COMM215

First the Foundation, then Innovation

LESSON 7
INTERVAL ESTIMATION

SAMIE L.S. LY

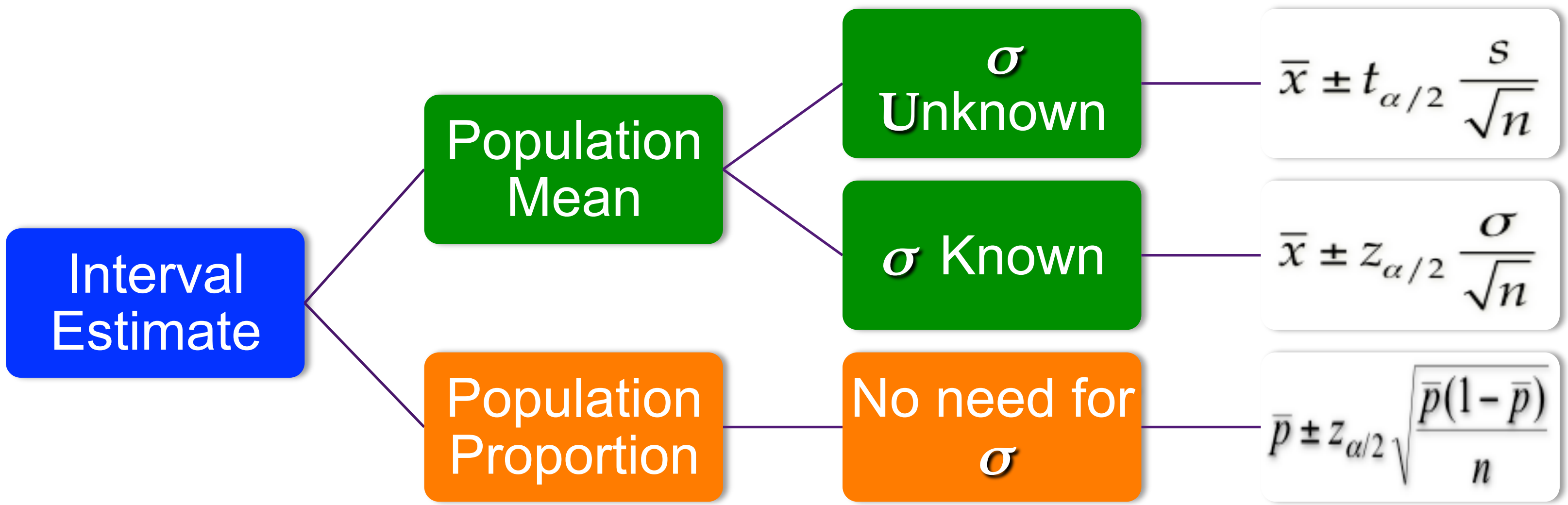
THIS WEEK'S PLAN

Part I: Theory + Practice (Interval Estimation)

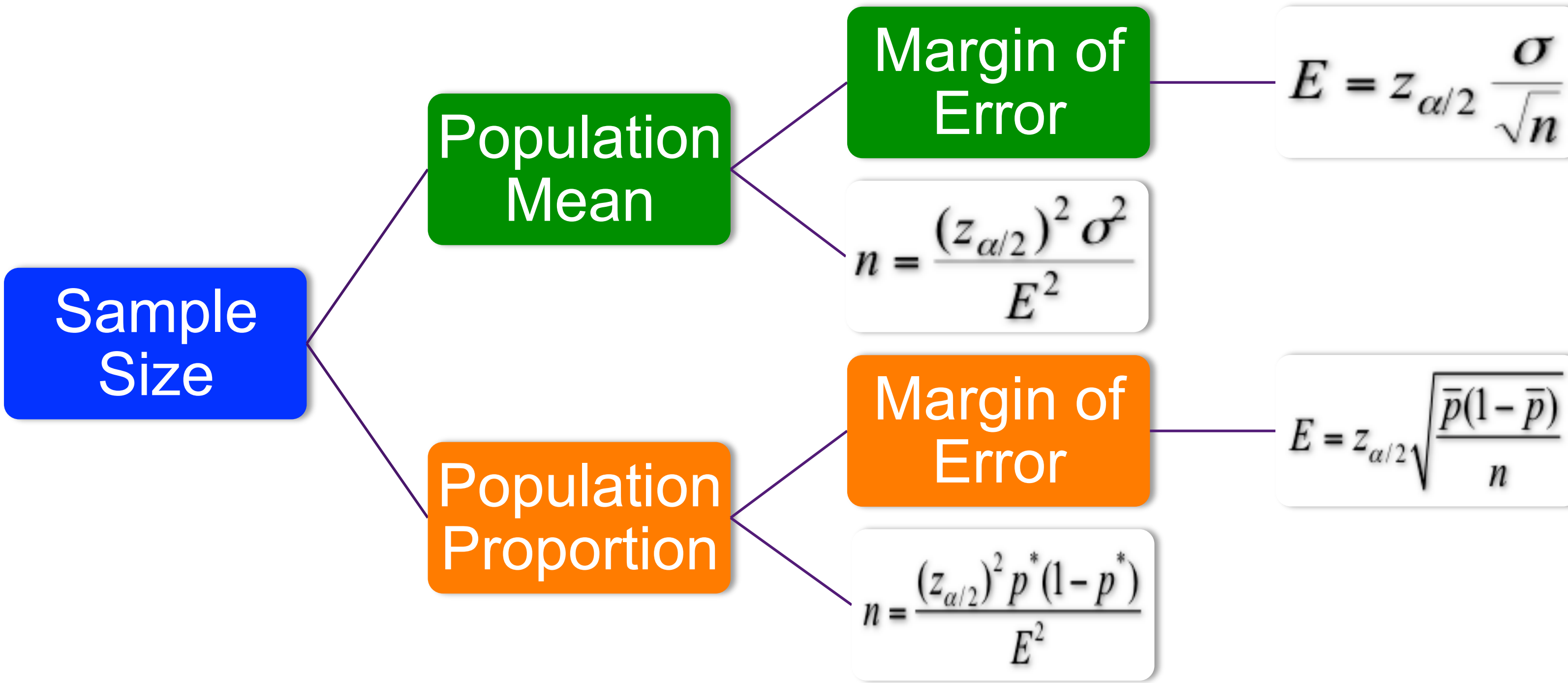
Part II: Theory + Practice (Interval Estimation)

