Confidence Intervals

- Point Estimates
- Confidence Intervals for the Mean
  - The Standard Deviation of a Population is Known
  - The Standard Deviation of a Population is Unknown
- Confidence Intervals for Proportions
- Confidence Intervals for Finite Populations

- Reading: Chapter 8 (except Section 8.5)

Point Estimates

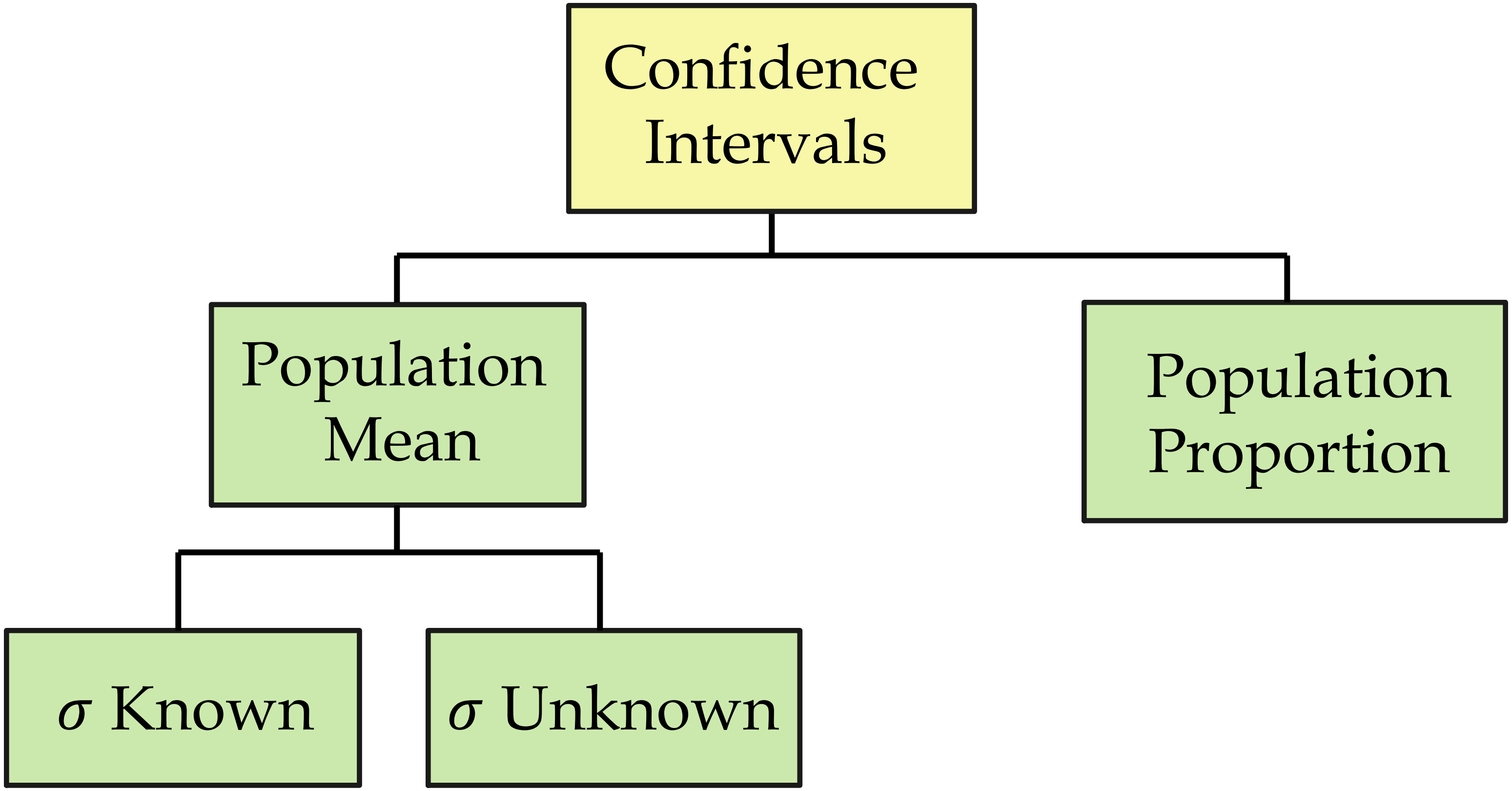
- In point estimation, we use the data from the sample to compute a value of a sample statistic that serves as an estimate of a population parameter

- **Estimator**
  - A *statistic* used to estimate a population parameter
  - For example, $\bar{x}$, the sample mean, is an estimator of the population mean μ
- **Estimate**
  - A *particular value* of the estimator obtained from the sample
  - For example, $\bar{x} = 125$, the mean for a particular sample, is an estimate of the population mean μ

Point Estimates

- A **point estimate** is a single value that best describes the population parameter of interest
- A **point estimator** is a more general term - it's a statistic we use to obtain the point estimate
  - We refer to
    - $\bar{x}$ as the *point estimator* of the population mean μ
    - s is the *point estimator* of the population standard deviation σ
    - $\bar{p}$ is the *point estimator* of the population proportion p
- An **interval estimate** provides a range of values that best describes the population parameter of interest

