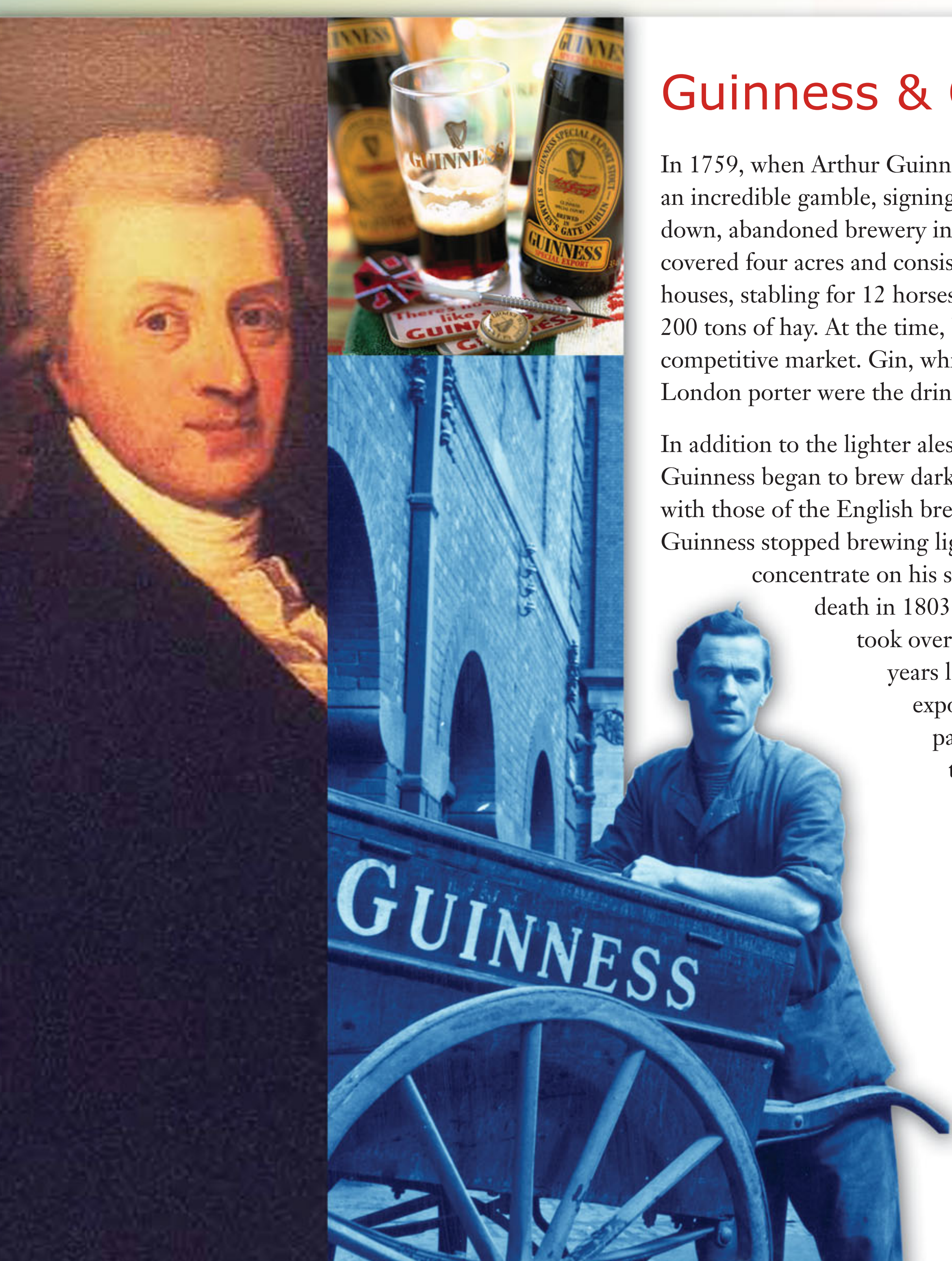
Confidence Intervals for Means

CHAPTER

12



Guinness & Co.

In 1759, when Arthur Guinness was 34 years old, he took an incredible gamble, signing a 9000-year lease on a run-down, abandoned brewery in Dublin. The brewery covered four acres and consisted of a mill, two malt houses, stabling for 12 horses, and a loft that could hold 200 tons of hay. At the time, brewing was a difficult and competitive market. Gin, whiskey, and the traditional London porter were the drinks of choice.

In addition to the lighter ales that Dublin was known for, Guinness began to brew dark porters to compete directly with those of the English brewers. Forty years later, Guinness stopped brewing light Dublin ales altogether to concentrate on his stouts and porters. Upon his death in 1803, his son Arthur Guinness II took over the business, and a few years later the company began to export Guinness stout to other parts of Europe. By the 1830s, the Guinness St. James's Gate Brewery had become the largest in Ireland. In 1886, the Guinness Brewery, with an annual production of 1.2 million barrels, was the first major brewery to be incorporated as a public company on the London Stock Exchange. During the 1890s, the company began to employ scientists. One of those, William S. Gosset, was hired as a chemist to test the quality of the brewing process. Gosset was not only an early pioneer of quality control methods in industry but a statistician whose work made modern statistical inference possible.[1]

As a chemist at the Guinness Brewery in Dublin, William S. Gosset was in charge of quality control. His job was to make sure that the stout (a thick, dark beer) leaving the brewery was of high enough quality to meet the standards of the brewery's many discerning customers. It's easy to imagine, when testing stout, why testing a large amount of stout might be undesirable, not to mention dangerous to one's health. So to test for quality Gosset often used a sample of only 3 or 4 observations per batch. But he noticed that with samples of this size, his tests for quality weren't quite right. He knew this because when the batches that he rejected were sent back to the laboratory for more extensive testing, too often the test results turned out to be wrong. As a practicing statistician, Gosset knew he had to be wrong *some* of the time, but he hated being wrong more often than the theory predicted. One result of Gosset's frustrations was the development of a test to handle small samples, the main subject of this chapter.

12.1 The Sampling Distribution for the Mean

You've learned how to create confidence intervals for proportions. Now we want to do the same thing for means. For proportions we found the confidence interval as

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

The ME was equal to a critical value, z^*, times $SE(\hat{p})$. Our confidence interval for means will look very similar:

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

And our ME will be a critical value times $SE(\bar{y})$. So let's put the pieces together. What the Central Limit Theorem told us back in Chapter 10 is exactly what we need.

The Central Limit Theorem

When a random sample is drawn from *any* population with mean μ and standard deviation σ, its sample mean, $\bar{y}$, has a sampling distribution whose *shape* is approximately Normal as long as the sample size is large enough. The larger the sample used, the more closely the Normal approximates the sampling distribution for the mean. The mean of the sampling distribution is μ, and its standard deviation is $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$.

[1]Source: Guinness & Co., www.guinness.com/global/story/history.

This gives us a sampling distribution and a standard deviation for the mean. All we need is a random sample of quantitative data and the true value of the population standard deviation σ.

But wait. That could be a problem. To compute $\sigma/\sqrt{n}$ we need to know σ. How are we supposed to know σ? Suppose we told you that for 25 young executives the mean value of their stock portfolios is $125,672. Would that tell you the value of σ? No, the standard deviation depends on how similarly the executives invest, not on how well they invested (the mean tells us that). But we need σ because it's the numerator of the standard deviation of the sample mean: $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$. So what can we do? The obvious answer is to use the sample standard deviation, s, from the data instead of σ. The result is the standard error: $SE(\bar{y}) = \frac{s}{\sqrt{n}}$.

A century ago, people just plugged the standard error into the Normal model, assuming it would work. And for large sample sizes it *did* work pretty well. But they began to notice problems with smaller samples. The extra variation in the standard error was wreaking havoc with the margins of error.

Gosset was the first to investigate this phenomenon. He realized that not only do we need to allow for the extra variation with larger margins of error, but we also need a new sampling distribution model. In fact, we need a whole *family* of models, depending on the sample size, n. These models are unimodal, symmetric, and bell-shaped, but the smaller our sample, the more we must stretch out the tails. Gosset's work transformed Statistics, but most people who use his work don't even know his name.

To find the sampling distribution of $\frac{\bar{y}}{s/\sqrt{n}}$, Gosset simulated it *by hand*. He drew paper slips of small samples from a hat *hundreds of times* and computed the means and standard deviations with a mechanically cranked calculator. Today you could repeat in seconds on a computer the experiment that took him over a year. Gosset's work was so meticulous that not only did he get the shape of the new histogram approximately right, but he even figured out the exact *formula* for it from his sample. The formula was not confirmed mathematically until years later by Sir Ronald Aylmer Fisher.

Gosset's *t*

Gosset made decisions about the stout's quality by using statistical inference. He knew that if he used a 95% confidence interval, he would fail to capture the true quality of the batch about 5% of the time. However, the lab told him that he was in fact rejecting about 15% of the good batches. Gosset knew something was wrong, and it bugged him.

Gosset took time off from his job to study the problem and earn a graduate degree in the emerging field of Statistics. He figured out that when he used the standard error $\frac{s}{\sqrt{n}}$, the shape of the sampling model was no longer Normal. He even figured out what the new model was and called it a t-distribution.

The Guinness Company didn't give Gosset a lot of support for his work. In fact, it had a policy against publishing results. Gosset had to convince the company that he was not publishing an industrial secret and (as part of getting permission to publish) had to use a pseudonym. The pseudonym he chose was "Student," and ever since, the model he found has been known as **Student's *t***.