- z-Based Confidence Intervals for a Population Mean: σ known
- t-Based Confidence Intervals for a Population Mean: σ unknown
- Sample Size Determination
- Confidence Intervals for a Population Proportion

LESSON 7 IN A NUTSHELL



LESSON 7 IN A NUTSHELL



POINT ESTIMATION $\overline{x}$

Point estimator is a sample statistic used to estimate a population parameter.

The sample mean $\overline{x}$ is a point estimator of the population mean μ.

The sample proportion $\overline{p}$ is the point estimator of the population proportion p.

PROBLEM # 7.1

A simple random sample of 8 employees is selected from a large firm. For the 8 employees, the number of days each was absent during the past month was found to be 0,2,4,2,1,7,3 and 2, respectively.

a. What is the point estimate for *μ*, the mean number of days absent for the firm's employees?

$$\bar{x} = \frac{\sum x}{n} = \frac{21}{8} = 2.625$$

b. What is the point estimate for *σ*², the variance of the number of days absent?

$$s^2 = \sum \frac{(x-\bar{x})^2}{n-1} = \frac{31.875}{7} = 4.554$$

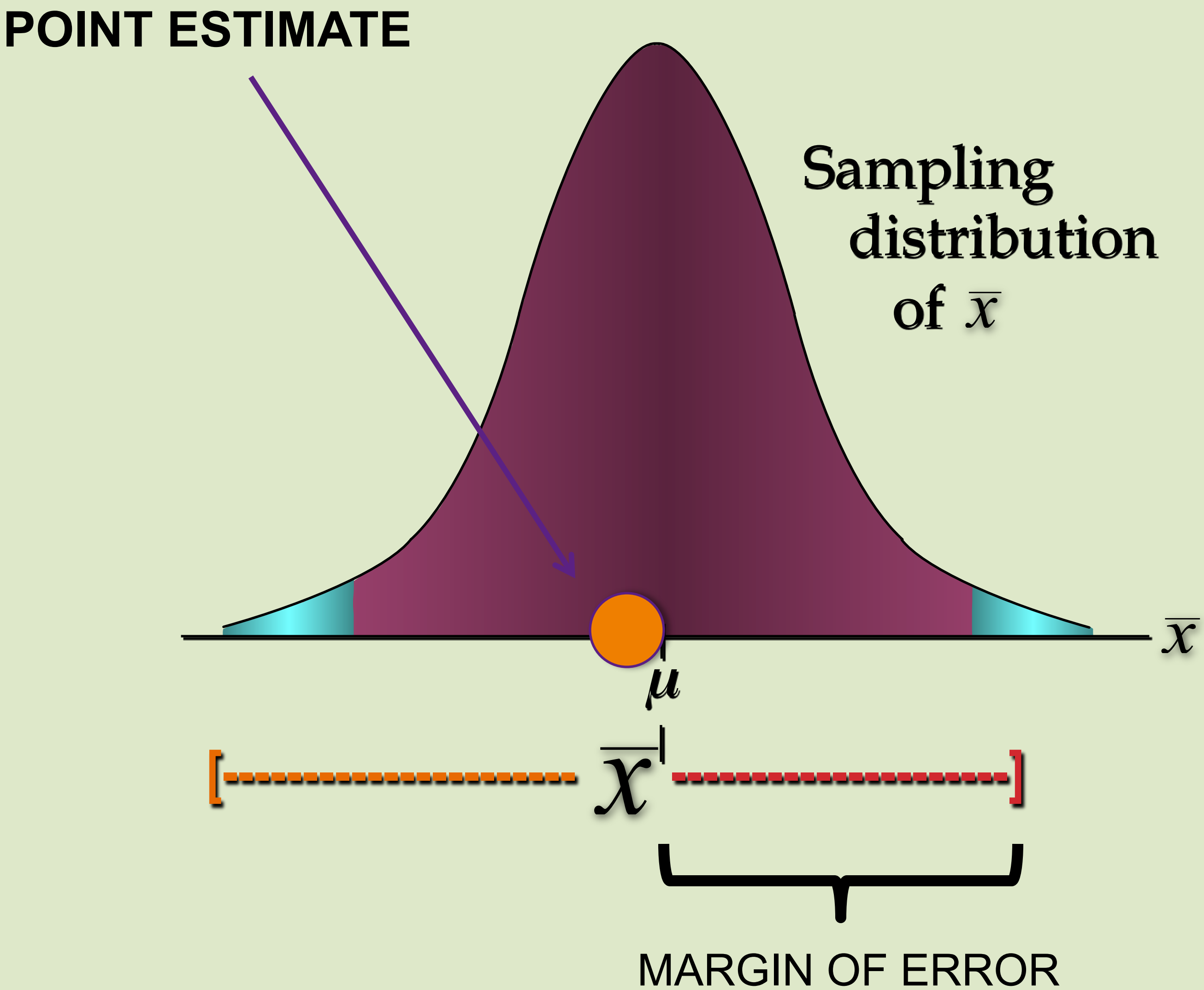
INTERVAL ESTIMATION

Very often, a point estimator cannot be expected to provide the exact value of the population parameter,

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

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

Bowerman, et al. (2017) pp. 348



Bowerman, et al. (2017) pp. 349

INTERVAL ESTIMATION

Interval estimator is affected by the sample size

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

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

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

PROBLEM # 7.2*

During the month of July, an auto manufacturer gives its production employees a vacation period so it can tool up for the new model run. In surveying a simple random sample of 200 production workers, the personnel director finds that 38% of them plan to vacation out of state for at least one week during this period. Is this a point estimate or an interval estimate? Explain.

