

# Lecture 2

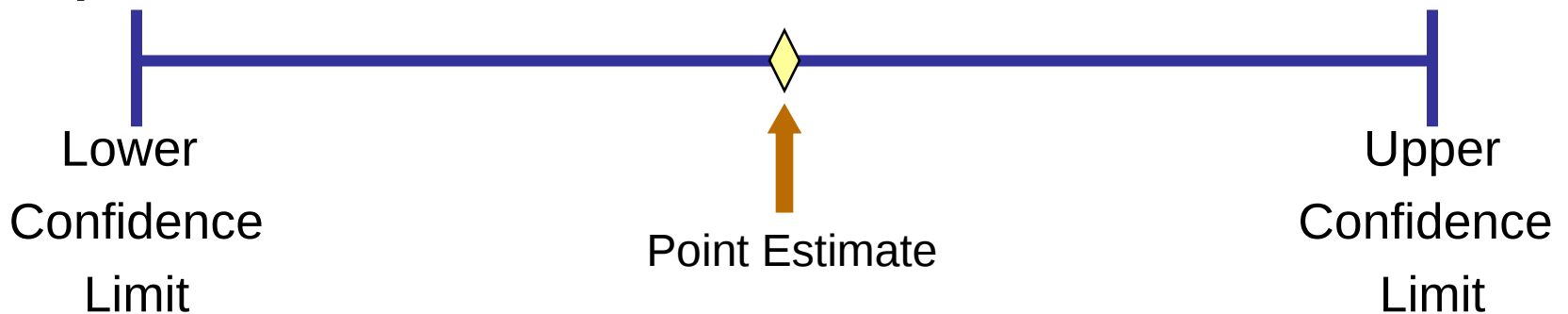
## Estimating Single Population Parameters

# 8.1 Point and Confidence Interval Estimates for a Population Mean

- **Point Estimate**
  - A single statistic, determined from a sample, that is used to estimate the corresponding population parameter
- **Sampling Error**
  - The difference between a measure (a statistic) computed from a sample and the corresponding measure (a parameter) computed from the population

# Confidence Interval

- An interval developed from sample values such that if all possible intervals of a given width were constructed, a percentage of these intervals, known as the confidence level, would include the true population parameter



# Point Estimates

- Population parameter can be estimated with sample statistic (point estimate)

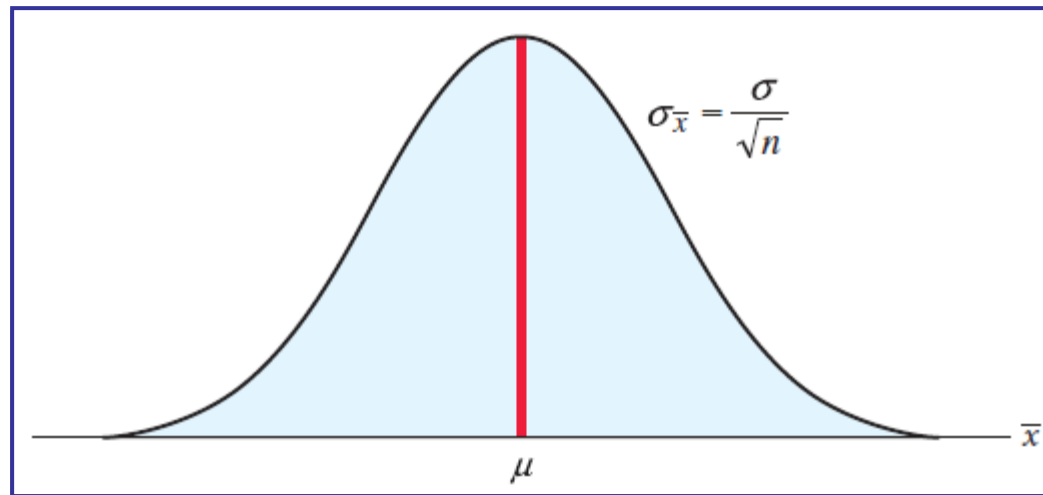
	Population	Sample
Mean	$\mu$	$\bar{x}$
Proportion	$p$	$\bar{p}$

# Confidence Interval Estimate for the Population Mean, $\sigma$ Known

- **Case 1:**
  - The simple random sample is drawn from a normal distribution
- **Case 2:**
  - The population does not have a normal distribution or the distribution of the population is not known
- In both these cases, the sampling distribution for  $\bar{x}$  is assumed to be **normally distributed**

