Business Statistics

Week 9

Confidence Interval Estimation

Agenda

Time	Activity
10 minutes	Point and Interval Estimate
30 minutes	Confidence Interval Estimate for the Mean (σ Known)
30 minutes	Confidence Interval Estimate for the Mean (σ Unknown)
30 minutes	Confidence Interval Estimate for the Proportion
30 minutes	Determining Sample Size
70 minutes	Exercise

Outline

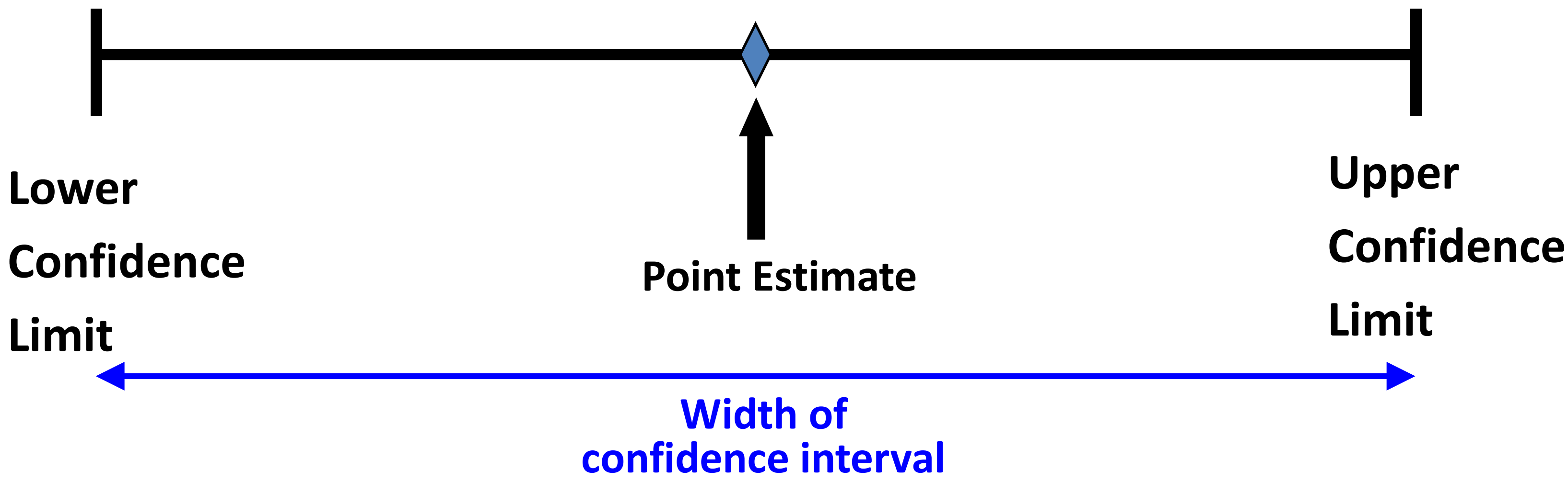
- Point and Interval Estimate
- Confidence Interval Estimate for the Mean (σ Known)
- Confidence Interval Estimate for the Mean (σ Unknown)
- Confidence Interval Estimate for the Proportion
- Determining Sample Size

What for?