Confidence Intervals



Confidence Intervals for the Mean

A **confidence interval for the mean** is an interval estimate around a *sample mean* that provides us with a range within which the true population mean is expected to lie

The confidence interval for the mean:

- is computed for a given **confidence level**
  - probability that the interval includes the population mean
- has an *upper* and a *lower confidence limits*, $UCL_{\bar{x}}$ and $LCL_{\bar{x}}$
  - These limits describe the range of values that, with a certain level of confidence, contains the actual population mean

Confidence Intervals for the Mean, σ Known

Assumptions:

- Sampling distribution of $\bar{x}$ is normal
  - The sample size is at least 30 ($n \geq 30$)
  - The population distribution is normal

Confidence Interval for the Mean (σ Known)

$$UCL_{\bar{x}} = \bar{x} + z_{\alpha/2}\sigma_{\bar{x}}$$
$$LCL_{\bar{x}} = \bar{x} - z_{\alpha/2}\sigma_{\bar{x}}$$

where $\bar{x}$ = the sample mean

$z_{\alpha/2}$ = the critical z-score

$\sigma_{\bar{x}}$ = the standard error of the mean

Confidence Intervals for the Mean, σ Known

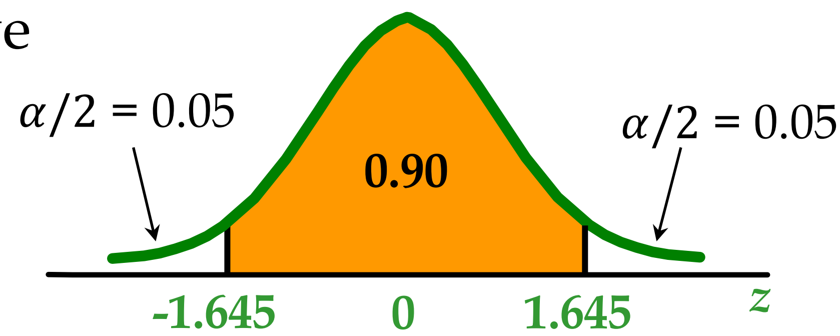
1. $z_{\alpha/2}$ is called the critical *z*-score
2. The variable α is known as the significance level

$$\boldsymbol{\alpha} = \mathbf{1 - Confidence\ Level}$$

3. $z_{\alpha/2}$ is the *z* value which corresponds to the area of $\alpha/2$ in the *right tail* of the standard normal distribution

Example: Suppose we construct the 90% confidence interval. Then, α = 1-0.9 = 0.10. The critical *z*-score $z_{\alpha/2} = z_{0.05}$ = 1.645 corresponds to the area of $\alpha/2$ = 0.05 on the right under the standard normal distribution curve

- *Hint* for calculations in Excel:
  - The area to the left of $z_{\alpha/2}$ is 0.95
  - The area to the left of -$z_{\alpha/2}$ is 0.05



Confidence Intervals for the Mean, σ Known

Example: U.S. consumers are increasingly viewing debit cards as a substitute for cash. The average amount spent annually on a debit card is \$7,790. Assume that this average was based on a sample of 100 consumers and that the population standard deviation is \$500. Construct the 90% confidence interval for the population mean amount spent annually on a debit card.

