Chapter 8

Confidence Interval Estimation

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

We have a problem…

- Glob Oil Inc. produces hydraulic fluids that are typically used in manufacturing equipment.
- It has been reported that one of its recent product, MBH-104, performs poorly when the temperature of the machineries increase.
- You are asked to find out two things:
  - The **mean** temperature of the fluid, when used in a given machinery
  - The **proportion** of products that their temperature increases more than a given number
- But here's the problem:

We have already shipped out a lot of orders, and we have only a limited amount of previously produced products in our inventory.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

We have a problem…

- What are we looking for?

  Two **population parameters**:

  - The population mean of the temperature
  - The population proportion of failed fluids

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

We have a problem…

- What are we looking for?

  Two **population parameters**:

  - The population mean of the temperature
  - The population proportion of failed fluids
- **But we have no access to the population**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

We have a problem…

- What are we looking for?

  Two **population parameters**:

  - The population mean of the temperature
  - The population proportion of failed fluids
- **But we have no access to the population**
- What are we going to employ?

  Two **sample statistics**:

  - The sample mean of the temperature
  - The sample proportion of the temperature

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

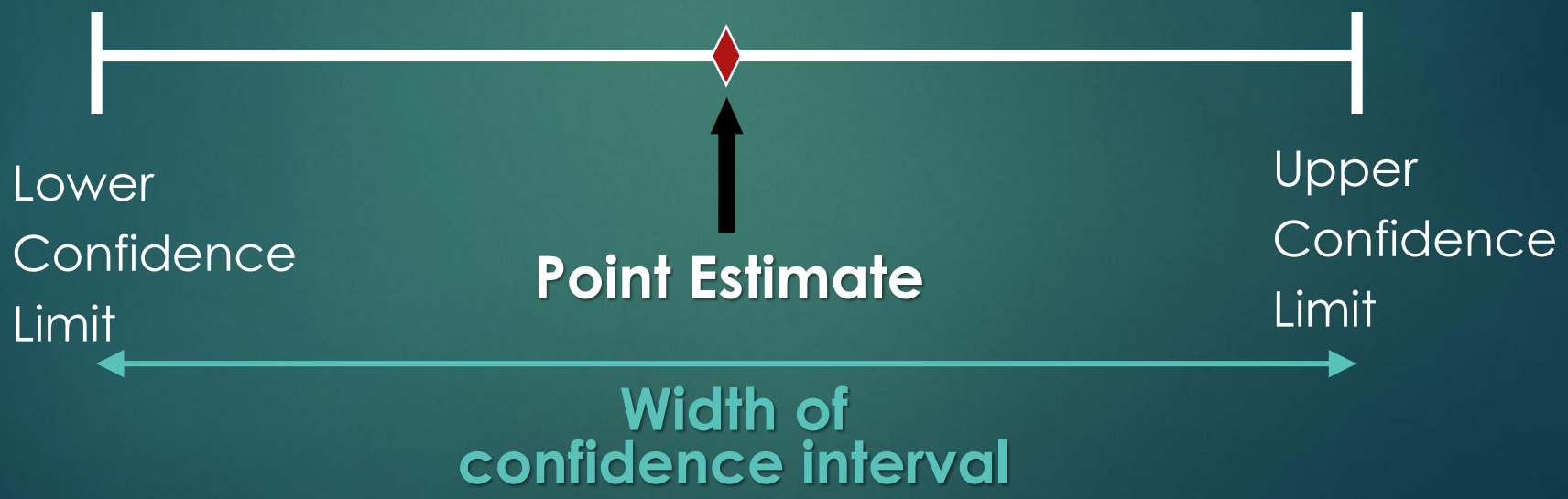
Let's get to work, then…

We need a Point Estimate along with a Confidence Interval

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Point and Interval Estimates

- A point estimate is a single number
- A confidence interval provides additional information about the variability of the estimate



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Point Estimates

We already know about the Point Estimates

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

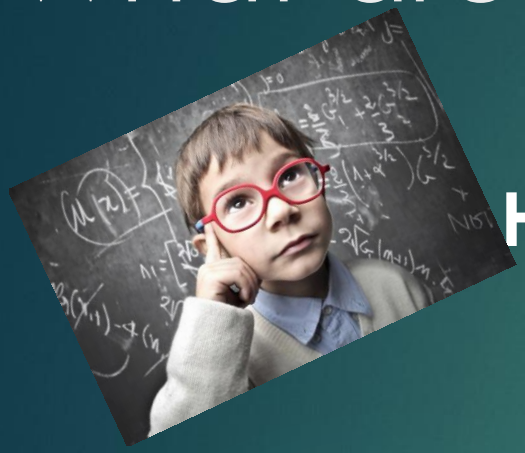
Copyright © 2015, 2012, 2009 Pearson Education, Inc.

What's Confidence Interval, then?

Now that we have a Point Estimate, let's figure out a Confidence Interval

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

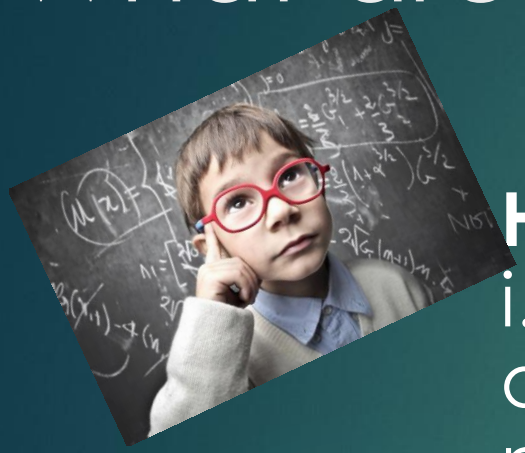
What are Confidence Intervals?



How accurate is the point estimate?

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

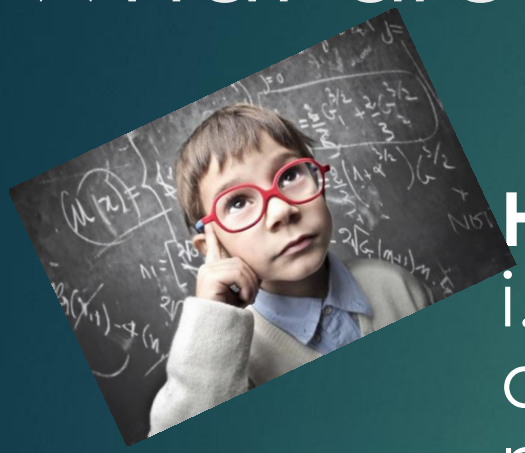
What are Confidence Intervals?



How accurate is the point estimate?
i.e., How much uncertainty is associated with a point estimate of a population parameter?

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

What are Confidence Intervals?



How accurate is the point estimate? i.e., How much uncertainty is associated with a point estimate of a population parameter?

- An **interval estimate** provides more information about a population characteristic than does a point estimate.

- Such interval estimates are called **confidence intervals**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

What are Confidence Intervals?

- An interval gives a **range of values**:
  - Taking into consideration **the variation in sample statistics** from sample to sample
    - Remember the different means that we obtained from different samples in Oddland?

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

What are Confidence Intervals?

- An interval gives a **range of values**:
  - Taking into consideration the variation in sample statistics from sample to sample
    - Remember the different means that we obtained from different samples in Oddland?
  - Based on **observations from only one sample**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

What are Confidence Intervals?

- An interval gives a **range of values**:
  - Taking into consideration the variation in sample statistics from sample to sample
    - Remember the different means that we obtained from different samples in Oddland?
  - Based on observations from only one sample
  - Gives information about **closeness to unknown population parameters**; i.e. the accuracy

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

What are Confidence Intervals?

- An interval gives a **range of values**:
  - Taking into consideration the variation in sample statistics from sample to sample
    - Remember the different means that we obtained from different samples in Oddland?
  - Based on observations from only one sample
  - Gives information about closeness to unknown population parameters; i.e. the accuracy
  - Stated in terms of **level of confidence**
    - e.g. 95% confident, 99% confident
    - Can never be 100% confident

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Remember this example from Ch. 7?

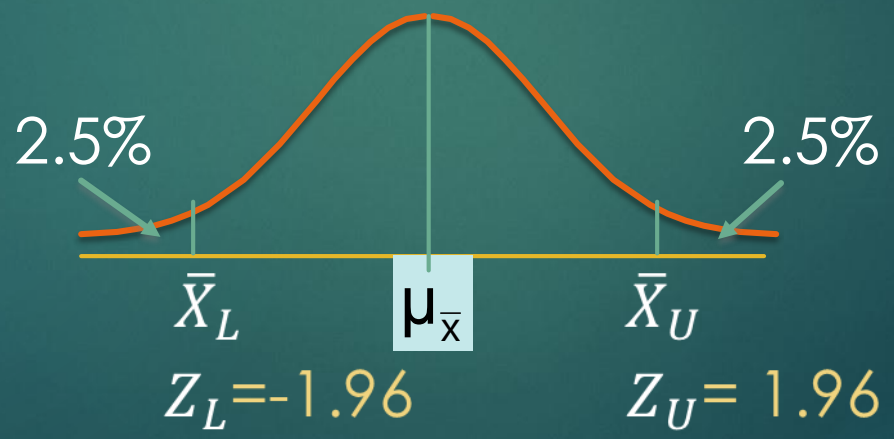
- A population is normally distributed
- **μ** is the population mean, and **σ** is the population standard deviation.
- We happen to know their values: $\mu = 368$, $\sigma = 15$

Find a **symmetrically distributed interval** around **μ** that will include 95% of the sample means when sample sizes are equal to 25: i.e., $n = 25$.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Remember this example from Ch. 7?

- Since the interval contains 95% of the sample means 5% of the sample means will be outside the interval
- Since the interval is symmetric 2.5% will be above the upper limit and 2.5% will be below the lower limit.
- Using Excel, the Z score with probability 2.5% (i.e., P=0.0250) is -1.96 and the Z score with probability 97.5% (P=0.9750) is 1.96. Thus, Z_L=-1.96 and Z_U=1.96



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Remember this example from Ch. 7?