Estimation

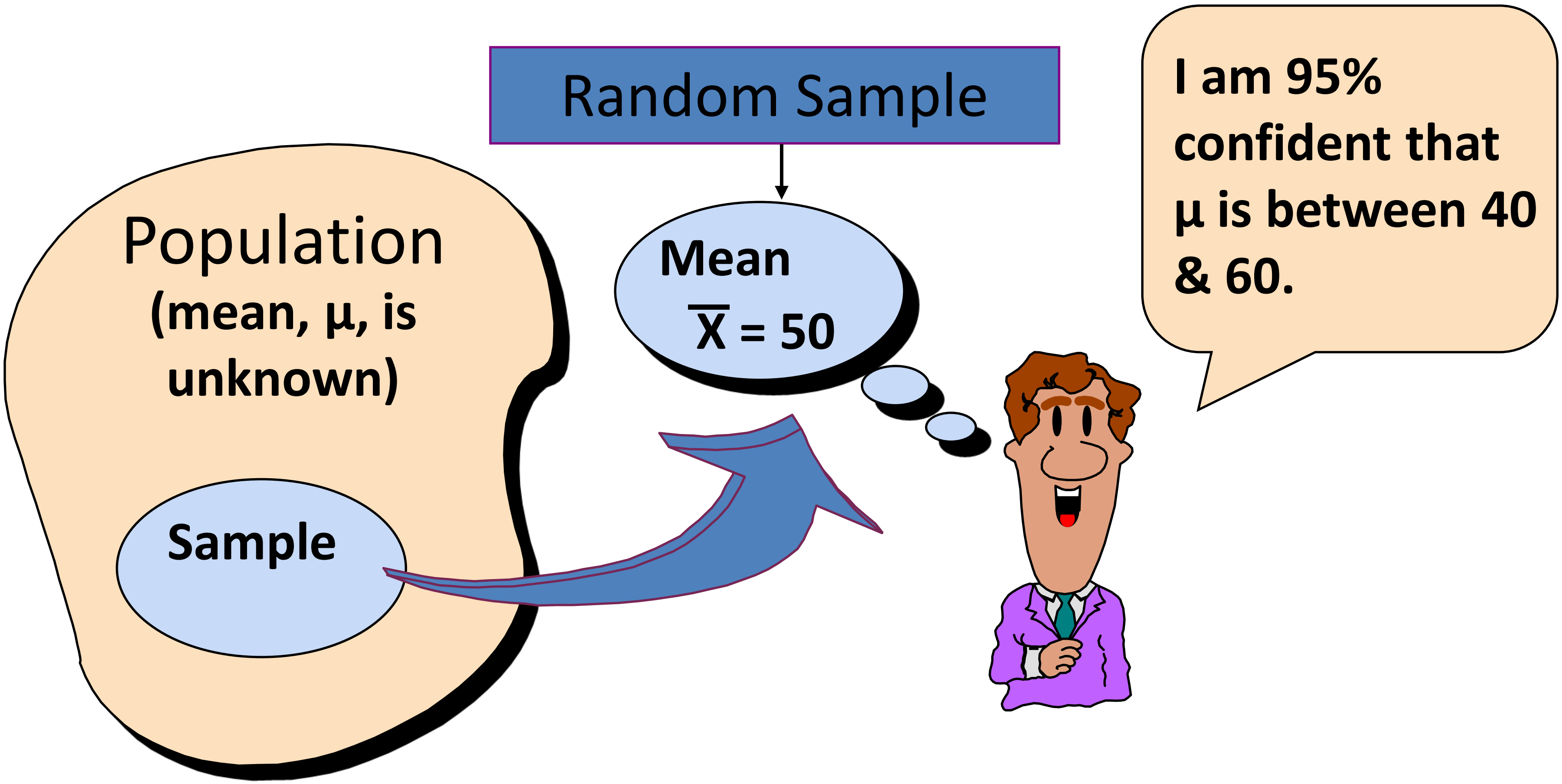
- A **point estimate** is the value of a single sample statistic, such as a sample mean.
- A **confidence interval estimate** is a range of numbers, called an interval, constructed AROUND THE POINT ESTIMATE.

Estimation



We can estimate a Population Parameter …		with a Sample Statistic (a Point Estimate)
Mean	μ	$\bar{X}$
Proportion	π	p

Estimation Process



General Formula

- The general formula for all confidence intervals is:

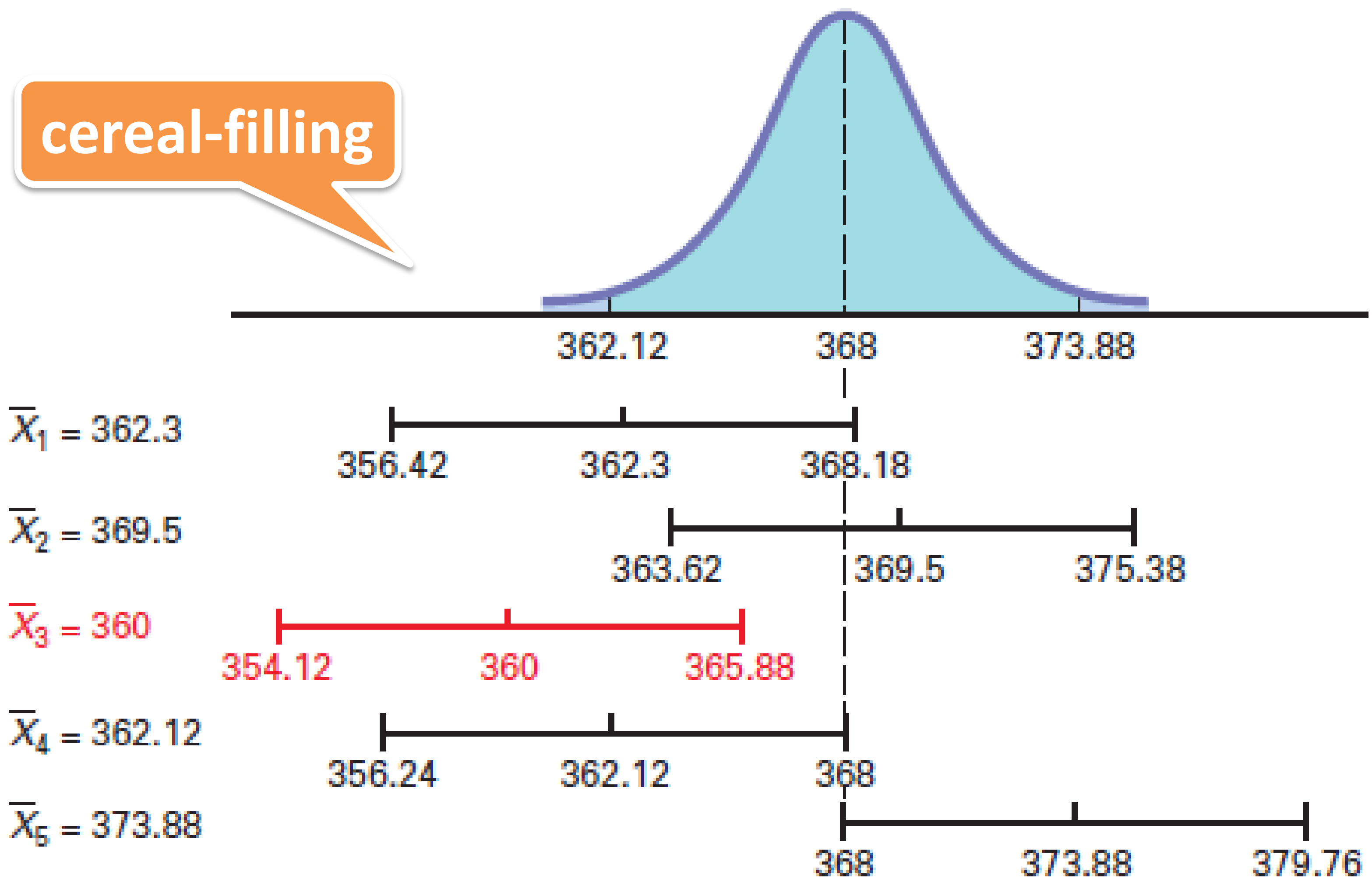
Point Estimate ± (Critical Value)(Standard Error)

Where:

- Point Estimate is the sample statistic estimating the population parameter of interest
- Critical Value is a table value based on the sampling distribution of the point estimate and the desired confidence level
- Standard Error is the standard deviation of the point estimate

CONFIDENCE INTERVAL ESTIMATE FOR THE MEAN (σ KNOWN)

Population Mean vs. Sample Mean



Level of Confidence

$$\left(1-\alpha\right)\times 100\%$$

Where:
α is the proportion in the tails of the distribution that is outside the confidence interval.