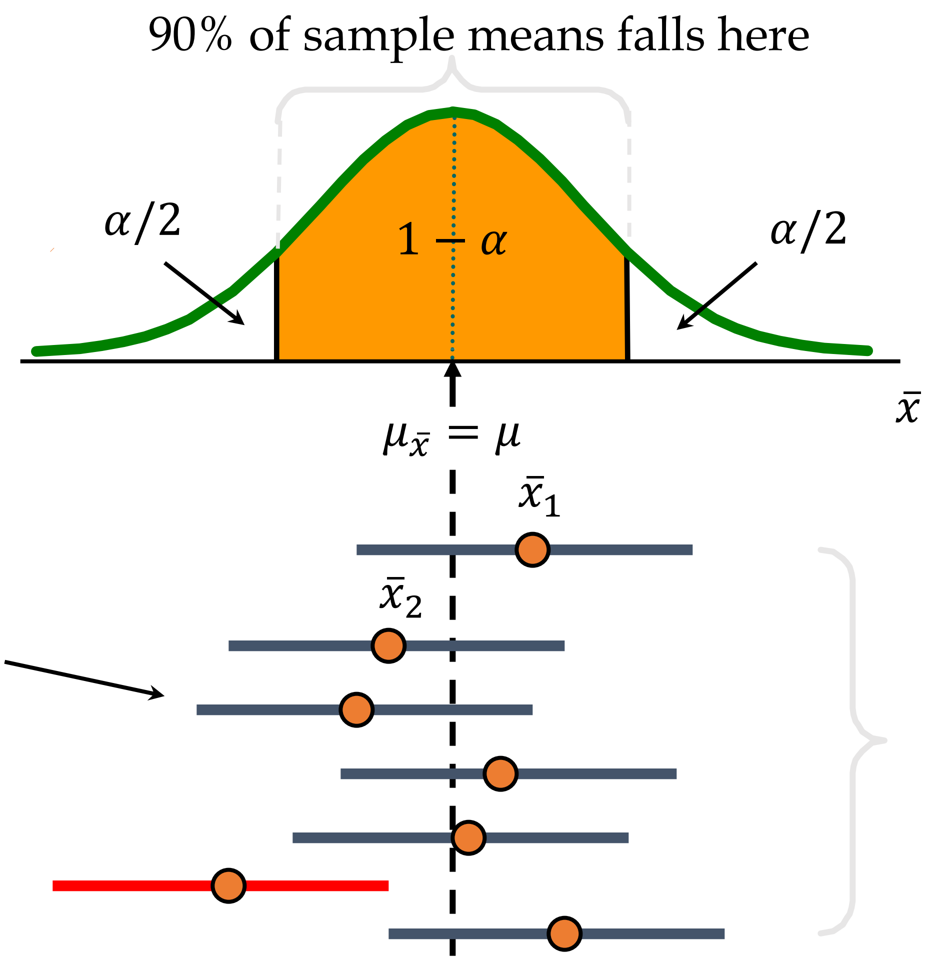
1. Find the standard error of the mean: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{\sqrt{100}} = 50$
2. Find $z_{\alpha/2}$ for the 90% confidence level: $z_{\alpha/2} = z_{0.05} = 1.645$
3. Calculate the interval endpoints:

$$UCL_{\bar{x}} = \bar{x} + z_{\alpha/2}\sigma_{\bar{x}} = 7{,}790 + 1.645 \times 50 = \$7{,}872.24$$

$$LCL_{\bar{x}} = \bar{x} - z_{\alpha/2}\sigma_{\bar{x}} = 7{,}790 - 1.645 \times 50 = \$7{,}707.76$$

Interpreting a Confidence Interval

We are 90% confident that the population mean amount spent annually on a debit card is between \$7,707.76 and \$7,872.24:



Each interval extends from

$$\bar{x} - z_{\alpha/2}\sigma_{\bar{x}}$$

to

$$\bar{x} + z_{\alpha/2}\sigma_{\bar{x}}$$

But $\bar{x}$ varies from sample to sample

For the 90% confidence, 90% of intervals constructed contain μ ;

10% do not

Interpreting a Confidence Interval

For our example:

We are 90% confident that the population mean amount spent annually on a debit card is between \$7,707.76 and \$7,872.24

- Although the population mean may or may not be in this interval, 90% of a large number of sample means drawn from the population will produce confidence intervals which include the population mean
- This makes us 90% confident that the interval we constructed contains the population mean

Caution: Confidence interval does not imply that there is a 90% *probability* that the population mean μ falls within the confidence interval

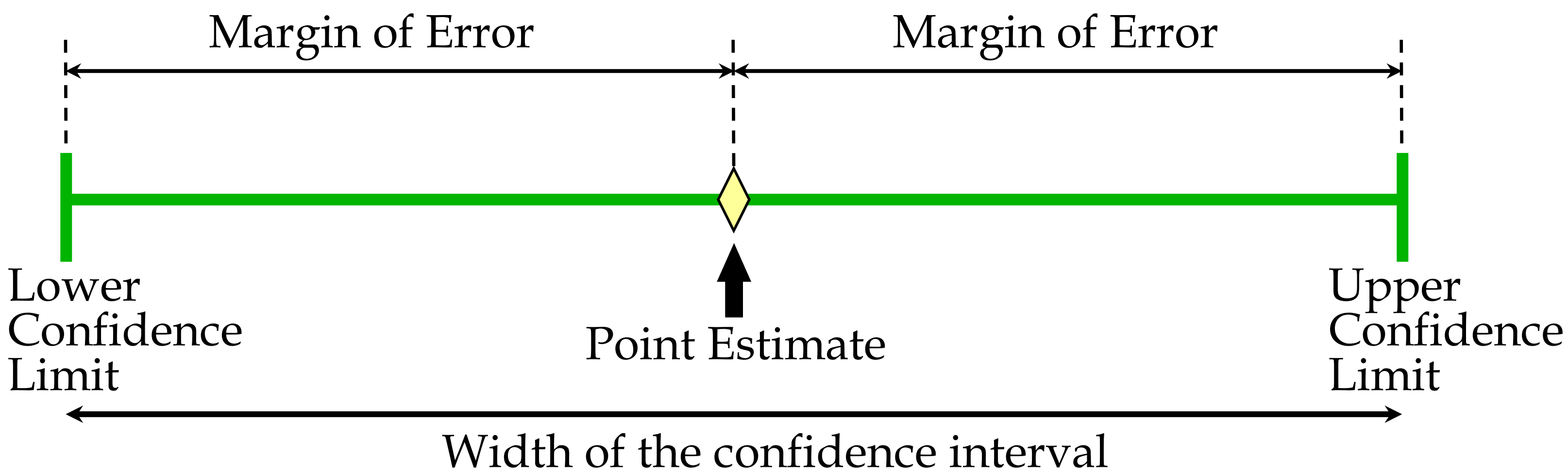
- μ is an unknown but constant number: it is either within the confidence interval or not
- We do not know the value of μ, but this does not make the population mean μ random

