Chapter 8
Interval Estimation

- Population Mean: σ Known
- Population Mean: σ Unknown
- Determining the Sample Size
- Population Proportion
- Big data and Interval estimation

Margin of Error and the Interval Estimate (1 of 2)

- A point estimator cannot be expected to provide the exact value of the population parameter.
- An interval estimate can be computed by adding and subtracting a margin of error to the point estimate.

$$\text{Point Estimate } +/- \text{ Margin of Error}$$

- The purpose of an interval estimate is to provide information about how close the point estimate is to the value of the parameter.

Margin of Error and the Interval Estimate (2 of 2)

The general form of an interval estimate of a population mean is

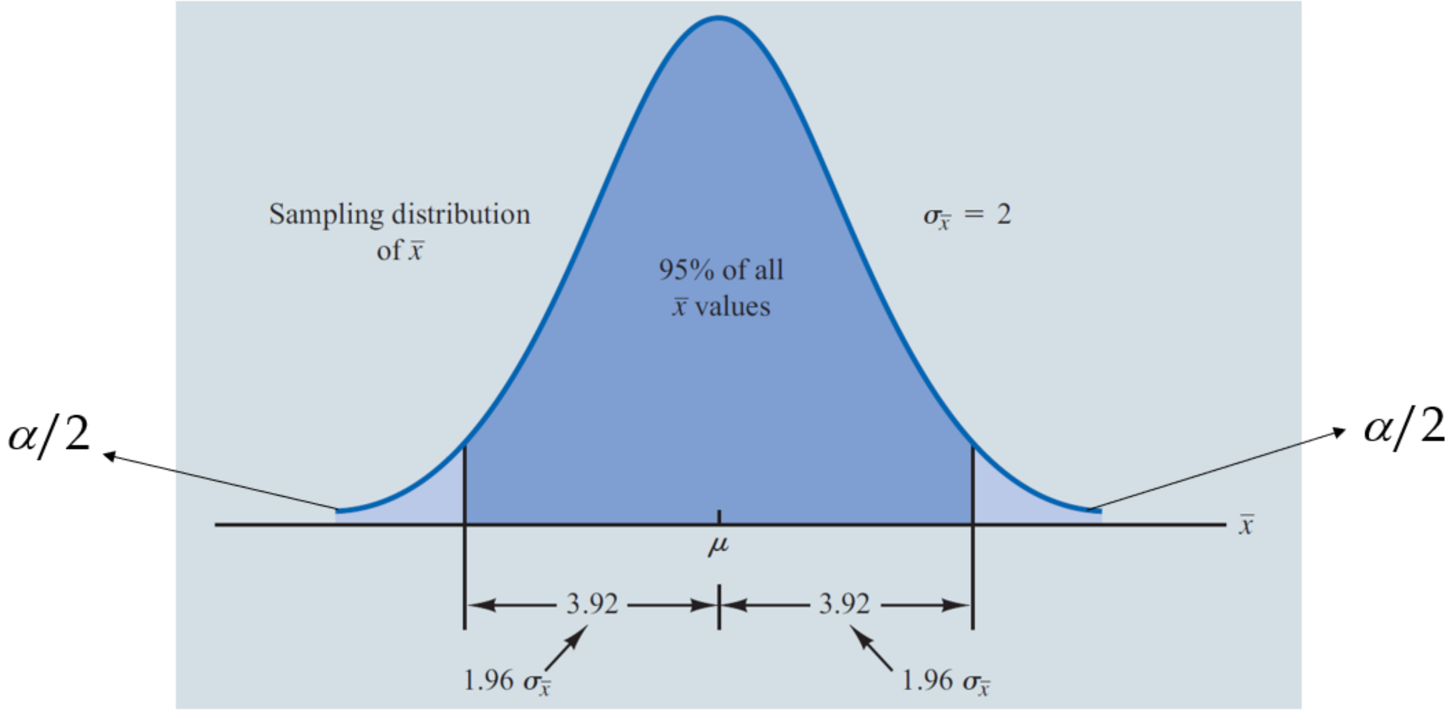
$$\bar{x} \pm \text{Margin of Error}$$

Interval Estimate of a Population Mean: σ Known (1 of 11)