Gosset's model is always bell-shaped, but the details change with the sample sizes (Figure 12.1). So the Student's t-models form a family of related distributions that depend on a parameter known as **degrees of freedom**. We often denote degrees of freedom as df and the model as t_{df}, with the numerical value of the degrees of freedom as a subscript.

Student's t-models are unimodal, symmetric, and bell-shaped, just like the Normal model. But t-models with only a few degrees of freedom have a narrower peak than the Normal model and have much fatter tails. (That's what makes the margin of error bigger.) As the degrees of freedom increase, the t-models look more and more like the Normal model. In fact, the t-model with infinite degrees of freedom is exactly Normal.[2] This is great news if you happen to have an infinite

[2]Formally, in the limit as the number of degrees of freedom goes to infinity.

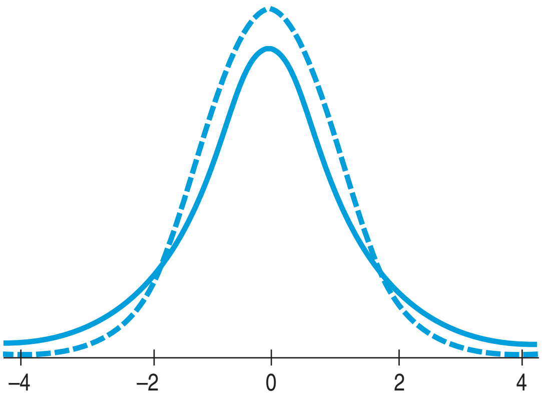


Figure 12.1 The t-model (solid curve) with 2 degrees of freedom has fatter tails than the Normal model (dashed curve). So the 68–95–99.7 Rule doesn't work for t-models with only a few degrees of freedom.

z or t?

If you know σ, use z. (That's rare!) Whenever you use s to estimate σ, use t.

number of data values. Unfortunately, that's not practical. Fortunately, above a few hundred degrees of freedom it's very hard to tell the difference. Of course, in the rare situation that we *know* σ, it would be foolish not to use that information. If we don't have to estimate σ, we can use the Normal model. Typically that value of σ would be based on (lots of) experience, or on a theoretical model. Usually, however, we estimate σ by s from the data and use the t-model.

Using a known standard deviation

Variation is inherent in manufacturing, even under the most tightly controlled processes. To ensure that parts do not vary too much, however, quality professionals monitor the processes by selecting samples at regular intervals (see Chapter 22 for more details). The mean performance of these samples is measured, and if it lies too far from the desired target mean, the process may be stopped until the underlying cause of the problem can be determined. In silicon wafer manufacturing, the thickness of the film is a crucial measurement. To assess a sample of wafers, quality engineers compare the mean thickness of the sample to the target mean. But, they don't estimate the standard deviation of the mean by using the standard error derived from the same sample. Instead they base the standard deviation of the mean on the historical process standard deviation, estimated from a vast collection of similar parts. In this case, the standard deviation can be treated as "known" and the normal model can be used for the sampling distribution instead of the t distribution.

12.2 A Confidence Interval for Means

Notation Alert!

Ever since Gosset, the letter t has been reserved in Statistics for his distribution.

To make confidence intervals, we need to use Gosset's model. Which one? Well, for means, it turns out the right value for degrees of freedom is df $= n - 1$.

Practical sampling distribution model for means

When certain conditions are met, the standardized sample mean,

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

follows a Student's t-model with $n - 1$ degrees of freedom. We find the standard error from:

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

When Gosset corrected the Normal model for the extra uncertainty, the margin of error got bigger, as you might have guessed. When you use Gosset's model instead of the Normal model, your confidence intervals will be just a bit wider. That's just the correction you need. By using the t-model, you've compensated for the extra variability in precisely the right way.

One-sample t-interval

When the assumptions and conditions are met, we are ready to find the **confidence interval for the population mean**, μ. The confidence interval is:

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

where the standard error of the mean is:

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

The critical value t^*_{n-1} depends on the particular confidence level, C, that you specify and on the number of degrees of freedom, $n - 1$, which we get from the sample size.

Finding t^*-Values

The Student's t-model is different for each value of degrees of freedom. We might print a table like Table Z (in Appendix D) for each degrees of freedom value, but that's a lot of pages and not likely to be a bestseller. One way to shorten the book is to limit ourselves to 80%, 90%, 95% and 99% confidence levels. So Statistics books usually have one table of t-model critical values for a selected set of confidence levels. This one does too; see Table T in Appendix D. (You can also find tables on the Internet.)

The t-tables run down the page for as many degrees of freedom as can fit, and, as you can see from Figure 12.2, they are much easier to use than the Normal tables. Then they get to the bottom of the page and run out of room. Of course, for *enough* degrees of freedom, the t-model gets closer and closer to the Normal, so the tables give a final row with the critical values from the Normal model and label it "∞ df."

Two tail probability	0.20	0.10	0.05
One tail probability	0.10	0.05	0.025
Table T Values of t_α — df			
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
Confidence levels	80%	90%	95%

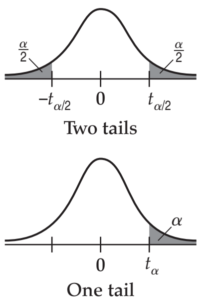


Figure 12.2 Part of Table T in Appendix D.

For Example — Finding a confidence interval for the mean

According to the Environmental Defense Fund, "Americans are eating more and more salmon, drawn to its rich taste and health benefits. Increasingly they are choosing *farmed* salmon because of its wide availability and low price. But in the last few years, farmed salmon has been surrounded by controversy over its health risks and the ecological impacts of salmon aquaculture operations. Studies have shown that some farmed salmon is relatively higher in contaminants like PCBs than wild salmon, and there is mounting concern over the industry's impact on wild salmon populations."

In a widely cited study of contaminants in farmed salmon, fish from many sources were analyzed for 14 organic contaminants.[3] One of those was the insecticide mirex, which has been shown to be carcinogenic and is suspected of being toxic to the liver, kidneys, and endocrine system. Summaries for 150 mirex concentrations (in parts per million) from a variety of farmed salmon sources were reported as:

$$n = 150;\quad \bar{y} = 0.0913 \text{ ppm};\quad s = 0.0495 \text{ ppm}$$

Question: The Environmental Protection Agency (EPA) recommends to recreational fishers as a "screening value" that mirex concentrations be no larger than 0.08 ppm. What does the 95% confidence interval say about that value?

Answer: Because $n = 150$, there are 149 df. From Table T in Appendix D, we find $t^*_{140,\,0.025} = 1.977$ (from technology, $t^*_{149,\,0.025} = 1.976$), so a 95% confidence interval can be found from:

$$\bar{y} \pm t^* \times SE(\bar{y}) = \bar{y} \pm 1.977 \times \frac{s}{\sqrt{n}} = 0.0913 \pm 1.977\frac{0.0495}{\sqrt{150}} = (0.0833, 0.0993)$$

If this sample is representative (as the authors claim it is), we can be 95% confident that it contains the true value of the mean mirex concentration. Because the interval from 0.0834 to 0.0992 ppm is entirely above the recommended value set by the EPA, we have reason to believe that the true mirex concentration exceeds the EPA guidelines.

12.3 Assumptions and Conditions

Gosset found the t-model by simulation. Years later, when Fisher showed mathematically that Gosset was right, he needed to make some assumptions to make the proof work. These are the assumptions we need in order to use the Student's t-models.

We Don't *Want* to Stop

We check conditions hoping that we can make a meaningful analysis of our data. The conditions serve as *disqualifiers*—we keep going unless there's a serious problem. If we find minor issues, we note them and express caution about our results. If the sample is not an SRS, but we believe it's representative of some populations, we limit our conclusions accordingly. If there are outliers, rather than stop, we perform the analysis both with and without them. If the sample looks bimodal, we try to analyze subgroups separately. Only when there's major trouble—like a strongly skewed small sample or an obviously non-representative sample—are we unable to proceed at all.

Independence Assumption

Independence Assumption: The data values should be independent. There's really no way to check independence of the data by looking at the sample, but we should think about whether the assumption is reasonable.

Randomization Condition: The data arise from a random sample or suitably randomized experiment. Randomly sampled data—and especially data from a Simple Random Sample (SRS)—are ideal.

When a sample is drawn without replacement, technically we ought to confirm that we haven't sampled a large fraction of the population, which would threaten the independence of our selections.

10% Condition: The sample size should be no more than 10% of the population. In practice, though, we often don't mention the 10% Condition when estimating means. Why not? When we made inferences about proportions, this condition was crucial because we usually had large samples. But for means our samples are

[3] Ronald A. Hites, Jeffery A. Foran, David O. Carpenter, M. Coreen Hamilton, Barbara A. Knuth, and Steven J. Schwager, "Global Assessment of Organic Contaminants in Farmed Salmon," *Science* 9 January 2004: Vol. 303, no. 5655, pp. 226–229.

generally smaller, so this problem arises only if we're sampling from a small population (and then there's a correction formula we could use).

Normal Population Assumption

Student's t-models won't work for data that are badly skewed. How skewed is too skewed? Well, formally, we assume that the data are from a population that follows a Normal model. Practically speaking, there's no way to be certain this is true.

And it's almost certainly *not* true. Models are idealized; real data are, well, real. The good news, however, is that even for small samples, it's sufficient to check a condition.

Nearly Normal Condition. The data come from a distribution that is unimodal and symmetric. This is a much more practical condition and one we can check by making a histogram.[4] For small samples, it can be hard to see any distribution shape in the histogram. Unfortunately, the condition matters most when it's hardest to check.

