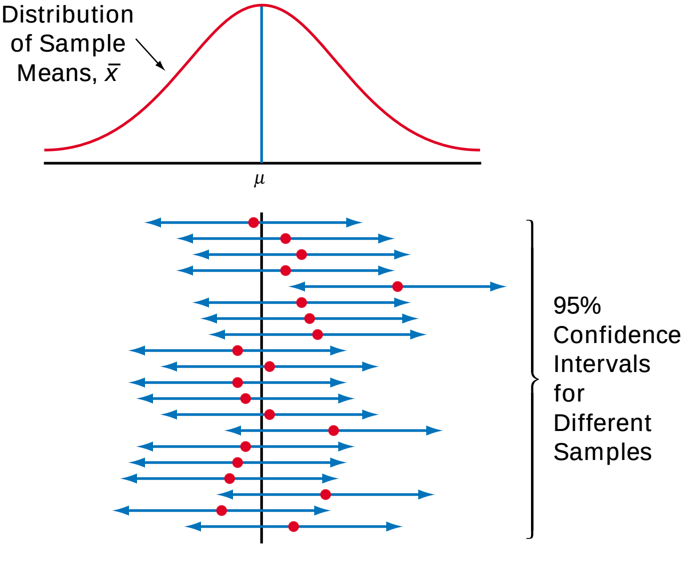
8.6 Confidence Intervals

Governments often commission polls to gauge support for new initiatives. The polling organization surveys a small number of people and estimates support in the entire population based on the sample results. Opinion polls printed in newspapers often include a note such as "These results are accurate to within ±3%, 19 times in 20." In a statement like this, the figure ±3% is a margin of error, and the phrase "19 times in 20" is a confidence level of 0.95 or 95%. The statement means that, if 43% of the sample supported an initiative, there is a 95% probability (you can be 95% confident) that between 40% and 46% of the population supports the initiative. The range 40% to 46% is an example of a confidence interval. The probability of error for this finding is $1 - 0.95 = 0.05$ or 5%.

In such surveys, you do not know the population mean, μ. However, you can determine **confidence intervals**, ranges of values within which μ is likely to fall. These intervals are centered on the sample mean, $\bar{x}$, and their widths depend on the confidence level, $1 - \alpha$. For example, a 95% confidence interval has a 0.95 probability of including μ. In a normal distribution, μ is as likely to lie above the confidence interval as below it, so

$$P\left(\bar{x} - z_{0.975}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{0.975}\frac{\sigma}{\sqrt{n}}\right) = 0.95,$$

where $P(Z < z_{0.975}) = 0.975$
$= 97.5\%$



A $(1 - \alpha)$ or $(1 - \alpha) \times 100\%$ confidence interval for μ, given a population standard deviation σ and a sample of size n with sample mean $\bar{x}$, represents the range of values

$$\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Project Prep

You will need to construct a confidence interval when you complete your probability distributions project.

The table gives a list of common confidence levels and their associated z-scores.

Confidence Level	Tail size, $\frac{\alpha}{2}$	z-score, $z_{\frac{\alpha}{2}}$
90%	0.05	1.645
95%	0.025	1.960
99%	0.005	2.576

Example 1 Drying Times

A paint manufacturer knows from experience that drying times for latex paints have a standard deviation of $\sigma = 10.5$ min. The manufacturer wants to use the slogan: "Dries in T minutes" on its advertising. Twenty test areas of equal size are painted and the mean drying time, $\bar{x}$, is found to be 75.4 min. Find a 95% confidence level for the actual mean drying time of the paint. What would be a reasonable value for T?