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow \bar{X} = \mu + Z\frac{\sigma}{\sqrt{n}}$$

- Calculating the lower limit of the interval

$$\bar{X}_L = \mu + Z_L \frac{\sigma}{\sqrt{n}} = 368 + (-1.96)\frac{15}{\sqrt{25}} = 362.12$$

- Calculating the upper limit of the interval

$$\bar{X}_U = \mu + Z_U \frac{\sigma}{\sqrt{n}} = 368 + (1.96)\frac{15}{\sqrt{25}} = 373.88$$

- **95% of all sample means of sample size 25 are between 362.12 and 373.88**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

What did we do in this example?

- We knew that the population distribution is normal
- We knew the population parameters.
  - i.e.; **μ** and **σ** were known to us.
- We used Central Limit Theorem
  - Using normal probabilities, we computed an interval around the population mean that includes 95% of the sample means
  - The above is equivalent to this:

    We found the percentage of sample means that are within certain distance from the population mean, as determined by $\bar{X}_L$ and $\bar{X}_U$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

In real life, however…

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

In real life, however…

- We often deal with one sample.
  - i.e., we are unable to draw the sampling distribution of the mean by taking different samples repeatedly.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

In real life, however…

- We often deal with one sample.
  - i.e., we are unable to draw the sampling distribution of the mean by taking different samples repeatedly.
- We have no idea… yes, no idea about the population parameters such as population mean or standard deviation. We just know that they exist
  - i.e., we cannot identify $\bar{X}_L$ and $\bar{X}_U$ even after the Z-value for given probability is known to us.
  - At best, for I=95% and n=100, we have :
  - $\bar{X}_L = \mu + (1.96)\frac{\sigma}{\sqrt{100}}$ and $\bar{X}_U = \mu - (1.96)\frac{\sigma}{\sqrt{100}}$
  - We don't know what **μ** and **σ** are.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

The interval of interest:

- We would like to identify

$$\mu \mp Z \frac{\sigma}{\sqrt{n}}$$

- In real life, we only have **Z** (given the width of the interval; e.g. 95%) and **n** (given the sample in hand). For instance, we have:

$$\mu \mp (1.96) \frac{\sigma}{\sqrt{100}}$$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Let's move forward by making an assumption:

- Though **we do not know μ**, let's assume that **we know σ**.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

The interval of interest:

- Thus, we would like to identify

$$\mu \mp Z \frac{\sigma}{\sqrt{n}}$$

- Under this assumption, we have **Z** (given the width of the interval; e.g. 95%), **n** (given the sample in hand), and **σ**. For instance, we have:

$$\mu \mp (1.96) \frac{15}{\sqrt{100}}$$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Let's move forward…

- What should we do about **μ**, the population mean, now that it is unknown?

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Let's move forward…

- What should we do about **μ**, the population mean, now that it is unknown?
- Why don't we use the sample mean, $\bar{X}$? After all, that is all that we have access to

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Let's move forward…

- Here's the sample in hand: n=25, $\bar{X}$=362.3
- The confidence level: 95%; i.e. Z=∓1.96

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Let's move forward…

- Here's the sample in hand: n=25, $\bar{X}$=362.3
- The confidence level: 95%; i.e. Z=∓1.96
- NOTE: We also know that the population is normal, so we are not really worried about the medium-size sample. The Central Limit Theorem Condition is met.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Let's move forward…

- Here's the sample in hand: n=25, $\bar{X}$=362.3
- The confidence level: 95%; i.e. Z=∓1.96
- NOTE: We also know that the population is normal, so we are not really worried about the medium-size sample. The Central Limit Theorem Condition is met.
- Thus, we use the following interval to estimate the population mean, μ:
- $Interval = \bar{X} \mp (1.96)\frac{15}{\sqrt{25}} = 362.3 \mp (1.96)\frac{15}{\sqrt{25}}$

$$356.42 \leq \mu \leq 368.18$$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Other possible intervals

- Since the true population mean is 368, the resulting interval ($356.42 \leq \mu \leq 368.18$) actually includes the true population parameter.
- What is we had other samples?
- Let's do a thought experiment:

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Other possible intervals

Sample #	$\overline{X}$	Lower Limit	Upper Limit	Contain μ?
1	362.30	356.42	368.18	Yes

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Other possible intervals

Sample #	$\overline{X}$	Lower Limit	Upper Limit	Contain μ?
1	362.30	356.42	368.18	Yes
2	369.50	363.62	375.38	Yes

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Other possible intervals

Sample #	$\overline{X}$	Lower Limit	Upper Limit	Contain μ?
1	362.30	356.42	368.18	Yes
2	369.50	363.62	375.38	Yes
3	360.00	354.12	365.88	No

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Other possible intervals

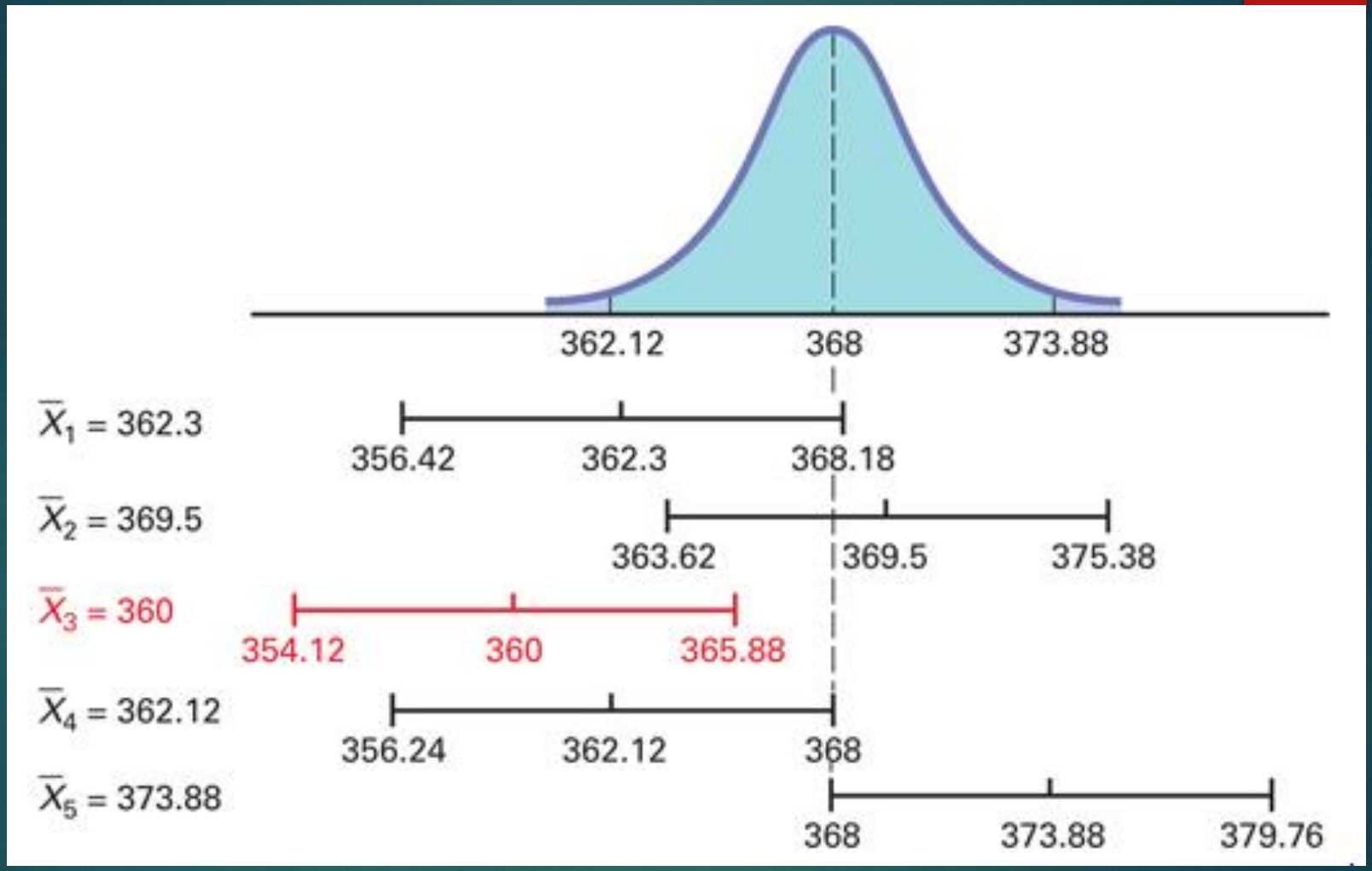
Sample #	$\overline{X}$	Lower Limit	Upper Limit	Contain μ?
1	362.30	356.42	368.18	Yes
2	369.50	363.62	375.38	Yes
3	360.00	354.12	365.88	No
4	362.12	356.24	368.00	Yes

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Other possible intervals

Sample #	$\overline{X}$	Lower Limit	Upper Limit	Contain μ?
1	362.30	356.42	368.18	Yes
2	369.50	363.62	375.38	Yes
3	360.00	354.12	365.88	No
4	362.12	356.24	368.00	Yes
5	373.88	368.00	379.76	Yes

Copyright © 2015, 2012, 2009 Pearson Education, Inc.