For very small samples ($n < 15$ or so), the data should follow a Normal model pretty closely. Of course, with so little data, it's rather hard to tell. But if you do find outliers or strong skewness, don't use these methods.

For moderate sample sizes (n between 15 and 40 or so), the t methods will work well as long as the data are unimodal and reasonably symmetric. Make a histogram to check.

When the sample size is larger than 40 or 50, the t methods are safe to use unless the data are extremely skewed. Make a histogram anyway. If you find outliers in the data and they aren't errors that are easy to fix, it's always a good idea to perform the analysis twice, once with and once without the outliers, even for large samples. The outliers may well hold additional information about the data, so they deserve special attention. If you find multiple modes, you may well have different groups that should be analyzed and understood separately.

If the data are extremely skewed, the mean may not be the most appropriate summary. But when our data consist of a collection of instances whose *total* is the business consequence—as when we add up the profits (or losses) from many transactions or the costs of many supplies—then the mean is just that total divided by n. And that's the value with a business consequence. Fortunately, in this instance, the Central Limit Theorem comes to our rescue. Even when we must sample from a very skewed distribution, the sampling distribution of our sample mean will be close to Normal, so we can use Student's t methods without much worry as long as the sample size is *large enough*.

How large is large enough? Here's the histogram of CEO compensations ($000) for Fortune 500 companies.

Notation Alert!

When we found critical values from a Normal model, we called them z^*. When we use a Student's t-model, we denote the critical values t^*.

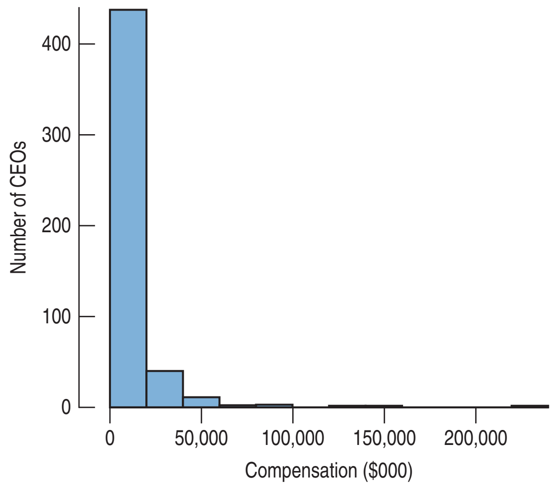


Figure 12.3 It's hard to imagine a distribution more skewed than these annual compensations from the Fortune 500 CEOs.

[4] Or we could check a normal probability plot.

Although this distribution is very skewed, the Central Limit Theorem will make the sampling distribution of the means of samples from this distribution more and more Normal as the sample size grows. Here's a histogram of the means of many samples of 100 CEOs:

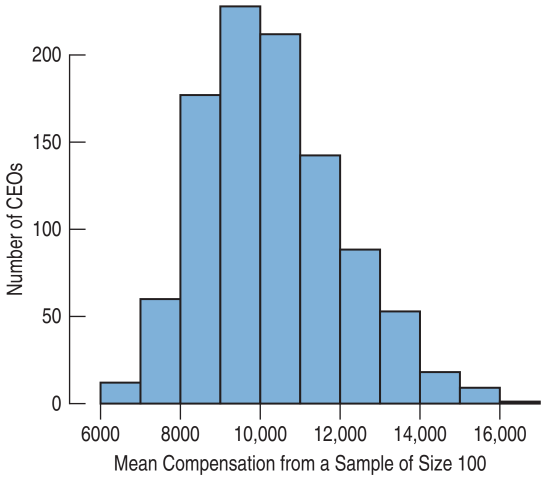
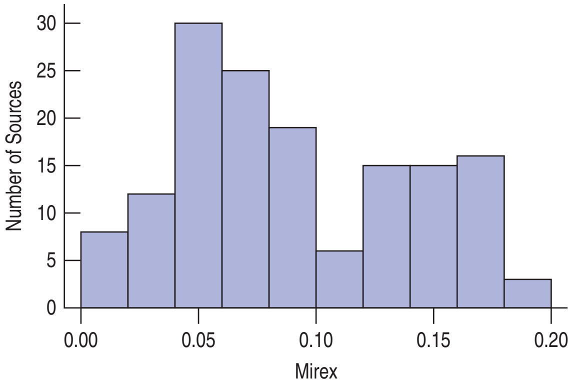


Figure 12.4 Even samples as small as 100 from the CEO data set produce means whose sampling distribution is nearly normal. Larger samples will have sampling distributions even more Normal.

Often, in modern business applications, we have samples of many hundreds, or thousands. We should still be on guard for outliers and multiple modes and we should be sure that the observations are independent. But if the mean is of interest, the Central Limit Theorem works quite well in ensuring that the sampling distribution of the mean will be close to the Normal for samples of this size.

For Example Checking the assumptions and conditions for a confidence interval for means

Researchers purchased whole farmed salmon from 51 farms in eight regions in six countries (see page 336). The histogram shows the concentrations of the insecticide mirex in the 150 samples of farmed salmon we examined in the previous example.



Question: Are the assumptions and conditions for making a confidence interval for the mean mirex concentration satisfied?

Answer:

✓ **Independence Assumption:** The fish were raised in many different places, and samples were purchased independently from several sources.

✓ **Randomization Condition:** The fish were selected randomly from those available for sale.

✓ **10% Condition:** There are lots of fish in the sea (and at the fish farms); 150 is certainly far fewer than 10% of the population.

✓ **Nearly Normal Condition:** The histogram of the data looks bimodal. While it might be interesting to learn the reason for that and possibly identify the subsets, we can proceed because the sample size is large.

It's okay to use these data about farm-raised salmon to make a confidence interval for the mean.

Just Checking

Every 10 years, the United States takes a census that tries to count every resident. In addition, the census collects information on a variety of economic and social questions. Businesses of all types use the census data to plan sales and marketing strategies and to understand the underlying demographics of the areas that they serve.

There are two census forms: the "short form," answered by most people, and the "long form," sent only to about one in six or seven households chosen at random. According to the Census Bureau (factfinder.census.gov), ". . . each estimate based on the long form responses has an associated confidence interval."

1 Why does the Census Bureau need a confidence interval for long-form information, but not for the questions that appear on both the long and short forms?

2 Why must the Census Bureau base these confidence intervals on t-models?

The Census Bureau goes on to say, "These confidence intervals are wider . . . for geographic areas with smaller populations and for characteristics that occur less frequently in the area being examined (such as the proportion of people in poverty in a middle-income neighborhood)."

3 Why is this so? For example, why should a confidence interval for the mean amount families spend monthly on housing be wider for a sparsely populated area of farms in the Midwest than for a densely populated area of an urban center? How does the formula for the one-sample t-interval show this will happen?

To deal with this problem, the Census Bureau reports long-form data only for ". . . geographic areas from which about two hundred or more long forms were completed—which are large enough to produce good quality estimates. If smaller weighting areas had been used, the confidence intervals around the estimates would have been significantly wider, rendering many estimates less useful."

4 Suppose the Census Bureau decided to report on areas from which only 50 long forms were completed. What effect would that have on a 95% confidence interval for, say, the mean cost of housing? Specifically, which values used in the formula for the margin of error would change? Which values would change a lot, and which values would change only slightly? Approximately how much wider would that confidence interval based on 50 forms be than the one based on 200 forms?

Guided Example Insurance Profits

Insurance companies take risks. When they insure a property or a life, they must price the policy in such a way that their expected profit enables them to survive. They can base their projections on actuarial tables, but the reality of the insurance business often demands that they discount policies to a variety of customers and situations. Managing this risk is made even more difficult by the fact that until the policy expires, the company won't know if they've made a profit, no matter what premium they charge.

A manager wanted to see how well one of her sales representatives was doing, so she selected 30 matured policies that had been sold by the sales rep and computed the (net) profit (premium charged minus paid claims), for each of the 30 policies.

The manager would like you, as a consultant, to construct a 95% confidence interval for the mean profit of the policies sold by this sales rep.

Profit (in $) from 30 policies		
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	−397.10
3255.60	1847.50	−397.31
3701.85	865.40	186.25
−803.35	1415.65	590.85
3865.90	2756.94	578.95

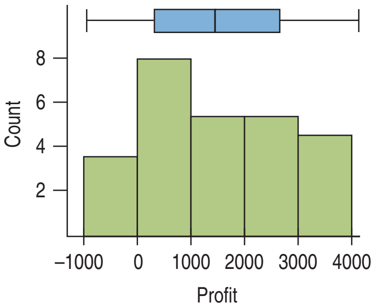
PLAN

Setup State what we want to know. Identify the variables and their context.

We wish to find a 95% confidence interval for the mean profit of policies sold by this sales rep. We have data for 30 matured policies.

Make a picture. Check the distribution shape and look for skewness, multiple modes, and outliers.

Here's a boxplot and histogram of these values.



The sample appears to be unimodal and fairly symmetric with profit values between −$1000 and $4000 and no outliers.

Model Think about the assumptions and check the conditions.

✓ **Independence Assumption**

This is a random sample so observations should be independent.

✓ **Randomization Condition**

This sample was selected randomly from the matured policies sold by the sales representative of the company.

✓ **Nearly Normal Condition**

The distribution of profits is unimodal and fairly symmetric without strong skewness.

State the sampling distribution model for the statistic.

We will use a Student's t-model with $n - 1 = 30 - 1 = 29$ degrees of freedom and find a one-sample t-interval for the mean.

DO

Mechanics Compute basic statistics and construct the confidence interval.