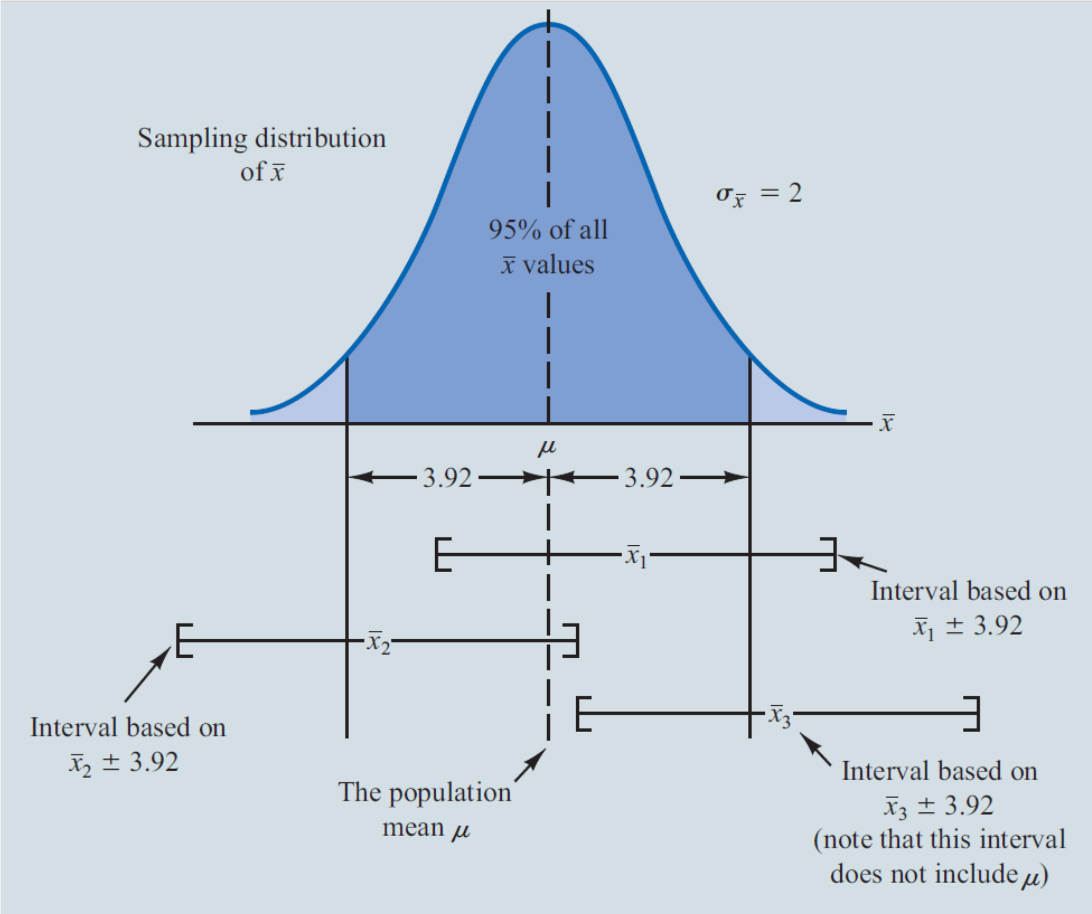
- In order to develop an interval estimate of a population mean, the margin of error must be computed using either:
  - the population standard deviation σ, or
  - the sample standard deviation s
- σ is rarely known exactly, but often a good estimate can be obtained based on historical data or other information.
- We refer to such cases as the <u>σ known</u> case.

Interval Estimate of a Population Mean: σ Known (2 of 11)

There is a $1-\alpha$ probability that the value of a sample mean will provide a margin of error of $z_{\alpha/2}\sigma_{\bar{x}}$ or less.



Interval Estimate of a Population Mean: σ Known (3 of 11)



Interval Estimate of a Population Mean: σ Known (4 of 11)

Interval Estimate of μ

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where: $\bar{x}$ is the sample mean

$1-\alpha$ is the confidence coefficient

$z_{\alpha/2}$ is the z value providing an area of $\alpha/2$ in the upper tail of the standard normal probability distribution

σ is the population standard deviation

n is the sample size

Interval Estimate of a Population Mean: σ Known (5 of 11)

Values of $z_{\alpha/2}$ for the Most Commonly Used Confidence Levels

Confidence level	α	$\alpha/2$	$z_{\alpha/2}$
90%	.1	.05	1.645
95%	.05	.025	1.960
99%	.01	.005	2.576

Meaning of Confidence

- Because 90% of all the intervals constructed using $\bar{x} \pm 1.645\sigma_{\bar{x}}$ will contain the population mean, we say we are 90% confident that the interval $\bar{x} \pm 1.645\sigma_{\bar{x}}$ includes the population mean μ.
- We say that this interval has been established at the 90% confidence level.
- The value .90 is referred to as the confidence coefficient.

Interval Estimate of a Population Mean: σ Known (6 of 11)

<u>Example</u>: Lloyds Department store

Each week Lloyds department store selects a simple random sample of 100 customers in order to learn about the amount spent per shopping trip. The historical data indicates that the population follows a normal distribution.

During most recent week, Lloyd's surveyed 100 customers (n = 100) and obtained a sample mean of $\bar{x} = \$82$.

Based on historical data, Lloyd's now assumes a known value of $\sigma = \$20$.

The confidence coefficient to be used in the interval estimate is .95.

Interval Estimate of a Population Mean: σ Known (7 of 11)

Example: Lloyds Department store

95% of the sample means that can be observed are within $\pm 1.96\sigma_{\bar{x}}$ of the population mean μ. The margin of error is:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.96\left(\frac{20}{\sqrt{100}}\right) = 3.92$$

Interval Estimate of a Population Mean: σ Known (8 of 11)

<u>Example</u>: Lloyds Department store

Interval estimate of μ is:

$$\$82 \pm \$3.92$$

or

$$\$78.08 \text{ to } \$85.29$$

We are <u>95% confident</u> that the interval contains the population mean.