This is a point estimate, since $\bar{p}$ = 0.38 is a single number that estimates the value of

the population parameter, p = the true proportion who vacation out of state for at least one week.

PROBLEM # 7.3

Differentiate between a point estimate and the interval estimate for a population parameter.

A point estimate is a single number that estimates the value of the population parameter, while an interval estimate includes a range of possible values which are likely to include the population parameter.

INTERVAL ESTIMATE

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

Point Estimate

Confidence Interval

Standard Deviation

Sample Size

Margin of Error

Bowerman, et al. (2017) pp. 353

CONFIDENCE INTERVAL

Before determining an Interval Estimate

We must know the confidence level.

Key word: **CONFIDENCE!**

How confident are you that the population mean will fall within this interval?

Are you 90% confident? 95%? or 99%?


INTERVAL EST. - BY SMIEVISION
WWW.TOONDOO.COM
MISS, CAN YOU GIVE ME AN INTERVAL ESTIMATE OF MY FINAL GRADE FOR THIS COURSE?
I CAN TELL YOU AT 99% CONFIDENCE INTERVAL YOUR GRADE WILL BE BETWEEN 10% AND 100%
ME TOO! ME TOO!

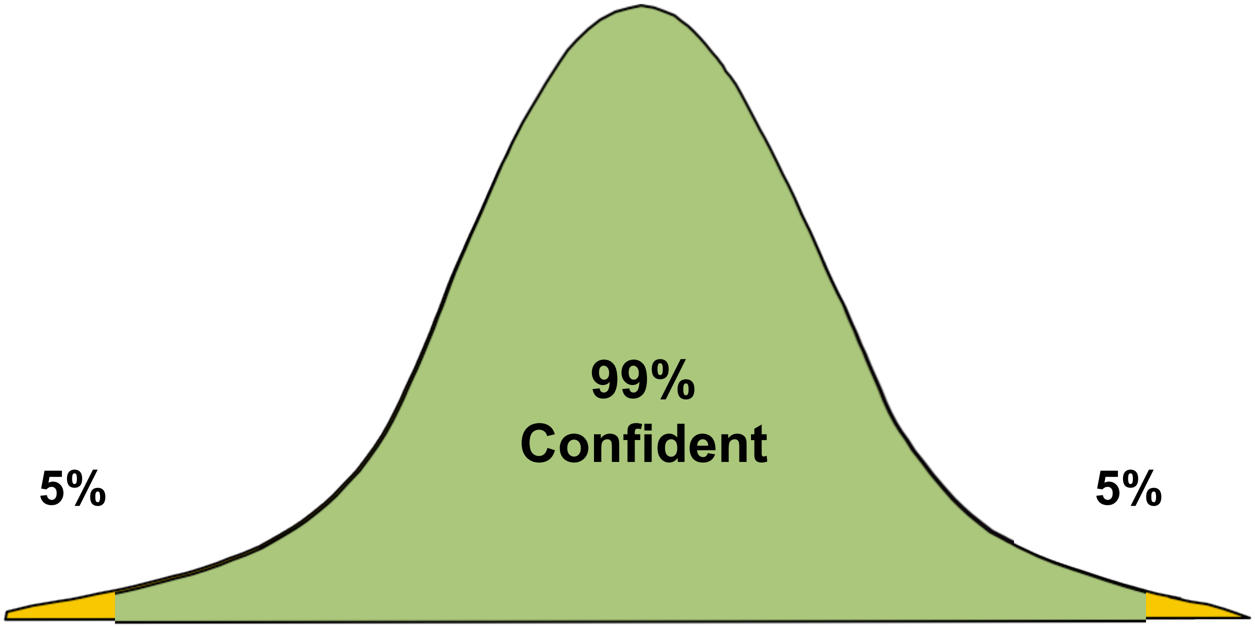
PROBLEM # 7.4

What is necessary for an interval estimate to be a confidence interval?

When the interval estimate is associated with a degree of confidence that it actually includes the population parameter, it is referred to as a confidence interval.

CONFIDENCE INTERVAL

Is the interval wider for 90%? or 99%?



Still 10% doubting

Bowerman, et al. (2017) pp. 352

COMM215
First the Foundation, then Innovation

PROBLEM # 7.5*

What role does the central limit theorem play in the construction of a confidence interval for the population mean?

If the population cannot be assumed to be normally distributed,when the sample size is at least 30 we can apply the central limit theorem in order for the sampling distribution of the sample mean to be approximately normal