Using software, we obtain the following basic statistics:

$$n = 30$$
$$\bar{y} = \$1438.90$$
$$s = \$1329.60$$

Remember that the standard error of the mean is equal to the standard deviation divided by the square root of n.

The standard error of the mean is:

$$SE(\bar{y}) = \frac{s}{\sqrt{n}} = \frac{1329.60}{\sqrt{30}} = \$242.75$$

The critical value we need to make a 95% confidence interval comes from a Student's t table, a computer program, or a calculator. We have $30 - 1 = 29$ degrees of freedom. The selected confidence level says that we want 95% of the probability to be caught in the middle, so we exclude 2.5% in *each* tail, for a total of 5%. The degrees of freedom and 2.5% tail probability are all we need to know to find the critical value. Here it's 2.045.

There are $30 - 1 = 29$ degrees of freedom. The manager has specified a 95% level of confidence, so the critical value (from table T) is 2.045.

The margin of error is:

$$ME = 2.045 \times SE(\bar{y})$$
$$= 2.045 \times 242.75$$
$$= \$496.42$$

The 95% confidence interval for the mean profit is:

$$\$1438.90 \pm \$496.42$$
$$= (\$942.48, \$1935.32)$$

REPORT

Conclusion Interpret the confidence interval in the proper context.

MEMO

Re: Profit from Policies

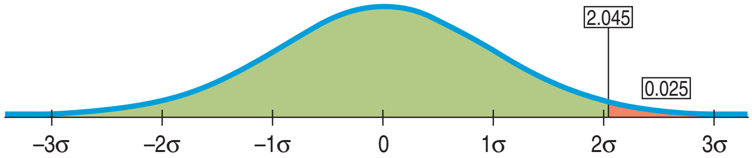
From our analysis of the selected policies, we are 95% confident that the true mean profit of policies sold by this sales rep is contained in the interval from \$942.48 to \$1935.32.

When we construct confidence intervals in this way, we expect 95% of them to cover the true mean and 5% to miss the true value. That's what "95% confident" means.

Caveat: Insurance losses are notoriously subject to outliers. One very large loss could influence the average profit substantially. However, there were no such cases in this data set.

Finding Student's *t* Critical Values

The critical value in the Guided Example was found in the Student's t Table in Appendix D. To find the critical value, locate the row of the table corresponding to the degrees of freedom and the column corresponding to the probability you want. Since a 95% confidence interval leaves 2.5% of the values on either side, we look for 0.025 at the top of the column or look for 95% confidence directly in the bottom row of the table. The value in the table at that intersection is the critical value we need. In the Guided Example, the number of degrees of freedom was $30 - 1 = 29$, so we located the value of 2.045.



Probability

	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

Figure 12.5 Using Table T to look up the critical value t^* for a 95% confidence level with 29 degrees of freedom.

12.4 Cautions about Interpreting Confidence Intervals

So What *Should* You Say?

Since 95% of random samples yield an interval that captures the true mean, you should say: "I am 95% confident that the interval from \$942.48 to \$1935.32 contains the mean profit of all policies sold by this sales representative." It's also okay to say something slightly less formal: "I am 95% confident that the mean profit for all policies sold by this sales rep is between \$942.48 and \$1935.32." Remember: *Your uncertainty is about the interval, not the true mean.* The interval varies randomly. The true mean profit is neither variable nor random—just unknown.

Confidence intervals for means offer new, tempting, wrong interpretations. Here are some ways to keep from going astray:

- ***Don't say,*** "*95% of all the policies* sold by this sales rep have profits between \$942.48 and \$1935.32." The confidence interval is about the *mean*, not about the measurements of individual policies.
- ***Don't say,*** "We are 95% confident that *a randomly selected policy* will have a net profit between \$942.48 and \$1935.32." This false interpretation is also about individual policies rather than about the *mean* of the policies. We are 95% confident that the *mean* profit of all (similar) policies sold by this sales rep is between \$942.48 and \$1935.32.
- ***Don't say,*** "The mean profit is \$1438.90 95% *of the time*." That's about means, but still wrong. It implies that the true mean varies, when in fact it is the confidence interval that would have been different had we gotten a different sample.
- Finally, ***don't say,*** "95% *of all samples* will have mean profits between \$942.48 and \$1935.32." That statement suggests that *this* interval somehow sets a standard for every other interval. In fact, this interval is no more (or less) likely to be correct than any other. You could say that 95% of all possible samples would produce intervals that contain the true mean profit. (The problem is that because we'll never know what the true mean profit is, we can't know if our sample was one of those 95%.)

Just Checking

In discussing estimates based on the long-form samples, the Census Bureau notes, "The disadvantage . . . is that . . . estimates of characteristics that are also reported on the short form will not match the [long-form estimates]."

The short-form estimates are values from a complete census, so they are the "true" values—something we don't usually have when we do inference.

5 Suppose we use long-form data to make 100 95% confidence intervals for the mean age of residents, one for each of 100 of the census-defined areas. How many of these 100 intervals should we expect will *fail* to include the true mean age (as determined from the complete short-form census data)?

12.5 Sample Size

How large a sample do we need? The simple answer is always "larger." But more data cost money, effort, and time. So how much is enough? Suppose your computer took an hour to download a movie you wanted to watch. You wouldn't be happy. Then you hear about a program that claims to download movies in under a half hour. You're interested enough to spend $29.95 for it, but only if it really delivers. So you get the free evaluation copy and test it by downloading a movie 10 times. Of course, the mean download time is not exactly 30 minutes as claimed. Observations vary. If the margin of error were 8 minutes, though, you'd probably be able to decide whether the software was worth the money. Doubling the sample size would require another 5 or so hours of testing and would reduce your margin of error to a bit under 6 minutes. You'd need to decide whether that's worth the effort.

As we make plans to collect data, we should have some idea of how small a margin of error is required to be able to draw a conclusion or detect a difference we want to see. If the size of the effect we're studying is large, then we may be able to tolerate a larger ME. If we need great precision, however, we'll want a smaller ME, and, of course, that means a larger sample size.

Armed with the ME and confidence level, we can find the sample size we'll need. Almost.

We know that for a mean, $ME = t^*_{n-1} \times SE(\bar{y})$ and that $SE(\bar{y}) = \frac{s}{\sqrt{n}}$, so we can determine the sample size by solving this equation for n:

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

The good news is that we have an equation; the bad news is that we won't know most of the values we need to compute it. When we thought about sample size for proportions, we ran into a similar problem. There we had to guess a working value for p to compute a sample size. Here, we need to know s. We don't know s until we get some data, but we want to calculate the sample size *before* collecting the data. We might be able to make a good guess, and that is often good enough for this purpose. If we have no idea what the standard deviation might be or if the sample size really matters (for example, because each additional individual is very expensive to sample or experiment on), it might be a good idea to run a small *pilot study* to get some feeling for the size of the standard deviation.

That's not all. Without knowing n, we don't know the degrees of freedom, and we can't find the critical value, t^*_{n-1}. One common approach is to use the corresponding z^* value from the Normal model. If you've chosen a 95% confidence level, then just use 2, following the 68–95–99.7 Rule, or 1.96 to be more precise. If your estimated sample size is 60 or more, it's probably okay—z^* was a good guess. If it's smaller than that, you may want to add a step, using z^* at first, finding n, and then replacing z^* with the corresponding t^*_{n-1} and calculating the sample size once more.

Sample size calculations are *never* exact. The margin of error you find *after* collecting the data won't match exactly the one you used to find n. The sample size formula depends on quantities that you won't have until you collect the data, but using it is an important first step. Before you collect data, it's always a good idea to know whether the sample size is large enough to give you a good chance of being able to tell you what you want to know.

By Hand

Sample size calculations

Let's give the sample size formula a spin. Suppose we want an ME of 8 minutes and we think the standard deviation of download times is about 10 minutes. Using a 95% confidence interval and $z^* = 1.96$, we solve for n:

$$8 = 1.96\frac{10}{\sqrt{n}}$$

$$\sqrt{n} = \frac{1.96 \times 10}{8} = 2.45$$

$$n = (2.45)^2 = 6.0025$$

That's a small sample size, so we use $(6 - 1) = 5$ degrees of freedom to substitute an appropriate t^* value. At 95%, $t_5^* = 2.571$. Now we can solve the equation one more time:

$$8 = 2.571\frac{10}{\sqrt{n}}$$

$$\sqrt{n} = \frac{2.571 \times 10}{8} \approx 3.214$$

$$n = (3.214)^2 \approx 10.33$$

To make sure the ME is no larger than you want, you should always round *up*, which gives $n = 11$ runs. So, to get an ME of 8 minutes, we should find the downloading times for $n = 11$ movies.

For Example **Finding the sample size for a confidence interval for means**

In the 150 samples of farmed salmon (see page 336), the mean concentration of mirex was 0.0913 ppm with a standard deviation of 0.0495 ppm. A 95% confidence interval for the mean mirex concentration was found to be: (0.0833, 0.0993).

Question: How large a sample would be needed to produce a 95% confidence interval with a margin of error of 0.004?

Answer: We will assume that the standard deviation is 0.0495 ppm. The margin of error is equal to the critical value times the standard error. Using z^*, we find:

$$0.004 = 1.96 \times \frac{0.0495}{\sqrt{n}}$$

Solving for n, we find:

$$\sqrt{n} = 1.96 \times \frac{0.0495}{0.004}$$

or

$$n = \left(1.96 \times \frac{0.0495}{0.004}\right)^2 = 588.3$$

The t^* critical value with 400 df is 1.966 instead of 1.960. Using that value, the margin of error is:

$$1.966 \times \frac{0.0495}{\sqrt{589}} = 0.00401$$

You could go back and use 1.966 instead of 1.960 in the equation for n, above, and you would find that n should be 592. That will give a margin of error of 0.004, but the uncertainty in the standard deviation is likely to make such differences unimportant.