Using Excel to construct a confidence interval - *σ* Known (1 of 2)

- Excel Formula Worksheet

	A	B	C	D
1	Amount		Interval Estimate of a Population Mean:	
2	72		σ Known Case	
3	91			
4	74		Sample Size	=COUNT(A1:A101)
5	115		Sample Mean	=AVERAGE(A1:A101)
6	71			
7	120		Population Standard Deviation	20
8	37		Confidence Coefficient	0.95
9	96		Level of Significance	=1-D8
10	91			
11	105		Margin of Error	=CONFIDENCE.NORM(D9,D7,D4)
12	104			
13	89		Point Estimate	=D5
14	70		Lower Limit	=D13-D11
15	125		Upper Limit	=D13+D11
16	43			
17	61			
100	71			
101	84			
102				

Note: Rows 18 to 99 are not shown.

Using Excel to construct a confidence interval - σ Known (2 of 2)

- Excel Value Worksheet

	A	B	C	D	E
1	Amount		Interval Estimate of a Population Mean:		
2	72		σ Known Case		
3	91				
4	74		Sample Size	100	
5	115		Sample Mean	82	
6	71				
7	120		Population Standard Deviation	20	
8	37		Confidence Coefficient	0.95	
9	96		Level of Significance	0.05	
10	91				
11	105		Margin of Error	3.92	
12	104				
13	89		Point Estimate	82	
14	70		Lower Limit	78.08	
15	125		Upper Limit	85.92	
16	43				
17	61				
100	71				
101	84				
102					

Note: Rows 18 to 99 are not shown.

Interval Estimate of a Population Mean: σ Known (9 of 11)

<u>Example</u>: Lloyds Department store

Confidence level	Margin of Error	Interval estimate
90%	3.92	78.08 – 85.92
95%	3.29	78.71 – 85.29
99%	5.15	76.85 – 87.15

In order to have a higher degree of confidence, the margin of error and thus the width of the confidence interval must be larger.

Interval Estimate of a Population Mean: σ Known (10 of 11)

Adequate Sample Size

- In most applications, a sample size of $n \geq 30$ is adequate.
- If the population distribution is highly skewed or contains outliers, a sample size of 50 or more is recommended.
- If the population is not normally distributed but is roughly symmetric, a sample size as small as 15 will suffice.
- If the population is believed to be at least approximately normal, a sample size of less than 15 can be used.

Interval Estimate of a Population Mean: σ Known (11 of 11)

- If an estimate of the population standard deviation σ cannot be developed prior to sampling, we use the sample standard deviation s to estimate σ.
- This is the σ unknown case.
- In this case, the interval estimate for μ is based on the t distribution.
- (We'll assume for now that the population is normally distributed.)

t Distribution (1 of 6)

- William Gosset, writing under the name “Student,” is the founder of the *t* distribution.
- Gosset was an Oxford graduate in mathematics and worked for the Guinness Brewery in Dublin.
- He developed the *t* distribution while working on small-scale materials and temperature experiments.

t Distribution (2 of 6)