Margin of Error

The **Margin of Error** $ME_{\bar{x}}$ is the amount added and subtracted to the point estimate to form the confidence interval:

$$UCL_{\bar{x}}, LCL_{\bar{x}} = \bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$$

$$= \bar{x} \pm ME_{\bar{x}} \quad \Longrightarrow \quad ME_{\bar{x}} = z_{\alpha/2}\sigma_{\bar{x}}$$



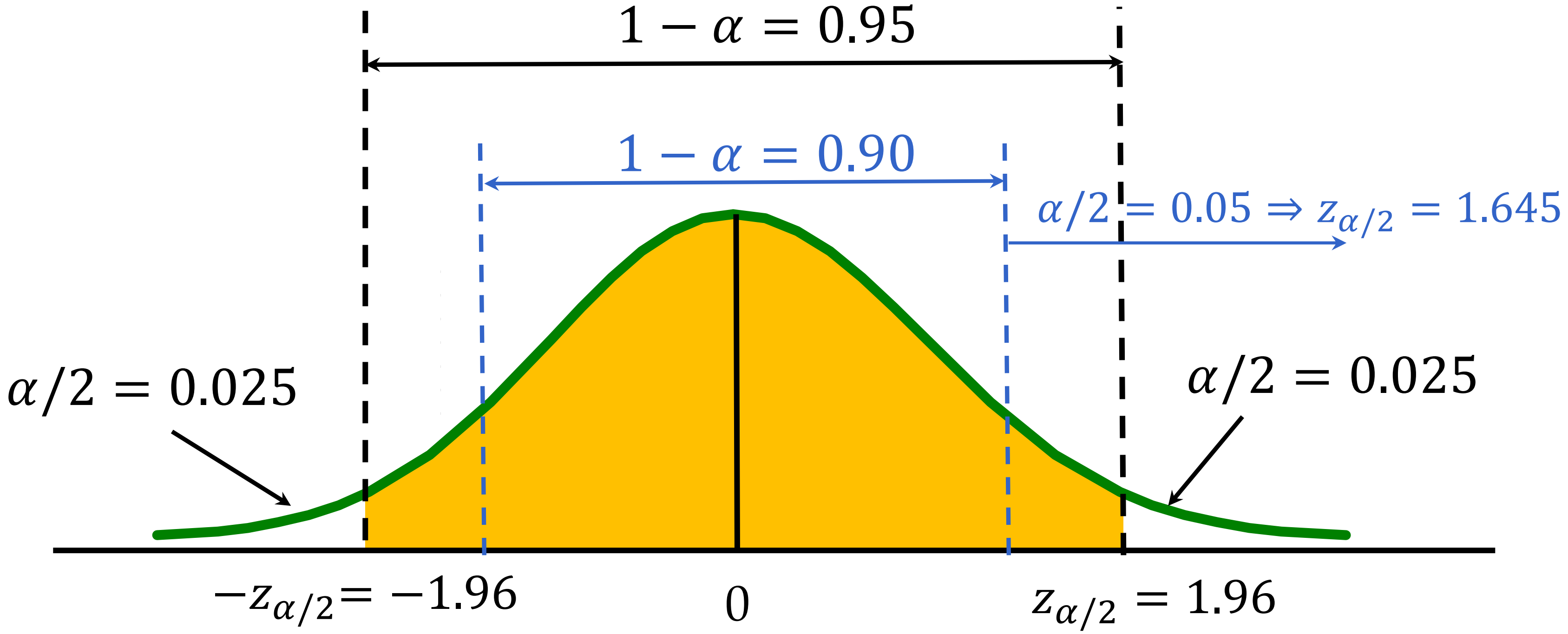
Changing the Sample Size

Increasing the sample size *n* while keeping the confidence level constant *reduces* the margin of error, resulting in a narrower (more precise) confidence interval

$$n \uparrow \ \Rightarrow \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \downarrow \Rightarrow ME_{\bar{x}} = z_{\alpha/2}\sigma_{\bar{x}} \ \downarrow$$

Changing the Confidence Level

Consider an increase in the confidence level from 90 to 95%:



$z = \pm 1.96$ encloses 95% of the area under the curve, with 2.5% in each tail

Changing the Confidence Level

As the confidence level ↑:

- $\Rightarrow$ $\alpha \downarrow$ (α = 1 – Confidence Level)
- $\Rightarrow$ The area on the right of the critical *z*-score ↓
- $\Rightarrow$ The critical *z*-score $z_{\alpha/2} \uparrow$
- $\Rightarrow$ $ME_{\bar{x}} \uparrow$ $(ME_{\bar{x}} = z_{\alpha/2}\sigma_{\bar{x}})$
- $\Rightarrow$ The confidence interval becomes wider when the confidence level increases

Changing the Confidence Level

z-scores for the most commonly used confidence levels are shown in this table:

Confidence Level: $100(1-\alpha)\%$	Significance Level: $100(\alpha)\%$	Critical z-score: $z_{\alpha/2}$
80%	20%	$z_{0.10} = \pm 1.28$
90%	10%	$z_{0.05} = \pm 1.645$
95%	5%	$z_{0.025} = \pm 1.96$
98%	2%	$z_{0.01} = \pm 2.33$
99%	1%	$z_{0.005} = \pm 2.575$

Confidence Intervals for the Mean with Small Samples when σ is Known

When the sample size n is less than 30 and σ is known, the population must be normally distributed to calculate a confidence interval

- With $n < 30$ the Central Limit Theorem cannot be applied, so we can't say the sampling distribution will be approximately normal…
- …but the sampling distribution is always normal (regardless of the sample size) if the population is normally distributed

Confidence Intervals for the Mean with Small Samples, σ is Known

Example: A sample of 25 cereal boxes of Granola Crunch yields a mean weight of 1.02 pounds of cereal per box. Construct a 95% confidence interval of the mean weight of all cereal boxes assuming *the weight is normally distributed* with a population standard deviation of 0.03 pounds.

1. Find the standard error of the mean: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.02}{\sqrt{25}} = 0.006$
2. Find $z_{\alpha/2}$ for the 95% confidence level : $z_{\alpha/2} = z_{0.025} = 1.96$
3. Calculate the interval endpoints:

$$UCL_{\bar{x}} = \bar{x} + z_{\alpha/2}\sigma_{\bar{x}} = 1.02 + 1.96 \times 0.006 = 1.032$$

$$LCL_{\bar{x}} = \bar{x} - z_{\alpha/2}\sigma_{\bar{x}} = 1.02 - 1.96 \times 0.006 = 1.008$$

4. Based on our sample, we are 95% confident that the mean weight of all cereal boxes is between 1.008 and 1.032 lbs.

Excel Exercise >>

Confidence Intervals for the Mean When σ is Unknown

When the population standard deviation σ is unknown, we substitute s, the **sample standard deviation**, in its place to estimate the standard error of the mean

Sample Standard Deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

where: $\bar{x}$ = The sample mean

n = The sample size

$(x_i - \bar{x})$ = The difference between each data value and the sample mean

Student's *t*-distribution

The Estimated Standard Error of the Mean:

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}$$

When the sample standard deviation s is used in place of the population standard deviation σ, the **Student's *t*-distribution** is used in place of the standard normal distribution to find the critical value

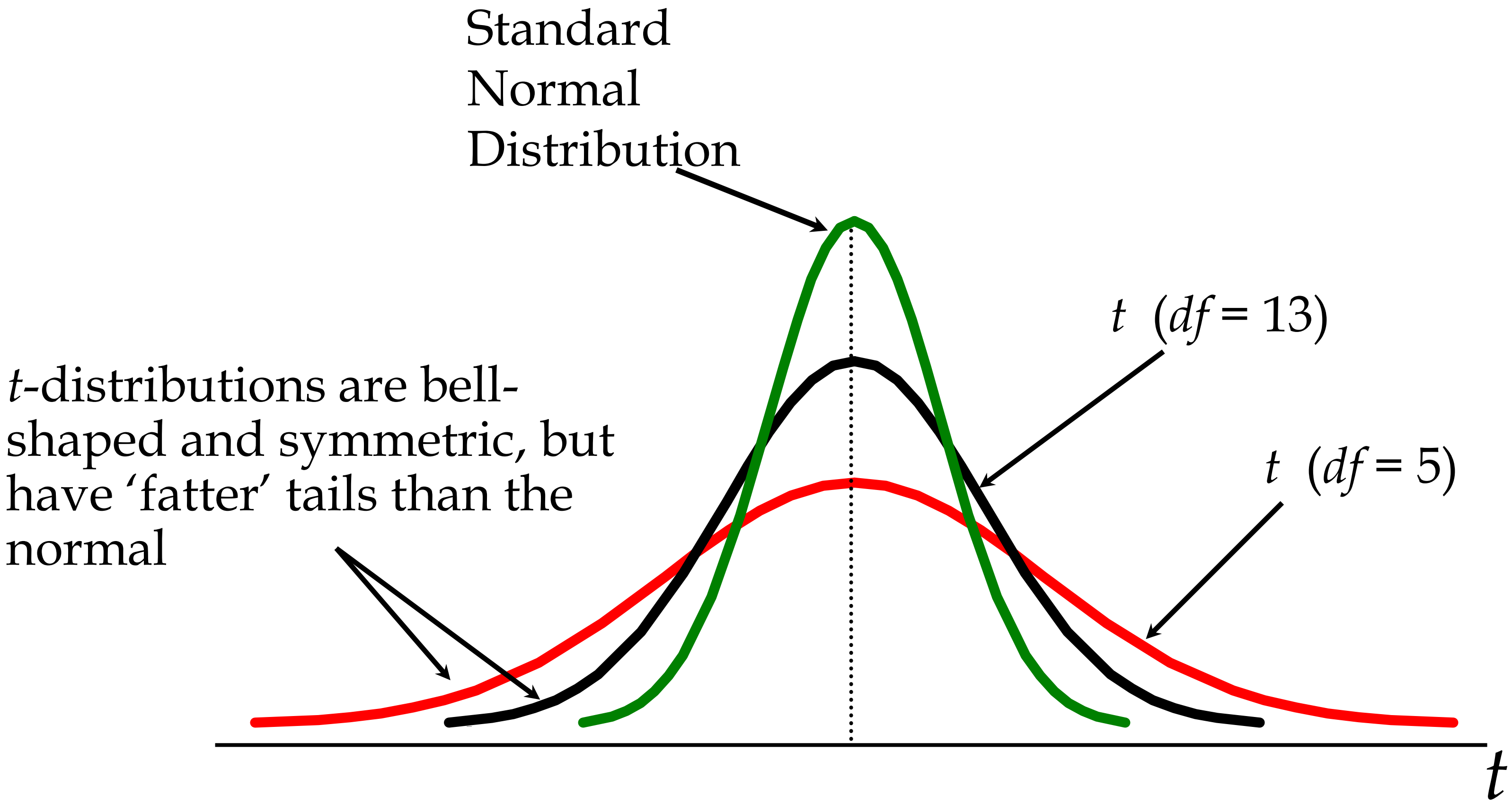
- The population is normally distributed
- Sample size $n \geq 30$