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

The confidence interval

- The 95% confidence interval for the population mean is the following:

$$356.42 \leq \mu \leq 368.18$$

- We can say:

WE ARE 95% CONFIDENT THAT THE POPULATION PARAMETER IS BTW. 356.42 and 368.18

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

The confidence interval

- In practice, you only take one sample of size **n**
- You do not know **μ**, so you do not know if the interval actually contains **μ**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

The confidence interval

- In practice, you only take one sample of size **n**
- You do not know **μ,** so you do not know if the interval actually contains **μ**
- However, if you:
  - take all possible samples of **n**
  - and compute their 95% confidence interval,

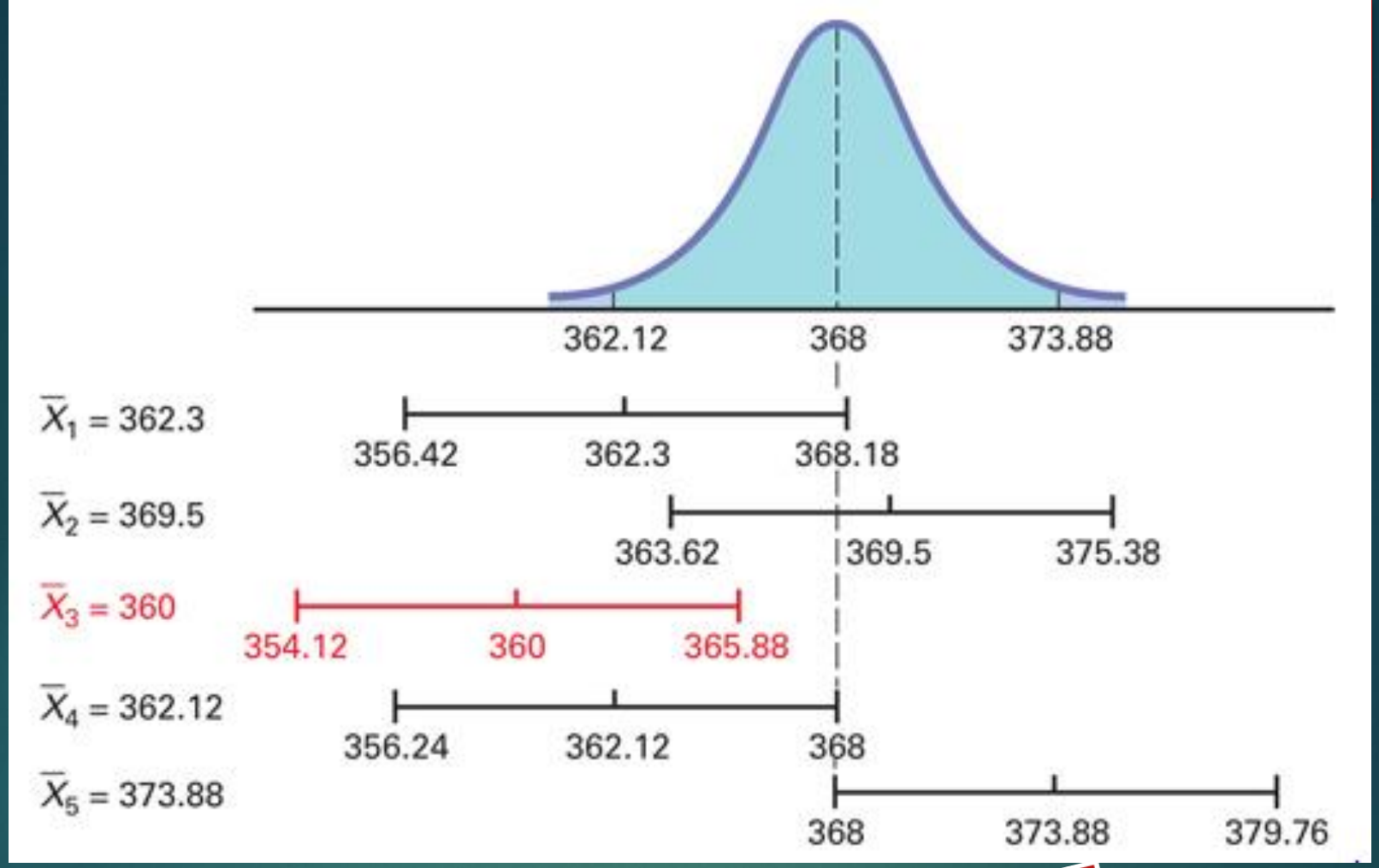
    e.g. $\bar{X} \mp (1.96)\frac{15}{\sqrt{25}}$,
- you do know that 95% of the intervals formed in this manner will contain **μ**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

The confidence interval

- Taking 100 samples and constructing the 95% confidence interval (for which Z-values are -1.96 and 1.96),:
  - $\bar{X}_1 \mp (1.96)\frac{15}{\sqrt{25}}$
  - $\bar{X}_2 \mp (1.96)\frac{15}{\sqrt{25}}$
  - $\bar{X}_3 \mp (1.96)\frac{15}{\sqrt{25}}$
  - …
  - $\bar{X}_{100} \mp (1.96)\frac{15}{\sqrt{25}}$
- we can say that 95 intervals (out of 100) are expected to include the population mean

Copyright © 2015, 2012, 2009 Pearson Education, Inc.



$$\text{Interval: } \mu \mp Z\frac{\sigma}{\sqrt{n}} \rightarrow \bar{X} \mp Z\frac{\sigma_{known}}{\sqrt{n}}$$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Population Mean= 69.637

Sample	n	Mean	Std Dev	Minimum	Median	Maximum	Range	95% Conf. Int.
S01	10	74.15	13.39	56.10	76.85	97.70	41.60	(67.70, 80.60)
S02	10	61.10	10.60	46.80	61.35	79.50	32.70	(54.65, 67.55)
S03	10	74.36	6.50	62.50	74.50	84.00	21.50	(67.91, 80.81)
S04	10	70.40	12.80	47.20	70.95	84.00	36.80	(63.95, 76.85)
S05	10	62.18	10.85	47.10	59.70	84.00	36.90	(55.73, 68.63)
S06	10	67.03	9.68	51.10	69.60	83.30	32.20	(60.58, 73.48)
S07	10	69.03	8.81	56.60	68.85	83.70	27.10	(62.58, 75.48)
S08	10	72.30	11.52	54.20	71.35	87.00	32.80	(65.85, 78.75)
S09	10	68.18	14.10	50.10	69.95	86.20	36.10	(61.73, 74.63)
S10	10	66.67	9.08	57.10	64.65	86.10	29.00	(60.22, 73.12)
S11	10	72.42	9.76	59.60	74.65	86.10	26.50	(65.97, 78.87)
S12	10	76.26	11.69	50.10	80.60	87.00	36.90	(69.81, 82.71)
S13	10	65.74	12.11	47.10	62.15	86.10	39.00	(59.29, 72.19)
S14	10	69.99	10.97	51.00	73.40	84.60	33.60	(63.54, 76.44)
S15	10	75.76	8.60	61.10	75.05	87.80	26.70	(69.31, 82.21)
S16	10	67.94	9.19	56.70	67.70	87.80	31.10	(61.49, 74.39)
S17	10	71.05	10.48	50.10	71.15	86.20	36.10	(64.60, 77.50)
S18	10	71.68	7.96	55.60	72.35	82.60	27.00	(65.23, 78.13)
S19	10	70.97	9.83	54.40	70.05	84.00	30.20	(64.52, 77.42)
S20	10	74.48	8.80	62.00	76.25	85.70	23.70	(68.03, 80.93)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

General Formula for Confidence Interval Estimations

- 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

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Level

- The **level of confidence** that the interval will contain the unknown population parameter
  - **e.g. With 95% confidence, we know that the population mean is included in the following interval: [356.42, 368.18]**
- A percentage (less than 100%)

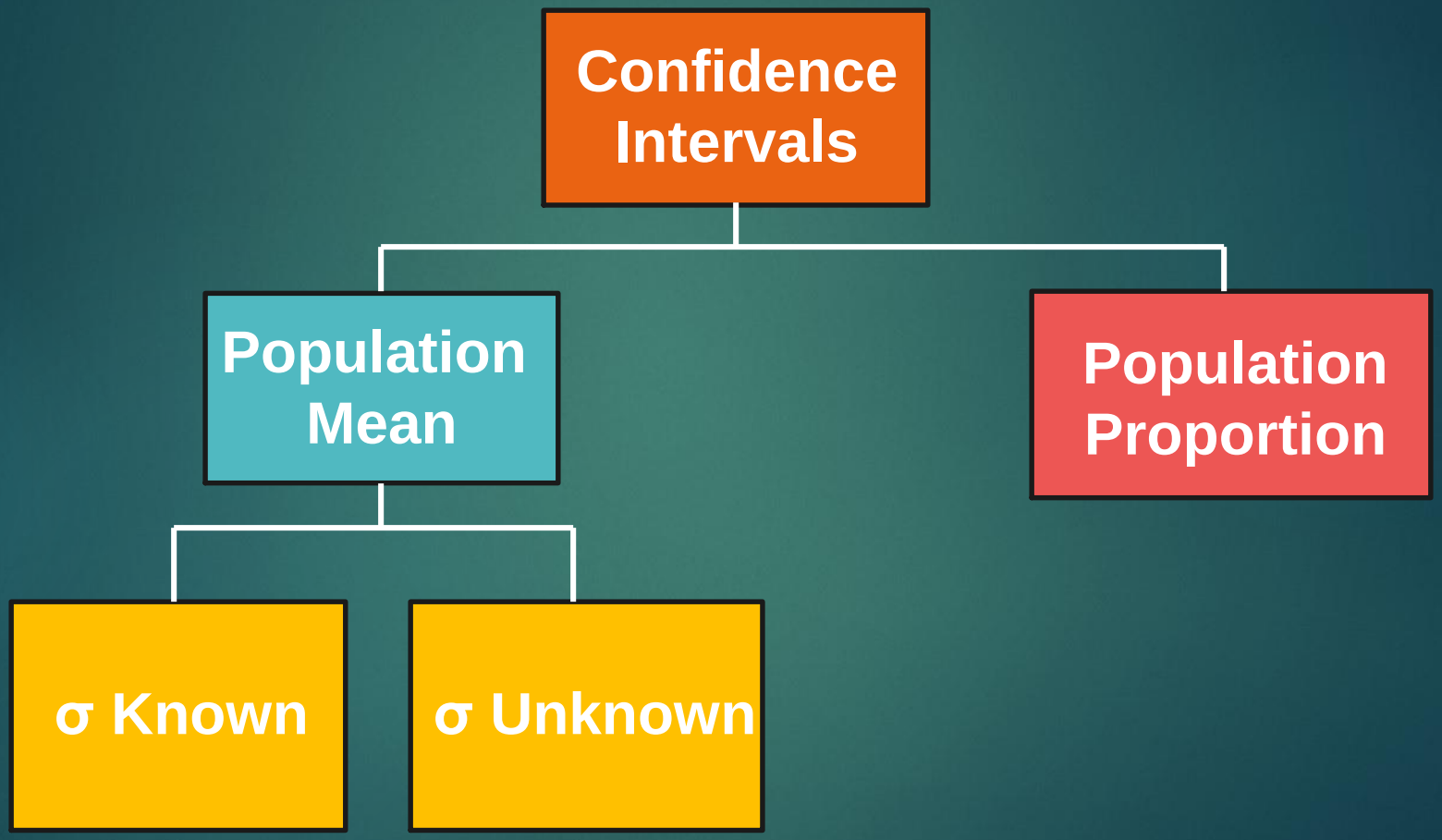
Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Level, (1-α)