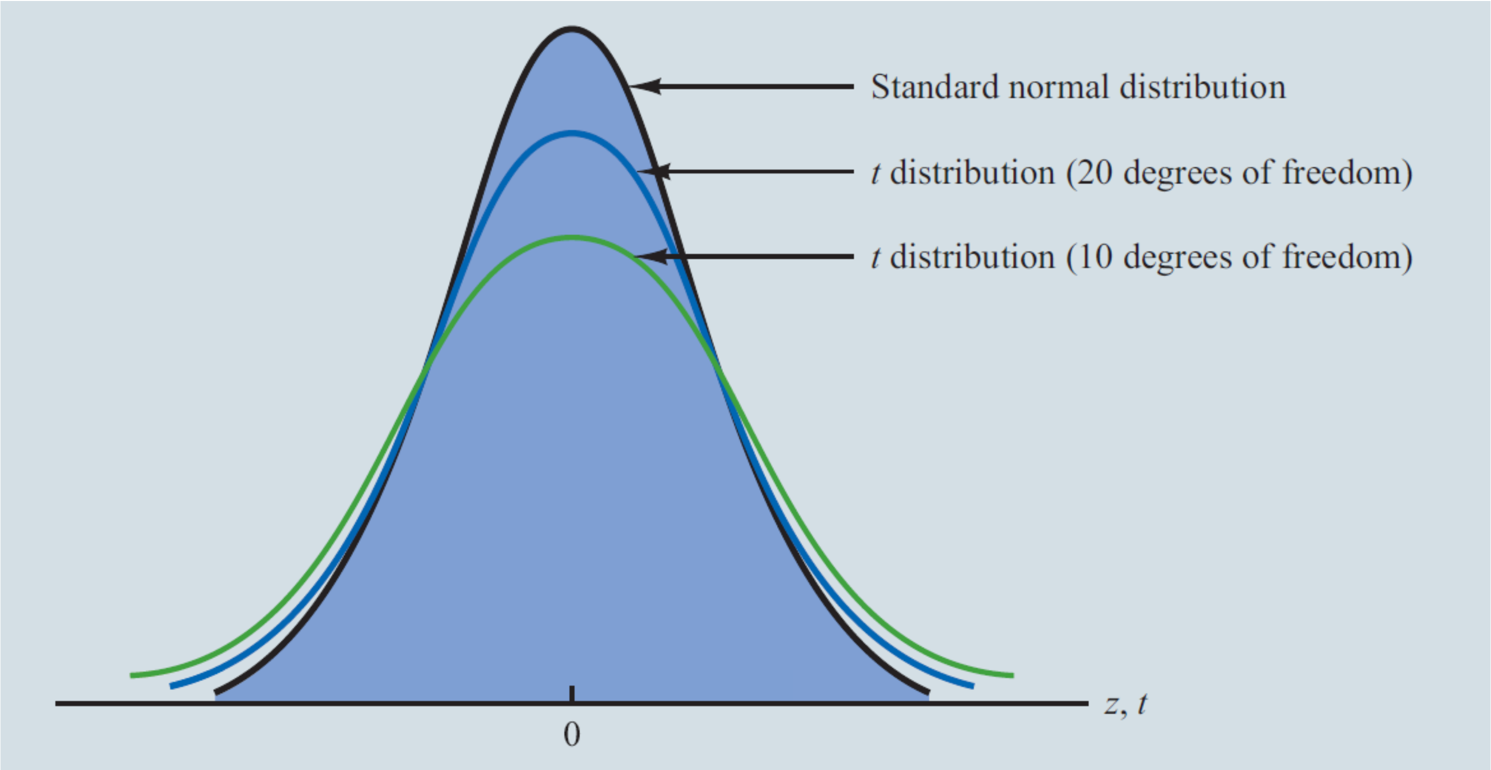
- The *t* distribution is a family of similar probability distributions.
- A specific *t* distribution depends on a parameter known as the degrees of freedom.
- Degrees of freedom refer to the number of independent pieces of information that go into the computation of *s*.

t Distribution (3 of 6)

- A *t* distribution with more degrees of freedom has less dispersion.
- As the degrees of freedom increase, the difference between the *t* distribution and the standard normal probability distribution becomes smaller and smaller.

t Distribution (4 of 6)

Comparison of the standard normal distribution with t distributions having 10 and 20 degrees of freedom.



t Distribution (5 of 6)

- For more than 100 degrees of freedom, the standard normal *z* value provides a good approximation to the *t* value.
- The standard normal *z* values can be found in the infinite degrees row (labeled ∞) of the *t* distribution table.

t Distribution (6 of 6)

Selected values from the *t* distribution table

Degrees of Freedom	Area in Upper Tail					
	.20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	.848	1.296	1.671	2.000	2.390	2.660
61	.848	1.296	1.670	2.000	2.389	2.659
62	.847	1.295	1.670	1.999	2.388	2.657
63	.847	1.295	1.669	1.998	2.387	2.656
64	.847	1.295	1.669	1.998	2.386	2.655
65	.847	1.295	1.669	1.997	2.385	2.654
66	.847	1.295	1.668	1.997	2.384	2.652
67	.847	1.294	1.668	1.996	2.383	2.651
68	.847	1.294	1.668	1.995	2.382	2.650
69	.847	1.294	1.667	1.995	2.382	2.649
⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	.846	1.291	1.662	1.987	2.368	2.632
91	.846	1.291	1.662	1.986	2.368	2.631
92	.846	1.291	1.662	1.986	2.368	2.630
93	.846	1.291	1.661	1.986	2.367	2.630
94	.845	1.291	1.661	1.986	2.367	2.629
95	.845	1.291	1.661	1.985	2.366	2.629
96	.845	1.290	1.661	1.985	2.366	2.628
97	.845	1.290	1.661	1.985	2.365	2.627
98	.845	1.290	1.661	1.984	2.365	2.627
99	.845	1.290	1.660	1.984	2.364	2.626
100	.845	1.290	1.660	1.984	2.364	2.626
∞	.842	1.282	1.645	1.960	2.326	2.576

CENGAGE

Interval Estimate of a Population Mean: σ Unknown (1 of 5)

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where:

$\bar{x}$ = the sample mean

$1-\alpha$ = the confidence coefficient

$t_{\alpha/2}$ = the t value providing an area of $\alpha/2$ in the upper tail of a t distribution with $n-1$ degrees of freedom

s = the sample standard deviation

n = the sample size

Interval Estimate of a Population Mean: σ Unknown (2 of 5)

Example: Credit card debt for the population of US households

The credit card balances of a sample of 70 households provided a mean credit card debt of $9,312 with a sample standard deviation of $4,007.