Student's t-distribution

The ***t*-distribution** is a continuous probability distribution with the following properties:

- It is bell-shaped and symmetric around the mean
- The area under the curve is equal to 1.0
- The shape of the curve depends on the **degrees of freedom** (df), $df = n - 1$
- The t-distribution is flatter and wider than the standard normal distribution
  - Therefore, the critical values from the t-distribution are greater than the critical z-scores for the same confidence level

Student's *t*-distribution



Student's *t*-distribution

The *t*-distribution is actually a family of distributions. As the number of degrees of freedom increases, the shape of the *t*-distribution becomes similar to the standard normal distribution

- With more than 100 degrees of freedom (a sample size *n* of more than 100), the two distributions are practically identical

Confidence Intervals for the Mean, σ Unknown

Assumptions:

- The sample size is at least 30 ($n \geq 30$)
- The population distribution is normal

Confidence Interval for the Mean (σ Unknown)

$$UCL_{\bar{x}} = \bar{x} + t_{\alpha/2}\hat{\sigma}_{\bar{x}}$$
$$LCL_{\bar{x}} = \bar{x} - t_{\alpha/2}\hat{\sigma}_{\bar{x}}$$

where $\bar{x}$ = the sample mean

$t_{\alpha/2}$ = the critical t-score

$\hat{\sigma}_{\bar{x}}$ = the estimated standard error of the mean

Confidence Intervals for the Mean, σ Unknown

Example: Suppose you want to estimate the mean mpg of ultra-green cars. From the sample of 25 cars, you obtain the mean of 96.2 mpg and the standard deviation of 10.7 mpg. Construct a 90% confidence interval for the population mean assuming that mpg follows a normal distribution.

1. The estimated standard error of the mean: $\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{10.7}{\sqrt{25}} = 2.14$
2. Find $t_{\alpha/2}$ for *df* = 25 – 1 = 24: $t_{\alpha/2} = t_{0.05} = 1.711$
3. Calculate the interval endpoints:

$$UCL_{\bar{x}} = \bar{x} + t_{\alpha/2}\hat{\sigma}_{\bar{x}} = 96.2 + 1.711 \times 2.14 = 99.86$$

$$LCL_{\bar{x}} = \bar{x} - t_{\alpha/2}\hat{\sigma}_{\bar{x}} = 96.2 - 1.711 \times 2.14 = 92.54$$

4. With 90% confidence, we can conclude that the mean mpg of ultra-green cars is between 92.54 mpg and 99.68 mpg

Excel Exercise >>

Summary

	σ is Known	σ is Unknown
Population (x) is normally distributed	$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$	$\bar{x} \pm t_{\alpha/2}\hat{\sigma}_{\bar{x}}$
Population (x) is NOT normally distributed, but Sample is large ($n \geq 30$)	$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$	$\bar{x} \pm t_{\alpha/2}\hat{\sigma}_{\bar{x}}$
Population (x) is NOT normally distributed, and Sample is small ($n < 30$)	Cannot be determined	Cannot be determined