12.6 Degrees of Freedom—Why $n - 1$?

The number of degrees of freedom $(n - 1)$ might have reminded you of the value we divide by to find the standard deviation of the data (since, after all, it's the same number). We promised back when we introduced that formula to say a bit more about why we divide by $n - 1$ rather than by n. The reason is closely tied to the reasoning of the t-distribution.

If only we knew the true population mean, μ, we would find the sample standard deviation using n instead of $n - 1$ as:

$$s = \sqrt{\frac{\sum(y - \mu)^2}{n}} \text{ and we'd call it } s.$$

We have to use $\bar{y}$ instead of μ, though, and that causes a problem. For any sample, $\bar{y}$ is as close to the data values as possible. Generally the population mean, μ, will be farther away. Think about it. GMAT scores have a population mean of 525. If you took a random sample of 5 students who took the test, their sample mean wouldn't be 525. The five data values will be closer to their own $\bar{y}$ than to 525. So if we use $\sum(y - \bar{y})^2$ instead of $\sum(y - \mu)^2$ in the equation to calculate s, our standard deviation estimate will be too small. The amazing mathematical fact is that we can compensate for the fact that $\sum(y - \bar{y})^2$ is too small just by dividing by $n - 1$ instead of by n. So that's all the $n - 1$ is doing in the denominator of s. We call $n - 1$ the degrees of freedom.

What Can Go Wrong?

First, you must decide when to use Student's t methods.

- **Don't confuse proportions and means.** When you treat your data as categorical, counting successes and summarizing with a sample proportion, make inferences using the Normal model methods. When you treat your data as quantitative, summarizing with a sample mean, make your inferences using Student's t methods.

Student's t methods work only when the Normal Population Assumption is true. Naturally, many of the ways things can go wrong turn out to be ways that the Normal Population Assumption can fail. It's always a good idea to look for the most common kinds of failure. It turns out that you can even fix some of them.

- **Beware of multimodality.** The Nearly Normal Condition clearly fails if a histogram of the data has two or more modes. When you see this, look for the possibility that your data come from two groups. If so, your best bet is to try to separate the data into groups. (Use the variables to help distinguish the modes, if possible. For example, if the modes seem to be composed mostly of men in one and women in the other, split the data according to the person's sex.) Then you can analyze each group separately.
- **Beware of skewed data.** Make a histogram of the data. If the data are severely skewed, you might try re-expressing the variable. Re-expressing may yield a distribution that is unimodal and symmetric, making it more appropriate for the inference methods for means. Re-expression cannot help if the sample distribution is not unimodal.

(continued)

What to Do with Outliers

As tempting as it is to get rid of annoying values, you can't just throw away outliers and not discuss them. It is not appropriate to lop off the highest or lowest values just to improve your results. The best strategy is to report the analysis with *and* without the outliers and comment on any differences.

- **Investigate outliers.** The Nearly Normal Condition also fails if the data have outliers. If you find outliers in the data, you need to investigate them. Sometimes, it's obvious that a data value is wrong and the justification for removing or correcting it is clear. When there's no clear justification for removing an outlier, you might want to run the analysis both with and without the outlier and note any differences in your conclusions. Any time data values are set aside, you *must* report on them individually. Often they will turn out to be the most informative part of your report on the data.[5]

  Of course, Normality issues aren't the only risks you face when doing inferences about means.

- **Watch out for bias.** Measurements of all kinds can be biased. If your observations differ from the true mean in a systematic way, your confidence interval may not capture the true mean. And there is no sample size that will save you. A bathroom scale that's 5 pounds off will be 5 pounds off even if you weigh yourself 100 times and take the average. We've seen several sources of bias in surveys, but measurements can be biased, too. Be sure to think about possible sources of bias in your measurements.

- **Make sure data are independent.** Student's *t* methods also require the sampled values to be mutually independent. We check for random sampling and the 10% Condition. You should also think hard about whether there are likely violations of independence in the data collection method. If there are, be very cautious about using these methods.

Ethics in Action

Recent reports have indicated that waiting times in hospital emergency rooms (ERs) across the United States are getting longer, with the average reported as 30 minutes in January 2008 (WashingtonPost.com). Several reasons have been cited for this rise in average ER waiting time including the closing of hospital emergency rooms in urban areas and problems with managing hospital flow. Tyler Hospital, located in rural Ohio, joined the Joint Commission's Continuous Service Readiness program and consequently agreed to monitor its ER waiting times. After collecting data for a random sample of 30 ER patients arriving at Tyler's ER during the last month, they found an average waiting time of 26 minutes with a standard deviation of 8.25 minutes. Further statistical analysis yielded a 95% confidence interval of 22.92 to 29.08 minutes, clear indication that Tyler's ER patients wait less than 30 minutes to see a doctor. Tyler's administration was not only pleased with the findings, but also sure that the Joint Commission would also be impressed. Their next step was to consider ways of including this message, "95% of Tyler's ER patients can expect to wait less than the national average to see a doctor," in their advertising and promotional materials.

ETHICAL ISSUE *Interpretation of the confidence interval is incorrect and misleading (related to Item C, ASA Ethical Guidelines). The confidence interval does not provide results for individual patients. So, it is incorrect to state that 95% of individual ER patients wait less (or can expect to wait less) than 30 minutes to see a doctor.*

ETHICAL SOLUTION *Interpret the results of the confidence interval correctly, in terms of the mean waiting time and not individual patients.*

[5] This suggestion may be controversial in some disciplines. Setting aside outliers is seen by some as unethical because the result is likely to be a narrower confidence interval or a smaller P-value. But an analysis of data with outliers left in place is *always* wrong. The outliers violate the Nearly Normal Condition and also the implicit assumption of a homogeneous population, so they invalidate inference procedures. An analysis of the nonoutlying points, along with a separate discussion of the outliers, is often much more informative, and can reveal important aspects of the data.

What Have We Learned?

Learning Objectives

- Know the sampling distribution of the mean.
  - To apply the Central Limit Theorem for the mean in practical applications, we must estimate the standard deviation. This *standard error* is

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

  - When we use the SE, the sampling distribution that allows for the additional uncertainty is Student's t.
- Construct confidence intervals for the true mean, μ.
  - A confidence interval for the mean has the form $\bar{y} \pm ME$.
  - The Margin of Error is $ME = t^*_{df} SE(\bar{y})$.
- Find t^* values by technology or from tables.
  - When constructing confidence intervals for means, the correct degrees of freedom is $n - 1$.
- Check the Assumptions and Conditions before using any sampling distribution for inference.
- Write clear summaries to interpret a confidence interval.

Terms

Degrees of freedom (df) A parameter of the Student's t-distribution that depends upon the sample size. Typically, more degrees of freedom reflects increasing information from the sample.

One-sample t-interval for the mean A one-sample t-interval for the population mean is:

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

The critical value t^*_{n-1} depends on the particular confidence level, C, that you specify and on the number of degrees of freedom, $n - 1$.

Student's t A family of distributions indexed by its degrees of freedom. The t-models are unimodal, symmetric, and bell-shaped, but generally have fatter tails and a narrower center than the Normal model. As the degrees of freedom increase, t-distributions approach the Normal model.

Technology Help: Inference for Means

Statistics packages offer convenient ways to make histograms of the data. That means you have no excuse for skipping the check that the data are nearly Normal.

Any standard statistics package can compute a confidence interval.

Inference results are sometimes reported in a table. You may have to read carefully to find the values you need. Often, confidence interval bounds are given together with related results for hypothesis tests (which we'll learn about in the next chapter). Here is an example of that kind of output (although no package we know gives results in exactly this form).

The commands to do inference for means on common statistics programs and calculators are not always obvious. (By contrast, the resulting output is usually clearly labeled and easy to read.) The guides for each program can help you start navigating.

Calculated mean, $\bar{y}$

```
Hypothesized value   30
Estimated mean       31.043478
DF                   22
Std Error             0.886
Alpha       0.05

            tTest                     tinterval
Statistic   1.178
Prob > |t|  0.2513     Upper 95%     32.880348
Prob > t    0.1257     Lower 95%     29.206608
Prob < t    0.8743
```

Confidence interval bounds

(continued)

EXCEL

XLSTAT

To find a confidence interval for a mean in Excel, you can set up the calculations using Excel's functions. For example, suppose you have 100 observations in cells A1:A100.

- In cell B2, enter "=AVERAGE(A1:A100)" to compute the sample mean.
- In cell B3, enter "=STDEV(A1:A11)/SQRT(100)" to compute the standard error of the mean.
- In cell B4, enter "=TINV(.05,99)" to compute t*.
- In cell B5, enter "=B2-B4*B3" as the lower end of the CI.
- In cell B6, enter "=B2+B4*B3" as the upper end of the CI.

JMP

- From the **Analyze** menu, select **Distribution**.
- For a confidence interval, scroll down to the "Moments" section to find the interval limits. (Be sure that your variables are "Continuous" type so that this section will be available.)

Comments

"Moment" is a fancy statistical term for means, standard deviations, and other related statistics.

MINITAB

- From the **Stat** menu, choose the **Basic Statistics** submenu.
- From that menu, choose 1-sample t . . . .
- Then fill in the dialog.

Comments

The dialog offers a clear choice between confidence interval and hypothesis test.

SPSS

- From the **Analyze** menu, choose the **Compare Means** submenu.
- From that, choose **One-Sample t-test** command.