Let us provide a 95% confidence interval estimate of the mean credit card debt for the population of US households. We will assume this population to be normally distributed.

Interval Estimate of a Population Mean: σ Unknown (3 of 5)

- At 95% confidence, $\alpha = .05$, and $\alpha/2 = .025$.

  $t_{.025}$ is based on $n - 1 = 70 - 1 = 69$ degrees of freedom.

Degrees of Freedom	Area in Upper Tail .20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	.848	1.296	1.671	2.000	2.390	2.660
61	.848	1.296	1.670	2.000	2.389	2.659
62	.847	1.295	1.670	1.999	2.388	2.657
63	.847	1.295	1.669	1.998	2.387	2.656
64	.847	1.295	1.669	1.998	2.386	2.655
65	.847	1.295	1.669	1.997	2.385	2.654
66	.847	1.295	1.668	1.997	2.384	2.652
67	.847	1.294	1.668	1.996	2.383	2.651
68	.847	1.294	1.668	1.995	2.382	2.650
69	.847	1.294	1.667	1.995	2.382	2.649

Interval Estimate of a Population Mean: σ Unknown (4 of 5)

Example: Credit card debt for the population of U.S. households

$$\bar{x} \pm t_{.025} \frac{s}{\sqrt{n}}$$

$$9{,}312 \pm 1.995 \frac{4007}{\sqrt{70}} = 9{,}312 \pm 955$$

We are 95% confident that the mean credit card debt for the population of U.S. households is between $8,357 and $10,267.

Using Excel's Descriptive Statistics Tool (1 of 3)

Steps

Step 1: Click the **Data** tab on the Ribbon

Step 2: In the **Analysis** group click **Data Analysis**

Step 3: Choose **Descriptive Statistics** from the list of Analysis Tools

Using Excel's Descriptive Statistics Tool (2 of 3)

Step 4: When the **Descriptive statistics** dialog box appears

- Enter **Input Range**
- Select **Grouped by columns**
- Select the check box for **Labels in First Row**
- Select **Output range:**
  - Enter C1 in the **Output Range** box
- Select the check box for **Summary Statistics**
- Select **Confidence Level for Mean**
  - Enter 95 in the **Confidence Level for Mean** box
- Click **OK**

Using Excel's Descriptive Statistics Tool (3 of 3)

- Excel Worksheets

95% confidence interval for credit card balances.

	A	B	C	D	E
1	NewBalance		NewBalance		
2	9430				
3	7535		Mean	9312	
4	4078		Standard Error	478.9281	
5	5604		Median	9466	
6	5179		Mode	13627	
7	4416		Standard Deviation	4007	
8	10676		Sample Variance	16056048	
9	1627		Kurtosis	-0.2960	
10	10112		Skewness	0.1879	
11	6567		Range	18648	
12	13627		Minimum	615	
13	18719		Maximum	19263	
14	14661		Sum	651840	
15	12195		Count	70	
16	10544		Confidence Level(95.0%)	955	
17	13659				
18	7061		Point Estimate	=D3	
19	6245		Lower Limit	=D18-D16	
20	13021		Upper Limit	=D3+D16	
70	9743				
71	10324				
72					

	A	B	C	D	E	F
1	NewBalance		NewBalance			
2	9430					
3	7535		Mean	9312	Point Estimate	
4	4078		Standard Error	478.9281		
5	5604		Median	9466		
6	5179		Mode	13627		
7	4416		Standard Deviation	4007		
8	10676		Sample Variance	16056048		
9	1627		Kurtosis	-0.2960		
10	10112		Skewness	0.1879		
11	6567		Range	18648		
12	13627		Minimum	615		
13	18719		Maximum	19263		
14	14661		Sum	651840		
15	12195		Count	70		
16	10544		Confidence Level(95.0%)	955	Margin of Error	
17	13659					
18	7061		Point Estimate	9312		
19	6245		Lower Limit	8357		
20	13021		Upper Limit	10267		
70	9743					
71	10324					
72						

CENGAGE

Interval Estimate of a Population Mean: σ Unknown (5 of 5)

