,447
	7	1,415	1,895	2,365
	8	1,397	1,860	2,306
	9	1,383	1,833	2,262
	10	1,372	1,812	2,228
	11	1,363	1,796	2,201
	12	1,356	1,782	2,179
	13	1,350	1,771	2,160
	14	1,345	1,761	2,145
	15	1,341	1,753	2,131
	16	1,337	1,746	2,120
	17	1,333	1,740	2,110
	18	1,330	1,734	2,101
	19	1,328	1,729	2,093
	⋮	⋮	⋮	⋮
	∞	1,282	1,645	1,960
Níveis de confiança		80%	90%	95%

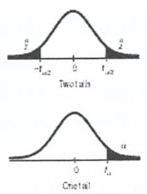


Figura 12.5 Parte da tabela T do apêndice C.

TESTE RÁPIDO

Na discussão sobre as estimativas baseadas em amostras de formulários completos, o Serviço de Recenseamento observa: "A desvantagem ... é que ... as estimativas das características também relatadas no formulário resumido não serão compatíveis com as [estimativas do formulário completo]".

As estimativas do formulário resumido são valores de um censo completo, assim, são valores "verdadeiros" – algo que geralmente não temos quando fazemos inferência.

5. Suponha que usemos os dados do formulário completo para fazer 100 intervalos de confiança de 95% para a média da idade dos residentes, uma para cada uma das 100 áreas definidas pelo censo. Quantos desses 100 intervalos devemos esperar que irão *deixar* de incluir a verdadeira média da idade (como foi determinada a partir dos dados do censo do formulário resumido)?

6. Com base na amostra do formulário completo, podemos testar a hipótese nula de que a média da renda familiar numa região era a mesma do censo anterior. Seria provável que o erro padrão para tal teste aumentasse ou diminuísse se usássemos uma área com mais respondentes de formulários completos?

> Para altos graus de liberdade, a forma dos modelos t de Student muda de modo mais gradual. A tabela T no Apêndice C inclui graus de liberdade entre 100 e 1000, assim, você pode determinar o valor-P para aproximadamente qualquer gl. Se os seus gl não estão listados, faça a abordagem da precaução usando o primeiro valor abaixo disponível ou use a tecnologia para encontrar o valor exato.

Por exemplo, suponha que obtemos o resultado de um teste t sobre valores de uma amostra com um alto grau de liberdade, com 19 gl, e queremos o valor-P da cauda superior. A partir da tabela, vemos que 1,639 está entre 1,328 e 1,729. Tudo o que podemos dizer é que o valor-P procurado está entre os valores-P desses dois valores críticos, assim, $0{,}05 < p < 0{,}10$.

Também podemos usar a tecnologia. As calculadoras ou programas estatísticos podem fornecer valores críticos para um modelo t para qualquer grau de liberdade e qualquer nível de confiança que você precise. Eles podem ir direto para os valores-P quando você testa uma hipótese. Com tabelas, podemos apenas aproximar os valores-P colocando-os entre duas das colunas. Geralmente, isso é bom o suficiente. Uma maior precisão não representará necessariamente um auxílio na tomada de uma boa decisão de negócios.

Precisaríamos ter executado um teste t de uma amostra para saber que não seria possível rejeitar a hipótese nula de que a média era de \$1500?

Afinal, vimos que o intervalo \$942,48 a \$1935,32 continha todos os valores plausíveis para o lucro médio com uma confiança de 95%. Visto que \$1500 era um desses valores plausíveis, não temos evidência para sugerir que a média não é \$1500.

Como queremos um teste unilateral, nosso nível α do intervalo de confiança de 95% seria de 0,025, correspondendo somente a um lado do intervalo de confiança. Se quisermos um nível α de 0,05, poderíamos olhar para o intervalo de confiança de 90%: (\$1022,26; \$1855,54). Visto que \$1500 também está nesse intervalo, poderíamos chegar à mesma conclusão, a de não ser possível rejeitar a hipótese de que a média é \$1500.

12.6 O tamanho da amostra

Qual é o tamanho da amostra que precisamos? A resposta simples é sempre "maior". Porém, mais dados demandam mais dinheiro, esforço e tempo. Assim, quanto é o suficiente? Suponha que o seu computador leva uma hora para baixar um filme que você quer assistir. Você está frustrado com a situação. Então, fica sabendo de um programa que afirma baixar filmes em menos de meia hora. Você está interessado o suficiente para gastar \$29,95 no programa, mas somente se ele realmente cumprir o que afirma. Assim, você consegue uma cópia grátis de avaliação e a testa, baixando um filme 10 vezes. É claro, a média do tempo de baixar o filme não é exatamente 20 minutos, como foi afirmado. As observações variam. Se a margem de erro foi de 8 minutos, entretanto, você provavelmente será capaz de decidir se o *software* valeu o dinheiro gasto. Dobrar o tamanho da amostra demandaria outras cinco ou mais horas de teste e reduziria sua margem de erro para pouco menos do que 6 minutos. Você precisaria decidir se vale o esforço.

Ao planejarmos a coleta de dados, devemos ter alguma ideia de quão pequena deve ser a margem de erro para sermos capazes de chegar a uma conclusão ou detectar uma diferença relevante. Se o tamanho do efeito que estamos estudando for grande, podemos tolerar uma ME maior. Se necessitarmos de uma grande precisão, contudo, devemos ter uma ME menor e, portanto, um tamanho maior da amostra.

Munidos com a ME e o nível de confiança, podemos encontrar o tamanho da amostra que precisaremos. Quase.

Sabemos que, para uma média, a $ME = t^*_{n-1} \times SE(\bar{y})$ e que $EP(\bar{y}) = \dfrac{s}{\sqrt{n}}$, logo podemos determinar o tamanho da amostra solucionando essa equação para n:

$$ME = t^*_{n-1} \times \frac{s}{\sqrt{n}}.$$

A boa notícia é que temos uma equação; a péssima notícia e que não sabemos a maioria dos valores que precisamos para calculá-la. Quando pensávamos sobre o tamanho da amostra para proporções, nos deparamos com um problema similar. Lá, tínhamos de estimar um valor para p a fim de calcular o tamanho da amostra. Aqui, precisamos conhecer s. Não conhecemos s até que consigamos alguns dados, mas queremos calcular o tamanho da amostra *antes* de coletar os dados. Somos capazes de estimar e

Cálculo manual do tamanho da amostra

Vamos transformar a fórmula do tamanho da amostra. Suponha que queremos uma ME de oito minutos e achamos que o desvio padrão do tempo de baixar os filmes é de aproximadamente 10 minutos. Usando um intervalo de confiança de 95% e $z^* = 1{,}96$, solucionamos para n.

$$8 = 1{,}96\,\frac{10}{\sqrt{n}}$$

$$\sqrt{n} = \frac{1{,}96 \times 10}{8} = 2{,}45$$

$$n = (2{,}45)^2 = 6{,}0025$$

É um tamanho de amostra pequeno, então usamos $(6 - 1)$ graus de liberdade para substituir um valor t^* apropriado. Com 95%, $t^*_5 = 2{,}571$. Agora podemos solucionar a equação mais uma vez:

$$8 = 2{,}571\,\frac{10}{\sqrt{n}}$$

$$\sqrt{n} = \frac{2{,}571 \times 10}{8} \approx 3{,}214$$

$$n = (3{,}214)^2 \approx 10{,}33$$

Para ter certeza de que a ME não é maior do que você quer, sempre arredonde o resultado para cima, o que resulta $n = 11$. Assim, para conseguir uma ME de oito minutos, devemos encontrar o tempo de *download* pra $n = 11$ filmes.

isso geralmente é bom o suficiente para esse objetivo. Se não tivermos ideia de qual será o desvio padrão ou se o tamanho da amostra realmente importa (por exemplo, pelo fato de que cada indivíduo adicional é muito caro para amostrar ou experimentar), uma alternativa é executar um *estudo piloto* pequeno para ter alguma percepção do tamanho do desvio padrão.

Isso não é tudo. Sem conhecer n, não sabemos os graus de liberdade e não podemos encontrar o valor crítico, t^*_{n-1}. Uma abordagem comum é usar o valor z^* correspondente do modelo Normal. Se você escolheu um nível de confiança de 95%, então use apenas 2, seguindo a Regra 68-95-99,7 ou 1,96 para ser mais preciso. Se o seu tamanho da amostra estimado é 60 ou mais, provavelmente isso está certo – z^* foi uma boa estimativa. Se for menor do que isso, você pode acrescentar uma etapa, usar z^* primeiro, encontrar n e então substituir z^* pelo t^*_{n-1} correspondente e calcular o tamanho da amostra mais uma vez.

Os cálculos do tamanho da amostra *nunca* são exatos. A margem de erro que você encontra *após* a coleta dos dados não será exatamente igual àquela que você usou para encontrar n. A fórmula do tamanho da amostra depende de valores que você não tem até que a coleta dos dados seja realizada, mas usá-la é uma primeira etapa importante. Antes de coletar os dados, é recomendável saber se o tamanho da amostra é grande o suficiente para que você seja capaz de descobrir o que quer saber.

*12.7 Graus de liberdade – por que $n - 1$?

O número dos graus de liberdade $(n - 1)$ pode lembrá-lo do valor que dividimos para encontrar o desvio padrão dos dados (visto que é, afinal, o mesmo número). Prometemos, quando introduzimos aquela fórmula, falar um pouco mais sobre o porquê da divisão por $n - 1$, em vez de n. A razão está intimamente ligada ao raciocínio sobre distribuição t.

Se soubéssemos a verdadeira média da população, μ, encontraríamos o desvio padrão da amostra usando n, em vez de $n - 1$, como:

$$s = \sqrt{\frac{\sum(y - \mu)^2}{n}} \text{ e a chamaríamos de } s.$$

Usamos $\bar{y}$, em vez de μ, o que causa um problema. Para qualquer amostra, $\bar{y}$ estará tão próximo quanto possível dos valores dos dados. Geralmente, a média da população, μ, estará mais longe. Pense sobre isso. Os escores do GMAT têm uma média populacional de 525. Se você coletar uma amostra aleatória de 5 alunos que fizeram o teste, a sua média da amostra não será 525. Os cinco valores dos dados estarão mais próximos do seu próprio $\bar{y}$ do que de 525. Assim, se usarmos $\sum(y - \bar{y})^2$ em vez de $\sum(y - \mu)^2$ na equação para calcular s, nossa estimativa do desvio padrão ficará, em geral, abaixo do verdadeiro valor, isto é, ela subestima o verdadeiro valor. O fato matemático surpreendente é que podemos compensar $\sum(y - \bar{y})^2$ por subestimar apenas dividindo por $n - 1$, em vez de n. Portanto, isso é o que o $n - 1$ está fazendo no denominador de s. Chamamos $n - 1$ de graus de liberdade.