Solution 1: Using Pencil and Paper

For a 95% confidence level, the acceptable probability of error, or significance level, is $\alpha = 5\%$, so $z_{\frac{\alpha}{2}} = z_{0.975} = 1.960$. Substituting into the formula gives

$$\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

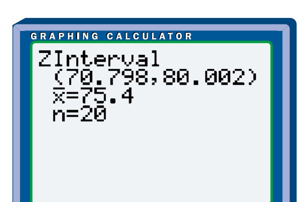
$$75.4 - (1.960)\left(\frac{10.5}{\sqrt{20}}\right) < \mu < 75.4 + (1.960)\left(\frac{10.5}{\sqrt{20}}\right)$$

$$70.8 < \mu < 80.0$$

So, the manufacturer can be 95% confident that the actual mean drying time for the paint lies between 70.8 min and 80.0 min. It would be reasonable to advertise: "Dries in 80 minutes."

Solution 2: Using a Graphing Calculator

Use the Z-Interval instruction in the STAT TESTS menu. Make sure the input is set to Stats. Enter the population parameter $\sigma = 10.5$ and the sample parameters $\bar{x} = 75.4$ and $n = 20$. Set the confidence level to 0.95. Select Calculate and press ENTER to find the required interval.



Often, you want to know the *proportion* of a population that have a particular opinion or characteristic. This proportion is simply p, the probability of success in the binomial distribution. When data are expressed in terms of proportions, the confidence interval formula becomes

$$\hat{p} - z_{\frac{\alpha}{2}}\frac{\sqrt{pq}}{\sqrt{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\frac{\sqrt{pq}}{\sqrt{n}},$$

where $\hat{p}$ is the proportion in the sample. This sample proportion is an estimate of the population proportion just as the sample mean, $\bar{x}$, is an estimate of the population mean, μ.

For many polls, the population proportion is not known. In fact, the purpose of the polls is to estimate this parameter. Since $p \doteq \hat{p}$ and $q \doteq 1 - \hat{p}$, you can *estimate* a confidence interval using the formula

$$\hat{p} - z_{\frac{\alpha}{2}}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

Example 2 Municipal Elections

Voter turnout in municipal elections is often very low. In a recent election, the mayor got 53% of the voters, but only about 1500 voters turned out.

a) Construct a 90% confidence interval for the proportion of people who support the mayor.

b) Comment on any assumptions you have to make for your calculation.

Solution

a) In this case, you want to find a confidence interval for a *proportion* of the population in a binomial distribution. Here, p is the proportion of the population who support the mayor, so you can use the election results to estimate this proportion:

$$\begin{aligned} p &\doteq \hat{p} \\ &= 0.53 \end{aligned} \quad \text{and} \quad \begin{aligned} q &= 1 - p \\ &\doteq 1 - 0.53 \\ &= 0.47 \end{aligned}$$

These estimated values gives a 90% confidence interval of

$$\hat{p} - z_{\frac{\alpha}{2}}\frac{\sqrt{pq}}{\sqrt{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\frac{\sqrt{pq}}{\sqrt{n}}$$

$$0.53 - 1.645\frac{\sqrt{0.53(0.47)}}{\sqrt{1500}} < p < 0.53 + 1.645\frac{\sqrt{0.53(0.47)}}{\sqrt{1500}}$$

$$0.51 < p < 0.55$$

So the mayor can be 90% confident of having the support of between 51% and 55% of the population.

b) You have to assume that the people who voted are representative of whole population. This assumption might not be valid because the people who take the trouble to vote are likely to be the ones most interested in municipal affairs.

Sample Sizes and Margin of Error

Given a sample of size n, from a normal population with standard deviation σ, you can use the sample mean to construct a confidence interval. You can express this confidence interval in terms of its central value and width. For example, suppose a sample of bolts has a 95% confidence interval of 7.51 mm < μ < 7.55 mm for the diameters. You can express this interval as a 95% confident estimate of 7.53 mm ± 0.02 mm for the mean diameter μ. Sometimes, however, a statistician might first decide on the confidence interval width, or **margin of error**, required, and then use this value to calculate the minimum sample size necessary to achieve this width. Opinion polls and other surveys are constructed in this way.