Comments

The commands suggest neither a single mean nor an interval. But the results provide both.

Brief CASE

Real Estate

A real estate agent is trying to understand the pricing of homes in her area, a region comprised of small to midsize towns and a small city. For each of 1200 homes recently sold in the region, the file **Real_Estate_sample1200** holds the following variables:

- *Sale Price* (in $)
- *Lot size* (size of the lot in acres)
- *Waterfront* (Yes, No)
- *Age* (in years)
- *Central Air* (Yes, No)
- *Fuel Type* (Wood, Oil, Gas, Electric, Propane, Solar, Other)
- *Condition* (1 to 5, 1 = Poor, 5 = Excellent)
- *Living Area* (living area in square feet)
- *Pct College* (% in zip code who attend a four-year college)

- *Full Baths* (number of full bathrooms)
- *Half Baths* (number of half bathrooms)
- *Bedrooms* (number of bedrooms)
- *Fireplaces* (number of fireplaces)

The agent has a family interested in a four bedroom house. Using confidence intervals, how should she advise the family on what the average price of a four bedroom house might be in this area? Compare that to a confidence interval for two bedroom homes. How does the presence of central air conditioning affect the mean price of houses in this area? Use confidence intervals and graphics to help answer that question.

Explore other questions that might be useful for the real estate agent in knowing how different categorical factors affect the sale price and write up a short report on your findings.

Donor Profiles

A philanthropic organization collects and buys data on their donor base. The full database contains about 4.5 million donors and over 400 variables collected on each, but the data set **Donor_Profiles** is a sample of 916 donors and includes the variables:

- *Age* (in years)
- *Homeowner* (H = Yes, U = Unknown)
- *Gender* (F = Female, M = Male, U = Unknown)
- *Wealth* (Ordered categories of total household wealth from 1 = Lowest to 9 = Highest)
- *Children* (Number of children)
- *Donated Last* (0 = Did not donate to last campaign, 1 = Did donate to last campaign)
- *Amt Donated Last* ($ amount of contribution to last campaign)

The analysts at the organization want to know how much people donate on average to campaigns, and what factors might influence that amount. Compare the confidence intervals for the mean *Amt Donated Last* by those known to own their homes with those whose homeowner status is unknown. Perform similar comparisons for *Gender* and two of the *Wealth* categories. Write up a short report using graphics and confidence intervals for what you have found. (Be careful not to make inferences directly about the differences between groups. We'll discuss that in Chapter 14. Your inference should be about single groups.)

(The distribution of *Amt Donated Last* is highly skewed to the right, and so the median might be thought to be the appropriate summary. But the median is $0.00 so the analysts must use the mean. From simulations, they have ascertained that the sampling distribution for the mean is unimodal and symmetric for samples larger than 250 or so. Note that small differences in the mean could result in millions of dollars of added revenue nationwide. The average cost of their solicitation is $0.67 per person to produce and mail.)

Exercises

SECTION 12.1

1. From Chapter 5 a survey of 25 randomly selected customers found the following ages (in years):

20	32	34	29	30
30	30	14	29	11
38	22	44	48	26
25	22	32	35	32
35	42	44	44	48

Recall that the mean was 31.84 years and the standard deviation was 9.84 years.

a) What is the standard error of the mean?
b) How would the standard error change if the same size had been 100 instead of 25? (Assume that the sample standard deviation didn't change.)

2. From Chapter 5, a random sample of 20 purchases showed the following amounts (in $):

39.05	2.73	32.92	47.51
37.91	34.35	64.48	51.96
56.95	81.58	47.80	11.72
21.57	40.83	38.24	32.98
75.16	74.30	47.54	65.62

Recall that the mean was $45.26 and the standard deviation was $20.67.

a) What is the standard error of the mean?
b) How would the standard error change if the same size had been 5 instead of 20? (Assume that the sample standard deviation didn't change.)

3. For the data in Exercise 1:

a) How many degrees of freedom does the t-statistic have?
b) How many degrees of freedom would the t-statistic have if the sample size had been 100?

4. For the data in Exercise 2:

a) How many degrees of freedom does the t-statistic have?
b) How many degrees of freedom would the t-statistic have if the sample size had been 5?

SECTION 12.2

5. Find the critical value t^* for:

a) a 95% confidence interval based on 24 df.
b) a 95% confidence interval based on 99 df.

6. Find the critical value t^* for:

a) a 90% confidence interval based on 19 df.
b) a 90% confidence interval based on 4 df.

7. For the ages in Exercise 1:

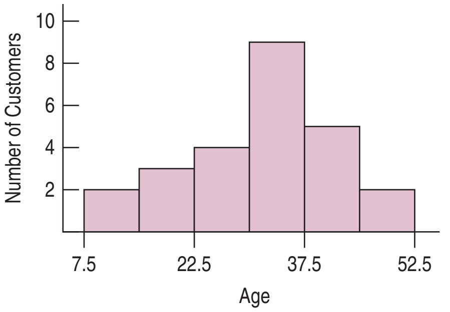
a) Construct a 95% confidence interval for the mean age of all customers, assuming that the assumptions and conditions for the confidence interval have been met.
b) How large is the margin of error?
c) How would the confidence interval change if you had assumed that the standard deviation was known to be 10.0 years?

8. For the purchase amounts in Exercise 2:

a) Construct a 90% confidence interval for the mean purchases of all customers, assuming that the assumptions and conditions for the confidence interval have been met.
b) How large is the margin of error?
c) How would the confidence interval change if you had assumed that the standard deviation was known to be $20?

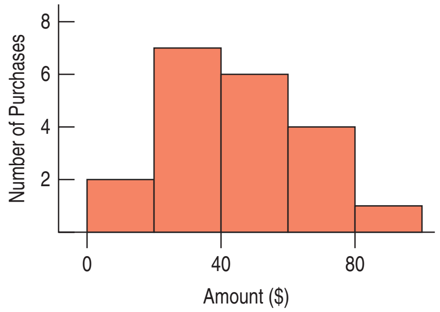
SECTION 12.3

9. For the confidence intervals of Exercise 7, a histogram of the data looks like this:



Check the assumptions and conditions for your inference.

10. For confidence intervals of Exercise 8, a histogram of the data looks like this:



Check the assumptions and conditions for your inference.

SECTION 12.5

11. For the confidence interval in Exercise 7:

a) How large would the sample size have to be to cut the margin of error in half?
b) About how large would the sample size have to be to cut the margin of error by a factor of 10?

12. For the confidence interval in Exercise 8:

a) To reduce the margin of error to about \$4, how large would the sample size have to be?
b) How large would the sample size have to be to reduce the margin of error to \$0.80?

CHAPTER EXERCISES

13. *t*-models. Using the t tables, software, or a calculator, estimate:

a) the critical value of t for a 90% confidence interval with df = 17.
b) the critical value of t for a 98% confidence interval with df = 88.

14. *t*-models, part 2. Using the t tables, software, or a calculator, estimate:

a) the critical value of t for a 95% confidence interval with df = 7.
b) the critical value of t for a 99% confidence interval with df = 102.

15. Confidence intervals. Describe how the width of a 95% confidence interval for a mean changes as the standard deviation (s) of a sample increases, assuming sample size remains the same.

16. Confidence intervals, part 2. Describe how the width of a 95% confidence interval for a mean changes as the sample size (n) increases, assuming the standard deviation remains the same.

17. Confidence intervals and sample size. A confidence interval for the price of gasoline from a random sample of 30 gas stations in a region gives the following statistics:

$$\bar{y} = \$4.49 \quad s = \$0.29$$

a) Find a 95% confidence interval for the mean price of regular gasoline in that region.
b) Find the 90% confidence interval for the mean.
c) If we had the same statistics from a sample of 60 stations, what would the 95% confidence interval be now?

18. Confidence intervals and sample size, part 2. A confidence interval for the price of gasoline from a random sample of 30 gas stations in a region gives the following statistics:

$$\bar{y} = \$4.49 \quad SE(\bar{y}) = \$0.06$$

a) Find a 95% confidence interval for the mean price of regular gasoline in that region.
b) Find the 90% confidence interval for the mean.
c) If we had the same statistics from a sample of 60 stations, what would the 95% confidence interval be now?

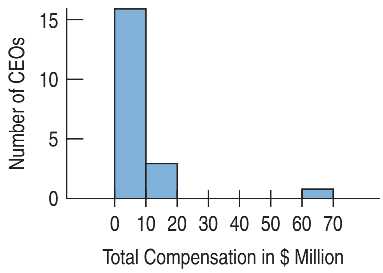
19. Marketing livestock feed. A feed supply company has developed a special feed supplement to see if it will promote weight gain in livestock. Their researchers report that the 77 cows studied gained an average of 56 pounds and that a 95% confidence interval for the mean weight gain this supplement produces has a margin of error of ±11 pounds. Staff in their marketing department wrote the following conclusions. Did anyone interpret the interval correctly? Explain any misinterpretations.

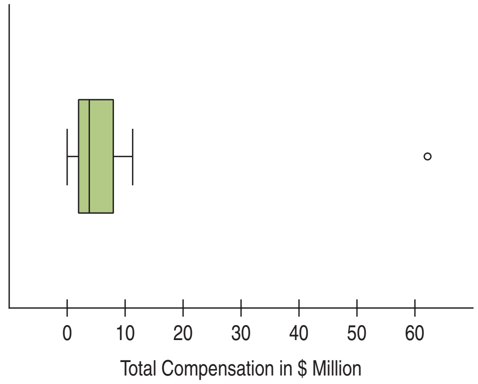
a) 95% of the cows studied gained between 45 and 67 pounds.
b) We're 95% sure that a cow fed this supplement will gain between 45 and 67 pounds.
c) We're 95% sure that the average weight gain among the cows in this study was between 45 and 67 pounds.
d) The average weight gain of cows fed this supplement is between 45 and 67 pounds 95% of the time.
e) If this supplement is tested on another sample of cows, there is a 95% chance that their average weight gain will be between 45 and 67 pounds.