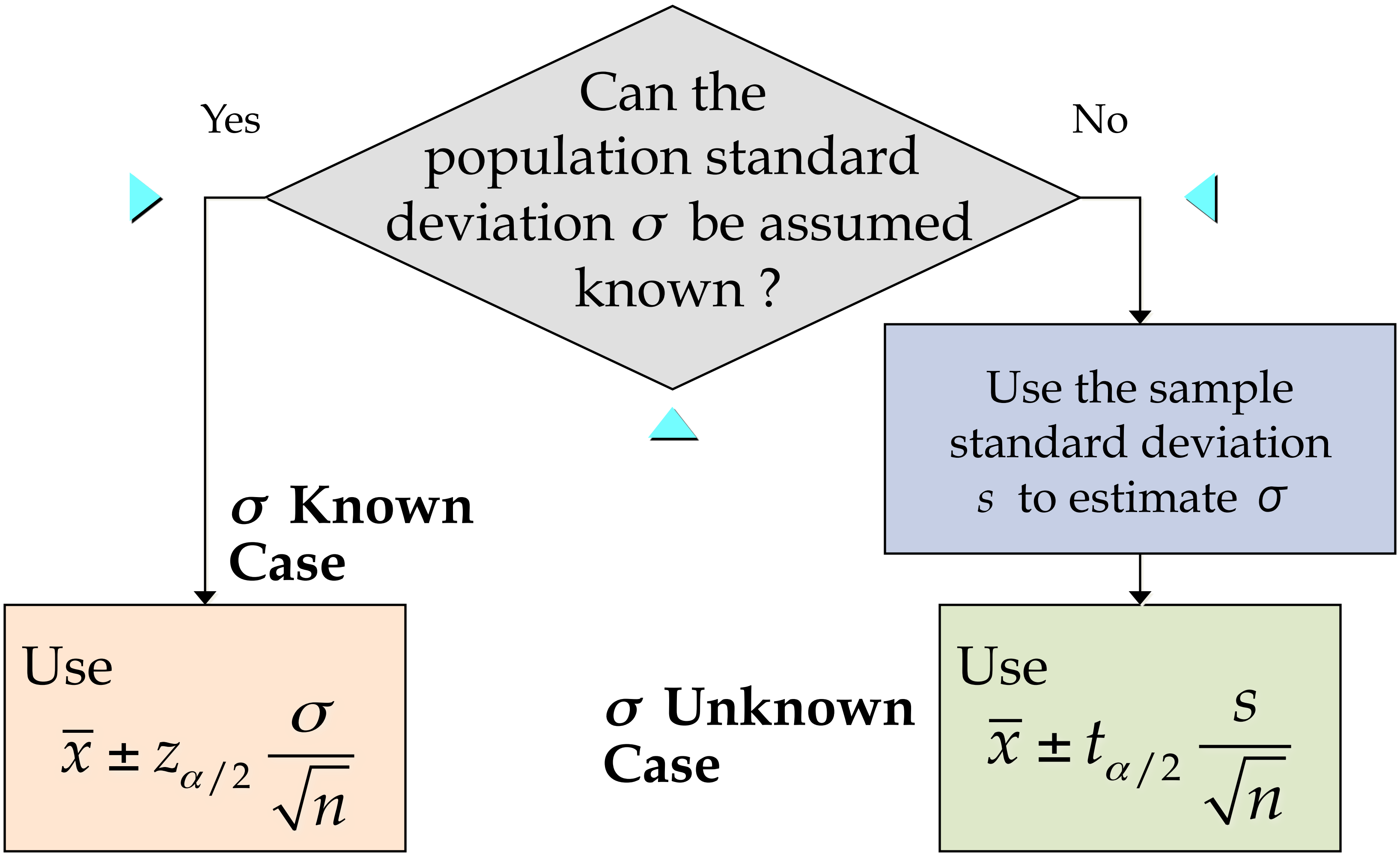
INTERVAL ESTIMATE OF A POPULATION MEAN

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.