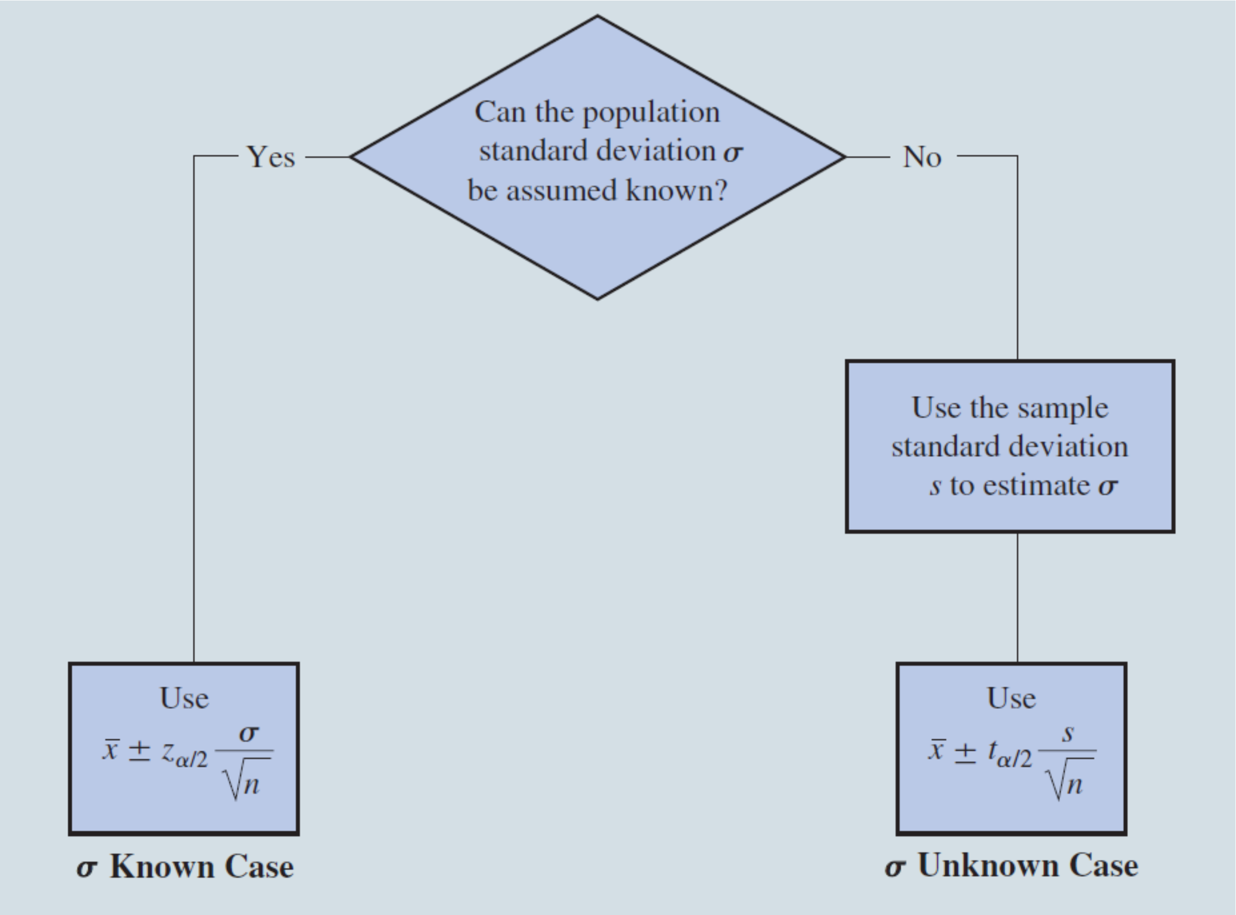
Adequate Sample Size

Usually, a sample size of $n \geq 30$ is adequate when using the expression

$\bar{x} \pm t_{\alpha/2}\, s/\sqrt{n}$ to develop an interval estimate of a population mean.

- If the population distribution is highly skewed or contains outliers, a sample size of 50 or more is recommended.
- If the population is not normally distributed but is roughly symmetric, a sample size as small as 15 will suffice.
- If the population is believed to be at least approximately normal, a sample size of less than 15 can be used.

Summary of Interval Estimation Procedures for a Population Mean



Sample Size for an Interval Estimate of a Population Mean (1 of 5)

- Let *E* = the desired margin of error.
- *E* is the amount added to and subtracted from the point estimate to obtain an interval estimate.
- If a desired margin of error is selected prior to sampling, the sample size necessary to satisfy the margin of error can be determined.

Sample Size for an Interval Estimate of a Population Mean (2 of 5)

- Margin of Error

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Necessary Sample Size

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

Sample Size for an Interval Estimate of a Population Mean (3 of 5)

- The Necessary Sample Size equation requires a value for the population standard deviation σ.
- If σ is unknown, a preliminary or planning value for σ can be used in the equation.

1. Use the estimate of the population standard deviation computed in a previous study.
2. Use a pilot study to select a preliminary study and use the sample standard deviation from the study.
3. Use judgment or a "best guess" for the value of σ.

Sample Size for an Interval Estimate of a Population Mean (4 of 5)

<u>**Example:**</u> Cost of renting Automobiles in United States

A previous study that investigated the cost of renting automobiles in the United States found a mean cost of approximately $55 per day for renting a midsize automobile with a standard deviation of $9.65.

Suppose the project director wants an estimate of the population mean daily rental cost such that there is a .95 probability that the sampling error is $2 or less.

How large of a sample size is needed to meet the required precision?