O QUE PODE DAR ERRADO?

Primeiro, você deve decidir quando usar os métodos do t de Student.

- **Não confunda proporções e médias.** Quando você tratar seus dados como categóricos, contando sucessos e resumindo com uma proporção da amostra, faça inferências usando os métodos do modelo Normal. Quando você tratar seus dados como quantitativos, resumindo com a média da amostra, faça inferências usando os métodos do t de Student.
- **Tenha cuidado na interpretação quando os intervalos de confiança são sobrepostos.** Se os intervalos de confiança das médias de dois grupos se sobreporem, não tire conclusões precipitadas de que as médias são iguais. Duas médias podem ser significativamente diferentes e mesmo assim seus intervalos de confiança se sobreporem. Aprenderemos, no próximo capítulo, a testar diretamente a diferença entre duas médias. Se os intervalos de confiança não se sobrepõem, podemos rejeitar a hipótese nula com segurança, mas os métodos do próximo capítulo são mais poderosos.

Os métodos do t de Student funcionam somente quando a suposição da população Normal for verdadeira. Naturalmente, muitos dos erros estão relacionados a falhas na Suposição da População Normal. É recomendável procurar pelos tipos mais comuns de falhas. No final das contas, você pode até mesmo consertar algumas delas.

- **Tenha cuidado com a multimodalidade.** A condição de Normalidade evidentemente falha se os dados têm duas ou mais modas. Quando você tiver essa situação, procure pela possibilidade de que seus dados tenham vindo de dois grupos. Se for esse o caso, sua melhor aposta é separar os dados em dois grupos. (Use as variáveis para ajudar a distinguir as modas, se possível. Por exemplo, se as modas parecem compostas, na sua maioria, de homens em um grupo e mulheres em outro, separe os dados de acordo com o sexo.) Depois você pode analisar cada grupo separadamente.
- **Tenha cuidado com dados assimétricos.** Faça um histograma dos dados. Se os dados forem muito assimétricos, você pode tentar transformar a variável. A transformação pode gerar uma distribuição unimodal e simétrica, tornando-a mais apropriada para os métodos de inferência para médias. A transformação não irá ajudar se a distribuição amostral não for unimodal.
- **Investigue os valores atípicos.** A Condição de Normalidade também falha se os dados tiverem valores atípicos. Se você encontrar valores atípicos nos dados, deve investigá-los. Às vezes, é óbvio que o valor dos dados está errado, e a justificativa para removê-los ou corrigi-los também está clara. Quando não existe uma justificativa clara para remover um valor atípico, você pode executar uma análise com ou sem um valor atípico e anotar todas as diferenças nas suas conclusões. Cada vez que os valores dos dados são desprezados, você *deve* relatá-los individualmente. Geralmente, eles são a parte mais informativa do seu relatório dos dados.[5]

Por mais tentador que seja se livrar dos valores inoportunos, você não pode simplesmente jogar fora os valores atípicos e não discuti-los. Não é correto remover os valores mais altos ou mais baixos apenas para melhorar seus resultados.

[5] Essa sugestão pode ser discutível em algumas disciplinas. Desprezar valores atípicos é visto por alguns como antiético, porque o resultado provavelmente é um intervalo de confiança menor ou um valor-P menor. No entanto, uma análise dos dados com os valores atípicos deixados no lugar está *sempre* errada. Os valores atípicos violam a Condição de Normalidade e também a suposição implícita de uma população homogênea, portanto, invalidam os procedimentos de inferência. Uma análise dos pontos não atípicos, junto com uma discussão separada dos valores atípicos, é geralmente muito mais informativa e pode revelar aspectos importantes dos dados.

É claro, os problemas de Normalidade não são os únicos riscos com os quais você se depara ao fazer inferências sobre as médias.

- **Tenha cuidado com a tendenciosidade.** As mensurações de todos os tipos podem ser tendenciosas. Se as suas observações diferem da média verdadeira de forma sistemática, seu intervalo de confiança pode não capturar a média verdadeira. E não existe um tamanho da amostra que irá salvá-lo. Uma balança de banheiro que pesa dois quilos a menos sempre pesará você dois quilos a menos mesmo que você se pese 100 vezes e faça a média. Vimos várias fontes de tendenciosidade nas pesquisas, mas as mensurações também podem ser tendenciosas. Certifique-se de considerar possíveis fontes de tendenciosidade nas suas mensurações.
- **Certifique-se de que os dados sejam independentes.** Os métodos t de Student também requerem que os valores da amostra sejam mutuamente independentes. Precisamos verificar se temos uma amostragem aleatória e ainda a condição dos 10%. Você também deve verificar se existem prováveis violações da independência no método de coleta de dados. Se existirem, tenha cautela no uso destes métodos.
- **Certifique-se de que os dados representam uma amostra aleatória apropriada.** Em condições ideais, todos os dados que analisamos foram coletados de uma amostra aleatória simples ou gerados por um experimento aleatório. Quando isso não acontecer, tenha cuidado com inferências obtidas deles. Você ainda pode calcular o intervalo de confiança corretamente ou fazer o cálculo correto do valor-P, mas isso não o isenta de cometer um erro sério na inferência.

ÉTICA EM AÇÃO

Relatórios recentes indicaram que o tempo de espera nas salas de emergência (SE) nos Estados Unidos está ficando mais longo, com uma média hoje de 30 minutos (WashingtonPost.com; Janeiro de 2008). Vários motivos foram citados para esse aumento no tempo médio de espera nas SE, incluindo o fechamento das salas de emergência nas áreas urbanas e problemas com o fluxo de gerenciamento do hospital. O Hospital Tyler, localizado na área rural de Ohio, recentemente se uniu ao programa Joint Comission's Continuous Service Readiness e concordou em monitorar seu tempo de espera na SE. Após a coleta de dados de uma amostra aleatória de 30 pacientes que chegaram à sala de emergência do Hospital Tyler no mês passado, foi encontrado um tempo médio de espera de 26 minutos, com um desvio padrão de 8,25 minutos. Análises estatísticas posteriores geraram um intervalo de confiança de 95% como tendo valores de 22,92 a 29,08 minutos, indicação clara de que os pacientes da sala de emergência do Hospital Tyler esperaram menos do que 30 minutos para serem atendidos por um médico. A administração do Tyler, satisfeita com os resultados, tinha certeza de que a Joint Comission ficaria impressionada. Seu próximo passo era pensar em maneiras de incluir a mensagem "95% dos pacientes da SE do Tyler podem esperar aguardar menos que a média nacional para ser atendidos por um médico" nos seus materiais de promoção e propaganda.

PROBLEMA ÉTICO *A interpretação do intervalo de confiança está errada e é enganosa (relacionado ao Item C, ASA Ethical Guidelines). O intervalo de confiança não fornece resultados para pacientes individuais. Portanto, é incorreto afirmar que 95% dos pacientes individuais da SE aguardam menos (ou podem esperar aguardar menos) do que 30 minutos para ver um médico.*

SOLUÇÃO ÉTICA *Interpretar os resultados do intervalo de confiança corretamente, em termos da média do tempo de espera e não dos pacientes individuais.*

O que aprendemos?

Primeiro aprendemos a criar intervalos de confiança e testar hipóteses sobre proporções. Agora, voltamos nossa atenção às médias e aprendemos que a inferência estatística para as médias se vale dos mesmos conceitos; somente os algoritmos do nosso modelo mudaram.

- Aprendemos que o que podemos afirmar sobre a média da população é inferido dos dados, usando a média e o desvio padrão de uma amostra aleatória representativa.
- Aprendemos a descrever a distribuição amostral das médias da amostra usando um novo modelo que selecionamos da família do t de Student com base nos nossos graus de liberdade.
- Aprendemos que nossa régua para mensurar a variabilidade das médias da amostra é o erro padrão:

$$EP(\bar{y}) = \frac{s}{\sqrt{n}}.$$

- Aprendemos a encontrar a margem de erro para um intervalo de confiança usando a régua do erro padrão e um valor crítico baseado no modelo t de Student.
- Também aprendemos a usar a régua para testar hipóteses sobre a média populacional.
- Mais importante, aprendemos que, no raciocínio da inferência, é necessário verificar se as suposições apropriadas foram satisfeitas e se a interpretação adequada dos intervalos de confiança e valores-P permanecem os mesmos a despeito de estarmos investigando médias ou proporções.

Termos

Graus de liberdade (gl) Parâmetro da distribuição t de Student que depende do tamanho da amostra. Normalmente, mais graus de liberdade refletem no aumento da informação amostral.

Intervalo t de uma amostra para a média Um intervalo t de uma amostra para a média é:

$$\bar{y} \pm t^*_{n-1} \times SE(\bar{y}), \text{ onde } EP(\bar{y}) = \frac{s}{\sqrt{n}}$$

O valor crítico t^*_{n-1} depende de um nível de confiança particular, C, que você especifica e do número dos graus de liberdade, $n - 1$.

t de Student Família de distribuições indexada pelos seus graus de liberdade. Os modelos t são unimodais, simétricos e com forma de sino, mas geralmente têm mais áreas nas caudas e menos nos centros do que o modelo Normal. À medida que os graus de liberdade aumentam, as distribuições t se aproximam do modelo Normal.

Teste t de uma amostra para a média O teste t de uma amostra para a média testa a hipótese H_0: $\mu = \mu_0$ usando a estatística

$$t_{n-1} = \frac{\bar{y} - \mu_0}{SE(\bar{y})}, \text{ onde } EP(\bar{y}) = \frac{s}{\sqrt{n}}.$$

Habilidades

PLANEJAR

- Ser capaz de declarar as suposições necessárias para os testes t na determinação dos intervalos de confiança.
- Saber examinar seus dados para violações das condições que fariam inferências insensatas ou inválidas sobre a média da população.

- Entender que um teste de hipótese pode ser executado com um intervalo de confiança escolhido adequadamente.

FAZER

- Saber calcular e interpretar um teste t para a média populacional usando um pacote estatístico ou trabalhando a partir de um resumo estatístico da amostra.
- Saber calcular e interpretar um intervalo de confiança para a média da população com base na distribuição t utilizando um pacote estatístico ou trabalhando a partir do resumo estatístico de uma amostra.