- It is conventional to state the confidence level using α in the following fashion:
  - Suppose confidence level = 95%
  - Also written **$(1 - \alpha) = 0.95$, so $\alpha = 0.05$**
    - 95% of all the confidence intervals that can be constructed will contain the unknown true parameter
    - Yet, a specific interval either will contain or will not contain the true parameter

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Intervals



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Interval for μ (σ Known)

- Assumptions
  - Population standard deviation **σ is known**
  - Population is normally distributed
  - If population is not normal, use large sample (n > 30)

- Confidence interval estimate:

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

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Interval for μ (σ Known)

- Assumptions
  - Population standard deviation **σ is known**
  - Population is normally distributed
  - If population is not normal, use large sample (n > 30)

- Confidence interval estimate:

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

Point Estimate ± (Critical Value)(Standard Error)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Finding the Critical Value, $Z_{a/2}$

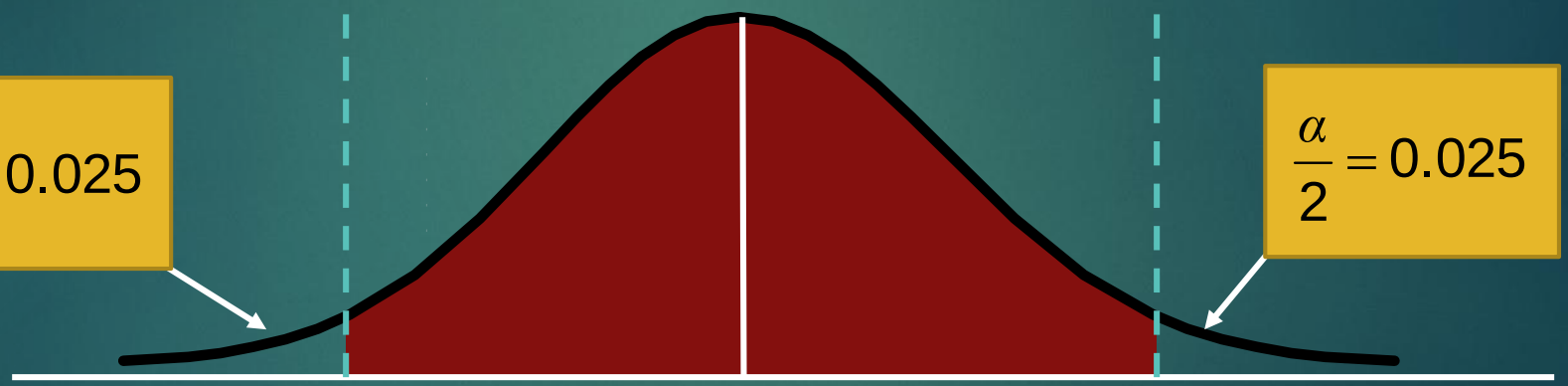
- Consider a 95% confidence interval:

$$Z_{\alpha/2} = \pm 1.96$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.025$$

$$\frac{\alpha}{2} = 0.025$$

$$\frac{\alpha}{2} = 0.025$$



Z units:	$Z_{\alpha/2} = -1.96$	**0**	$Z_{\alpha/2} = 1.96$
X units:	**Lower Confidence Limit**	**Point Estimate**	**Upper Confidence Limit**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Common Levels of Confidence

- Commonly used confidence levels are 90%, 95%, and 99%

Confidence Level	Confidence Coefficient, $1-\alpha$	$Z_{\alpha/2}$ value
80%	0.80	1.28
90%	0.90	1.645
95%	0.95	1.96
98%	0.98	2.33
99%	0.99	2.58
99.8%	0.998	3.08
99.9%	0.999	3.27

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Common Levels of Confidence

- Commonly used confidence levels are:

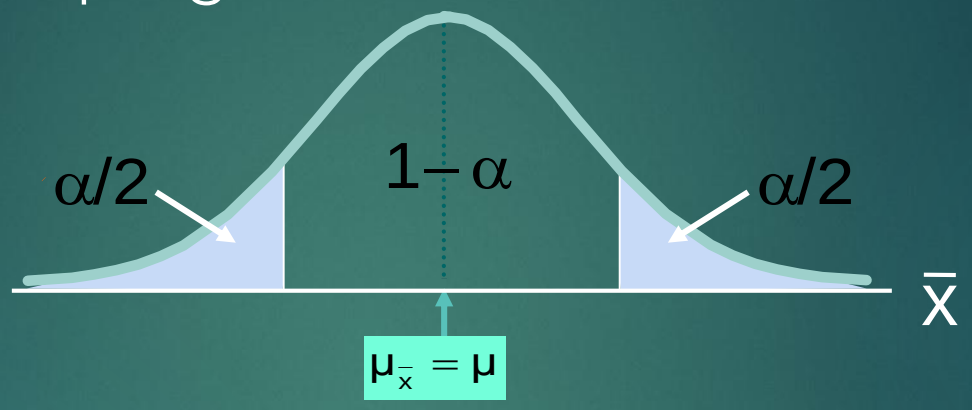
  90%, 95%, and 99%

- Confidence level 90% → Critical Value: $Z_{0.5} = 1.645$
- Confidence level 95% → Critical Value: $Z_{0.25} = 1.96$
- Confidence level 99% → Critical Value: $Z_{0.05} = 2.58$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Intervals and Level of Confidence

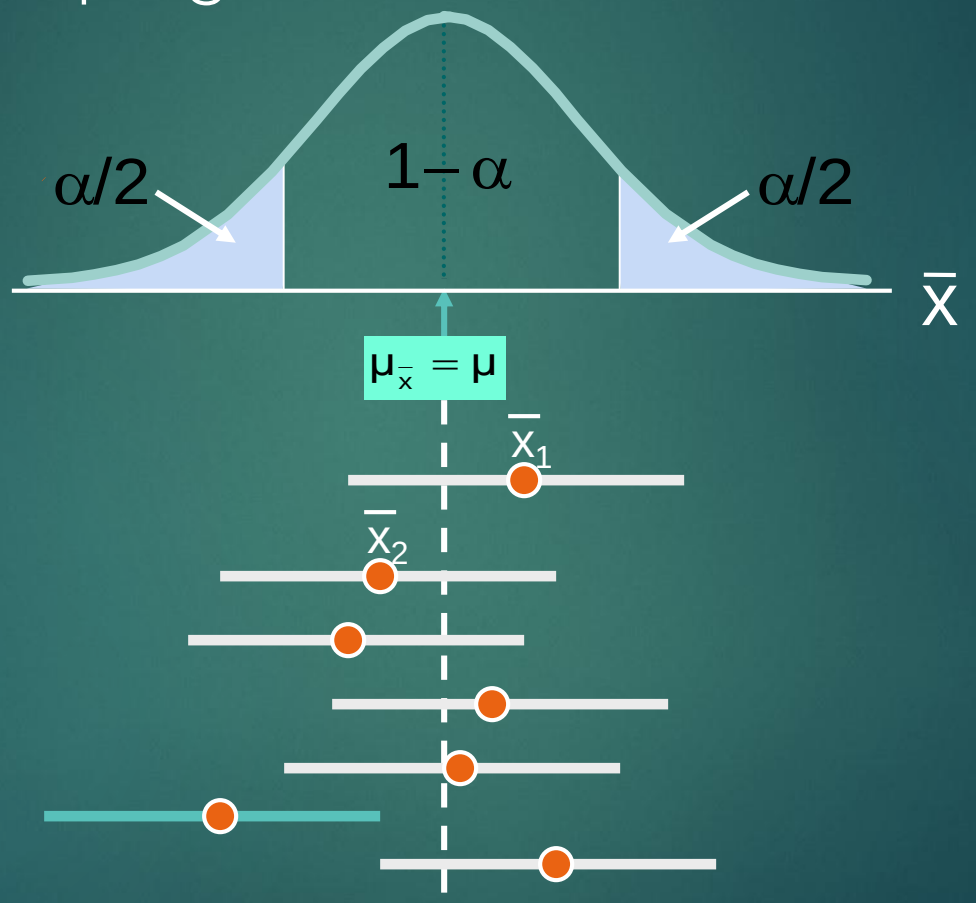
Sampling Distribution of the Mean



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Intervals and Level of Confidence