Assumptions

- Population standard deviation (σ) is known
- Population is normally distributed
- If population is not normal, use large sample

CONFIDENCE INTERVAL FOR THE MEAN (σ KNOWN)

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

or

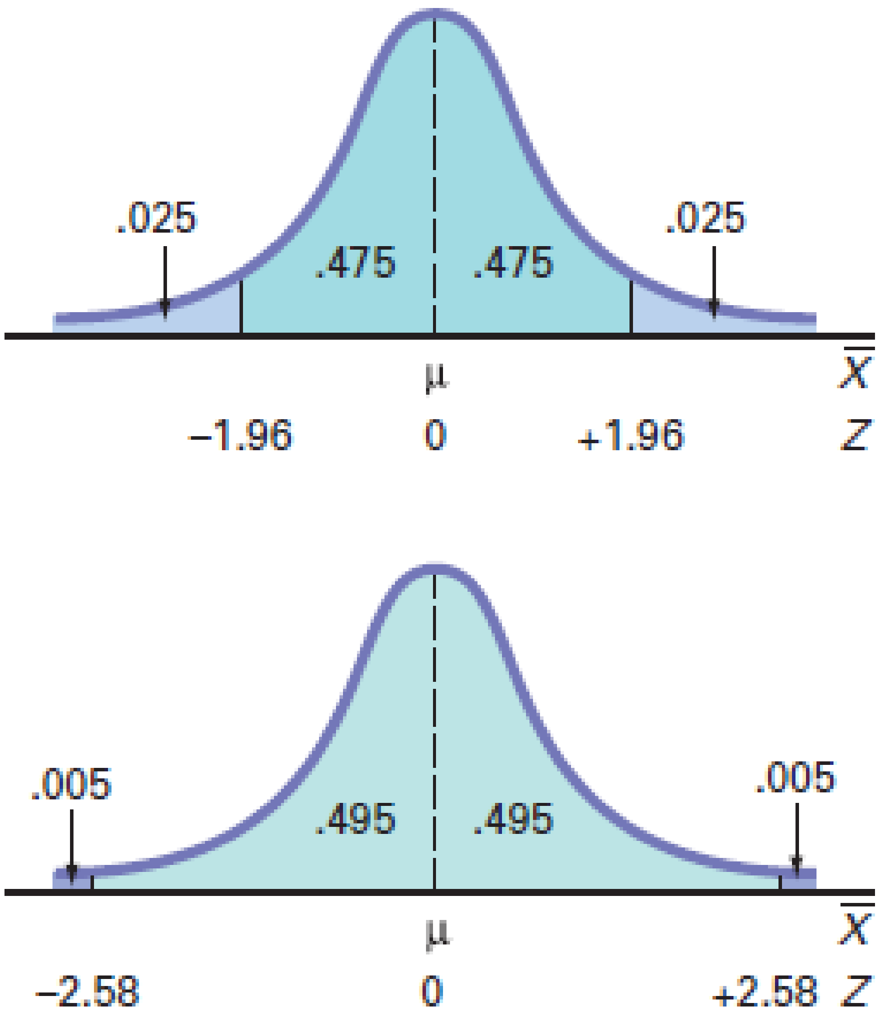
$$\overline{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Critical Value

- The value of $Z_{\alpha/2}$ needed for constructing a confidence interval is called the **critical value** for the distribution.

LOC	α	$\alpha/2$	$Z_{\alpha/2}$
99%	0.01	0.005	2.58
95%	0.05	0.025	1.96
90%	0.1	0.05	1.645

Increasing Level of Confidence



Increasing Level of Confidence

- Any increase in the level of confidence is achieved only by widening (and making less precise) the confidence interval.
- There is no "free lunch" here. You would have more confidence that the population mean is within a broader range of values; however, this might make the interpretation of the confidence interval less useful.

Do You Ever Truly Know σ?

- Probably not!
- In virtually all real world business situations, σ is not known.
- If there is a situation where σ is known then μ is also known (since to calculate σ you need to know μ.)
- If you truly know μ there would be no need to gather a sample to estimate it.

EXAMPLE

Paper Manufacture (cont’d)

A paper manufacturer has a production process that operates continuously throughout an entire production shift. The paper is expected to have a mean length of 11 inches, and the standard deviation of the length is 0.02 inch. At periodic intervals, a sample is selected to determine whether the mean paper length is still equal to 11 inches or whether something has gone wrong in the production process to change the length of the paper produced. You select a random sample of 100 sheets, and the mean paper length is 10.998 inches.

Paper Manufacture

a. Construct a 95% confidence interval estimate for the population mean paper length.
b. Construct a 99% confidence interval estimate for the population mean paper length.