RELATAR

- Ser capaz de explicar o significado de um intervalo de confiança para a média da população. Deixar claro que a aleatoriedade associada ao nível de confiança é uma declaração sobre os limites do intervalo e não sobre o valor do parâmetro da população.
- Entender que um intervalo de confiança de 95% não captura 95% dos valores da amostra.
- Ser capaz de interpretar o resultado de um teste de uma hipótese sobre a média da população.
- Saber que não "aceitamos" a hipótese nula se não formos capazes de rejeitá-la. Afirmamos que não foi possível rejeitá-la.
- Entender que o valor-P de um teste não fornece a probabilidade de que a hipótese nula esteja correta.

Ajuda tecnológica: Inferência de médias

Os pacotes estatísticos oferecem formas convenientes para fazer histogramas dos dados. Isso significa que você não tem desculpa para não verificar se os dados são aproximadamente Normais.

Qualquer pacote estatístico padrão pode executar um teste de hipóteses. Veja um exemplo de uma saída genérica de um pacote (embora nenhum pacote forneça os resultados exatamente dessa forma).

Hipótese nula

Hipótese alternativa

```
Teste H₀: μ(velocidade) = 30 vs Hₐ: μ(velocidade) > 30
Média da amostra = 31,04378
t = 1,178 com 22 gl
valor-P = 0,1257
```

A estatística-t (e os graus de liberdade)

O valor-P geralmente é dado por último

O pacote calcula a média e o desvio padrão da amostra da variável e encontra o valor-P na distribuição t, com base no número apropriado de graus de liberdade. Todos os pacotes estatísticos modernos apresentam os valores-P. O pacote pode também fornecer informações adicionais, como a média da amostra, o desvio padrão da amostra, o valor da estatística-t e os graus de liberdade. Eles são úteis para interpretar o valor-P resultante e indicar a diferença entre um resultado realmente significativo de um que é apenas estatisticamente significativo.

Os pacotes estatísticos que calculam o desvio padrão estimado da distribuição amostral geralmente o rotulam como "erro padrão" ou "EP".

Os resultados de inferência também são, às vezes, apresentados em tabelas. Você deverá procurar cuidadosamente os valores que precisa. Em geral, os resultados do teste e os limites do intervalo de confiança correspondente são fornecidos juntos. E, muitas vezes, você deve ler cuidadosamente para encontrar a hipótese alternativa. Veja a seguir um exemplo desse tipo de saída.

```
Valor hipotético      30
Média estimada        31,043478
GL                    22
Erro padrão            0,886
Alfa          0,05

                 Teste t                  Intervalo t
Estatística      1,178
Prob > |t|       0,2513         Superior 95%   32,880348
Prob > t         0,1257         Inferior 95%   29,206608
Prob < t         0,8743
```

μ_0

Média calculada $\bar{y}$

O nível alfa geralmente é 0,05. Alguns pacotes permitem escolher o valor α.

t-statistic

Unilateral $H_A: \mu > 30$

Bilateral (perceba o |t|)

Valores-P para alternativa

Intervalo de confiança correspondente

Unilateral $H_A: \mu < 30$

Os comandos para fazer inferência para médias em programas estatísticos comuns e calculadoras nem sempre são óbvios (já a saída resultante é geralmente identificada com clareza e fácil de ler.) Os guias de cada programa podem ajudá-lo a começar a navegar.

Excel

Especifique as fórmulas. Encontre t^* com a função INVT (alfa, gl). O suplemento DDXL oferece os testes t e os intervalos de confiança.

Comentários

O Excel não tem funções prontas para encontrar os valores-P.

Minitab

Do menu **Stat**, escolha o submenu **Basic Statistics** e depois **1-sample t** (*t* de uma amostra) Preencha a caixa de diálogo.

Comentários

A caixa de diálogo oferece uma escolha clara entre um intervalo de confiança e um teste.

SPSS

Do menu **Analyze**, escolha o submenu **Compare means** (comparar médias). O SPSS não realiza testes de hipóteses para proporções. A partir daí, escolha o comando **One-Sample t-test** (teste *t* de uma amostra).

Comentários

Os comandos não sugerem um teste nem um intervalo. Porém, os resultados fornecem tanto um quanto o outro.

JMP

Do menu **Analyze**, selecione **Distribuition**. Para um intervalo de confiança, vá até a seção "Moments" para encontrar os limites do intervalo (certifique-se de que suas variáveis sejam do tipo contínuas para que essa seção esteja disponível). Para um teste de hipóteses, clique no triângulo vermelho ao lado do nome da variável e escolha **Test Mean** (Teste para a média) do menu. Preencha a caixa de diálogo resultante.

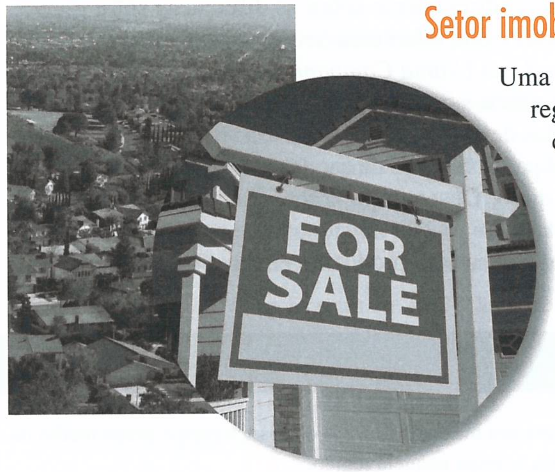
Comentário

"Momento" é um termo estatístico extravagante para médias, desvios padrão e outras estatísticas relacionadas.

DATA DESK

Selecione as variáveis.
Do menu **Calc**, escolha **Estimate** para um intervalo de confiança ou **Test** para um teste de hipóteses. Selecione o intervalo ou teste a partir do menu suspenso e faça outras escolhas na caixa de diálogo resultante.

Projetos de estudo de pequenos casos



Setor imobiliário

Uma corretora de imóveis busca entender o preço das casas na sua área, uma região formada por cidades pequenas e médias. Para cada uma das 1200 casas recentemente vendidas na região, o arquivo **ch12_MCSP_Real_Estate** tem as seguintes variáveis:

- *Preço de venda* (em $)
- *Tamanho do lote* (em acres)
- *De frente para água* (Sim, Não)
- *Idade* (em anos)
- *Ar central* (Sim, Não)
- *Tipo de combustível* (Madeira, Óleo, Gás, Elétrico, Propano, Solar, Outro)
- *Condição* (1 a 5, 1 = Ruim, 5 = Excelente)
- *Área habitável* (em pés quadrados)
- *Faculdade* (% de habitantes no código da área que frequentam uma faculdade de quatro anos)
- *Banheiros* (número de banheiros completos)
- *Lavabos* (número de lavabos)
- *Quartos* (número de quartos)
- *Lareiras* (número de lareiras)

A corretora tem uma família interessada numa casa com quatro quartos. Usando intervalos de confiança, como ela poderia informar a família sobre o preço médio de uma casa de quatro quartos nessa área? Compare esse valor com o intervalo de confiança para casas com dois quartos. Como a presença de ar condicionado central afeta o preço médio das casas nessa área? Utilize intervalos de confiança e gráficos para ajudar a responder essa questão.

Explore outras questões que possam ser úteis para a corretora de imóveis, sabendo como fatores categóricos diferentes afetam o preço de venda, e escreva um pequeno relatório sobre a sua conclusão.

Perfil de doadores

Uma organização filantrópica coleta e compra dados para sua base de doadores. O banco de dados completo contém aproximadamente 4,5 milhões de doadores e mais de 400 variáveis coletadas para cada um, mas o conjunto de dados **ch_12_MCSP_Donor_Profiles** é uma amostra de 916 doadores e inclui as seguintes variáveis:

- *Idade* (em anos)
- *Proprietário da casa* (H = Sim, U = Desconhecido)
- *Gênero* (F = Feminino, M = Masculino, U = Desconhecido)

- *Riqueza* (Categorias ordenadas do total de domicílios ricos de 1 = menos ricos a 9 = mais ricos)
- *Filhos* (Número de filhos)
- *Última doação* (0 = Não fez doação na última campanha, 1 = Doou na última campanha)
- *Quantia doada na última campanha* ($ quantia da contribuição da última campanha)

Os analistas da organização querem saber quanto, em média, as pessoas doam às campanhas e quais os fatores que podem influenciar essa quantia. Compare os intervalos de confiança para o valor médio da *Quantia Doada na Última Campanha* daqueles que possuem casas próprias com aqueles que o *status* de proprietário é desconhecido. Execute comparações similares para *Gênero* e para duas categorias de *Riqueza*. Escreva um breve relatório usando gráficos e intervalos de confiança para relatar o que você encontrou. (Não faça inferências diretamente sobre as diferenças entre grupos. Discutiremos isso no próximo capítulo. Sua inferência deve ser sobre grupos únicos.)

(A distribuição da *Quantia Doada na Última Campanha* tem uma grande assimetria à direita, por isso a mediana poderia ser considerada o resumo apropriado. Mas como a mediana é $0,00, os analistas devem usar a média. Para simulações, eles verificaram que a distribuição amostral para a média é unimodal e simétrica para amostras maiores do que 250 ou em torno disso. Observe que diferenças pequenas na média podem resultar em milhões de dólares de receita adicionada de todo o país. O custo médio da solicitação é de $0,67 por pessoa para produzir e enviar a correspondência.)

EXERCÍCIOS

1. Modelos *t*. Usando as tabelas *t*, um *software* ou uma calculadora, determine:

a) o valor crítico de *t* para um intervalo de confiança de 90% com um gl = 17.

b) o valor crítico de *t* para um intervalo de confiança de 98% com um gl = 88.

c) o valor-P para $t \geq 2{,}09$ com 4 graus de liberdade.

d) o valor-P para $|t| > 1{,}78$ com 22 graus de liberdade.

2. Modelos *t*, parte 2. Usando as tabelas *t*, um *software* ou uma calculadora, estime:

a) o valor crítico de *t* para um intervalo de confiança de 95% com um gl = 7.

b) o valor crítico de *t* para um intervalo de confiança de 99% com um gl = 102.

c) o valor-P para $t \leq 2{,}19$ com 41 graus de liberdade.

d) o valor-P para $|t| > 2{,}33$ com 12 graus de liberdade.

3. Intervalos de confiança. Descreva como a amplitude de um intervalo de confiança de 95% para uma média se altera à medida que o desvio padrão, *s*, aumenta, assumindo que o tamanho da amostra permaneça o mesmo.

4. Intervalos de confiança, parte 2. Descreva como a amplitude de um intervalo de confiança de 95% para uma média se altera à medida que o tamanho da amostra, *n*, aumenta, assumindo que o desvio padrão permanece o mesmo.