# Confidence Interval Estimate for the Population Mean, $\sigma$ Known

- **Standard Error**
  - A value that measures the spread of the sample means around the population mean
  - The standard error is reduced when the sample size is increased

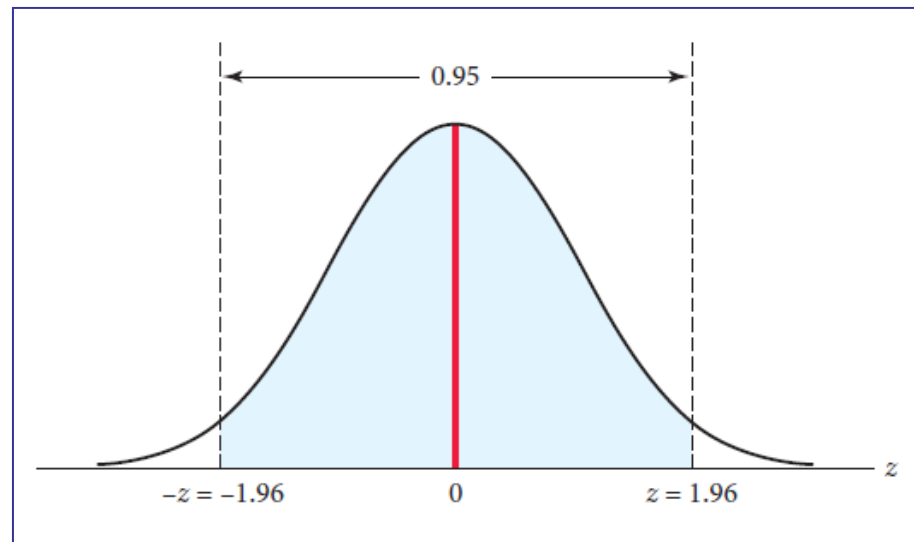


# Confidence Level

- The percentage of sample means in the interval formed by a specified distance above and below  $\mu$  can be calculated
- This percentage corresponds to the probability that the sample mean will be in the specified interval
- **Confidence Level:** The percentage of all possible confidence intervals that will contain the true population parameter

# Critical Value

- $P(-1.96 \leq z \leq 1.96) = 0.4750 + 0.4750 = 0.95$
- 95% of all sample means will fall in the range:  
$$\mu - 1.96 \frac{\sigma}{\sqrt{n}} \quad \text{-----} \quad \mu + 1.96 \frac{\sigma}{\sqrt{n}}$$
- $z = 1.96$  is referred to as the **critical value**





# Confidence Interval Calculation

- General Format:

Point estimate  $\pm$  (Critical value)(Standard error)

- Confidence Interval estimate for  $\mu$ ,  
 $\sigma$  known

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

$z$  - Critical value from the standard normal table for a specified confidence level

$\sigma$  - Population standard deviation

$n$  - Sample size

# Critical Values for Confidence Levels

Confidence Level	Critical Value
80%	$Z = 1.28$
90%	$Z = 1.645$
95%	$Z = 1.96$
99%	$Z = 2.575$

Critical values can be found using the standard normal table, or using Excel's **NORM.S.INV** function

# Confidence Interval Estimate for the Population Mean, $\sigma$ Known

- **Step 1:** Define the population of interest and select a simple random sample of size  $n$
- **Step 2:** Specify the confidence level
- **Step 3:** Compute the sample mean
- **Step 4:** Determine the standard error of the sampling distribution
- **Step 5:** Determine the critical value,  $z$ , from the standard normal table.
- **Step 6:** Compute the confidence interval estimate

# Margin of Error

- A measure of how close we expect the point estimate to be to the population parameter with the specified level of confidence

$$e = z \frac{\sigma}{\sqrt{n}}$$

$e$  - Margin of error

$z$  - Critical value

$\frac{\sigma}{\sqrt{n}}$  - Standard error of the sample distribution

- Lowering the confidence level is one way to reduce the margin of error
- The margin of error can be reduced by increasing the sample size.