Sampling Distribution of the Mean

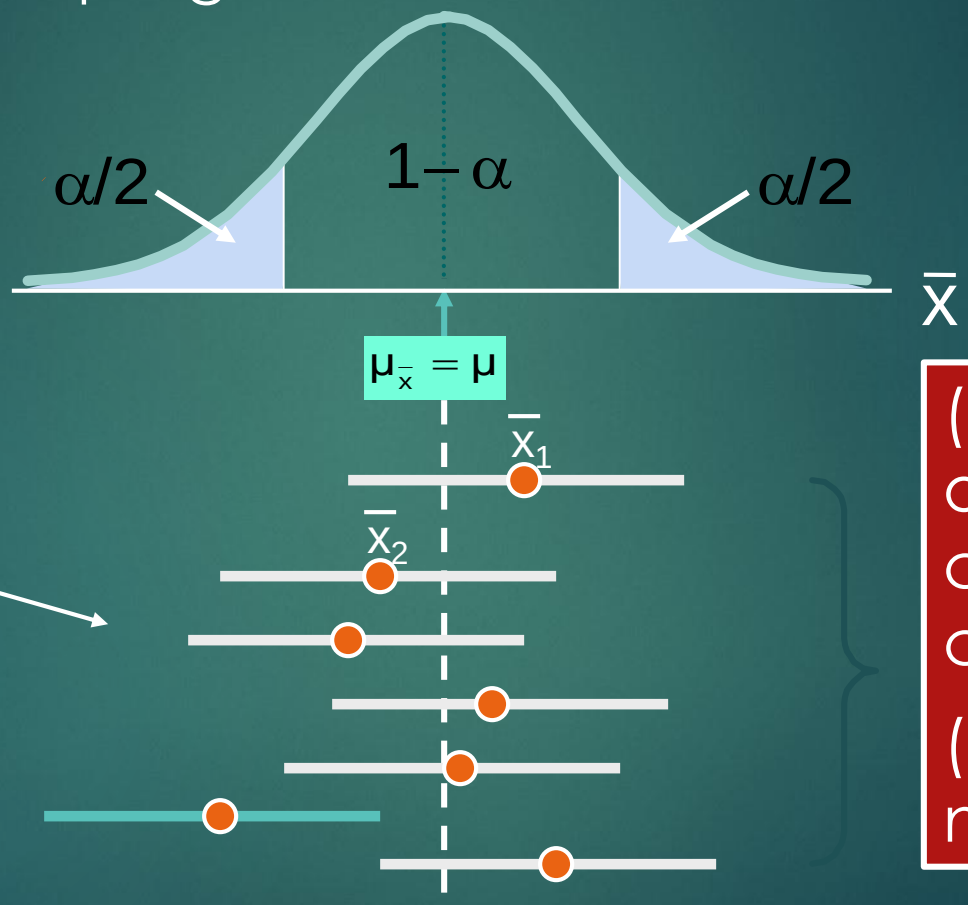


Confidence Intervals

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Intervals and Level of Confidence

Sampling Distribution of the Mean



Intervals extend from

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

to

$$\bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

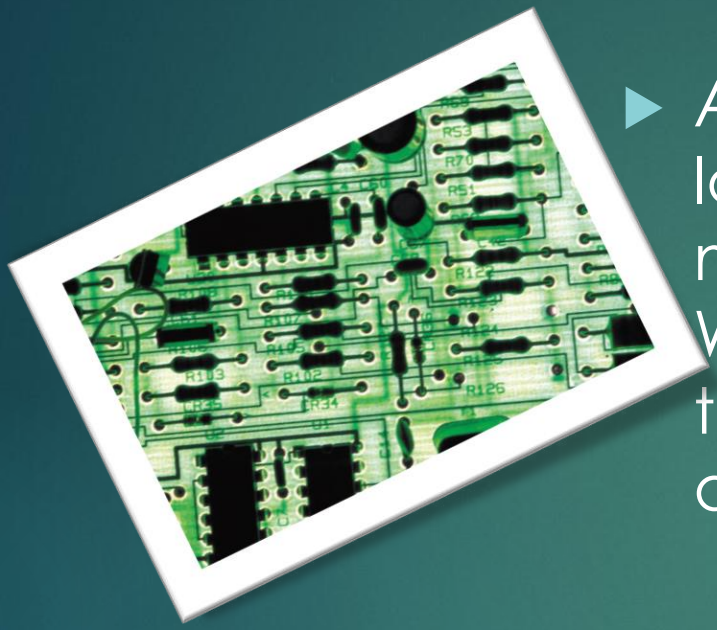
(1-α)100% of intervals constructed contain μ;

(α)100% do not.

Confidence Intervals

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

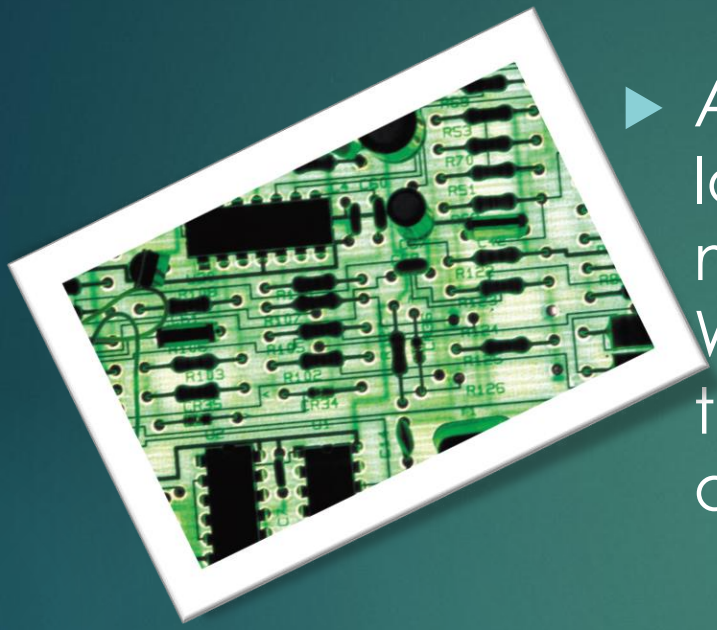
Example



- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is 0.35 ohms.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example



- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is 0.35 ohms.

- **Determine a 95% confidence interval for the true mean resistance of the population. (Reminder: $Z_{0.25} = 1.96$)**

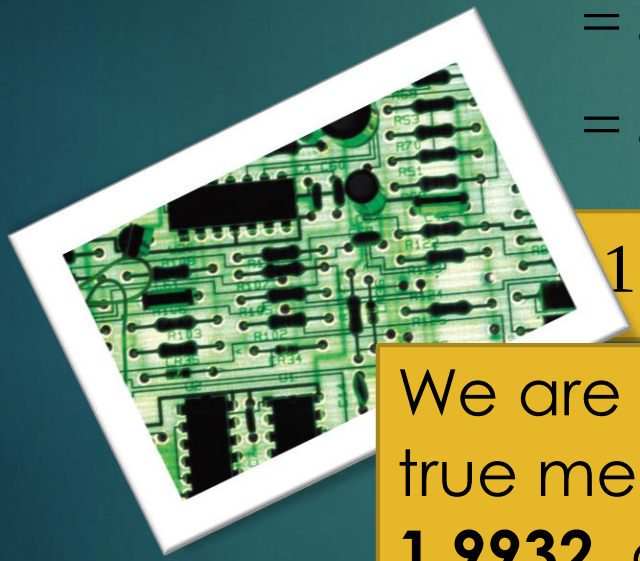
Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

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

$$= 2.20 \pm 1.96(0.35/\sqrt{11})$$

$$= 2.20 \pm 0.2068$$



$$1.9932 \le \mu \le 2.4068$$

We are 95% confident that the true mean resistance is between **1.9932** and **2.4068** ohms.

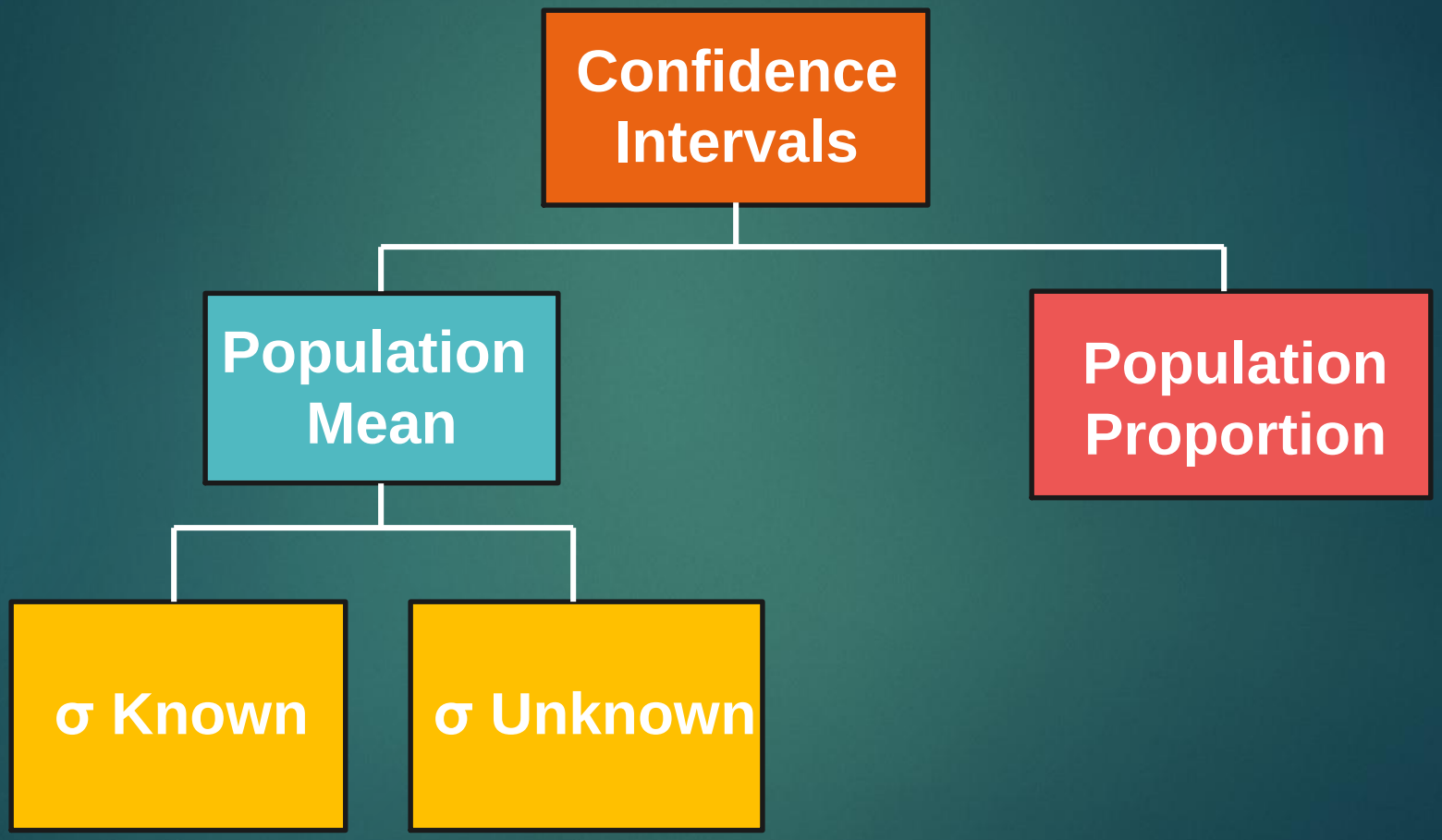
Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Assignment:

Homework 8.1.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Intervals



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Do You Ever Truly Know σ?