5. Intervalos de confiança e tamanho da amostra. Um intervalo de confiança para o preço da gasolina de uma amostra aleatória de 30 postos numa região fornece as seguintes estatísticas:

$$\bar{y} = \$4{,}49 \quad s = \$0{,}29$$

a) Encontre um intervalo de 95% confiança para o preço médio da gasolina comum naquela região.

b) Encontre um intervalo de 90% confiança para a média.

c) Se tivéssemos as mesmas estatísticas para uma amostra de 60 postos de gasolina, qual seria, agora, o intervalo de 95%?

6. Intervalos de confiança e tamanho da amostra, parte 2. Um intervalo de confiança para o preço da gasolina de uma amostra aleatória de 30 postos numa região fornece as seguintes estatísticas:

$$\bar{y} = \$4{,}49 \quad EP(\bar{y}) = \$0{,}06$$

a) Encontre um intervalo de confiança de 95% para o preço médio da gasolina comum naquela região.

b) Encontre um intervalo de confiança 90% para a média.

c) Se tivéssemos as mesmas estatísticas para uma amostra de 60 postos de gasolina, qual seria, agora, o intervalo de 95%?

7. *Marketing* de ração para animais de fazenda. Uma empresa de ração desenvolveu um suplemento alimentar com o objetivo de promover o ganho de peso em animais criados em fazenda. Seus analistas relatam que as 77 vacas estudadas ganharam uma média de 56 libras e que um intervalo de confiança de 95% para o aumento de peso esse suplemento produz tem uma margem de erro de ± 11 libras. A equipe do departamento de *marketing* escreveu as seguintes conclusões. Algum dos analistas interpretou o intervalo de confiança corretamente? Explique as interpretações equivocadas.

a) 95% das vacas estudadas ganharam entre 45 e 67 libras.

b) Estamos 95% certos de que uma vaca alimentada com esse suplemento irá ganhar entre 45 e 67 libras.

c) Estamos 95% certos de que o aumento de peso médio entre as vacas desse estudo estava entre 45 e 67 libras.

d) O aumento de peso médio das vacas alimentadas com esse suplemento está entre 45 e 67 libras 95% do tempo.

e) Se esse suplemento for testado numa outra amostra de vacas, existe uma chance de 95% de que o aumento médio de peso estará entre 45 e 67 libras.

8. Custos da refeição. Uma empresa está interessada em estimar os custos do almoço em sua cafeteria. Após fazer um levantamento de dados com seus empregados, a equipe calculou que um intervalo de confiança de 95% para a quantia média gasta com almoços num período de seis meses é de ($780; $920). Agora, a organização está tentando escrever seu relatório e considerando as seguintes interpretações. Comente cada uma delas.

a) 95% de todos os empregados pagam entre $780 e $920 pelo almoço.

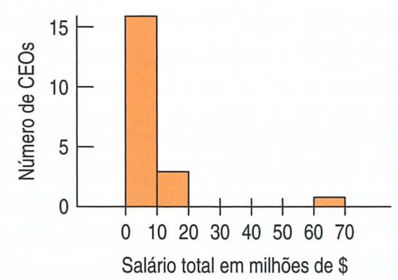
b) 95% de todos os empregados amostrados pagam entre $780 e $920 pelo almoço.

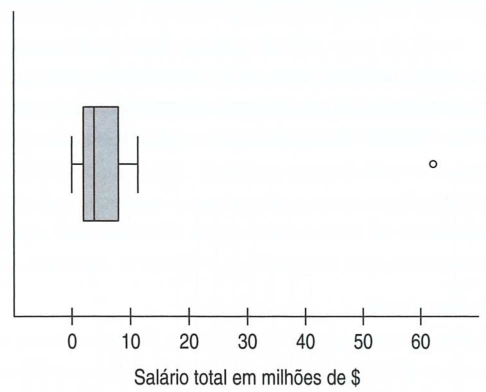
c) Estamos 95% certos de que os empregados dessa amostra pagam em média entre $780 e $920 pelo almoço.

d) 95% de todos os empregados amostrados terão um custo médio para o almoço entre $780 e $920.

e) Estamos 95% certos de que a quantia média que todos os empregados pagam pelo almoço está entre $780 e $920.

9. Salários do CEO. Uma amostra de 20 CEOs da *Forbes 500* mostra que os salários anuais totais abrangem um mínimo de $0,1 a $62,24 milhões. A média para esses CEOs é de $7946,00 milhões. O histograma e o diagrama de caixa e bigodes são:

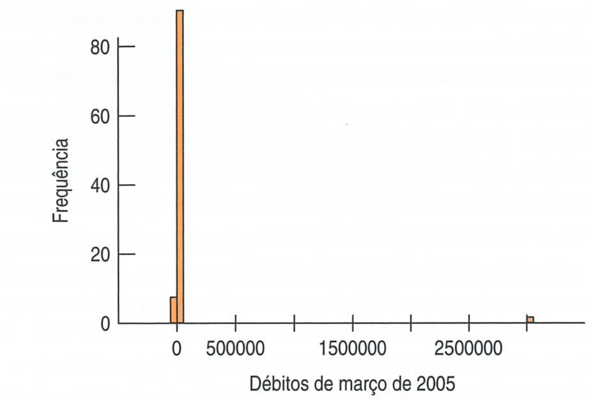


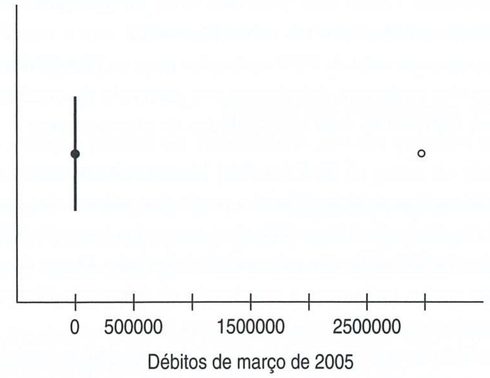


Com base nesses dados, um programa de computador calculou que um intervalo de confiança de 95% para o salário médio anual de todos os CEOs da *Forbes 500* é (1,69; 14,20) em milhões de $. Por que você deve estar hesitante em confiar nesse intervalo de confiança?

10. Débitos no cartão de crédito. Uma empresa de cartão de crédito toma uma amostra aleatória de 100 proprietários de cartão de crédito para ver quanto eles gastaram no cartão no mês passado.

Um histograma e um diagrama de caixa e bigodes são:





Um programa de computador calculou que o intervalo de 95% confiança para a quantia média gasta em março de 2005 é (-$28366,84; $90691,49). Explique por que os analistas não acharam o intervalo de confiança útil e mostre o que deu errado.

11. Área de estacionamento. Esperando atrair mais compradores ao centro, uma cidade constrói uma nova garagem pública no distrito comercial central. A cidade planeja pagar pela estrutura com a cobrança do estacionamento. Para uma amostra aleatória de 44 dias de semana, a renda média diária do estacionamento foi de $126, com um desvio padrão de $15.

a) Que suposição você deve fazer para usar essas estatísticas para inferência?

b) Encontre um intervalo de confiança de 90% para a renda média diária que esse estacionamento irá gerar.

c) Explique, no contexto, o que esse intervalo de confiança significa.

d) Explique o que 90% de confiança quer dizer nesse contexto.

e) O consultor que aconselhou a cidade nesse projeto previu que o estacionamento geraria uma renda média diária de $128. Com base no seu intervalo de confiança, o que você acha da previsão do consultor? Por quê?

12. Moradia. O ano de 2008 foi difícil para a economia. Houve um grande número de execução de hipotecas de casas de família. Numa grande comunidade, os corretores de imóveis amostraram aleatoriamente 36 lances de compradores potenciais para determinar a perda média no valor das casas. A amostra revelou que a perda média foi de $11560 com um desvio padrão de $1500.

a) Que suposições e condições devem ser verificadas antes de encontrar o intervalo de confiança? Como você as verificaria?
b) Encontre um intervalo de 95% confiança para a perda média do valor das casas.
c) Interprete esse intervalo e explique o que 95% de confiança significa.
d) Suponha que a perda média, nacional, do valor das casas nessa época foi de $10000. Você acha que a perda na comunidade amostrada difere significativamente da média nacional? Explique.

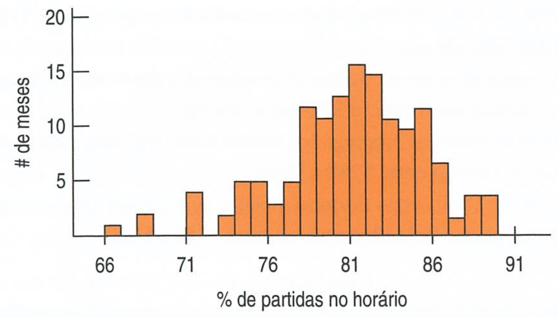
13. Área de estacionamento, parte 2. Suponha que com a finalidade de planejar o orçamento, a cidade do Exercício 11 precisa de uma estimativa melhor da renda média diária gerada pela cobrança do estacionamento.
a) Alguém sugere que a cidade use seus dados para criar um intervalo de 95% confiança em vez do intervalo de 90% criado inicialmente. Como esse intervalo seria melhor para a cidade? (Você não precisa criar um novo intervalo.)
b) Como um intervalo de 95% seria pior para os planejadores?
c) Como eles poderiam determinar um intervalo de confiança que melhor servisse às suas necessidades de planejamento?

14. Moradia, parte 2. No Exercício 12, encontramos um intervalo de 95% de confiança para estimar a perda dos valores das casas.
a) Suponha que o desvio padrão para as perdas fosse de $3000 em vez dos $1500 utilizado para aquele intervalo. O que um desvio padrão maior faria com a amplitude do intervalo de confiança (supondo o mesmo nível de confiança)?
b) Seu colega de aula sugere que a margem de erro do intervalo poderia ser reduzida se o nível de confiança fosse mudado para 90% em vez de 95%. Você concorda com essa afirmação? Por quê?
c) Em vez de mudar o nível de confiança, seria estatisticamente mais apropriado coletar uma amostra maior?

15. Orçamento estadual. Os estados que contam com os impostos sobre as vendas como receita para financiar a educação, segurança pública e outros programas, geralmente acabam com superávit no orçamento durante os períodos de crescimento (quando as pessoas gastam mais em bens de consumo) e déficit no orçamento durante a recessão (quando as pessoas gastam menos em bens de consumo). Cinquenta e um pequenos varejistas de um estado com uma economia em crescimento foram recentemente amostrados. A amostra mostrou um aumento médio de $2350 na receita dos impostos sobre as vendas adicionais coletadas pelo varejista em comparação ao semestre anterior. O desvio padrão da amostra é $425.
a) Encontre um intervalo de 95% de confiança para o aumento médio na receita dos impostos sobre as vendas.
b) Que suposições você fez sobre essa inferência? Você acha que as condições apropriadas foram satisfeitas?
c) Explique o que o seu intervalo significa e forneça um exemplo do que ele não significa.