20. Meal costs. A company is interested in estimating the costs of lunch in their cafeteria. After surveying employees, the staff calculated that a 95% confidence interval for the mean amount of money spent for lunch over a period of six months is (\$780, \$920). Now the organization is trying to write its report and considering the following interpretations. Comment on each.

a) 95% of all employees pay between \$780 and \$920 for lunch.
b) 95% of the sampled employees paid between \$780 and \$920 for lunch.
c) We're 95% sure that employees in this sample averaged between \$780 and \$920 for lunch.
d) 95% of all samples of employees will have average lunch costs between \$780 and \$920.
e) We're 95% sure that the average amount all employees pay for lunch is between \$780 and \$920.

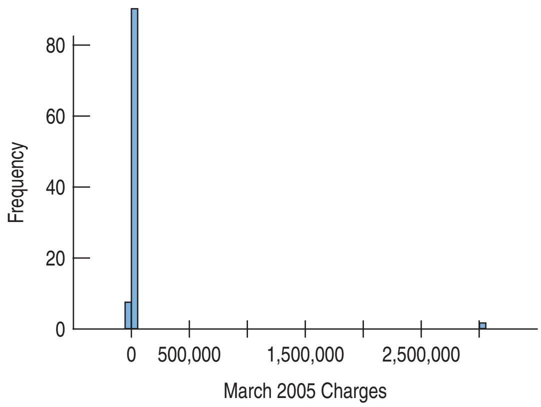
21. CEO compensation. A sample of 20 CEOs from the Forbes 500 shows total annual compensations ranging from a minimum of \$0.1 to \$62.24 million. The average for these 20 CEOs is \$7.946 million. The histogram and boxplot are as follows:

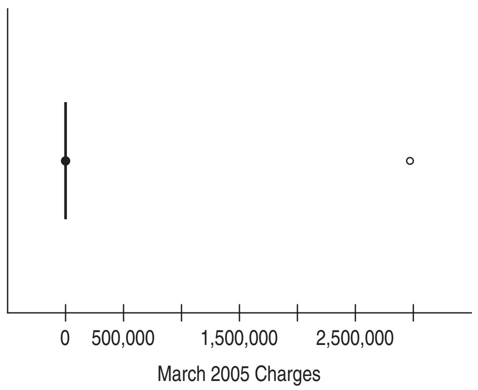




Based on these data, a computer program found that a confidence interval for the mean annual compensation of all Forbes 500 CEOs is (1.69, 14.20) $M. Why should you be hesitant to trust this confidence interval?

22. Credit card charges. A credit card company takes a random sample of 100 cardholders to see how much they charged on their card last month. A histogram and boxplot are as follows:





A computer program found that the 95% confidence interval for the mean amount spent in March 2005 is (−$28,366.84, $90,691.49). Explain why the analysts didn't find the confidence interval useful, and explain what went wrong.

23. Parking. Hoping to lure more shoppers downtown, a city builds a new public parking garage in the central business district. The city plans to pay for the structure through parking fees. For a random sample of 44 weekdays, daily fees collected averaged $126, with a standard deviation of $15.

a) What assumptions must you make in order to use these statistics for inference?
b) Find a 90% confidence interval for the mean daily income this parking garage will generate.
c) Explain in context what this confidence interval means.
d) Explain what 90% confidence means in this context.
e) The consultant who advised the city on this project predicted that parking revenues would average $128 per day. Based on your confidence interval, what do you think of the consultant's prediction? Why?

24. Housing 2008 was a difficult year for the economy. There were a large number of foreclosures of family homes. In one large community, realtors randomly sampled 36 bids from potential buyers to determine the average loss in home value. The sample showed the average loss was $11,560 with a standard deviation of $1500.

a) What assumptions and conditions must be checked before finding a confidence interval? How would you check them?
b) Find a 95% confidence interval for the mean loss in value per home.
c) Interpret this interval and explain what 95% confidence means.
d) Suppose nationally, the average loss in home values at this time was $10,000. Do you think the loss in the sampled community differs significantly from the national average? Explain.

25. Parking, part 2. Suppose that for budget planning purposes the city in Exercise 23 needs a better estimate of the mean daily income from parking fees.

a) Someone suggests that the city use its data to create a 95% confidence interval instead of the 90% interval first created. How would this interval be better for the city? (You need not actually create the new interval.)
b) How would the 95% confidence interval be worse for the planners?
c) How could they achieve a confidence interval estimate that would better serve their planning needs?

26. Housing, part 2. In Exercise 24, we found a 95% confidence interval to estimate the loss in home values.

a) Suppose the standard deviation of the losses was $3000 instead of the $1500 used for that interval. What would the larger standard deviation do to the width of the confidence interval (assuming the same level of confidence)?
b) Your classmate suggests that the margin of error in the interval could be reduced if the confidence level were changed to 90% instead of 95%. Do you agree with this statement? Why or why not?
c) Instead of changing the level of confidence, would it be more statistically appropriate to draw a bigger sample?

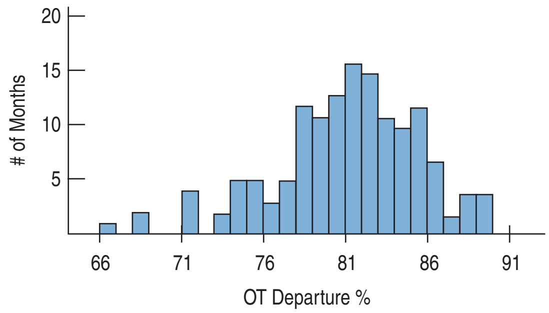
27. State budgets. States that rely on sales tax for revenue to fund education, public safety, and other programs often end up with budget surpluses during economic growth periods (when people spend more on consumer goods) and budget deficits during recessions (when people spend less on consumer goods). Fifty-one small retailers in a state with a growing economy were recently sampled. The sample showed a mean increase of \$2350 in additional sales tax revenue collected per retailer compared to the previous quarter. The sample standard deviation = \$425.

a) Find a 95% confidence interval for the mean increase in sales tax revenue.
b) What assumptions have you made in this inference? Do you think the appropriate conditions have been satisfied?
c) Explain what your interval means and provide an example of what it does not mean.

28. State budgets, part 2. Suppose the state in Exercise 27 sampled 16 small retailers instead of 51, and for the sample of 16, the sample mean increase again equaled \$2350 in additional sales tax revenue collected per retailer compared to the previous quarter. Also assume the sample standard deviation = \$425.

a) What is the standard error of the mean increase in sales tax revenue collected?
b) What happens to the accuracy of the estimate when the interval is constructed using the smaller sample size?
c) Find and interpret a 95% confidence interval.
d) How does the margin of error for the interval constructed in Exercise 27 compare with the margin of error constructed in this exercise? Explain statistically how sample size changes the accuracy of the constructed interval. Which sample would you prefer if you were a state budget planner? Why?

29. Departures. What are the chances your flight will leave on time? The U.S. Bureau of Transportation Statistics of the Department of Transportation publishes information about airline performance. Here are a histogram and summary statistics for the percentage of flights departing on time each month from 1995 through 2006.

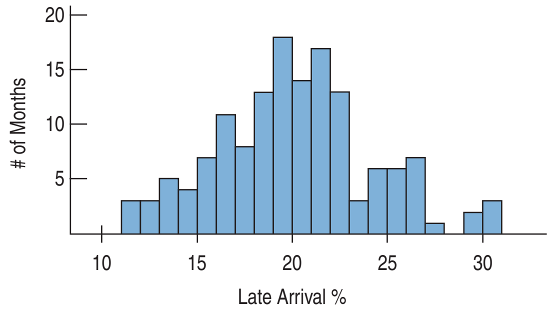


n	144
$\bar{y}$	81.1838
s	4.47094

There is no evidence of a trend over time. (The correlation of On Time Departure% with time is $r = -0.016$.)

a) Check the assumptions and conditions for inference.
b) Find a 90% confidence interval for the true percentage of flights that depart on time.
c) Interpret this interval for a traveler planning to fly.

30. Late arrivals. Will your flight get you to your destination on time? The U.S. Bureau of Transportation Statistics reported the percentage of flights that were late each month from 1995 through 2006. Here's a histogram, along with some summary statistics:



n	144
$\bar{y}$	20.0757
s	4.08837

We can consider these data to be a representative sample of all months. There is no evidence of a time trend.

a) Check the assumptions and conditions for inference about the mean.
b) Find a 99% confidence interval for the true percentage of flights that arrive late.
c) Interpret this interval for a traveler planning to fly.

T **31. Computer lab fees.** The technology committee has stated that the average time spent by students per lab visit has increased, and the increase supports the need for increased lab fees. To substantiate this claim, the committee randomly samples 12 student lab visits and notes the amount of time spent using the computer. The times in minutes are as follows:

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

a) Plot the data. Are any of the observations outliers? Explain.

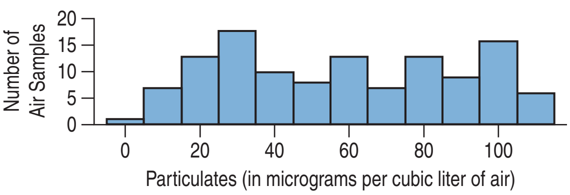
b) The previous mean amount of time spent using the lab computer was 55 minutes. Find a 95% confidence interval for the true mean. What do you conclude about the claim? If there are outliers, find intervals with and without the outliers present.