Summary of Interval Estimation Procedures for a Population Mean



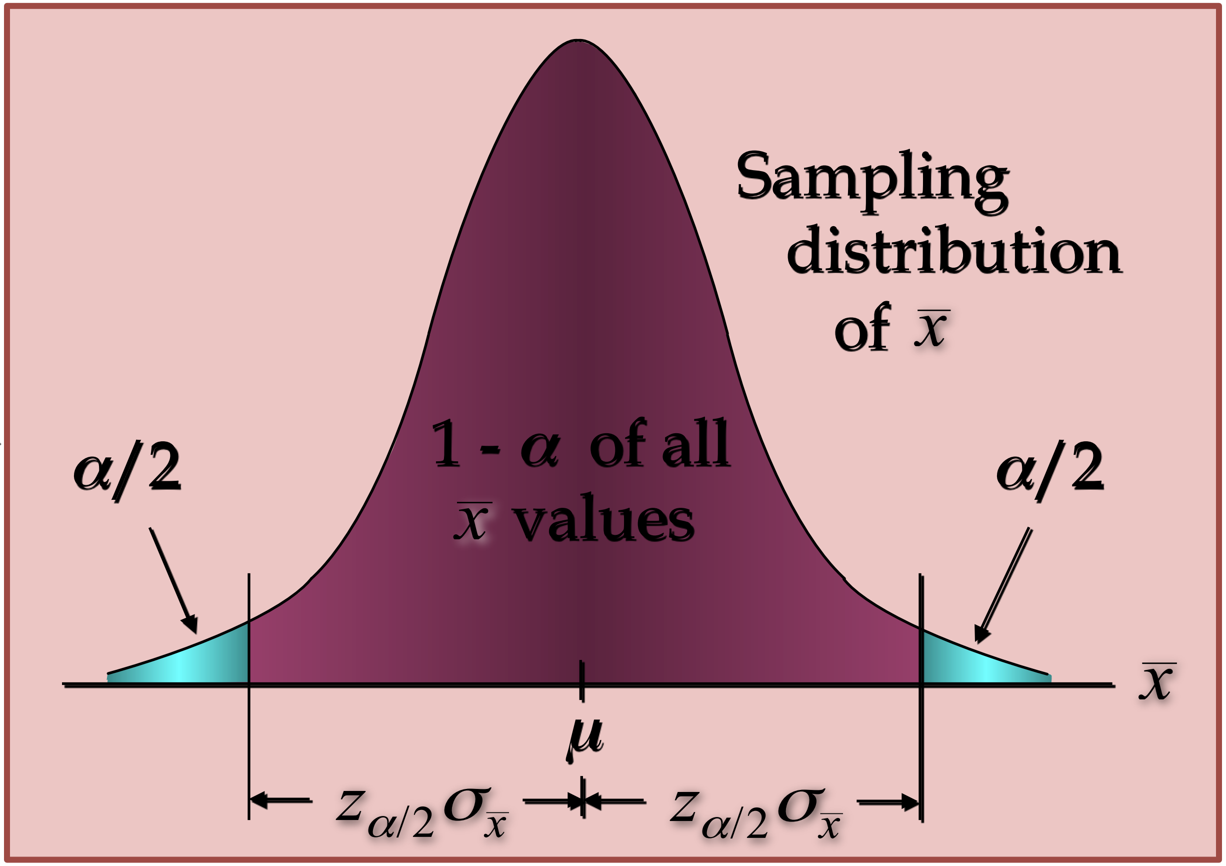
POPULATION MEAN: σ KNOWN

Interval Estimate

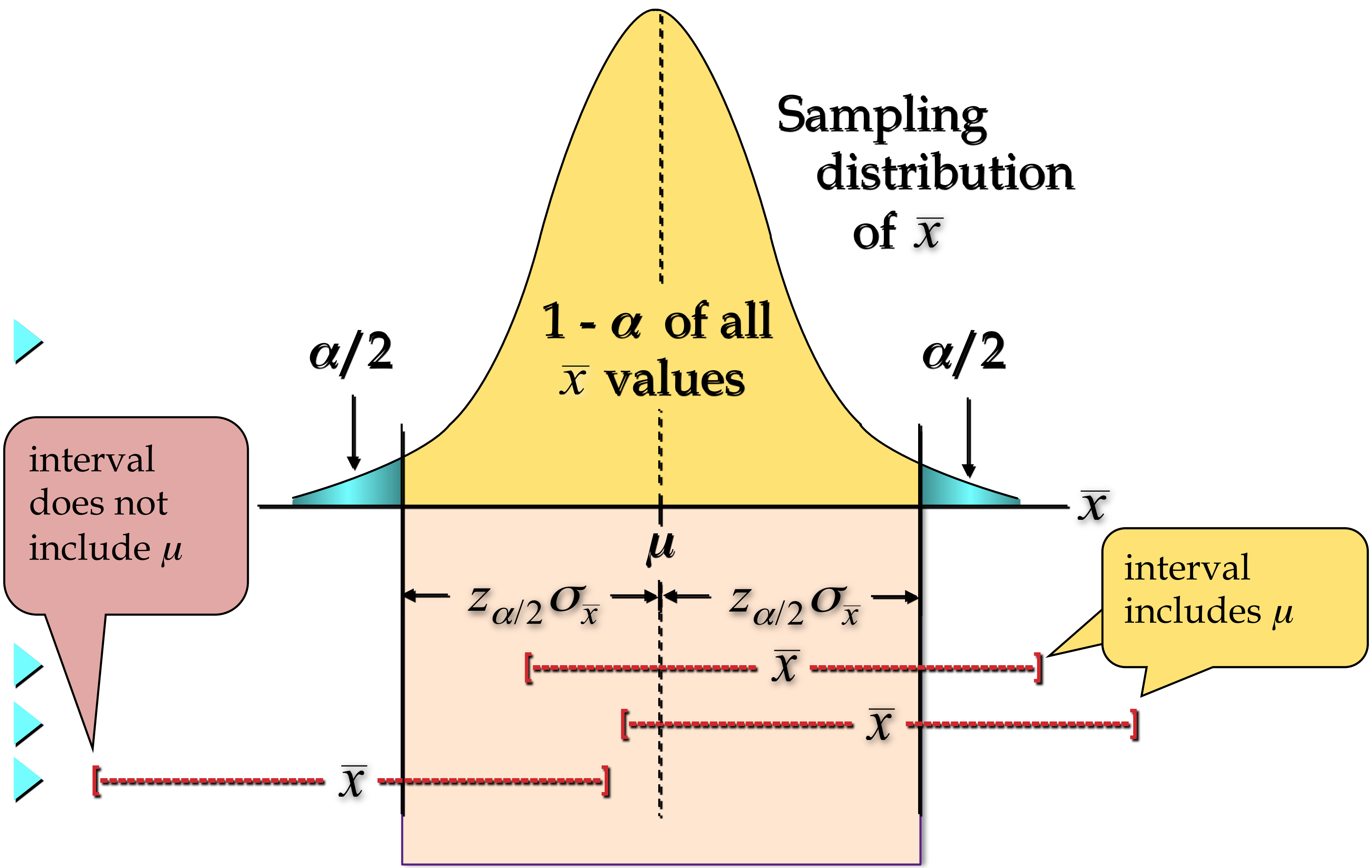
Margin of Error

INTERVAL ESTIMATE OF A POPULATION MEAN: σ KNOWN

There is a 1 − α 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



INTERVAL ESTIMATE OF A POPULATION MEAN: σ KNOWN

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

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

Confidence Level	α	$\alpha/2$	Table Look-up Area	$z_{\alpha/2}$
90%	.10	.05	.9500	1.645
95%	.05	.025	.9750	1.960
99%	.01	.005	.9950	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

- Adequate Sample Size

  - In most applications, a sample size of n = 30 is adequate.

  - If the population distribution is highly skewed or contains outliers, a sample size of 50 or more is recommended.

PROBLEM # 7.6

In using the standard normal distribution to construct a confidence interval for the population mean, which two assumptions are necessary if the sample size is less than 30?

In this case, we need to assume that

1. the population is **normally distributed**
2. the population **standard deviation is known**.

PROBLEM # 7.7

A simple random sample of 30 has been collected from a population for which it is known that σ = 10.0. The sample mean has been calculated as 240.0. Construct and interpret the 90% and 95% confidence intervals for the population mean.

PROBLEM # 7.7

A simple random sample of 30 has been collected from a population for which it is known that *σ* = 10.0. The sample mean has been calculated as 240.0. Construct and interpret the 90% and 95% confidence intervals for the population mean.

a. For a confidence level of 90%, z = 1.645. (In the normal distribution, 90% of the area falls between z = -1.645 and z = 1.645.) The 90% confidence interval for μ is:

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 240 \pm 1.645\frac{10}{\sqrt{30}} = 240 \pm 3.003$$, or between 236.997 and 243.003

COMM215 First the Foundation, then Innovation

PROBLEM # 7.7

A simple random sample of 30 has been collected from a population for which it is known that σ = 10.0. The sample mean has been calculated as 240.0. Construct and interpret the 90% and 95% confidence intervals for the population mean.

b*. For a confidence level of 95%, z = 1.96. (In the normal distribution, 95% of the area falls between z = -1.96 and z = 1.96.) The 95% confidence interval for μ is:

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 240 \pm 1.96\frac{10}{\sqrt{30}} = 240 \pm 3.578$$, or between 236.422 and 243.578

COMM215 First the Foundation, then Innovation

PROBLEM # 7.7

We could also obtain these confidence intervals by using Excel worksheet.