The width of the confidence interval $\bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ is $w = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$. Solving this equation for the sample size, n, gives

$$n = \left(\frac{2z_{\frac{\alpha}{2}}\sigma}{w}\right)^2$$

If the pollsters know (or have a good estimate of) the population standard deviation, σ, they can use this formula to find the sample size they require for a survey to have a specified margin of error.

Example 3 Sample Size for Quality-Control Testing

Suppose the diameters of the bolts mentioned above have a standard deviation of $\sigma = 0.1$. How large a sample would you need to be 90% confident that the mean diameter is 7.53 mm ±0.01 mm?

Solution

For a 90% confidence level, $z_{\frac{\alpha}{2}} = 1.645$. Substituting the known values into the equation for n gives

$$n = \left(\frac{2z_{\frac{\alpha}{2}}\sigma}{w}\right)^2$$
$$= \left(\frac{2(1.645)(0.1)}{(0.02)}\right)^2$$
$$\doteq 271$$

You would need a sample of about 270 bolts.

You can use a similar method to find the sample sizes required for surveys involving population proportions. The margin of error for a proportion is $w = 2z_{\frac{\alpha}{2}}\frac{\sqrt{pq}}{\sqrt{n}}$, so $n = \left(\frac{2z_{\frac{\alpha}{2}}\sqrt{pq}}{w}\right)^2$. This formula simplifies to

$$n = 4pq\left(\frac{z_{\frac{\alpha}{2}}}{w}\right)^2$$

Example 4 Sample Size for a Poll

A recent survey indicated that 82% of secondary-school students graduate within five years of entering grade 9. This result is considered accurate within plus or minus 3%, 19 times in 20. Estimate the sample size in this survey.

Solution

The result describes a confidence interval with a margin of error w = 6%, and the confidence level is "19 times out of 20" or 95%, giving $\alpha = 0.05$.

Here, as in Example 2, you have a binomial distribution with data expressed as proportions. You can use the survey results to estimate p. Since 82% of the students in the survey graduated,

$p \doteq \hat{p}$ and $q = 1 - p$
$= 0.82$ $\quad\doteq 1 - 0.82$
$\quad = 0.18$

Substituting into the formula for n,

$$n = 4pq\left(\frac{z_{\frac{\alpha}{2}}}{w}\right)^2$$
$$\doteq 4(0.82)(0.18)\left(\frac{1.960}{0.06}\right)^2$$
$$\doteq 630$$

So, to obtain the stated level of accuracy and confidence, approximately 630 people would need to be surveyed.

WEB CONNECTION

www.mcgrawhill.ca/links/MDM12

To learn more about confidence intervals, visit the above web site and follow the links.

Key Concepts

- A $P\%$ confidence interval for normally distributed data is given by $\bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$. Here, α is the acceptable probability of error, and $\alpha = (100 - P)\%$.
- A $P\%$ confidence interval for a population proportion, based on binomial data, is given by $\hat{p} \pm z_{\frac{\alpha}{2}}\frac{\sqrt{pq}}{\sqrt{n}}$.
- For a specified margin of error w, the required sample size is $n = \left(\frac{2z_{\frac{\alpha}{2}}\sigma}{w}\right)^2$.

Communicate Your Understanding

1. How does the population distribution affect the distribution of the sample means?
2. a) Why is the z-score for 97.5% used to construct a 95% confidence interval? Support your answer with a sketched distribution.
   b) Given that $z_{0.975} = 1.96$, approximately how many standard deviations wide is a 95% confidence interval?
3. To obtain the desired margin of error, an investigator must sample 2000 people. List at least three possible problems the investigator may encounter.
4. Interpret the following headline using confidence intervals: "4 out of every 13 Canadians think that the government should subsidize professional sports teams in Canada. These results are considered accurate within plus or minus 4%, nine times out of ten."