CONFIDENCE INTERVAL ESTIMATE FOR THE MEAN (σ UNKNOWN)

Confidence Interval For The Mean (σ Unknown)

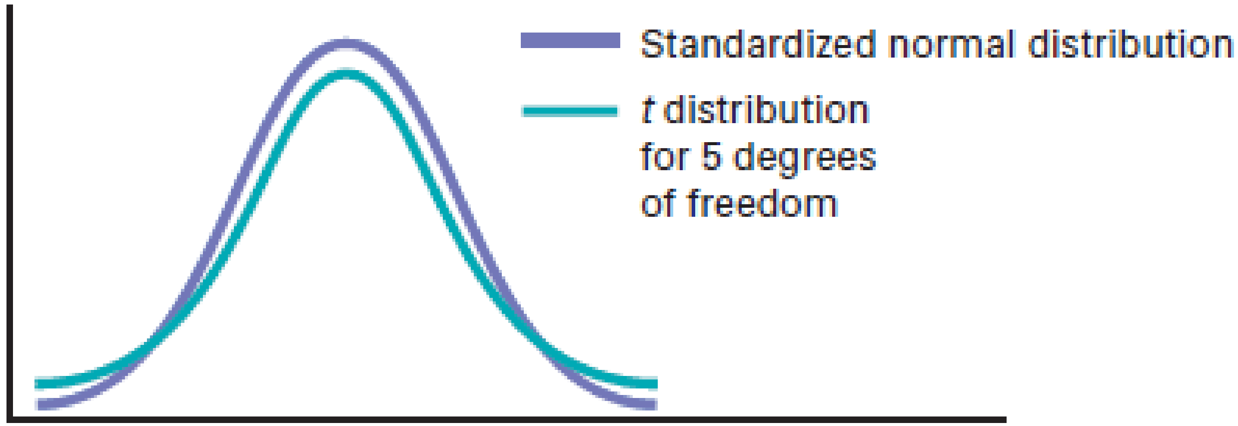
$$\overline{X} \pm t_{\alpha/2,df} \frac{S}{\sqrt{n}}$$

or

$$\overline{X} - t_{\alpha/2,df} S \leq \mu \leq \overline{X} + t_{\alpha/2,df} \frac{S}{\sqrt{n}}$$

Student's t Distribution

- William S. Gosset
- Assumes that the random variable *X is normally distributed*
- The t Table



Student's t Distribution

- The t is a family of distributions
- The $t_{\alpha/2}$ value depends on degrees of freedom (d.f.)
  - Number of observations that are free to vary after sample mean has been calculated

$$d.f. = n - 1$$

Degrees of Freedom (df)

Idea: Number of observations that are free to vary after sample mean has been calculated

Example: Suppose the mean of 3 numbers is 8.0

Let $X_1 = 7$

Let $X_2 = 8$

What is X_3?

If the mean of these three values is 8.0, then X_3 must be 9 (i.e., X_3 is not free to vary)

Here, n = 3, so degrees of freedom = n – 1 = 3 – 1 = 2

(2 values can be any numbers, but the third is not free to vary for a given mean)

EXAMPLE

Electric Insulators (cont'd)

A manufacturing company produces electric insulators. If the insulators break when in use, a short circuit is likely. To test the strength of the insulators, you carry out destructive testing to determine how much force is required to break the insulators. You measure force by observing how many pounds are applied to the insulator before it breaks. You collect the data by selecting 30 insulators to be used in the experiment.

Electric Insulators

You organize the data collected in a worksheet:

1,870	1,728	1,656	1,610	1,634	1,784	1,522	1,696	1,592	1,662
1,866	1,764	1,734	1,662	1,734	1,774	1,550	1,756	1,762	1,866
1,820	1,744	1,788	1,688	1,810	1,752	1,680	1,810	1,652	1,736

To analyze the data, you need to construct a 95% confidence interval estimate for the population mean force required to break the insulator