Confidence Intervals for Proportions

The **confidence interval for the proportion** is an interval estimate around a *sample proportion* that provides us with a range of values within which the true population proportion is expected to lie

Sample Proportion:

$$\bar{p} = \frac{x}{n}$$

Standard Error of the Proportion:

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

where p = population proportion

x = the number of observations of interest

n = the sample size

Confidence Intervals for Proportions

Since the population proportion p is unknown, it is estimated using the sample proportion, $\bar{p}$

The Estimated Standard Error of the Proportion:

$$\hat{\sigma}_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

where $\bar{p}$ = the sample proportion

n = the sample size

Confidence Intervals for Proportions

Proportion data follow the binomial distribution, which can be approximated by the normal distribution under the following conditions:

$$np \geq 5 \quad \text{and} \quad n(1 - p) \geq 5$$

where p = the population proportion

n = the sample size

- p is not known $\Rightarrow$ Use $\bar{p}$ to check these conditions

Confidence Intervals for Proportions

Confidence Interval for a Proportion:

$$UCL_p = \bar{p} + z_{\alpha/2}\hat{\sigma}_{\bar{p}}$$
$$LCL_p = \bar{p} - z_{\alpha/2}\hat{\sigma}_{\bar{p}}$$

where $\bar{p}$ = the sample proportion
$z_{\alpha/2}$ = the critical *z*-score
$\hat{\sigma}_{\bar{p}}$ = estimated standard error of the proportion

Margin of Error for a Confidence Interval for the Proportion

$$UCL_p, LCL_p = \bar{p} \pm z_{\alpha/2}\hat{\sigma}_{\bar{p}}$$

$$= \bar{p} \pm ME_p \quad \Rightarrow \quad ME_p = z_{\alpha/2}\hat{\sigma}_{\bar{p}}$$

Confidence Intervals for Proportions

Example: From a random sample of U.S. citizens, 22 of 100 people are found to have blue eyes. Calculate a 98% confidence interval for the population proportion of blue eyes for U.S. citizens.

1. Calculate the sample proportion and the approximate standard error of the proportion:

$$\bar{p} = \frac{x}{n} = \frac{22}{100} = 0.22 \qquad \hat{\sigma}_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{0.22\,(1-0.22)}{100}} = 0.041$$

2. Find $z_{\alpha/2}$ for 98% confidence level: $z_{\alpha/2} = z_{0.01} = 2.33$

3. Calculate the interval endpoints:

$$UCL_p = \bar{p} + z_{\alpha/2}\hat{\sigma}_{\bar{p}} = 0.22 + 2.33 \times 0.041 = 0.316$$

$$LCL_p = \bar{p} - z_{\alpha/2}\hat{\sigma}_{\bar{p}} = 0.22 - 2.33 \times 0.041 = 0.124$$

Excel Exercise >>

Confidence Intervals for Finite Populations

When determining a confidence interval with a finite population, we need to adjust the standard error by using the **finite population correction factor**

$$\sqrt{\frac{N-n}{N-1}}$$

Confidence Intervals for Finite Populations

Confidence Interval for the Mean of a Finite Population (σ Known)

$$UCL_{\bar{x}} = \bar{x} + z_{\alpha/2}\sigma_{\bar{x}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

$$LCL_{\bar{x}} = \bar{x} - z_{\alpha/2}\sigma_{\bar{x}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

where $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Confidence Intervals for Finite Populations

Confidence Interval for the Mean of a Finite Population (σ Unknown)

$$UCL_{\bar{x}} = \bar{x} + t_{\alpha/2}\hat{\sigma}_{\bar{x}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

$$LCL_{\bar{x}} = \bar{x} - t_{\alpha/2}\hat{\sigma}_{\bar{x}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

where $\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}$

Confidence Intervals for Finite Populations

Confidence Interval for a Proportion of a Finite Population

$$UCL_p = \bar{p} + z_{\alpha/2}\hat{\sigma}_{\bar{p}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

$$LCL_p = \bar{p} - z_{\alpha/2}\hat{\sigma}_{\bar{p}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

where $\hat{\sigma}_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$

Excel Exercise >>

Excel Time: Exercise 8.9

Banking fees have received much attention during the recent economic recession as banks look for ways to recover from crisis. A sample of 30 customers paid an average fee of $12.55 per month on their interest-bearing checking accounts. Assume the population standard deviation is $1.75.

1. What is the margin of error for the 95% confidence interval for the average fee in the population?
   - Use NORM.INV() or NORM.S.INV()
   - Use CONFIDENCE.NORM()
2. Construct a 95% confidence interval for the average fee.
3. How would the confidence interval change if the confidence level increased, for example, to 98%?