	A	B	C	D
1	**Confidence interval for the population mean,**			
2	**using the z distribution and known**			
3	**(or assumed) pop. std. deviation, sigma:**			
4				
5	Sample size, n:			30
6	Sample mean, xbar:			240.000
7	Known or assumed pop. sigma:			10.0000
8	Standard error of xbar:			1.82574
9				
10	Confidence level desired:			0.90
11	alpha = (1 - conf. level desired):			0.10
12	z value for desired conf. int.:			1.6449
13	z times standard error of xbar:			3.003
14				
15	Lower confidence limit:			236.997
16	Upper confidence limit:			243.003

	A	B	C	D
1	**Confidence interval for the population mean,**			
2	**using the z distribution and known**			
3	**(or assumed) pop. std. deviation, sigma:**			
4				
5	Sample size, n:			30
6	Sample mean, xbar:			240.000
7	Known or assumed pop. sigma:			10.0000
8	Standard error of xbar:			1.82574
9				
10	Confidence level desired:			0.95
11	alpha = (1 - conf. level desired):			0.05
12	z value for desired conf. int.:			1.9600
13	z times standard error of xbar:			3.578
14				
15	Lower confidence limit:			236.422
16	Upper confidence limit:			243.578

- **t-Based Confidence Intervals for a Population Mean: σ unknown**
- **Sample Size Determination**
- **Confidence Intervals for a Population Proportion**

POPULATION MEAN: σ UNKNOWN

[t-distribution Presentation]

http://prezi.com/pvxqngtd_9qw/understanding-the-t-table/

INTERVAL ESTIMATE OF A POPULATION MEAN: σ UNKNOWN

- 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.)

Bowerman, et al. (2017) pp. 355

INTERVAL ESTIMATE OF A POPULATION MEAN: σ UNKNOWN

Interval Estimate

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

where: $1 - \alpha$ = 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

PROBLEM #7.8

When the t-distribution is used in constructing a confidence interval based on a sample size of less than 30, what assumption must be made about the shape of the underlying population?

When $n < 30$, we must assume that the population is approximately normally distributed.

PROBLEM # 7.9

In using the t distribution table, what value of *t* would correspond to an upper-tail area of 0.025 for 19 degrees of freedom?

Referring to the 0.025 column and the d.f. = 19 row of the t table, the value of t corresponding to an upper tail area of 0.025 is t = 2.093.

PROBLEM # 7.10

A consumer magazine has contacted a simple random sample of 33 owners of a certain model of automobile and asked each owner how many defects has to be corrected within the first 2 months of ownership. The average number of defects was $\bar{x}$ =3.7, with a standard deviation of 1.8 defects.

a. Use the t distribution to construct a 95% confidence interval for μ = the average number of defects for this model.

b. Use the z distribution to construct a 95% confidence interval for μ = the average number of defects for this model.

c. Given that the population standard deviation is not known, which of these two confidence intervals should be used as the interval estimate for μ?

PROBLEM # 7.10

A consumer magazine has contacted a simple random sample of 33 owners of a certain model of automobile and asked each owner how many defects has to be corrected within the first 2 months of ownership. The average number of defects was $\bar{x}$ =3.7, with a standard deviation of 1.8 defects.

a. Use the t distribution to construct a 95% confidence interval for μ = the average number of defects for this model.

$$\bar{x} \pm t \frac{s}{\sqrt{n}} = 3.7 \pm 2.037 \frac{1.8}{\sqrt{33}} = 3.7 \pm 0.638,$$

or between 3.062 and 4.338.

PROBLEM # 7.10

A consumer magazine has contacted a simple random sample of 33 owners of a certain model of automobile and asked each owner how many defects has to be corrected within the first 2 months of ownership. The average number of defects was $\bar{x}$=3.7, with a standard deviation of 1.8 defects.

a. Use the t distribution to construct a 95% confidence interval for μ = the average number of defects for this model.

For a confidence level of 95%, the right-tail area of interest is (1 - 0.95)/2 = 0.025 with d.f. = n - 1 = 33 - 1 = 32. Referring to the 0.025 column and the d.f. = 32 row of the t table, t = 2.037. The 95% confidence interval for μ is:

$$\bar{x} \pm t \frac{s}{\sqrt{n}} = 3.7 \pm 2.037 \frac{1.8}{\sqrt{33}} = 3.7 \pm 0.638$$, or between 3.062 and 4.338.

PROBLEM # 7.10

A consumer magazine has contacted a simple random sample of 33 owners of a certain model of automobile and asked each owner how many defects has to be corrected within the first 2 months of ownership. The average number of defects was $\bar{x}$=3.7, with a standard deviation of 1.8 defects.

b. Use the z distribution to construct a 95% confidence interval for μ = the average number of defects for this model.

$$\bar{x} \pm z\frac{s}{\sqrt{n}} = 3.7 \pm 1.96\frac{1.8}{\sqrt{33}} = 3.7 \pm 0.614,$$

or between 3.086 and 4.314.

COMM215
First the Foundation, then Innovation

PROBLEM # 7.10

A consumer magazine has contacted a simple random sample of 33 owners of a certain model of automobile and asked each owner how many defects has to be corrected within the first 2 months of ownership. The average number of defects was $\bar{x}$=3.7, with a standard deviation of 1.8 defects.

b. Use the z distribution to construct a 95% confidence interval for μ = the average number of defects for this model.