- Probably not!

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Do You Ever Truly Know σ?

- Probably not!
- If there is a situation where σ is known then μ is also known (since to calculate σ you need to know μ.)

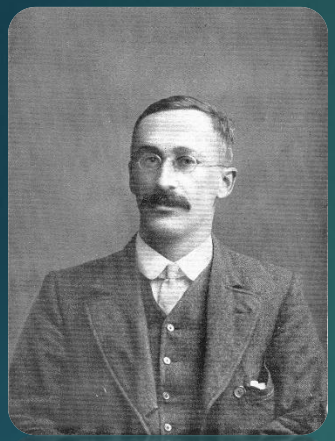
Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Do You Ever Truly Know σ?

- Probably not!
- 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.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Do You Ever Truly Know σ?



- William Gosset was a statistician, who was working for Guinness back in early twentieth century.
- He was working with small samples, and he didn't know the population mean and standard deviation
- He developed a probability distribution, which is known as **Student's *t* Distribution**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Interval for μ (σ Unknown)

- If the population standard deviation σ is unknown, we can substitute the sample standard deviation, S
- This introduces extra uncertainty, since sample standard deviation varies from sample to sample – like the sample mean.
- So we use the *t*-distribution, which takes this uncertainty into account, instead of the normal distribution

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Student's *t* Distribution

- **X** is a random variable, which is distributed **normally**.
- Using the observations in the sample, we can form the following sample statistic:

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

- Gosset showed that this sample statistics has a ***t*-distribution** with a **degree of freedom: n-1**

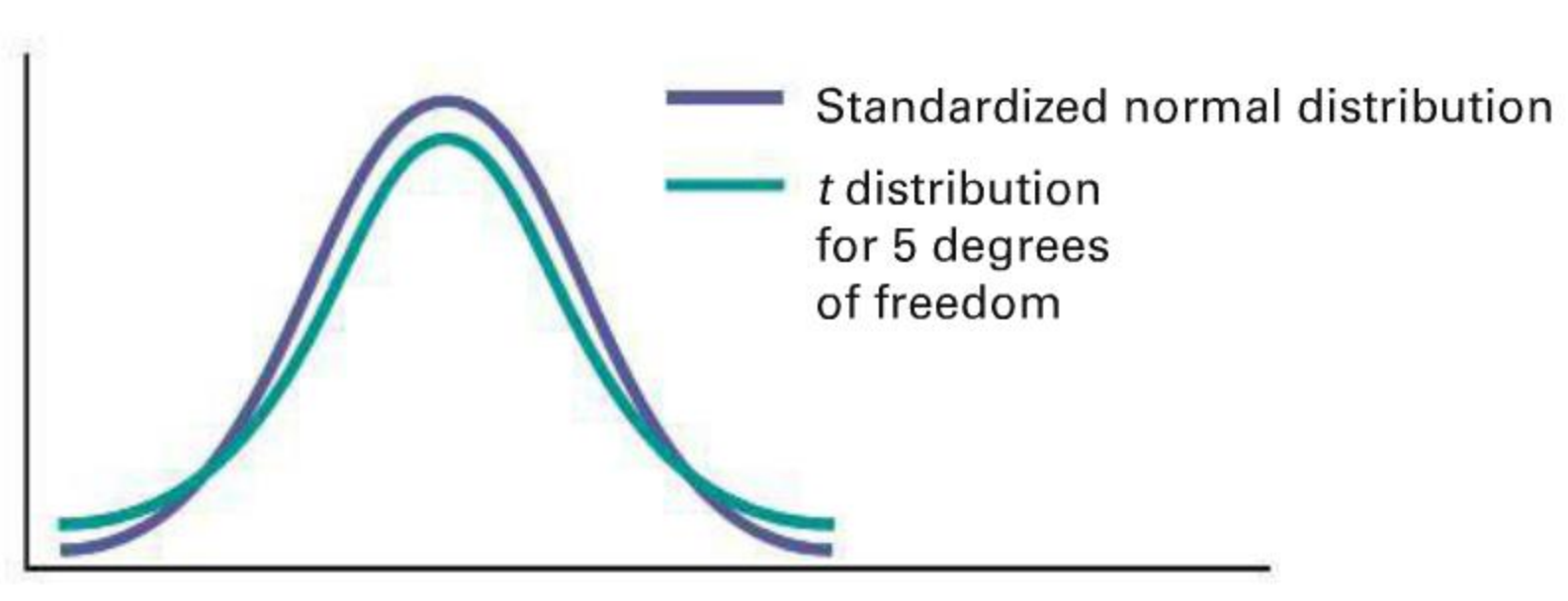
Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Student's *t* Distribution

- Almost like a normal distribution, Student's *t* Distribution is **bell-shaped** and **symmetrical**.
- **Mean, Median, and Mode** are identical in this distribution – **and they are equal to zero**.
- It has **relatively fat tails**. Since **S** is an additional sample statistic, which varies from sample to sample, the Student's t Distribution has more observations in the tails and less in the center, compared to a Standardized Normal Distribution.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Student's *t* Distribution



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

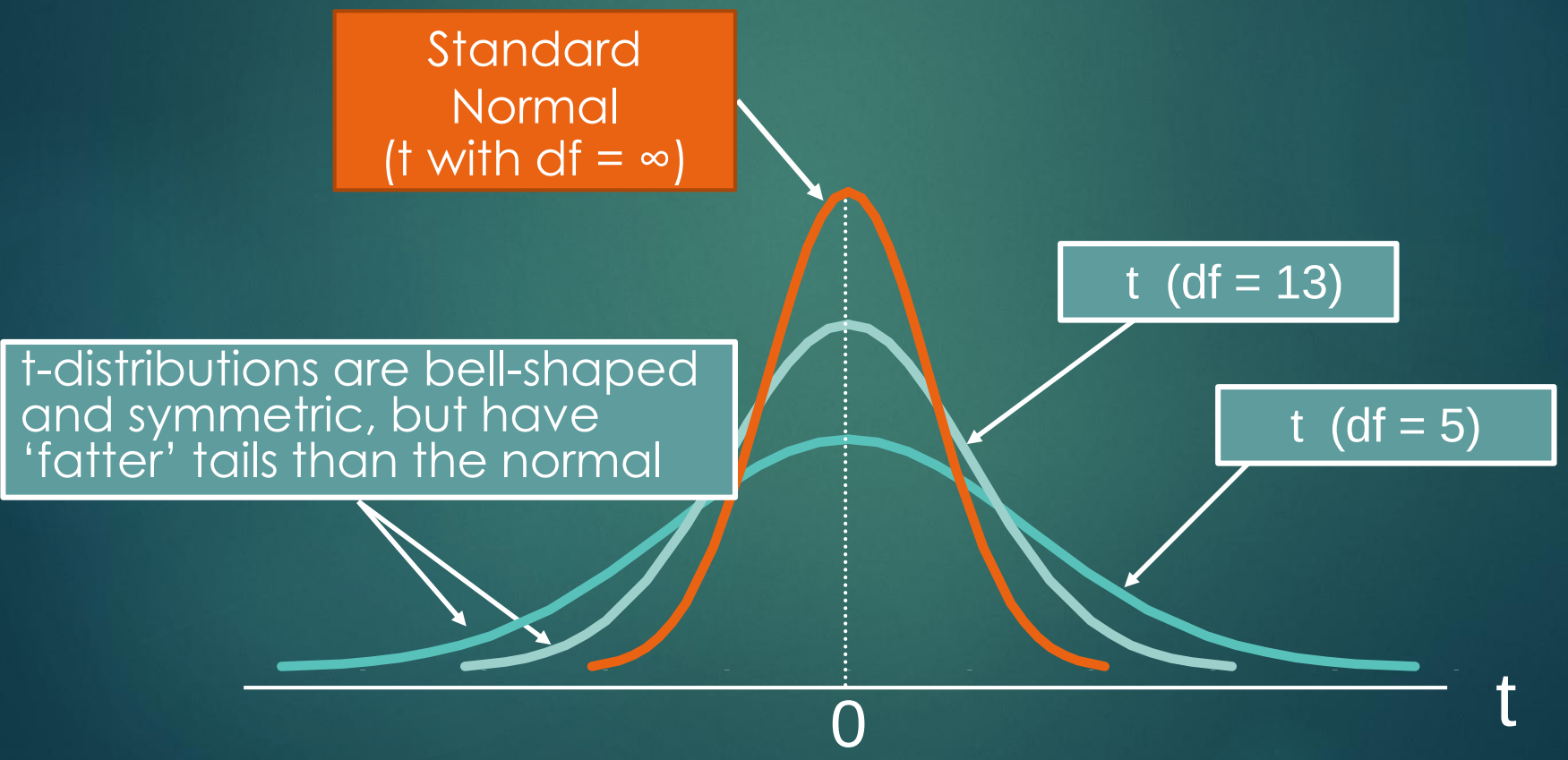
Student's *t* Distribution

- As the sample size increases (which causes an increase in degree of freedom), Student's *t* Distribution becomes more like Standardized Normal Distribution.
- Why? Because S can approximate σ better.
- How large should the sample be? About 120.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Student's *t* Distribution

t ⟶ **Z** as **n** increases



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Student's *t* Distribution

- Is the Student's t Distribution useful **if X is not distributed normally**?
  - As long as **the sample size is large enough** and **the population is not very skewed**, we can safely use Student's t Distribution to estimate the confidence interval around the population mean.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Student's *t* Distribution

How can we identify the t-values (i.e. the critical values)?