INTERVAL ESTIMATE FOR THE PROPORTION

Confidence Interval For Proportion

$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

or

$$p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

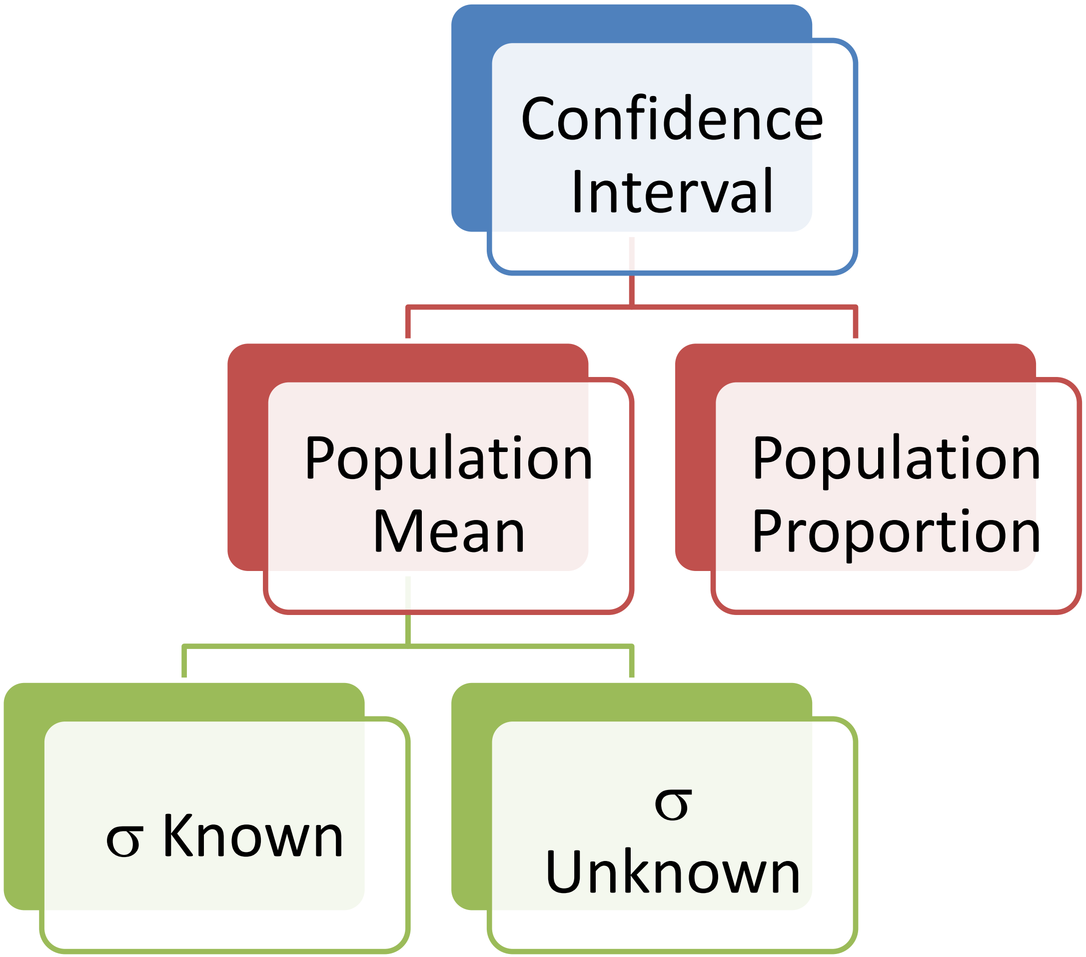
Note: must have np > 5 and n(1-p) > 5

EXAMPLE

Newspaper

The operations manager at a large newspaper wants to estimate the proportion of newspapers printed that have a nonconforming attribute. You collect the data by selecting a random sample n=200 of newspapers from all the newspapers printed during a single day. You organize the results, which show that 35 newspapers contain some type of nonconformance, in a worksheet. To analyze the data, you need to construct and interpret a 90% confidence interval for the proportion of newspapers printed during the day that have a nonconforming attribute.

Summary on Confidence Interval



DETERMINING SAMPLE SIZE

Sample Size For The Mean

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

Sampling Error (e)

Sample Size For The Mean

$$e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$n = \frac{Z_{\alpha/2}^2 \sigma^2}{e^2}$$

Sample Size For The Population

$$e = Z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

$$n = \frac{Z_{\alpha/2}^{2}\pi(1-\pi)}{e^{2}}$$

EXAMPLE

Electric Insulators

Returning to Electric Insulators exercise before, suppose you want to estimate, with 95% confidence, the population mean force required to break the insulator to within ± 25 pounds. On the basis of a study conducted the previous year, you believe that the standard deviation is 100 pounds. Determine the sample size needed.