16. Orçamento estadual, parte 2. Suponha que o estado do Exercício 15 amostrou 16 pequenos varejistas, em vez de 51, e para a amostra de 16 o aumento amostral médio novamente se igualou a $2350 na arrecadação fiscal das vendas adicionais por varejista comparado ao semestre anterior. Também assuma que o desvio padrão da amostra é $425.
a) Qual é o erro padrão do aumento médio na arrecadação fiscal das vendas coletadas?
b) O que acontece com a precisão da estimativa quando o intervalo é construído usando um tamanho de amostra menor?
c) Encontre e interprete um intervalo de 95% de confiança.
d) Como a margem de erro para o intervalo construído no Exercício 15 se compara à margem de erro obtida neste exercício? Explique estatisticamente como o tamanho da amostra muda a precisão do intervalo construído. Que amostra você preferiria se você fosse um planejador do orçamento do estado? Por quê?

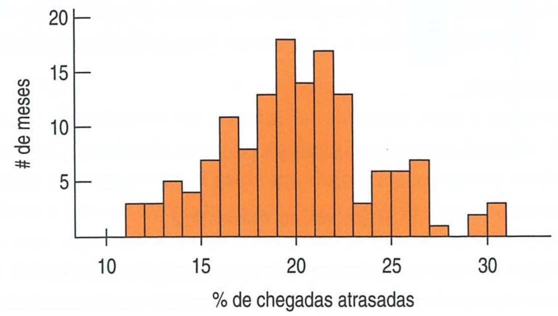
17. Partidas. Quais são as chances de seu voo partir na hora? A Agência de Estatística do Transporte do Departamento de Transportes dos Estados Unidos publica informações sobre o desempenho das companhias aéreas. Veja um histograma e um resumo estatístico para o percentual de voos que partiram no horário a cada mês, de 1995 até 2006.



n	144
$\bar{y}$	81,1838
s	4,47094

Não existe evidência de uma tendência sobre o horário (a correlação do percentual de saídas no horário com o tempo é r =-0,016).
a) Verifique as suposições e condições para a inferência.
b) Encontre um intervalo de 90% de confiança para o verdadeiro percentual de voos que partem no horário.
c) Interprete esse intervalo para um viajante que está planejando voar.

18. Chegadas atrasadas. O seu voo vai deixá-lo no seu destino na hora? A Agência de Estatística do Transporte dos Estados Unidos relatou o percentual de voos atrasados mensalmente no período de 1995 a 2006. Veja um histograma junto com um resumo estatístico.



n	144
$\bar{y}$	20,0757
s	4,08837

Podemos considerar esses dados uma amostra representativa de todos os meses. Não existe evidência de uma tendência no tempo (r = –0,07).
a) Verifique as suposições e condições para inferência sobre a média.
b) Encontre um intervalo de 99% de confiança para o verdadeiro percentual de voos que chegam atrasados.
c) Interprete esse intervalo para um viajante que está planejando voar.

19. Comércio na Internet. Uma pesquisadora de mercado de uma grande empresa do setor de vestuário que depende de vendas por catálogo decidiu investigar se o montante das vendas *on-line* mudou. Ela compara a média mensal de vendas *on-line* dos últimos meses com um número histórico para a média mensal de vendas para compras *on-line*. Ela encontra um valor-P de 0,01. Explique, nesse contexto, o que 1% significa.

20. Desempenhos padrão. A Associação de Golfe dos Estados Unidos (USGA) determina padrões de desempenho para as bolas de golfe. Por exemplo, a velocidade inicial da bola não deve exceder a 250 pés por segundo quando mensurada por um aparelho aprovado pela USGA. Suponha que um fabricante introduza um novo tipo de bola e forneça uma amostra aleatoriamente selecionada de bolas para o teste. Com base na média da velocidade da amostra, a USGA encontra um valor-P de 0,34. Explique, nesse contexto, o que o percentual de 34% representa.

21. Pagamentos do seguro social. A média mensal dos benefícios do seguro social para viúvos e viúvas em 2005 foi de $967 (*Statistical Abstract of the United States* – Agência do Censo dos Estados Unidos). Os pagamentos variam de região para região. No Texas, a média mensal do benefício se igualou a $940. Um advogado da assistência social de uma área rural do Texas acredita que a média mensal de 2005 do benefício do seguro social para viúvos e viúvas difere significativamente da média geral do Texas. Para testar essa suposição, o advogado amostrou aleatoriamente 100 pagamentos mensais dos benefícios para viúvas/viúvos e encontrou a média da amostra = $915 com um desvio padrão de $90.
a) Determine e interprete uma estimativa de um intervalo de 95% de confiança para a média amostral de 2005 do benefício do seguro social para viúvos/viúvas da área rural do Texas.
b) Em um teste de hipóteses executado para determinar se a média da área rural era diferente, o teste foi rejeitado com um valor-P = 0,007 (usando os mesmos resultados da amostra acima e um nível de significância = 0,05). Explique como o intervalo de confiança construído na parte **a** é consistente com os resultados do teste de hipótese. Sua discussão deve incluir o nível de confiança, os limites do intervalo, o valor-P e a decisão do teste de hipóteses.

22. Pagamentos do seguro social, parte 2. Num município vizinho, um jornal escreveu que os benefícios do seguro social para viúvos/viúvas ficaram significativamente mais baixos que nos demais municípios do Texas. O jornal relatou que a média do benefício de 2005 para viúvos e viúvas era de $900, com base numa amostra aleatória de 100 e um desvio padrão de $90.
a) Determine e interprete uma estimativa de um intervalo de 95% de confiança para a média amostral dos benefícios do seguro social dos viúvo/viúvas nesse município.
b) A média do benefício do seguro social desse município é diferente da média de $940 do Texas?

23. Segurança da TV. O fabricante de uma estante de metal para televisores precisa ter certeza de que seu produto não irá falhar sob o peso de um aparelho de TV normal. Visto que alguns aparelhos grandes pesam aproximadamente 300 libras, os inspetores de segurança da empresa estabeleceram um padrão para assegurar que as estantes suportem 500 libras em média. Seus inspetores regularmente sujeitam uma amostra aleatória de estantes a um aumento de peso até que elas falhem. Eles testam a hipótese H_0: μ = 500 contra H_A: $\mu > 500$, usando um nível de significância $\alpha = 0{,}01$. Se a amostra das estantes falha nesse teste de segurança, os inspetores não irão certificar o produto para venda ao público em geral.
a) Esse é um teste de cauda superior ou inferior? Nesse contexto do problema, por que você acha que isso é importante?
b) Explique o que irá acontecer se os inspetores cometerem um erro do Tipo I.
c) Explique o que irá acontecer se os inspetores cometerem um erro do Tipo II.

24. Controle de qualidade. Durante um angiograma, os problemas do coração podem ser examinados por um pequeno tubo (cateter) introduzido no coração por uma veia da perna do paciente. É importante que a empresa que manufatura o cateter mantenha um diâmetro de 2,00 mm (o desvio padrão é muito pequeno). Cada dia, a equipe de controle de qualidade faz várias mensurações para testar H_0: μ = 2,00 contra H_A: $\mu \neq 2{,}00$ com um nível de significância de α = 0,05. Se eles descobrirem um problema, irão parar o processo de manufatura até que ele seja corrigido.
a) Esse é um teste unilateral ou bilateral? Nesse contexto, por que você acha que isso é importante?
b) Explique, nesse contexto, o que acontece se a equipe de controle de qualidade comete um erro do Tipo I.
c) Explique, nesse contexto, o que acontece se a equipe de controle de qualidade comete um erro do Tipo II.

25. Segurança da TV, revisitada. O fabricante das estantes de metal para televisores do Exercício 23 quer revisar seu teste de segurança.
a) Se os advogados da empresa estão preocupados em ser processados por vender produtos sem segurança, eles deveriam aumentar ou diminuir o valor de α? Explique.
b) Nesse contexto, o que o poder do teste significa?
c) Se a empresa quer aumentar o poder do teste, quais opções ela tem? Explique as vantagens e desvantagens de cada opção.

26. Controle de qualidade, parte 2. A empresa do cateter do Exercício 24 está revisando seu procedimento de teste.
a) Suponha que o nível de significância é alterado para α = 0,01. A probabilidade do erro do Tipo II irá aumentar, diminuir ou permanecerá a mesma?
b) O que é pretendido pelo poder do teste que a empresa conduz?
c) Suponha que o processo de manufatura está falhando nos ajustes apropriados. À medida que a média real do diâmetro dos cateteres produzidos está cada vez mais acima dos 2,00 mm desejados, o poder do teste de qualidade irá aumentar, diminuir ou permanecer o mesmo?
d) O que a empresa pode fazer para melhorar o poder do teste?

27. Comércio na Internet, parte 2. A média de idade dos consumidores *on-line* alguns anos atrás era de 23,3 anos. À medida que indivíduos mais velhos adquirem confiança na Internet, acredita-se que a média da idade aumente. Queremos testar essa crença.

a) Escreva a hipótese apropriada.
b) Planejamos testar a hipótese nula selecionando uma amostra aleatória de 40 indivíduos que tenham feito uma compra *on-line* durante 2007. Você acha que as suposições necessárias para inferência foram satisfeitas? Explique.
c) Os compradores *on-line* na nossa amostra têm uma idade média de 24,2 anos, com um desvio padrão de 5,3 anos. Qual é o valor-P para esse resultado?
d) Explique (no contexto) o que esse valor-P significa.
e) Qual é a sua conclusão?

28. Economia de combustível. Uma empresa com uma frota grande de carros quer manter o custo de gasolina baixo e estipula o objetivo de alcançar um consumo médio para a frota de pelo menos 26 milhas por galão. Para analisar se o objetivo está sendo alcançado, eles verificam o consumo da gasolina de 50 viagens escolhidas ao acaso, encontrando uma média de 25,02 mpg e um desvio padrão de 4,83 mpg. Isso é uma forte evidência de que eles falharam no objetivo de economizar combustível?

a) Escreva a hipótese apropriada.
b) As condições necessárias para executar a inferência foram satisfeitas?
c) Teste a hipótese e encontre o valor-P.
d) Explique o que o valor-P significa nesse contexto.
e) Formule uma conclusão apropriada.