Practice

A

1. Construct the following confidence intervals.

	Confidence Level	$\bar{x}$	σ	Sample Size
a)	90%	15	3	36
b)	95%	30	10	75
c)	99%	6.4	2.5	60
d)	90%	30.6	8.7	120
e)	95%	41.8	12.6	325
f)	99%	4.25	0.86	44

2. Interpret each of the following statements using confidence intervals.
   a) In a recent survey, 42% of high school graduates indicated that they expected to earn over $100 000 per year by the time they retire. This survey is considered accurate within plus or minus 3%, 19 times in 20.
   b) A survey done by the incumbent MP indicated that 48% of decided voters said they would vote for him again in the next election. The result is considered accurate within plus or minus 5%, nine times in ten.

c) According to a market research firm, 28% of teenagers will purchase the latest CD by the rock band Drench. The result is considered accurate within ±4%, 11 out of 15 times.

Apply, Solve, Communicate

B

3. **Application** A large water pipeline is being constructed to link a town with a fresh water aquifer. A construction supervisor measured the diameters of 40 pipe segments and found that the mean diameter was 25.5 cm. In the past, pipe manufactured by the same company have had a standard deviation of 7 mm. Determine a 95% confidence interval for the mean diameter of the pipe segments.

4. **Application** A study of 55 patients with low-back pain reported that the mean duration of the pain was 17.6 months, with a standard deviation of 5.1 months. Assuming that the duration of this problem is normally distributed in the population, determine a 99% confidence interval for the mean duration of low-back pain in the population.

5. The Statsville school board surveyed 70 parents on the question: "Should school uniforms be instituted at your school?" 28% of the respondents answered "Yes." Construct a 90% confidence interval for the proportion of Statsville parents who want school uniforms.

6. The most popular name for pleasure boats is "Serenity." A survey of 200 000 boat owners found that 12% of their boats were named "Serenity." Determine a 90% confidence interval for the proportion of pleasure boats carrying this name.

7. The football coach wants an estimate of the physical-fitness level of 44 players trying out for the varsity team. He counted the number of sit-ups done in 2 min by each player. Here are the results:

38	95	86	63	68	73	26	43
90	30	71	100	92	57	71	67
47	56	68	61	61	92	83	50
66	51	87	64	80	58	60	103
14	39	88	75	60	87	70	66
95	26	75	61				

Construct a 90% confidence interval for the mean number of sit-ups that varsity football players are capable of performing.

8. The students in a school environment club are concerned that recycling efforts are failing in smaller communities. The amount of waste recycled in 33 towns with under 5000 households is given below:

12%	10%	12%	11%	13%	12%	18%
10%	33%	3%	10%	15%	12%	18%
20%	24%	18%	5%	13%	12%	14%
25%	17%	22%	11%	12%	26%	20%
17%	22%	11%	20%	30%		

a) Construct a 95% confidence interval for the mean percent of waste recycled in towns with under 5000 households.

b) **Communication** Write a letter to the mayors of towns with under 5000 households outlining the results.

9. A market-research firm asked 300 people about their shampoo-purchasing habits. Fifty-five people said they bought S'No Flakes. Determine a 95% confidence interval for the percent of people who purchase S'No Flakes.

10. A manufacturing company wants to estimate the average number of sick days its employees take per year. A pilot study found the standard deviation to be 2.5 days. How large a sample must be taken to obtain an estimate with a maximum error of 0.5 day and a 90% confidence level?

11. An industrial-safety inspector wishes to estimate the average noise level, in decibels (dB), on a factory floor. She knows that the standard deviation is 8 dB. She wants to be 90% confident that the estimate is correct to within ±2 dB. How many noise-level measurements should she take?

12. An ergonomics advisor wants to estimate the percent of computer workers who experience carpal-tunnel syndrome. An initial survey of 50 workers found three cases of the syndrome. To be 99% confident of an accuracy of ±2%, how many workers must the advisor survey?

13. **a)** Obtain a survey result from your local newspaper that contains accuracy information. Determine the sample size. State explicitly any assumptions you needed to make.

b) Try to find a survey result from your local newspaper that gives the sample size. Use this information to estimate the standard deviation for the population surveyed.