Office Worker

You want to have 90% confidence of estimating the proportion of office workers who respond to e-mail within an hour to within ± 0.05. Because you have not previously undertaken such a study, there is no information available from past data. Determine the sample size needed.

EXERCISE

8.9 (cont'd)

The manager of a paint supply store wants to estimate the actual amount of paint contained in 1-gallon cans purchased from a nationally known manufacturer. The manufacturer's specifications state that the standard deviation of the amount of paint is equal to 0.02 gallon. A random sample of 50 cans is selected, and the sample mean amount of paint per 1-gallon can is 0.995 gallon.

8.9

a. Construct a 99% confidence interval estimate for the population mean amount of paint included in a 1-gallon can.
b. On the basis of these results, do you think that the manager has a right to complain to the manufacturer? Why?
c. Must you assume that the population amount of paint per can is normally distributed here? Explain.
d. Construct a 95% confidence interval estimate. How does this change your answer to (b)?

8.15

A stationery store wants to estimate the mean retail value of greeting cards that it has in its inventory. A random sample of 100 greeting cards indicates a mean value of \$2.55 and a standard deviation of \$0.44.

a. Assuming a normal distribution, construct a 95% confidence interval estimate for the mean value of all greeting cards in the store's inventory.
b. Suppose there are 2,500 greeting cards in the store's inventory. How are the results in (a) useful in assisting the store owner to estimate the total value of the inventory?

8.16 (cont’d)

Southside Hospital in Bay Shore, New York, commonly conducts stress tests to study the heart muscle after a person has a heart attack. Members of the diagnostic imaging department conducted a quality improvement project with the objective of reducing the turnaround time for stress tests. Turnaround time is defined as the time from when a test is ordered to when the radiologist signs off on the test results. Initially, the mean turnaround time for a stress test was 68 hours..

8.16

After incorporating changes into the stress-test process, the quality improvement team collected a sample of 50 turnaround times. In this sample, the mean turnaround time was 32 hours, with a standard deviation of 9 hours

a. Construct a 95% confidence interval estimate for the population mean turnaround time.
b. Interpret the interval constructed in (a).
c. Do you think the quality improvement project was a success?

8.28 (cont’d)

The telephone company has the business objective of wanting to estimate the proportion of households that would purchase an additional telephone line if it were made available at a substantially reduced installation cost. Data are collected from a random sample of 500 households. The results indicate that 135 of the households would purchase the additional telephone line at a reduced installation cost.

8.28

a. Construct a 99% confidence interval estimate for the population proportion of households that would purchase the additional telephone line.

b. How would the manager in charge of promotional programs concerning residential customers use the results in (a)?

8.38

A survey is planned to determine the mean annual family medical expenses of employees of a large company. The management of the company wishes to be 95% confident that the sample mean is correct to within ±$50 of the population mean annual family medical expenses. A previous study indicates that the standard deviation is approximately $400.

a. How large a sample is necessary?
b. If management wants to be correct to within ±$25, how many employees need to be selected?

8.46 (cont'd)

A survey of 1,000 adults was conducted in March 2009 concerning "green practices." In response to the question of what was the most beneficial thing to do for the environment, 28% said buying renewable energy, 19% said using greener transportation, and 7% said selecting minimal or reduced packaging. Construct a 95% confidence interval estimate of the population proportion of who said that the most beneficial thing to do for the environment was

8.46

a. buy renewable energy.
b. use greener transportation.
c. select minimal or reduced packaging.
d. You have been asked to update the results of this study. Determine the sample size necessary to estimate, with 95% confidence, the population proportions in (a) through (c) to within ±0.02

8.47

In a study of 500 executives, 315 stated that their company informally monitored social networking sites to stay on top of information related to their company.

a. Construct a 95% confidence interval for the proportion of companies that informally monitored social networking sites to stay on top of information related to their company.
b. Interpret the interval constructed in (a).
c. If you wanted to conduct a follow-up study to estimate the population proportion of companies that informally monitored social networking sites to stay on top of information related to their company to within ±0.01 with 95% confidence, how many executives would you survey?

See you next week…

THANK YOU