29. Preço da competitividade. A SLIX está desenvolvendo uma cera de fluorocarbono de alto desempenho para competições de esqui *cross country*, projetada para ser usada sob uma enorme variedade de condições. Para justificar o preço que o *marketing* deseja, a cera deve ser muito rápida. Especificamente, o tempo médio para terminar o seu teste padrão de percurso deve ser menor que 55 segundos para o ex-campeão olímpico que agora é seu consultor. Para testá-la, o consultor irá esquiar o percurso 8 vezes.

a) Os tempos do campeão são 56,3; 65,9; 50,5; 52,4; 46,5; 57,8; 52,2 e 43,2 segundos para completar percurso de teste. Eles devem comercializar a cera? Explique.
b) Suponha que a empresa decida não comercializar a cera após o teste, mas que de fato a cera diminui o tempo médio do campeão para menos de 55 segundos. Que tipo de erro foi cometido? Explique o impacto de tal erro para a empresa.

30. Pipoca. A Pop's Popcorn Inc. precisa determinar o poder ótimo e o ajuste de tempo para sua nova pipoca de micro-ondas com sabor de alcaçuz. Ela quer encontrar uma combinação de poder e tempo que apresente uma pipoca de alta qualidade com menos de 10% dos grãos sem estourar, em média – um valor que sua pesquisa de mercado indica ser uma exigência dos seus consumidores. O seu departamento de pesquisa faz experiências com vários ajustes e determina que o poder de 9 com 4 minutos é ótimo. Os seus testes confirmam que esse ajuste satisfaz a condição de menos 10%. A empresa muda as instruções da caixa e promove uma nova garantia de devolução do dinheiro para menos de 10% de grãos não estourados.

a) Se, de fato, o ajuste resulta em mais de 10% de grãos não estourados, que tipo de erro foi cometido? Quais serão as consequências para a empresa?
b) Para reduzir o risco de cometer um erro, o presidente (o próprio Pop) pede para serem testadas mais 8 caixas de pipocas (selecionadas ao acaso) no ajuste especificado. Eles encontram o seguinte percentual de grãos não estourados: 7; 13,2; 10; 6; 7,8; 2,8; 2,2; 5,2. Isso fornece evidência de que o ajuste satisfaz o objetivo de menos de 10% de grãos não estourados? Explique.

31. Declarações falsas? Um fabricante declara que um novo projeto para um telefone sem fio aumentou o alcance para 150 pés, permitindo a vários consumidores usar o telefone por toda a casa e pátio. Um laboratório de testes independente encontra que uma amostra aleatória de 44 desses telefones funcionou acima de uma distância média de 142 pés, com um desvio padrão de 12 pés. Existe evidência de que a declaração do fabricante é falsa?

32. Declarações falsas, parte 2. Os fabricantes do Abolator, um aparelho portátil para exercícios vendido por $149,95, afirmam que usar sua máquina meros 6 minutos por dia resultará em uma perda média de peso de 8 libras durante a primeira semana. Uma organização de consumidores recrutou 30 voluntários para usar o produto de acordo com as recomendações do fabricante e encontrou uma perda média de peso de 4,7 libras com um desvio padrão de 6,1 libras. Existe evidência de que as declarações dos fabricantes do Abolator sejam falsas?

33. Chips Ahoy. Em 1998, a companhia Nabisco anunciou o "Desafio dos 1000 Chips", afirmando que cada pacote de 18 onças dos biscoitos Chips Ahoy continha pelo menos 1000 pedacinhos de chocolate. Os dedicados estudantes de estatística da Air Force Academy (Academia da Força Aérea) (não é brincadeira) compraram alguns pacotes de biscoitos aleatoriamente selecionados e contaram os pedacinhos de chocolate. Alguns dos seus dados são apresentados abaixo. (*Chance*, 12, n.1, 1999)

1219 1214 1087 1200 1419 1121 1325 1345
1244 1258 1356 1132 1191 1270 1295 1135

a) Verifique as suposições e condições para inferência. Comente qualquer preocupação que você tenha.
b) Crie um intervalo de 95% de confiança para o número médio de pedacinhos de chocolate nos pacotes dos biscoitos Chip Ahoy.
c) O que essa evidência indica sobre a afirmação da Nabisco? Use o seu intervalo de confiança para testar uma hipótese apropriada e declarar suas conclusões.

34. Relatórios dos consumidores. O *Consumer Reports* testou 14 marcas de iogurte de baunilha e encontrou os seguintes números de calorias por porção: 160, 200, 220, 230, 120, 180, 140, 130, 170, 190, 80, 120, 100 e 170.

a) Verifique as suposições e condições para inferência.
b) Crie um intervalo de confiança de 95% para a caloria média contida num iogurte de baunilha.
c) Um guia de dieta afirma que você terá 120 calorias numa porção de iogurte de baunilha. O que essa evidência indica? Use seu intervalo de confiança para testar uma hipótese apropriada e apresente sua conclusão.

35. Investimento. O estilo do investimento tem uma função na construção de fundos mútuos. Muitas ações individuais podem ser classificadas em dois grupos distintos: Crescimento e Valor. Uma ação de Crescimento tem alto potencial de rendimento e geralmente paga pouco ou não paga dividendos aos acionistas. Já as ações de Valor são comumente vistas como estáveis ou mais conservadoras, com um potencial de rendimento mais baixo. Uma família está tentando decidir em que tipo de fundos investir. Um consultor independente afirma que os Fundos Mútuos de Valor forneceram um retorno anual maior que 8% nos últimos cinco anos. Logo abaixo está um resumo estatístico para o período de 5 anos de uma amostra aleatória de fundos de Valor.

Variável	N	Média	Média EP	Desvio padrão
Retorno em 5 anos	35	8,418	0.493	2,916

Mínimo	Q1	Mediana	Q3	Máximo
2,190	6,040	7,980	10,840	14,320

Teste a hipótese de que o retorno médio de 5 anos para fundos de valor é maior que 8%, assumindo um nível de significância de 5%. O que a evidência indica sobre a afirmação do gerente de portfólio de que o retorno anual do período de 5 anos era maior que 8%? Formule sua conclusão.

36. Produção. Um fabricante de pneus estuda a adoção de um novo padrão de banda de rodagem de pneus para qualquer condição de tempo. Os testes indicaram que esses pneus irão fornecer melhor milhagem por gasolina e vida longa da banda de rodagem. O último teste é para a eficácia dos freios. A empresa espera que o pneu permita que um carro andando a 60 mph pare completamente numa média de 125 pés após os freios serem usados. Eles adotarão a nova banda de rodagem, a não ser que tenham forte evidência de que os pneus não satisfaçam o objetivo. As distâncias (em pés) para 10 paradas numa pista de teste foram 129, 128, 130, 132,135, 123, 102, 125, 128, e 130. A empresa deve adotar o novo padrão da banda de rodagem? Teste uma hipótese apropriada e declare sua conclusão. Explique como você lida com o valor atípico e por que você fez a sua recomendação.

37. Cobranças. As companhias de cartão de crédito perdem dinheiro com os proprietários que não fazem seus pagamentos mínimos. Eles usam uma variedade de métodos para incentivar os devedores a pagar o saldo dos seus cartões como cartas, telefonemas e, finalmente, contratando uma agência de cobrança. Para justificar o custo da utilização de uma agência de cobrança, a agência deve cobrar uma média de pelo menos $200 por cliente. Após um período experimental, durante o qual a agência tentou cobrar de uma amostra aleatória de 100 inadimplentes, o intervalo de confiança de 90% da quantia média resgatada foi relatado como estando no intervalo ($190,25; $250,75). A partir disso, qual(is) recomendação(ões) você faria para a empresa de cartão de crédito sobre o uso da agência de cobrança?

38. Brindes. Uma organização filantrópica envia "brindes" às pessoas da sua lista de endereços, na esperança de que elas respondam enviando uma doação. Os brindes mais comuns incluem etiquetas de endereçamentos postais, cartões comemorativos ou postais. A organização quer testar um novo brinde que custa $0,50 por item para produção e envio. Eles o enviam a uma amostra "pequena" de 2000 clientes e encontram um intervalo de 90% de confiança das doações médias de ($0,489; $0,879). Como um consultor, que recomendação(ões) você faria à organização sobre o uso desse brinde?

39. Cobrança, parte 2. O proprietário da agência de cobrança do Exercício 37 tem certeza de que pode cobrar mais que $200 por cliente, em média. Ele insiste que a empresa de cartão de crédito execute um teste maior. Você acha que um teste maior pode ajudar a empresa a tomar uma decisão melhor? Explique.

40. Brindes, parte 2. A organização filantrópica do Exercício 38 decidiu seguir adiante com o novo brinde. No envio 98000 prospectos, a nova correspondência gerou uma média de $0,78. Se eles tivessem decido, com base na sua experiência inicial, *não* usar esse brinde, que tipo de erro eles teriam cometido? Quais aspectos da sua experiência inicial sugeriram (na sua posição de consultor) que um teste maior seria proveitoso?

41. Pilhas. Uma empresa de pilhas declara que suas pilhas duram uma média de 100 horas, caso sujeitas ao uso normal. Existem várias reclamações de que as pilhas não duram tanto, assim, uma agência de testes independente decidiu testá-las. Das 16 pilhas testadas, o tempo médio de vida útil foi de 97 horas, com um desvio padrão de 12 horas.

a) Quais são as hipóteses nula e alternativa?
b) Um defensor do consumidor (que não conhece estatística) afirma que aquelas 97 horas são muito menos que as 100 horas anunciadas, por isso devemos rejeitar a afirmação da empresa. Explique para ele o problema de fazer isso.
c) Que suposições devemos fazer para proceder com a inferência?
d) A um nível de 5% de significância, o que você conclui?
e) Suponha que, na verdade, a vida média útil das pilhas da empresa seja somente de 98 horas. Foi cometido um erro na parte **d**? Se foi, de que tipo?

42. Criação de animais de estimação. Animais de estimação de luxo são um grande negócio, assim, um jovem empresário (com 12 anos de idade) decidiu criar *golden hamsters*, um tipo muito conhecido das lojas de animais de estimação e colecionadores (por incrível que pareça, quase todos os *golden hamsters* em cativeiro são descendentes de uma ninhada encontrada na Síria em 1930). Das 47 ninhadas recentes, houve uma média de 7,27 filhotes de *hamsters,* com um desvio padrão de 2,5 *hamsters* por ninhada.

a) Encontre e interprete um intervalo de 90% de confiança para o tamanho médio da ninhada.
b) Quão menor ou maior seria a margem de erro para um intervalo de confiança de 99%? Explique.
c) Com base nessas estatísticas, quantas ninhadas precisaríamos para estimar o tamanho médio da ninhada com erro de um filhote com 95% de confiança?