Sample Size for an Interval Estimate of a Population Mean (5 of 5)

Example: Cost of renting Automobiles in United States

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 2$$

At 95% confidence, $z_{.025} = 1.96$. Recall that $\sigma = 9.65$.

$$n = \frac{\left(z_{\alpha/2}\right)^2 \sigma^2}{E^2} = \frac{(1.96)^2(9.65)^2}{(2)^2} = 89.43 \approx 90$$

The sample size needs to be at least 90 mid size automobile rentals in order to satisfy the project director’s $2 margin-of-error requirement.

Interval Estimate of a Population Proportion (1 of 6)

The general form of an interval estimate of a population proportion is:

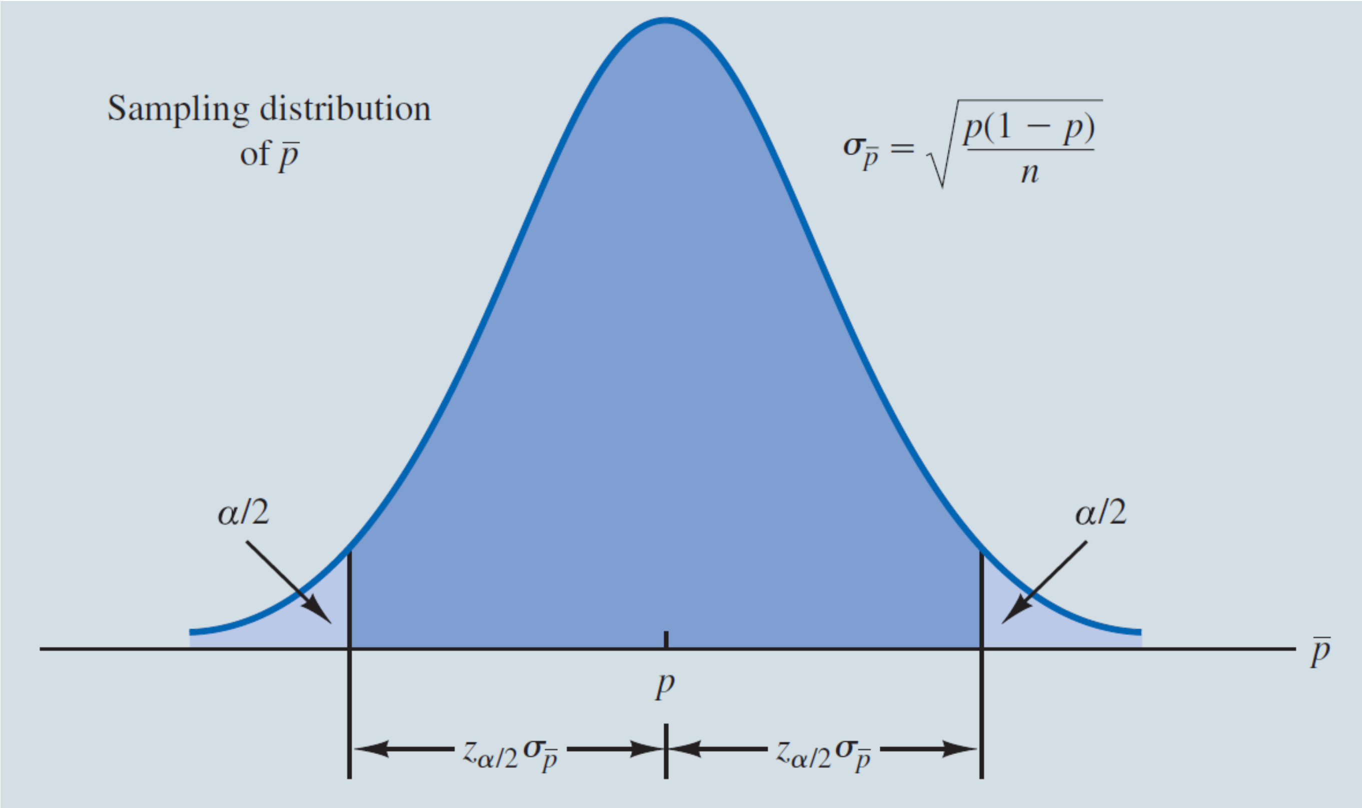
$$\bar{p} \pm \text{Margin of Error}$$

Interval Estimate of a Population Proportion (2 of 6)

- The sampling distribution of $\bar{p}$ plays a key role in computing the margin of error for this interval estimate.
- The sampling distribution of $\bar{p}$ can be approximated by a normal distribution whenever $np \geq 5$ and $n(1 - p) \geq 5$.

Interval Estimate of a Population Proportion (3 of 6)

Normal Approximation of Sampling Distribution of $\bar{p}$



Interval Estimate of a Population Proportion (4 of 6)

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

where : $1-\alpha$ is the confidence coefficient,

$z_{\alpha/2}$ is the z value providing an area of $\alpha/2$ in the upper tail of the standard normal probability distribution, and

$\bar{p}$ is the sample proportion

Interval Estimate of a Population Proportion (5 of 6)

Example: Survey of women golfers

A national survey of 900 women golfers was conducted to learn how women golfers view their treatment at golf courses in United States. The survey found that 396 of the women golfers were satisfied with the availability of tee times.