For a confidence level of 95%, z = 1.96 (in the standard normal distribution, 95% of the area is between z = -1.96 and z = 1.96). The 95% confidence interval for μ is:

$$\bar{x} \pm z\frac{s}{\sqrt{n}} = 3.7 \pm 1.96\frac{1.8}{\sqrt{33}} = 3.7 \pm 0.614$$

, or between 3.086 and 4.314.

PROBLEM # 7.10

A consumer magazine has contacted a simple random sample of 33 owners of a certain model of automobile and asked each owner how many defects has to be corrected within the first 2 months of ownership. The average number of defects was $\bar{x}$ =3.7, with a standard deviation of 1.8 defects.

c. Given that the population standard deviation is not known, which of these two confidence intervals should be used as the interval estimate for *μ*?

If σ is not known, the t distribution should be used in constructing a 95% confidence interval for μ. Therefore, the confidence interval found in part a. is the correct one.

Sample Size Determination

Confidence Intervals for a Population Proportion

SAMPLE SIZE

POPULATION MEAN

SAMPLE SIZE FOR AN INTERVAL ESTIMATE OF A POPULATION MEAN

Let E = the desired margin of error

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

If a desired margin of error is selected prior to sampling, the sample size necessary to satisfy the margin of error can be determined.

Bowerman, et al. (2017) pp. 364

SAMPLE SIZE FOR AN INTERVAL ESTIMATE OF A POPULATION MEAN

- Margin of Error

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

- Necessary Sample Size

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

Bowerman, et al. (2017) pp. 365

Sample Size

You need

population standard deviation σ

if unknown?

Sample Size for an Interval Estimate of a Population Mean*

- 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.

PROBLEM # 7.12

From past experience, a package-filling machine has been found to have a process **standard deviation of 0.65 ounces** of product weight. A simple random simple is to be selected from the machine's output for the purpose of determining the average weight of product being packed by the machine. **For 95% confidence that the sample mean will not differ from the actual population mean by more than 0.1 ounces,** what sample size is required?

$$n = \frac{z^2\sigma^2}{e^2} = \frac{1.96^2(0.65)^2}{0.10^2} = 162.31,$$

rounded up to 163

PROBLEM # 7.12

From past experience, a package-filling machine has been found to have a process **standard deviation of 0.65 ounces** of product weight. A simple random simple is to be selected from the machine's output for the purpose of determining the average weight of product being packed by the machine. **For 95% confidence that the sample mean will not differ from the actual population mean by more than 0.1 ounces,** what sample size is required?

For the 95% level of confidence, z = 1.96. The maximum likely error is e = 0.10 and the estimated process standard deviation is σ = 0.65. The required sample size is:

$$n = \frac{z^2\sigma^2}{e^2} = \frac{1.96^2(0.65)^2}{0.10^2} = 162.31$$, rounded up to 163

SAMPLE SIZE

POPULATION PROPORTION

Confidence Intervals for a Population Proportion

INTERVAL ESTIMATE OF A POPULATION PROPORTION

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

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

Bowerman, et al. (2017) pp. 367

INTERVAL ESTIMATE OF A POPULATION PROPORTION

- 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 $n(1 - p) \geq 5$.