43. Produção de peixes. O salmão cultivado é muito mais barato de produzir do que o salmão pescado em seu estado natural, mas os clientes estão preocupados com vários problemas recentemente divulgados sobre o cultivo do salmão, incluindo o tipo de alimento que eles recebem e as contaminações encontradas na sua carne (acesse www.healthcastle.com). Entre as contaminações, estão os componentes como policloretos de bifenila (PCBs), dioxinas, toxafeno, dieldrina, hexaclorobenzeno, lindano, heptacloro, cis-nonacloro, trans-nonacloro, gama-clorodano, alfa-clorodano, Mirex, endrina e DDT.

O EPA recomenda que o salmão contenha no máximo 0,08 ppm do inseticida Mirex. Um grupo de ambientalistas local está considerando um boicote ao salmão se ele exceder a 0,08 ppm. Um intervalo de 95% de confiança, obtido de uma amostra aleatória de salmão cultivado em 150 fazendas diferentes (*Science*, 9, Janeiro 2004) é de (0,0834 a 0,0992) ppm. Os dados eram unimodais e simétricos, sem valores atípicos.

a) Existe evidência de que as fazendas estão produzindo o salmão com uma contaminação média de Mirex maior que a quantia recomendada pelo EPA? Sua explicação deve discutir o nível de confiança, o valor-P e a decisão.

b) Discuta os dois tipos de erros que podem ser cometidos, num contexto de uma decisão de negócios, na proibição da venda de salmão cultivado.

44. Velocidade da transferência de arquivos. Uma estudante recentemente comprou um programa antivírus para ajudar a aumentar o desempenho do seu computador. Antes de instalar o programa, sua velocidade média de transferência de arquivos era de 480 kbps (kilobits por segundo). O programa custou $29,95 e ela quer saber se a velocidade de transferência aumentou.

Ela fez um teste de velocidade de *download* (www.bandwidthplace.com) 16 vezes e encontrou um intervalo de 90% de confiança de (482,6; 505,9) kbps. Ela fez uma tabela com as 16 velocidades e encontrou dados unimodais e aproximadamente simétricos, sem valores atípicos óbvios.

a) Existe evidência para sugerir que a velocidade da transferência dos arquivos do computador seja maior que 480 kbps? Sua explicação deve discutir o nível de confiança, o valor-P e a decisão.

b) Se ela descobrir, mais tarde, que a média é, na verdade, 465 kbps, que tipo de erro ela cometeu?

c) Você acha que a decisão de comprar o *software* foi correta com base nos dados?

45. Taxa para os laboratórios. O comitê de tecnologia declarou que o tempo médio gasto pelos estudantes no laboratório aumentou e o crescimento representa a necessidade de inflação nas taxas para o laboratório. Para comprovar essa afirmação, o comitê amostra aleatoriamente 12 visitas dos estudantes ao laboratório e anota o tempo gasto usando o computador. Os tempos em minutos são:

Tempo	Tempo
52	74
57	53
54	136
76	73
62	8
52	62

a) Faça um gráfico dos dados. Algumas das observações são valores atípicos? Explique.

b) O período médio de tempo anterior gasto usando o laboratório de computadores era de 55 minutos. Teste a hipótese de que a média é maior agora do que os 55 minutos anteriores utilizando $\alpha = 0{,}05$. Qual é a sua conclusão?

c) Se existirem valores atípicos, elimine-os e teste novamente a hipótese. Qual é a sua conclusão?

d) Discuta as implicações estatísticas de eliminar os valores atípicos. Por que alguns pesquisadores podem discordar da remoção dos valores atípicos do conjunto dos dados?

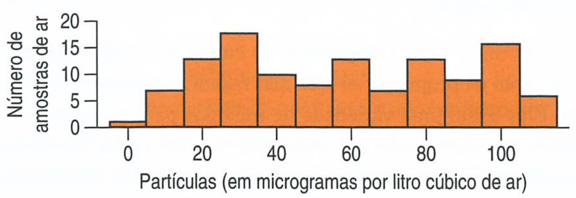
46. Baterias dos telefones celulares. Uma empresa que produz telefones celulares afirma que sua bateria padrão dura, em média, mais que as outras baterias no mercado. Para apoiar essa afirmação, a empresa publica um anúncio relatando os resultados de um experimento recente mostrando que, sob uso normal, suas baterias duram pelo menos 35 horas. Para investigar essa afirmação, um grupo de defesa do consumidor pediu à empresa os dados brutos. A empresa envia ao grupo os seguintes resultados:

35, 34, 32, 31, 34, 34, 32, 33, 35, 55, 32, 31

a) Teste uma hipótese apropriada e declare sua conclusão.

b) Explique como você lidou com o valor atípico e por que você fez a recomendação.

47. Crescimento e poluição do ar. Os funcionários do governo têm dificuldade em atrair novos negócios para comunidades com reputações duvidosas. Nevada tem sido um dos estados com o crescimento mais rápido do país nos últimos anos. Acompanhando esse crescimento rápido estão projetos enormes de novas construções. Visto que Nevada tem um clima seco, a construção cria uma visível poluição do ar. Níveis altos de poluição podem descrever um quadro pouco atraente para a área e podem também resultar em multas cobradas pelo governo federal. Como foi requerido pelos regulamentos do governo, os pesquisadores continuamente monitoram os níveis de poluição. Em testes mais recentes, 121 amostras do ar foram coletadas. Os níveis das partículas de pó devem ser relatados para as agências reguladoras federais. No relatório enviado à agência federal, foi observado que o nível médio das partículas = 57,6 microgramas/litro cúbico de ar e que o intervalo de confiança estimado foi de (52,06 mg a 63,07 mg). Um gráfico da distribuição das quantidades das substâncias particuladas também foi incluído e é apresentado abaixo.

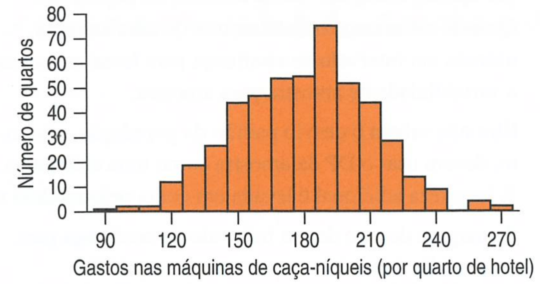


a) Discuta as suposições e condições para usar os métodos de inferência do t de Student com esses dados.

b) Você acha que o intervalo de confiança observado no relatório é válido? Explique brevemente.

48. Rendimento das convenções. Antigamente, Nevada era o único estado norte-americano em que o jogo era permitido. Embora o jogo continue a ser uma das maiores indústrias em Nevada, a proliferação da legalização do jogo em outras áreas do país exigiu que o estado e os governos locais analisassem outras possibilidades de crescimento. As autoridades do turismo em muitas cidades de Nevada ativamente recrutam convenções nacionais que trazem milhares de visitantes ao estado. Vários dados demográficos e econômicos

são coletados a partir de levantamentos de dados feitos com os participantes da convenção. Uma estatística de interesse é a quantia que os visitantes gastam nas máquinas de caça-níqueis. O estado de Nevada geralmente divulga o gasto com os caça-níqueis como a quantia gasta por quarto de hotel. Um levantamento de dados recente de 500 visitantes perguntou quanto eles gastavam com o jogo. O gasto médio por quarto foi de $180.



Os cassinos irão usar a informação relatada do levantamento de dados para estimar o gasto em caça-níqueis por quarto de hotel. Você acha que as estimativas produzidas pelo levantamento de dados irão representar os gastos de modo preciso? Explique usando as estatísticas relatadas e mostradas no gráfico.

49. Velocidade do trânsito. Os departamentos de polícia geralmente tentam controlar a velocidade do trânsito colocando sinalizadores de velocidade nas estradas, os quais avisam aos motoristas quão rápido eles estão dirigindo. Especialistas em segurança no trânsito determinam onde os sinalizadores devem ser colocados. Num teste recente, a polícia relatou a velocidade média registrada pelos carros andando numa rua movimentada próxima a uma escola de Ensino Fundamental. Para uma amostra de 25 velocidades, foi determinado que o valor médio acima do limite para os 25 valores registrados foi de 11,6 mph, com um desvio padrão de 8 mph. O intervalo de confiança de 95% estimado para essa amostra é de 8,30 mph a 14,90 mph.

a) Qual é a margem de erro para esse problema?

b) Os pesquisadores comentaram que o intervalo era muito abrangente. Explique especificamente o que deve ser feito para reduzir a margem de erro para menos que ± 2 mph.

50. Velocidade do trânsito II. Os sinalizadores de velocidade devem mensurar precisamente para maximizar a eficácia no controle da velocidade do trânsito. A precisão dos sinalizadores de velocidade será testada antes de colocá-los nas ruas da cidade. Para assegurar que as taxas de erro são estimadas com eficácia, os pesquisadores querem coletar uma amostra grande o suficiente a fim de garantir intervalos das estimativas precisos de quanto os sinalizadores de velocidade estão próximo da mensuração real das velocidades. Especificamente, os pesquisadores querem que uma margem de erro para uma única medida de velocidade seja no máximo $\pm 1{,}5$ mph.

a) Discuta como os pesquisadores podem obter uma estimativa razoável do desvio padrão do erro das velocidades mensuradas.

b) Suponha que o desvio padrão para o erro nas velocidades mensuradas seja igual a 4 mph. Com 95% de confiança, qual tamanho da amostra deveria ser coletado para assegurar que a margem de erro não seja maior que $\pm$ 1,0 mph?

51. Auditoria dos impostos I. Contadores públicos certificados geralmente são solicitados a comparecer com seus clientes se a receita federal fizer a auditoria do retorno dos impostos do seu cliente. Algumas firmas de contabilidade fornecem ao cliente a opção de pagar uma taxa quando o imposto de renda estiver pronto, o que garante a assessoria fiscal e o apoio do contador se o cliente sofrer auditoria. A taxa é cobrada como um seguro e é menor que a quantia que seria exigida se o cliente precisasse da assistência da firma em uma auditoria posteriormente. Uma grande firma de contabilidade quer determinar que taxa cobrar para os retornos do próximo ano. Nos anos anteriores, o custo médio real da assistência numa sessão de auditoria era de $650. Para determinar se esse custo mudou, a firma amostra aleatoriamente 32 taxas de clientes de auditoria. O custo médio da amostra da auditoria era de $680, com um desvio padrão de $75.

a) Estime, por meio de intervalo de 95% de confiança, o custo médio da auditoria.

b) Execute o teste apropriado para determinar se o custo médio da auditoria é diferente agora do que da média histórica de $650. Use um nível de significância de 0,05.

c) Comente como a estimativa do intervalo de confiança sustenta os resultados do teste de hipóteses.