Suppose one wants to develop a 95% confidence interval estimate for the proportion of the population of women golfers satisfied with the availability of tee times.

Interval Estimate of a Population Proportion (6 of 6)

Example: Survey of women golfers

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

$$\text{where}: n = 900, \quad \bar{p} = 396/900 = .44, \quad z_{\alpha/2} = 1.96$$

$$.44 \pm 1.96\sqrt{\frac{.44(1-.44)}{900}} = .44 \pm .0324$$

Survey results enable us to state with 95% confidence that between 40.76% and 47.24% of all women golfers are satisfied with the availability of tee times.

Using Excel to construct a confidence interval

- Excel Formula and Value Worksheet

	A	B	C	D	E
1	Response		Interval Estimate of a Population Proportion		
2	Yes				
3	No		Sample Size	=COUNTA(A2:A901)	
4	Yes		Response of Interest	Yes	
5	Yes		Count for Response	=COUNTIF(A2:A901,D4)	
6	No		Sample Proportion	=D5/D3	
7	No				
8	No		Confidence Coefficient	0.95	
9	Yes		Level of Significance (alpha)	=1-D8	
10	Yes		*z* Value	=NORM.S.INV(1-D9/2)	
11	Yes				
12	No		Standard Error	=SQRT(D6*(1-D6)/D3)	
13	No		Margin of Error	=D10*D12	
14	Yes				
15	No		Point Estimate	=D6	
16	No		Lower Limit	=D15-D13	
17	Yes		Upper Limit	=D15+D13	
18	No				
900	Yes				
901	Yes				
902					

	A	B	C	D	E	F	G
1	Response		Interval Estimate of a Population Proportion				
2	Yes						
3	No		Sample Size	900	Enter Yes as the Response of Interest		
4	Yes		Response of Interest	Yes			
5	Yes		Count for Response	396			
6	No		Sample Proportion	0.44			
7	No						
8	No		Confidence Coefficient	0.95			
9	Yes		Level of Significance	0.05			
10	Yes		*z* Value	1.96			
11	Yes						
12	No		Standard Error	0.0165			
13	No		Margin of Error	0.0324			
14	Yes						
15	No		Point Estimate	0.44			
16	No		Lower Limit	0.4076			
17	Yes		Upper Limit	0.4724			
18	No						
900	Yes						
901	Yes						
902							

Sample Size for an Interval Estimate of a Population Proportion (1 of 5)

Margin of Error

$$E = z_{\alpha/2}\sqrt{\frac{\bar{p}\left(1-\bar{p}\right)}{n}}$$

Solving for the necessary sample size n, we get

$$n = \frac{\left(z_{\alpha/2}\right)^2 \bar{p}(1-\bar{p})}{E^2}$$

However, $\bar{p}$ will not be known until after we have selected the sample. We will use the planning value p^* for $\bar{p}$.

Sample Size for an Interval Estimate of a Population Proportion (2 of 5)

Necessary Sample Size

$$n = \frac{\left(z_{\alpha/2}\right)^2 p^*(1-p^*)}{E^2}$$

The planning value *p** can be chosen by one of the following procedures:

1. Using the sample proportion from a previous sample of the same or similar units.
2. Selecting a preliminary sample and using the sample proportion from this sample.
3. Using judgment or a “best guess” for a *p** value.
4. Otherwise, use .50 as the *p** value.

Sample Size for an Interval Estimate of a Population Proportion (3 of 5)

Example: Survey of women golfers

Suppose the survey director wants to estimate the population proportion with a margin of error of .025 at 95% confidence.

How large of a sample size is needed to meet the required precision? (A previous sample of similar units yielded .44 for the sample proportion.)

Sample Size for an Interval Estimate of a Population Proportion (4 of 5)

<u>Example:</u> Survey of women golfers

$$E = z_{\alpha/2}\sqrt{\frac{p^*\left(1-p^*\right)}{n}} = .025$$

At 95% confidence, $z_{.0125} = 1.96$. Recall that $p^* = .44$.

$$n = \frac{\left(z_{\alpha/2}\right)^2 p^*(1-p^*)}{E^2} = \frac{(1.96)^2(.44)(.56)}{(.025)^2} = 1{,}514.5$$

A sample of size 1515 is needed to reach a desired precision of ± .025 at 95% confidence.

Sample Size for an Interval Estimate of a Population Proportion (5 of 5)

Note: We used .44 as the best estimate of p in the preceding expression. If no information is available about p, then .5 is often assumed because it provides the highest possible sample size. If we had used $p = .5$, the recommended n would have been 1537.

Implications of Big Data

- As the sample size becomes extremely large, the margin of error becomes extremely small and resulting confidence intervals become extremely narrow.
- No interval estimate will accurately reflect the parameter being estimated unless the sample is relatively free of nonsampling error.
- Statistical inference along with information collected from other sources can help in making the most informed decision.