T **32. Cell phone batteries.** A company that produces cell phones claims its standard phone battery lasts longer on average than other batteries in the market. To support this claim, the company publishes an ad reporting the results of a recent experiment showing that under normal usage, their batteries last at least 35 hours. To investigate this claim, a consumer advocacy group asked the company for the raw data. The company sends the group the following results:

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

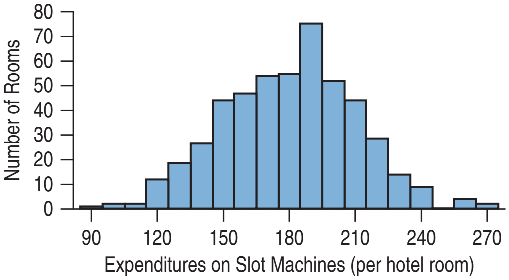
Find a 95% confidence interval and state your conclusion. Explain how you dealt with the outlier, and why.

33. Growth and air pollution. Government officials have difficulty attracting new business to communities with troubled reputations. Nevada has been one of the fastest growing states in the country for a number of years. Accompanying the rapid growth are massive new construction projects. Since Nevada has a dry climate, the construction creates visible dust pollution. High pollution levels may paint a less than attractive picture of the area, and can also result in fines levied by the federal government. As required by government regulation, researchers continually monitor pollution levels. In the most recent test of pollution levels, 121 air samples were collected. The dust particulate levels must be reported to the federal regulatory agencies. In the report sent to the federal agency, it was noted that the mean particulate level = 57.6 micrograms/cubic liter of air, and the 95% confidence interval estimate is (52.06 mg to 63.07 mg). A graph of the distribution of the particulate amounts was also included and is shown below.



a) Discuss the assumptions and conditions for using Student's t inference methods with these data.
b) Do you think the confidence interval noted in the report is valid? Briefly explain why or why not.

34. Convention revenues. At one time, Nevada was the only U.S. state that allowed gambling. Although gambling continues to be one of the major industries in Nevada, the proliferation of legalized gambling in other areas of the country has required state and local governments to look at other growth possibilities. The convention and visitor's authorities in many Nevada cities actively recruit national conventions that bring thousands of visitors to the state. Various demographic and economic data are collected from surveys given to convention attendees. One statistic of interest is the amount visitors spend on slot machine gambling. Nevada often reports the slot machine expenditure as amount spent per hotel guest room. A recent survey of 500 visitors asked how much they spent on gambling. The average expenditure per room was $180.



Casinos will use the information reported in the survey to estimate slot machine expenditure per hotel room. Do you think the estimates produced by the survey will accurately represent expenditures? Explain using the statistics reported and graph shown.

35. Traffic speed. Police departments often try to control traffic speed by placing speed-measuring machines on roads that tell motorists how fast they are driving. Traffic safety experts must determine where machines should be placed. In one recent test, police recorded the average speed clocked by cars driving on one busy street close to an elementary school. For a sample of 25 speeds, it was determined that the average amount over the speed limit for the 25 clocked speeds was 11.6 mph with a standard deviation of 8 mph. The 95% confidence interval estimate for this sample is 8.30 mph to 14.90 mph.

a) What is the margin of error for this problem?
b) The researchers commented that the interval was too wide. Explain specifically what should be done to reduce the margin of error to no more than ± 2 mph.

36. Traffic speed, part 2. The speed-measuring machines must measure accurately to maximize effectiveness in slowing traffic. The accuracy of the machines will be tested before placement on city streets. To ensure that error rates are estimated accurately, the researchers want to take a large enough sample to ensure usable and accurate interval estimates of how much the machines may be off in measuring actual speeds. Specially, the researchers want the margin of error for a single speed measurement to be no more than ± 1.5 mph.

a) Discuss how the researchers may obtain a reasonable estimate of the standard deviation of error in the measured speeds.

b) Suppose the standard deviation for the error in the measured speeds equals 4 mph. At 95% confidence, what sample size should be taken to ensure that the margin of error is no larger than ± 1.0 mph?

37. Tax audits. Certified public accountants are often required to appear with clients if the IRS audits the client's tax return. Some accounting firms give the client an option to pay a fee when the tax return is completed that guarantees tax advice and support from the accountant if the client were audited. The fee is charged up front like an insurance premium and is less than the amount that would be charged if the client were later audited and then decided to ask the firm for assistance during the audit. A large accounting firm is trying to determine what fee to charge for next year's returns. In previous years, the actual mean cost to the firm for attending a client audit session was \$650. To determine if this cost has changed, the firm randomly samples 32 client audit fees. The sample mean audit cost was \$680 with a standard deviation of \$75.

a) Develop a 95% confidence interval estimate for the mean audit cost.
b) Based on your confidence interval, what do you think of the claim that the mean cost has changed?

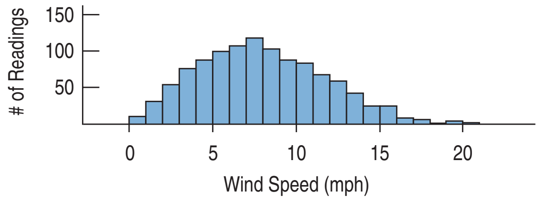
T **38. Tax audits, part 2.** While reviewing the sample of audit fees, a senior accountant for the firm notes that the fee charged by the firm's accountants depends on the complexity of the return. A comparison of actual charges therefore might not provide the information needed to set next year's fees. To better understand the fee structure, the senior accountant requests a new sample that measures the time the accountants spent on the audit. Last year, the average hours charged per client audit was 3.25 hours. A new sample of 10 audit times shows the following times in hours:

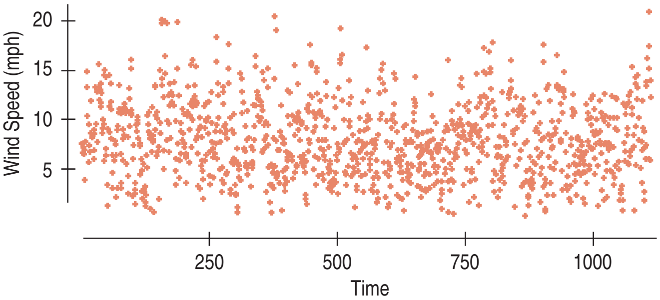
4.2, 3.7, 4.8, 2.9, 3.1, 4.5, 4.2, 4.1, 5.0, 3.4

a) Assume the conditions necessary for inference are met. Find a 90% confidence interval estimate for the mean audit time.
b) Based on your answer to part a, comment on the claim that the mean fees have increased.

T **39. Wind power.** Should you generate electricity with your own personal wind turbine? That depends on whether you have enough wind on your site. To produce enough energy, your site should have an annual average wind speed of at least 8 miles per hour, according to the Wind Energy Association. One candidate site was monitored for a year, with wind speeds recorded every 6 hours. A total of 1114 readings of wind speed averaged 8.019 mph with a standard deviation of 3.813 mph. You've been asked to make a statistical report to help the landowner decide whether to place a wind turbine at this site.

a) Discuss the assumptions and conditions for using Student's t inference methods with these data. Here are some plots that may help you decide whether the methods can be used:

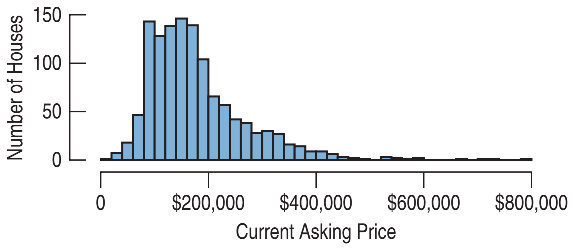




b) What would you tell the landowner about whether this site is suitable for a small wind turbine? Explain

40. Real estate crash? After the sub-prime crisis of late 2007, real estate prices fell almost everywhere in the U.S. In 2006–2007 before the crisis, the average selling price of homes in a region in upstate New York was \$191,300. A real estate agency wants to know how much the prices have fallen since then. They collect a sample of 1231 homes in the region and find the average asking price to be \$178,613.50 with a standard deviation of \$92,701.56. You have been retained by the real estate agency to report on the current situation.

a) Discuss the assumptions and conditions for using t-methods for inference with these data. Here are some plots that may help you decide what to do.



b) What would you report to the real estate agency about the current situation?

Just Checking Answers

1 Questions on the short form are answered by everyone in the population. This is a census, so means or proportions *are* the true population values. The long forms are just given to a sample of the population. When we estimate parameters from a sample, we use a confidence interval to take sample-to-sample variability into account.

2 They don't know the population standard deviation, so they must use the sample SD as an estimate. The additional uncertainty is taken into account by t-models.

3 The margin of error for a confidence interval for a mean depends, in part, on the standard error:

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

Since n is in the denominator, smaller sample sizes generally lead to larger SEs and correspondingly wider intervals. Because long forms are sampled at the same rate of one in every six or seven households throughout the country, samples will be smaller in less populous areas and result in wider confidence intervals.

4 The critical values for t with fewer degrees of freedom would be slightly larger. The $\sqrt{n}$ part of the standard error changes a lot, making the SE much larger. Both would increase the margin of error. The smaller sample is one fourth as large, so the confidence interval would be roughly twice as wide.

5 We expect 95% of such intervals to cover the true value, so 5 of the 100 intervals might be expected to miss.