52. Auditoria dos impostos II. Durante a revisão da amostra das taxas de auditoria, um contador sênior da firma observa que as taxas cobradas pelos contadores dependem da complexidade do retorno. Uma comparação das taxas atuais, portanto, talvez não forneça a informação necessária para ajustar as taxas para o próximo ano. Para entender melhor a estrutura da taxa, o contador sênior solicita uma nova amostra que mensure o tempo que o contador gasta na auditoria. No ano passado, a média de horas cobradas por cliente auditado foi de 3,25 horas. Uma nova amostra de 10 tempos auditados mostra os seguintes tempos em horas:

4,2, 3,7, 4,8, 2,9, 3,1, 4,5, 4,2, 4,1, 5,0, 3,4

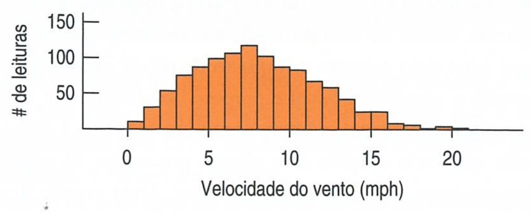
a) Assuma que as condições necessárias para inferência foram satisfeitas. Encontre uma estimativa por um intervalo de 90% de confiança para o tempo médio de auditoria.

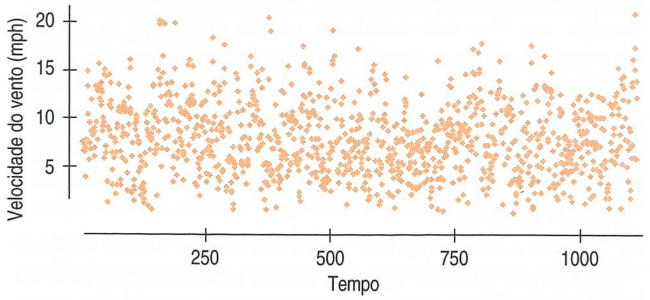
b) Execute o teste apropriado para determinar se o tempo médio de auditoria para as auditorias deste ano é significativamente diferente das 3,25 horas do ano passado. Use $\alpha = 0{,}10$.

c) Comente como o intervalo de confiança estimado sustenta os resultados do teste de hipóteses.

53. Poder do vento. Você deveria gerar eletricidade com sua própria turbina eólica? Depende da quantidade de vento na sua área. Para produzir energia suficiente, sua área deve ter uma velocidade média anual de vento de, pelo menos, 8 milhas por hora, de acordo com a Associação de Energia Eólica. Uma área foi monitorada por um ano, com as velocidades do vento registradas a cada 6 horas. De um total de 1114 leituras da velocidade do vento, apurou-se uma média de 3,813 mph. Você foi solicitado a fazer um relatório estatístico para ajudar o proprietário a decidir se coloca uma turbina eólica na sua propriedade.

a) Discuta as suposições e condições para o uso dos métodos de inferência com o t de Student para esses dados. Eis alguns diagramas que podem ajudá-lo a decidir se os métodos podem ser usados.

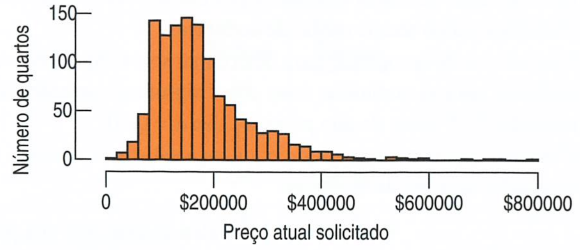




b) O que você diria ao proprietário sobre se essa área é adequada para a instalação de uma pequena turbina eólica? Explique.

53. Queda do setor imobiliário? Após a crise do *sub-prime* no final de 2007, os preços do setor imobiliário caíram em quase todos os Estados Unidos. Antes da crise, em 2006-2007, o preço médio de venda de casas numa região do estado de Nova York era de \$191300. Uma agência imobiliária quer saber quanto os preços caíram desde então. Ela coleta uma amostra de 1231 casas na região e descobre que o preço médio é \$178613,50, com um desvio padrão de \$92701,56. Você foi contratado pela agência imobiliária para fazer um relatório da situação atual.

a) Discuta as suposições e condições para o uso dos métodos t para inferência com esses dados. Aqui estão alguns diagramas que podem ajudá-lo a decidir o que fazer.



b) O que você relataria para a agência imobiliária sobre a situação atual?

RESPOSTAS DO TESTE RÁPIDO

1 As perguntas do formulário resumido são respondidas por cada um da população. Isso representa um censo, assim, as médias ou proporções *são* os valores verdadeiros da população. Os formulários completos são apenas entregues a uma amostra da população. Quando estimamos os parâmetros de uma amostra, usamos um intervalo de confiança para levar em conta a variabilidade de amostra para amostra.

2 Eles não sabem o desvio padrão da população, portanto, devem usar o DP da amostra como uma estimativa. A incerteza adicional é levada em conta pelo modelo t.

3 A margem de erro de um intervalo de confiança para uma média depende, em parte, do erro padrão:

$$EP(\bar{y}) = \frac{s}{\sqrt{n}}$$

Como n está no denominador, tamanhos de amostras menores geralmente levam a EPs maiores e, correspondentemente, a intervalos mais amplos. Visto que os formulários completos são amostrados na mesma taxa de um em cada seis ou sete domicílios por todo o país, as amostras serão menores em áreas menos populosas e resultam em intervalos de confiança mais amplos.

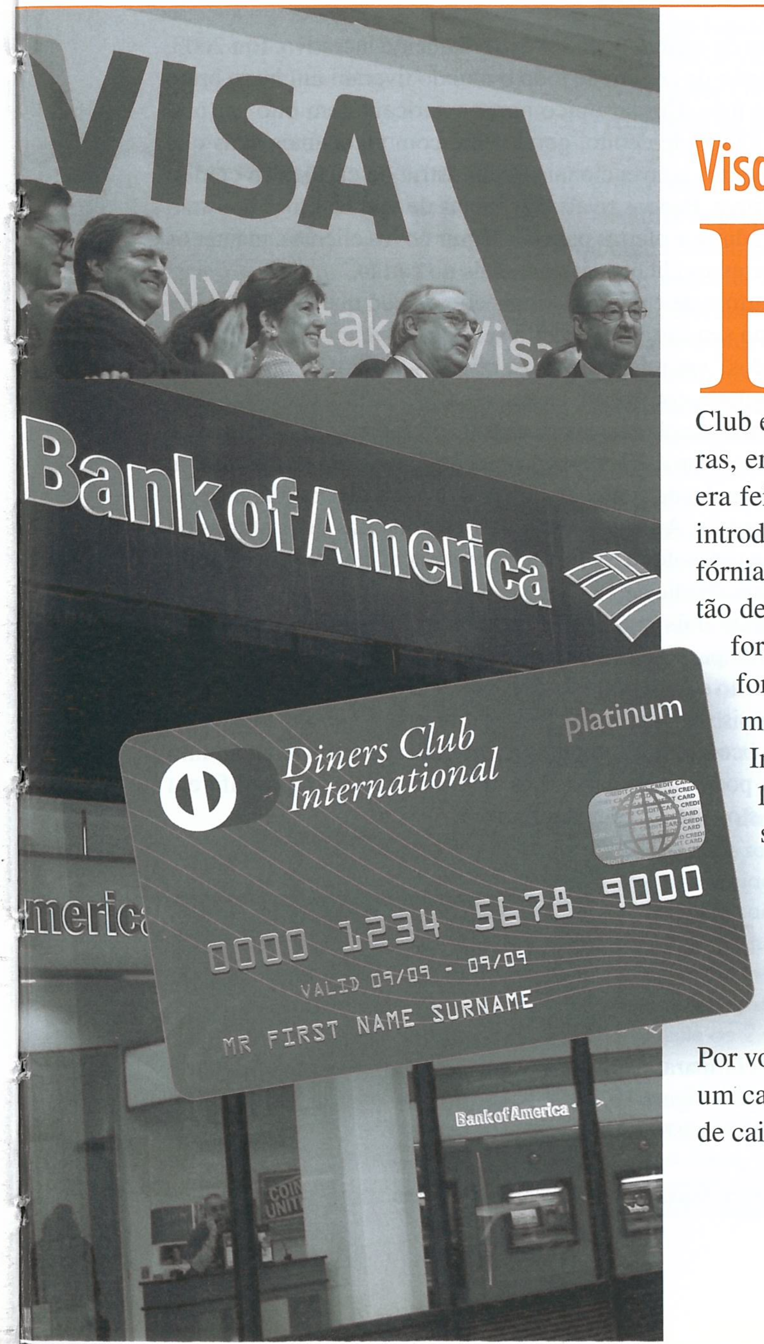
4 Os valores críticos para t com poucos graus de liberdade seriam levemente maiores. A parte $\sqrt{n}$ do erro padrão muda bastante, tornando o EP muito maior. Ambos aumentariam a margem de erro. A menor amostra é um quarto da maior, por isso o intervalo de confiança terá aproximadamente o dobro do tamanho.

5 Esperamos que 95% desses intervalos cubram o valor verdadeiro, então 5 dos 100 intervalos podem ser perdidos.

6 Seria provável que o erro padrão diminuísse se tivéssemos um tamanho da amostra maior.

Capítulo 13

Comparando Duas Médias



Visa Global Organization

Hoje, mais de um bilhão de pessoas e 24 milhões de comerciantes usam o cartão Visa em 170 países no mundo inteiro. No entanto, por volta de 1950, quando a ideia de transações sem dinheiro começava a se consolidar, somente o Diners Club e alguns varejistas, particularmente companhias petrolíferas, emitiam cartões de crédito. A grande maioria das compras era feita com dinheiro e cheques pessoais. O Bank of America introduziu seu programa do Bank Americard em Fresno, Califórnia, em 1958, e a American Express emitiu seu primeiro cartão de plástico em 1959. O cartão de crédito realmente ganhou força uma década mais tarde, quando um grupo de bancos formou um empreendimento conjunto para criar um sistema centralizado de pagamento. O National BankAmericard, Inc. (NBI) tomou posse do sistema de cartões de crédito em 1970 e, por questões de simplicidade e *marketing*, mudou seu nome para Visa em 1976 (o nome Visa é pronunciado quase da mesma forma em todas as línguas). Naquele ano, a Visa processou 679000 transações – volume processado em média a cada quatro minutos hoje. Com os avanços tecnológicos, a indústria do cartão de crédito mudou. Por volta de 1986, era possível usar um cartão Visa para sacar dinheiro de caixas automáticos.