	Upper Tail Area		
df	.10	**.05**	.025
1	3.078	6.314	12.706
2	1.886	2.920	4.303
3	1.638	2.353	3.182

This is part of a table for Critical Values of *t*. You can find it on PP. 746-7.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

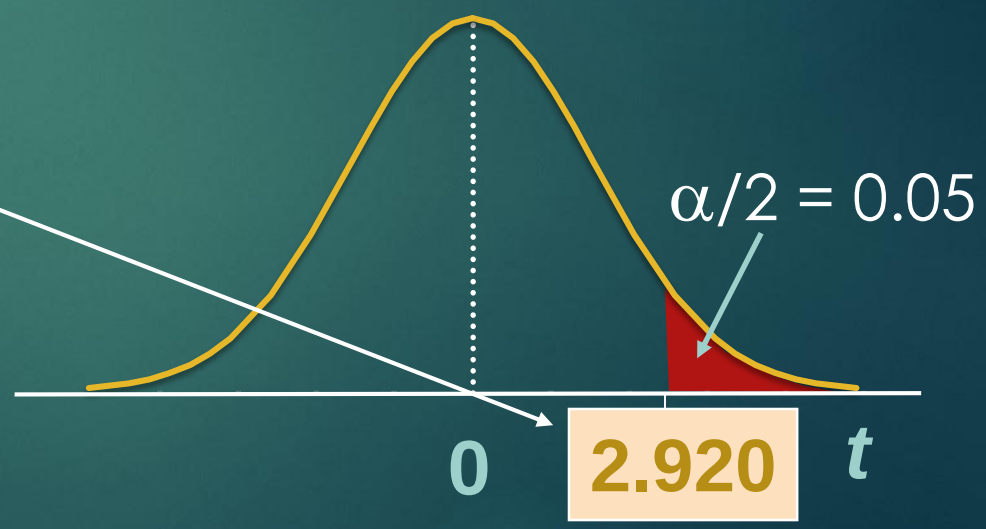
Student's *t* Distribution

How can we identify the t-values (i.e. the critical values)?

	Upper Tail Area		
df	.10	**.05**	.025
1	3.078	6.314	12.706
2	1.886	**2.920**	4.303
3	1.638	2.353	3.182

The body of the table contains t values, not probabilities

$n = 3$
$df = n - 1 = 2$
$\alpha = 0.10$
$\alpha/2 = 0.05$



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Selected t-distribution values

Confidence Level	t (10 d.f.)	t (20 d.f.)	t (30 d.f.)	Z (∞ d.f.)
0.80	1.372	1.325	1.310	1.28
0.90	1.812	1.725	1.697	1.645
0.95	2.228	2.086	2.042	1.96
0.99	3.169	2.845	2.750	2.58

t $\longrightarrow$ **Z** as **n** increases

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Student's *t* Distribution

- What is **degree of freedom (df)**?
  - It refers to **the number of sample values that are free to vary** when a certain sample statistic (like sample mean) is computed.
  - To compute **S** in a sample of size **n**, we need to compute $\bar{X}$.
  - **Given the mean ($\bar{X}$), how many values do you need to know before you can determine the remainder of the values? That is your degree of freedom.**
  - $\bar{X}$ is known. Once we know **n-1** values, we can determine the last value.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

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 $\mathbf{X_3}$?

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)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Interval for μ (σ Unknown)

- Assumptions
  - Population standard deviation is unknown
  - Population is normally distributed
  - If population is not normal, use large sample (n > 30)
- Use Student's *t* Distribution
- Confidence Interval Estimate:

$$\boxed{\bar{X}} \pm \boxed{t_{\alpha/2}} \boxed{\frac{S}{\sqrt{n}}}$$

Point Estimate ± (Critical Value)(Standard Error)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Interval for μ (σ Unknown)

- The $t_{a/2}$ is the critical value of the *t*-distribution with n -1 degrees of freedom and an area of a/2 in each tail

- Confidence Interval Estimate:

$$\boxed{\bar{X}} \pm \boxed{t_{\alpha/2}} \boxed{\frac{S}{\sqrt{n}}}$$

Point Estimate ± (Critical Value)(Standard Error)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- A random sample of n = 25 has $\bar{X}$ = 50 and S = 8. Form a 95% confidence interval for μ

  - d.f. = n – 1 = 24, so $t_{\alpha/2} = t_{0.025} = 2.0639$

The confidence interval is

$$\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}} = 50 \pm (2.0639) \frac{8}{\sqrt{25}}$$

$$46.698 \le \mu \le 53.302$$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Note:

- In this example, Interpreting this interval requires the assumption that the population you are sampling from **is approximately a normal distribution (especially since n is only 25).**
- This condition can be checked by creating a:
  - Normal probability plot or
  - Boxplot

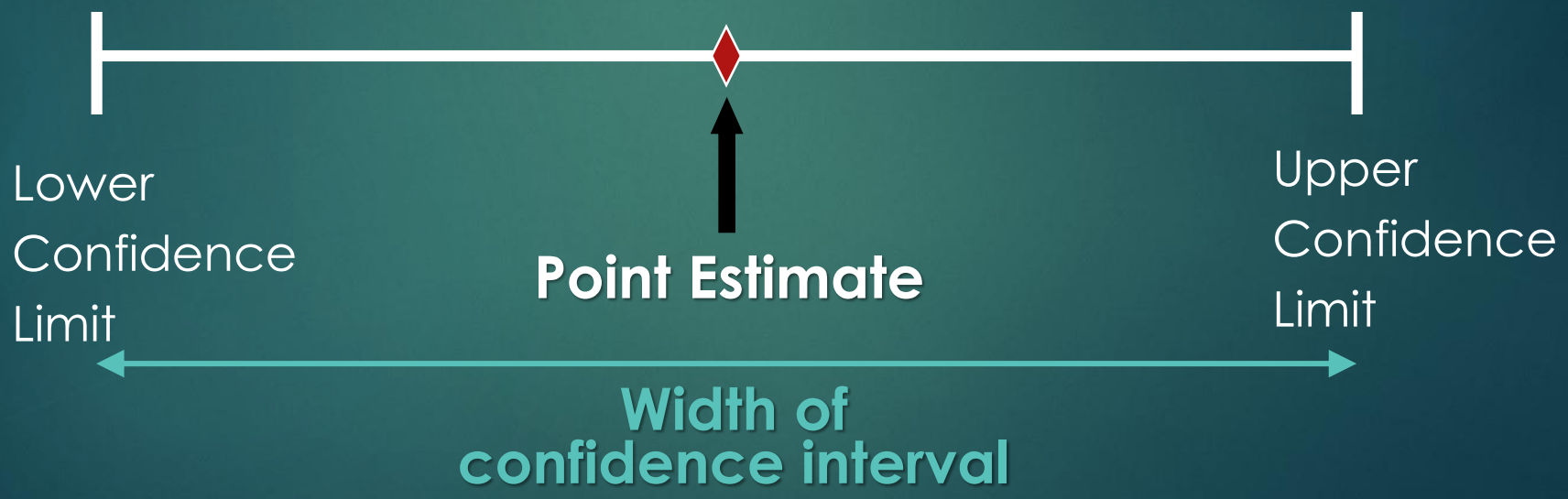
Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Let us watch another StatTalk video. This one is titled "Sampling and Parameters".

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Reminder: Point and Interval Estimates

- A point estimate is a single number
- A confidence interval provides additional information about the variability of the estimate



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Reminder: Confidence Interval for μ with **known σ**

- Assumptions
  - Population standard deviation **σ is known**
  - Population is normally distributed
  - If population is not normal, use large sample (n > 30)

- Confidence interval estimate:

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

Point Estimate ± (Critical Value)(Standard Error)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Reminder: Confidence Interval for μ with **unknown σ**

- Assumptions
  - Population standard deviation is unknown
  - Population is normally distributed
  - If population is not normal, use large sample (n > 30)
- Use Student's *t* Distribution
- Confidence Interval Estimate:

$$\boxed{\bar{X}} \pm \boxed{t_{\alpha/2}} \boxed{\frac{S}{\sqrt{n}}}$$

Point Estimate ± (Critical Value)(Standard Error)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- A survey of non-profit organizations showed that online fundraising has increased over the last year. Based on a random sample of 55 non-profit organizations, the mean of a one-time donation is $75, with standard deviation of $9.
  - **Construct a 95% confidence interval estimate for the population mean of one-time donations**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- Step 1) Write down the general formula for Confidence Interval Estimations:

Point Estimate ± (Critical Value)(Standard Error)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- Step 1) Write down the general formula for Confidence Interval Estimations:

Point Estimate ± (Critical Value)(Standard Error)

- Step 2) Identify the Point Estimate
  - In this case, the sample mean is $75.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- Step 1) Write down the general formula for Confidence Interval Estimations:

Point Estimate ± (Critical Value)(Standard Error)

- Step 2) Identify the Point Estimate
- Step 3) Find the Critical Value

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- Step 1) Write down the general formula for Confidence Interval Estimations:

Point Estimate ± (Critical Value)(Standard Error)

- Step 2) Identify the Point Estimate
- Step 3) Find the Critical Value
  - Step 3.1.) Identify the TYPE of Standard Error Available to you.
  - Step 3.2.) If the **Population Standard Error** is available, use **Standardized Normal Distribution**. If **Sample Standard Error** is available, use **Student's t Distribution.**
  - Step 3.3.) Identify the Critical Value given Step 3.2.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- Step 1) Write down the general formula for Confidence Interval Estimations:

Point Estimate ± (Critical Value)(Standard Error)