Chapter Problem

14. Take 12 random samples of ten data from the earthquake table on page 411. Use these samples to estimate a 90% confidence interval for the mean number of major earthquakes you should expect this year.

15. **Inquiry/Problem Solving** The table below gives a sample of population growth rates of wolves in Algonquin Park.

Year	Population Growth Rate
1989–90	–0.67
1990–91	0.12
1991–92	0.26
1992–93	–0.36
1993–94	–0.13
1994–95	0.24
1995–96	–0.02
1996–97	–0.17
1997–98	0.27
1998–99	–0.65

a) Use these samples to estimate the mean and standard deviation for the growth rate of the wolf population in Algonquin Park. Explain your results.

b) Assuming that the growth rates are normally distributed, estimate the probability that the growth rate for the wolf population is less than zero.

c) A population is in danger of extinction if its population growth rate is –0.05 or less. A study based on these samples claimed that there is a 71% probability that this wolf population is in danger of extinction. Is the study correct?

d) Construct a 90% confidence interval for the true population growth rate of wolves in Algonquin Park.

16. A social scientist wants to estimate the average salary of office managers in a large city. She wants to be 95% confident that her estimate is correct. Assume that the salaries are normally distributed and that $\sigma = \$1050$. How large a sample must she take to obtain the desired information and be accurate within \$200?

17. **Communication** An opinion pollster determines that a sample of 1500 people should give a margin of error of 3% at the 95% confidence level 19 times in 20. The pollster decides that an efficient way to find a representative sample of 1500 people is to conduct the poll at Pearson International Airport. Discuss, in terms of techniques such as stratified sampling, how representative the poll will be.

ACHIEVEMENT CHECK

Knowledge / Understanding	Thinking / Inquiry / Problem Solving	Communication	Application

18. Emilio has played ten rounds of golf at the Statsville course this season. His mean score is 80 and the standard deviation is 4. Assume Emilio's golf scores are normally distributed.

a) Find the 95% confidence interval of Emilio's mean golf score.

b) Predict how the confidence interval would change if the standard deviation of the golf scores was 8 instead of 4. Explain your reasoning.

c) Find the 95% confidence interval of Emilio's mean golf score if the standard deviation was 8 instead of 4. Does the answer support your prediction? Explain.

d) Emilio's most recent golf score at the course is 75. He claims that his game has improved and this latest score should determine whether he qualifies for entry in the Statsville tournament. Should the tournament organizers accept his claim? Justify your answer mathematically.

C

19. Given a sample of size n with mean $\bar{x}$, the population mean μ can be estimated via the 95% confidence interval defined by the probability

$$P\left(\bar{x} - z_{0.975}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{0.975}\frac{\sigma}{\sqrt{n}}\right) = 0.95 \quad (1)$$

If a value μ_0 is assumed for μ, a 5% significance level hypothesis test H_0: $\mu = \mu_0$: against H_1: $\mu \neq \mu_0$ can be performed on the sample mean $\bar{x}$, using the probability

$$P\left(\mu_0 - z_{0.975}\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu_0 + z_{0.975}\frac{\sigma}{\sqrt{n}}\right) = 0.95 \quad (2)$$

a) Show that the probability in equation (1) can be rearranged as

$$P\left(|\bar{x} - \mu| < z_{0.975}\frac{\sigma}{\sqrt{n}}\right)$$

b) Show that the probability in equation (2) can also be rearranged into a similar form.

c) Use parts a) and b) to prove that, if μ_0 is in the 95% confidence interval defined by equation (1), then H_0 will be accepted at the 5% significance level, but if μ_0 is not in this confidence interval, H_0 will be rejected and H_1 accepted.

20. Suppose you are designing a poll on a subject for which you have no information on what people's opinions are likely to be. What sample size should you use to ensure a 90% confidence level that your results are accurate to plus or minus 2%? Explain your reasoning.