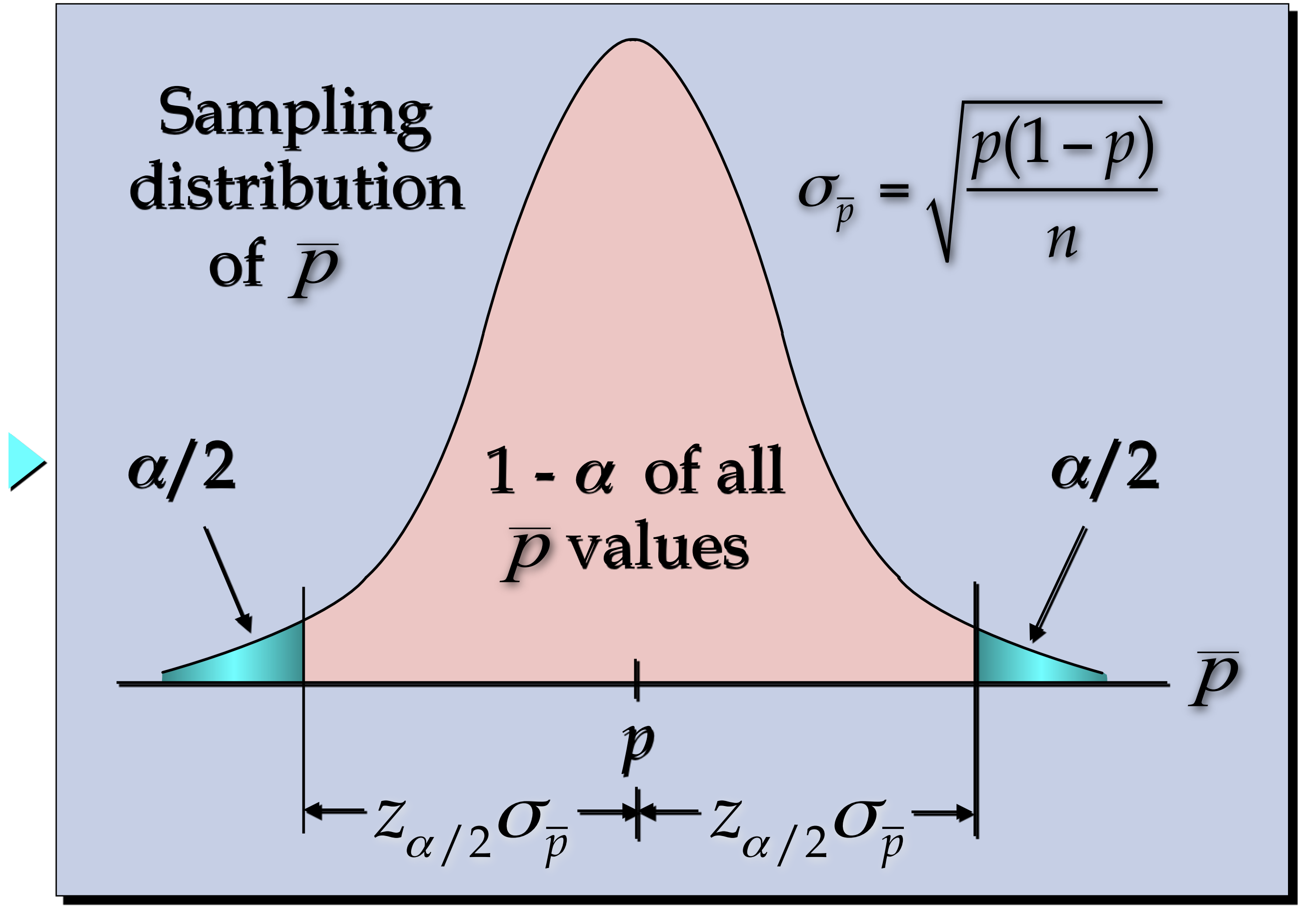
PROBLEM #7.13

Under what conditions is it appropriate to use the normal approximation to the binomial distribution in constructing the confidence interval for the population proportion?

The approximation is satisfactory whenever np and n(1 - p) are both ≥ 5. However, the approximation is better for large values of n and whenever p is closer to 0.5.

INTERVAL ESTIMATE OF A POPULATION PROPORTION

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



INTERVAL ESTIMATE OF A POPULATION PROPORTION

Interval Estimate

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

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

$\bar{p}$ is the sample proportion

Bowerman, et al. (2017) pp. 367

COMM215 First the Foundation, then Innovation

PROBLEM # 7.14

It has been estimated that 48% of U.S. households headed by persons in the 35-44 age group own mutual funds. Assuming this finding to be based on a simple random sample of 1000 households headed by persons in this age group, construct a 95% confidence interval for p= the population proportion of such households that own mutual funds. Source: Investment Company Institute, Investment Company Fact Book 2008, p.72.

$$p \pm z\sqrt{\frac{p(1-p)}{n}} = 0.48 \pm 1.96\sqrt{\frac{0.48(1-0.48)}{1000}} = 0.48 \pm 0.031,$$

or from 0.449 to 0.511

SAMPLE SIZE FOR AN INTERVAL ESTIMATE OF A POPULATION PROPORTION

- Margin of Error

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

Solving for the necessary sample size, we get

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

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

- Necessary Sample Size

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

The planning value p^* can be chosen by:

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

PROBLEM # 7.15

The Chevrolet dealers of a large county are conducting a study to determine the proportion of car owners in the county who are considering the purchase of a new car within the next year. If the population proportion is believed to be no more than 0.15, how many owners must be included in a simple random sample if the dealers want to be 90% confident that the maximum likely error will be no more than 0.02?

$$n = \frac{z^2 p(1-p)}{e^2} = \frac{1.645^2 (0.15)(1-0.15)}{0.02^2} = 862.55,$$

rounded up to 863

COMM215
First the Foundation, then Innovation

PROBLEM # 7.15

The Chevrolet dealers of a large county are conducting a study to determine the proportion of car owners in the county who are considering the purchase of a new car within the next year. If the population proportion is believed to be no more than 0.15, how many owners must be included in a simple random sample if the dealers want to be 90% confident that the maximum likely error will be no more than 0.02?

For the 90% level of confidence, $z = 1.645$. The maximum likely error is $e = 0.02$ and we will estimate the population proportion with $p = 0.15$.
The number of owners who must be included in the sample is:

$$n = \frac{z^2 p(1-p)}{e^2} = \frac{1.645^2(0.15)(1-0.15)}{0.02^2} = 862.55$$, rounded up to 863

PROBLEM # 7.16

Refer to Problem 7.15, suppose that (unknown to the dealers) the actual population proportion is really 0.35. If they use their estimated value ($p \leq 0.15$) in determining the sample size and then conduct the study, will their maximum likely error be greater than, equal to, or less than 0.02? Why?

$$e = z\sqrt{\frac{p(1-p)}{n}} = 1.645\sqrt{\frac{0.35(1-0.35)}{863}} = 0.027,$$

the new maximum likely error

COMM215 First the Foundation, then Innovation

PROBLEM # 7.16

Refer to Problem 7.15, suppose that (unknown to the dealers) the actual population proportion is really 0.35. If they use their estimated value ($p \leq 0.15$) in determining the sample size and then conduct the study, will their maximum likely error be greater than, equal to, or less than 0.02? Why?

The maximum likely error will be greater than 0.02.
This is because when $p = 0.35$ a larger sample size is needed than when $p = 0.15$.

$$e = z\sqrt{\frac{p(1-p)}{n}} = 1.645\sqrt{\frac{0.35(1-0.35)}{863}} = 0.027$$, the new maximum likely error

COMM215
First the Foundation, then Innovation

PROBLEM # 7-11*

For df=25, determine the value of A that corresponds to each of the following probabilities:

a. $P(t \geq A) = 0.025$

$P(t \geq A) = 0.025$. From the 0.025 column and the d.f. = 25 row of the t table, A = 2.060.

b. $P(t \leq A) = 0.10$

$P(t \leq A) = 0.10$. Referring to the 0.10 column and the d.f. = 25 row of the t table, the value of t corresponding to a right-tail area of 0.10 is t = 1.316. Since the curve is symmetrical, the value of t for a left-tail area of 0.10 is A = -1.316.

c. $P(-A \leq t \leq A) = 0.98$

$P(-A \leq t \leq A) = 0.98$. In this case, each tail will have an area of (1 - 0.98)/2 = 0.01. Referring to the 0.01 column and the d.f. = 25 row of the t table, A = 2.485.