Back <<

Excel Time: Margin of Error (σ Known)

Excel's **CONFIDENCE.NORM()** function calculates the margin of error for confidence intervals assuming infinite population:

=CONFIDENCE.NORM (alpha, standard_dev, size)

where:

alpha (α) = The significance level of the confidence interval

standard_dev (σ) = The standard deviation of the population

size (n) = Sample size

Alternatively, one can calculate the margin of error step-by-step and call NORM.INV() or NORM.S.INV() to find $z_{\alpha/2}$

Excel Time: Finding the Critical *z*-score

Recall that NORM.INV() and NORM.S.INV() calculate inverse of the normal cumulative probability and may result in positive or negative values

- These functions calculate *x* or *z* value for a given area under the probability density function on the left

At the same time, the critical *z*-score is a **positive** number by definition

⇒ One way to find the critical *z*-score with NORM.INV() and NORM.S.INV() in Excel:

1. Specify the area on the left of the negative of the critical value (use mean=0 and SD=1). Remember to divide the significance level by 2
2. Since the standard normal distribution is symmetric around zero, take an absolute value of the calculated number
3. For example, =ABS(NORM.INV(0.025,0,1)) would give the critical *z*-score for 95% confidence interval

Excel Time: Confidence Intervals for the Mean (σ Known)

This picture provides an example of how CONFIDENCE.NORM() and NORM.INV()/ NORM.S.INV() functions can be used to calculate the margin of error:

	A	B	C
1	**Exercise 8.9**		
2	**Given**		
3	x_bar	12.55	
4	Population SD	1.75	
5	Sample Size, *n*	30	
6	Confidence Level	0.95	
7			
8	**Solution**		
9	Siginificance Level, alpha	0.05	=1-B6
10	SE	0.3195	=B4/SQRT(B5)
11	Critical *z* -score	1.9600	=ABS(NORM.INV(B9/2,0,1))
12	ME	0.6262	=B11*B10
13	ME (using Excel function)	0.6262	=CONFIDENCE.NORM(B9,B4,B5)
14	UCL	13.1762	=B3+B12
15	LCL	11.9238	=B3-B12

Notes:

1. The significance level α is divided by 2 inside NORM.INV();
2. The critical z-score is a positive value $\Rightarrow$ Absolute value of the result returned by the NORM.INV() is taken (ABS() function)

Excel Time: Exercise 8.26

The following data shows the number of hours per day 12 adults spent in front of screens watching television-related content:

1.3 4.9 4.2 4.8 7.6 6.9 5.4 2.2 5.3 1.8 2.4 8.3

1. What assumptions need to be made about this population to estimate the confidence interval for the average number of hours per day adults spend in front of screens watching TV?
2. Construct a 95% confidence interval for the average number of hours/day adults watch TV:
   - Use T.INV.2T()
   - Use CONFIDENCE.T()

Back <<

Excel Time: Critical *t*-score

Excel's **T.INV.2T()** function returns the critical *t*-score for a *specified significance level* (i.e. area in both tails) and has the following characteristics:

=T.INV.2T (alpha, degrees of freedom)

where:

alpha (α) = The significance level for the confidence interval

degrees of freedom = n - 1

n = Sample size

In Exercise 8.26:

	A	B	C
1	$t_{0.025,11}$	2.2010	=T.INV.2T(0.05,11)

Excel Time: Margin of Error (σ Unknown)

Excel's function **CONFIDENCE.T()** provides the margin of error for a confidence interval for the mean when the population standard deviation is **unknown** assuming finite population

= CONFIDENCE.T(alpha, standard_dev, size)

where:

alpha (α) = The significance level for the confidence interval
standard_dev = The standard deviation of the sample
size = Sample size n

Excel Time: Exercise 8.34

The IRS reported that 197 tax returns were filed electronically in a random sample of 225 tax returns in 2018.

Construct a 95% confidence interval for the proportion of taxpayers who filed electronically in 2018:

- Use NORM.INV() or NORM.S.INV()

Back <<

Excel Time: Exercise 8.53

The University of Delaware would like to estimate the proportion of fans who purchase concessions at the first basketball game of the new season. The basketball facility has a capacity of 3,500 people and is routinely sold out. It was discovered that a total of 260 fans of a random sample of 400 purchased concessions during the game.

Construct a 97% confidence interval for the actual proportion of fans who purchased concessions during the game.

- Use NORM.INV() or NORM.S.INV()

Back <<

Excel Time: Exercise 8.75 (Extra Practice)

A local church has a congregation composed of 362 members. The church's administration would like to find out how the congregation would react if the church moved its 11:00AM service to 10:00AM. A random sample of 115 members was drawn, 46 of whom were if favor of the change.

Construct a 98% confidence interval for the actual proportion of members in favor of the change.

Excel Time: Exercise 8.62 (Extra Practice)

Aetna, a major health insurance company, would like to estimate the average wait time for a patient seeking emergency room services. A random sample of 33 emergency room patients had an average wait time of 222 minutes with a sample standard deviation of 76 minutes. Determine the confidence interval to estimate the average wait time for the following confidence levels:

- 90%
- 95%
- 99%
- How does the confidence level impact the precision of the estimate?