- Step 2) Identify the Point Estimate
- Step 3) Find the Critical Value
  - In this case, the Sample Standard Error is available. Thus, we employ Student's t Distribution
    - The degree of freedom is 54, and $\alpha/2$ is 0.025. Thus, the critical value is equal to 2.0049
    - (For more, refer to Table E3 in your textbook)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- Step 1) Write down the general formula for Confidence Interval Estimations:

Point Estimate ± (Critical Value)(Standard Error)

- Step 2) Identify the Point Estimate
- Step 3) Find the Critical Value
- Step 4) Identify the Standard Error
  - Step 3.1.) Identify the TYPE of Standard Error Available to you. Then, the Standard Error is as follows:
  - Step 3.2.1.) For **Population Standard Error:** $\sigma/\sqrt{n}$
  - Step 3.2.2.) For **Sample Standard Error:** $S/\sqrt{n}$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- Step 1) Write down the general formula for Confidence Interval Estimations:

Point Estimate ± (Critical Value)(Standard Error)

- Step 2) Identify the Point Estimate
- Step 3) Find the Critical Value
- Step 4) Identify the Standard Error
- Step 5) Estimate the interval using what you found in Steps 2-4.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- Step 1) Write down the general formula for Confidence Interval Estimations:

Point Estimate ± (Critical Value)(Standard Error)

- Step 2) Identify the Point Estimate: **$75**
- Step 3) Find the Critical Value: **2.0049**
- Step 4) Identify the Standard Error: $9/\sqrt{55}$
- Step 5) Estimate the interval using what you found in Steps 2-4: $75 \pm 2.0049 \times 9/\sqrt{55}$
- Thus: **$72.6 ≤ Population Mean ≤ $77.4**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example

- Customers Report Ratings has reported the MPG for 25 sedans in 2013. The data is included below.
- Construct a 95% confidence interval estimate for the population mean MPG of 2013 family sedans.

38	24	29
26	26	25
30	25	33
26	23	21
25	25	21
27	26	UNIT: MPG
22	31	Mean
27	26	27.12
39	37	SD
24	22	5.027

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

- Step 1) Write down the general formula for Confidence Interval Estimations:

Point Estimate ± (Critical Value)(Standard Error)

- Step 2) Identify the Point Estimate
- Step 3) Find the Critical Value (It's about **2** in this case)
- Step 4) Identify the Standard Error
- Step 5) Estimate the interval using what you found in Steps 2-4.

38	24	29
26	26	25
30	25	33
26	23	21
25	25	21
27	26	UNIT: MPG
22	31	**Mean**
27	26	**27.12**
39	37	**SD**
24	22	**5.027**

Example

- Step 1) Write down the general formula for Confidence Interval Estimations:

Point Estimate ± (Critical Value)(Standard Error)

- Step 2) Identify the Point Estimate: **27.12 MPG**
- Step 3) Find the Critical Value: **2.0639**
- Step 4) Identify the Standard Error: $5.027/\sqrt{25}$
- Step 5) Estimate the interval using what you found in Steps 2-4: $27.12 \pm 2.0639 \times 5.027/\sqrt{25}$
- Thus: **25.045 MPG ≤ Population Mean ≤ 29.195 MPG**

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Finding the Critical t Value using Excel

- Make use of the following function:

=T.INV.2T(1-ConfidenceLevel, df)

Example:

=T.INV.2T(1-95%, 24) yields 2.0639

For more, see Confidence_Interval_Estimation_UnknownSigma on the D2L.

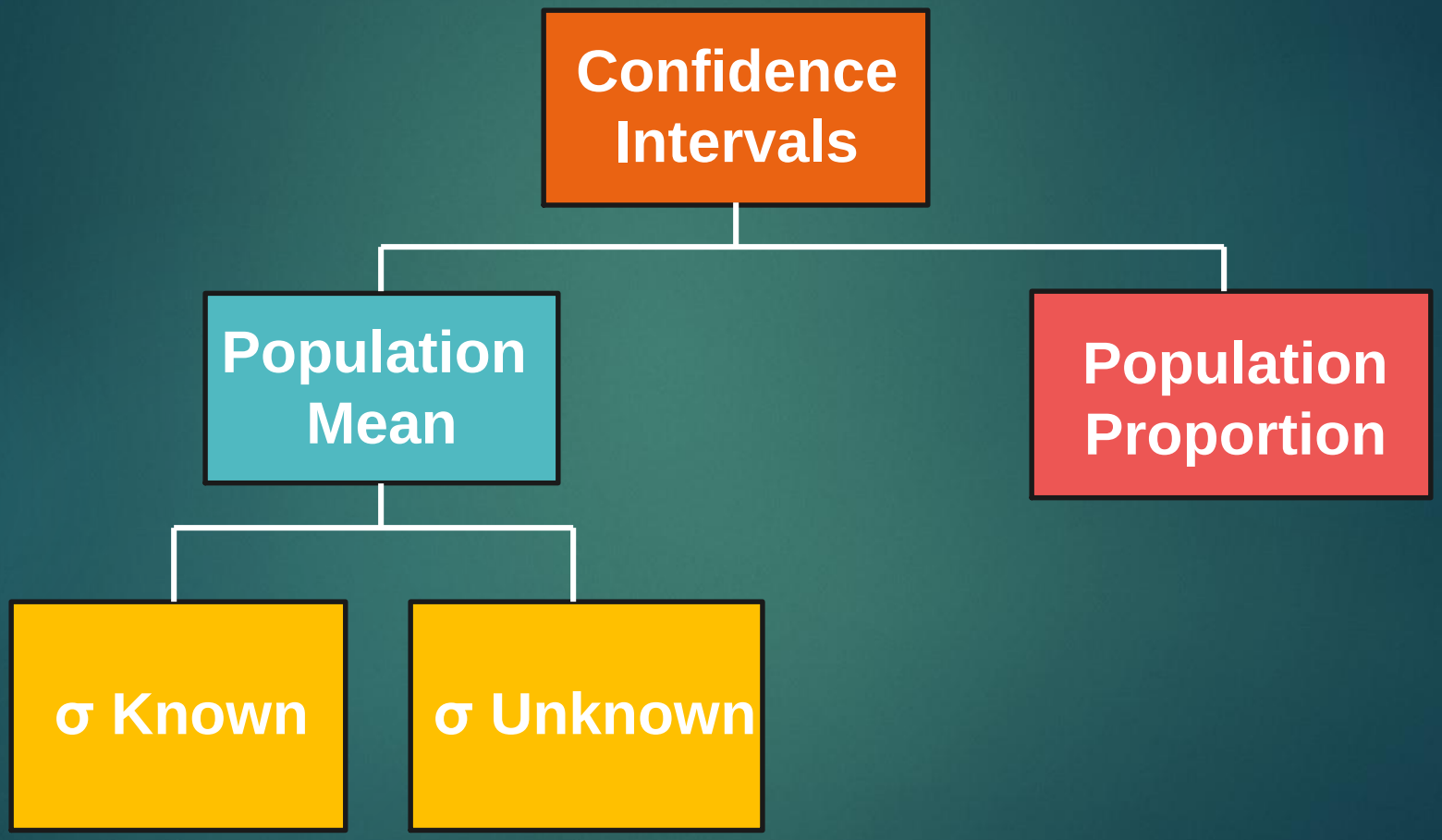
Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Assignment:

Homework 8.2.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Intervals



Copyright © 2015, 2012, 2009 Pearson Education, Inc.

- In a survey of 3,773 travelers, 1,207 said that *reputation* is very important when they choose an airline to fly with.

(Source: USA Today, January 18, 2013)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

- In a survey of 3,773 travelers, 1,207 said that *reputation* is very important when they choose an airline to fly with.

(Source: USA Today, January 18, 2013)

- Given the above survey result, what is the **relative size of the population of travelers** who consider *reputation* as a very important factor in choosing airline?

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Intervals for the Population Proportion, π

- An interval estimate for the population proportion (π) can be calculated by adding an allowance for uncertainty to the sample proportion (p)

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Intervals for the Population Proportion, π

- Recall that the distribution of the sample proportion is approximately normal if the sample size is large, with standard deviation

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

- We will estimate this with sample data:

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

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Confidence Interval Endpoints

- Upper and lower confidence limits for the population proportion are calculated with the formula

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

Point Estimate ± (Critical Value)(Standard Error)

- where
  - $Z_{\alpha/2}$ is the standard normal value for the level of confidence desired
  - p is the sample proportion
  - n is the sample size
- Note: must have np > 5 and n(1-p) > 5

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example:

- Refer to the Airline Survey example.
- Given the survey result, estimate a 95% confidence interval around the population proportion of those who consider reputation a very important factor.
- p: 1,207 individuals (out of 3,773) responded positively – thus:

$$p = \frac{1207}{3773}$$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example:

- Refer to the Airline Survey example.
- Given the survey result, estimate a 95% confidence interval around the population proportion of those who consider reputation a very important factor.

$$p = \frac{1207}{3773}$$

$$Interval: p \mp Z_{0.025}\sqrt{\frac{p(1-p)}{n}}$$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

Example:

- Refer to the Airline Survey example.
- Given the survey result, estimate a 95% confidence interval around the population proportion of those who consider reputation a very important factor.

$$p = \frac{1207}{3773}$$

$$Interval: p \mp Z_{0.025}\sqrt{\frac{p(1-p)}{n}} = 0.32 \mp 1.96 \times \sqrt{\frac{0.32 \times 0.68}{3773}}$$

Copyright © 2015, 2012, 2009 Pearson Education, Inc.

$$p = \frac{1207}{3773}$$

$$Interval: p \mp Z_{0.025}\sqrt{\frac{p(1-p)}{n}} = 0.32 \mp 1.96 \times \sqrt{\frac{0.32 \times 0.68}{3773}}$$

0.305 ≤ π ≤ 0.335

With 95% confidence, we can say that between 30.5% and 33.5% of travelers consider reputation an important factor in their decision when they choose an airline.

Assignment:

Homework 8.3.

Copyright © 2015, 2012, 2009 Pearson Education, Inc.