# Impact of Changing the Confidence Level - Example

- Each year, the recycling company must apply for a new contract with the state. The contract is in part based on the pounds of recycled materials collected. Part of the analysis is to estimate of the mean pounds of recycled material. The city has asked for both 99% and 90% confidence interval estimates for the mean.
- **Step 1:** Population recycling company customers,  $n = 100$
- **Step 2:** Confidence levels: 99% and 90%
- **Step 3:**  $\bar{x} = 40.78$
- **Step 4:**  $\sigma = 12.6$ ; standard error =  $\frac{12.6}{\sqrt{100}} = 1.26$
- **Step 5:** 99% -  $z = 2.575$       90% -  $z = 1.645$

• **Step 6:**

	99%	90%
	40.78 ± 3.24	40.78 ± 2.07
	37.54 ————— 40.22	38.71 ————— 42.85

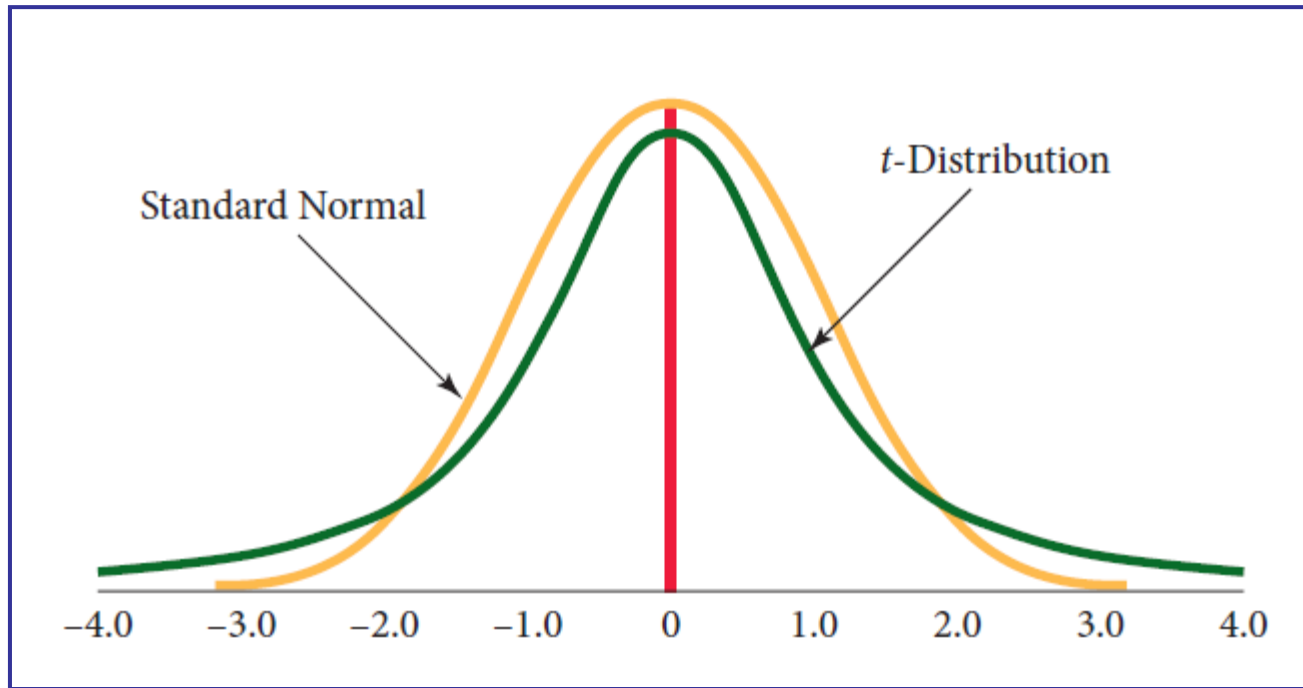
# Confidence Interval Estimate for the Population Mean, $\sigma$ Unknown

- In most cases, if the population mean is unknown, the population standard deviation is unknown too
- This introduces extra uncertainty, since sample standard deviation varies from sample to sample
- Confidence interval estimation process needs to be modified

# Student's $t$ -Distribution

- When  $\sigma$  is not known, the critical value is a  $t$ -value taken from a family of distributions called the **Student's  $t$ -distributions**
- It is bell shaped and symmetrical
- Has greater area of tails than standard normal distribution
- $t$ -distribution is defined by its degrees of freedom
- As the degrees of freedom increase, the  $t$ -distribution approaches the normal distribution

# Student's $t$ -Distribution



The  $t$ -distribution is based on the assumption that the population is normally distributed



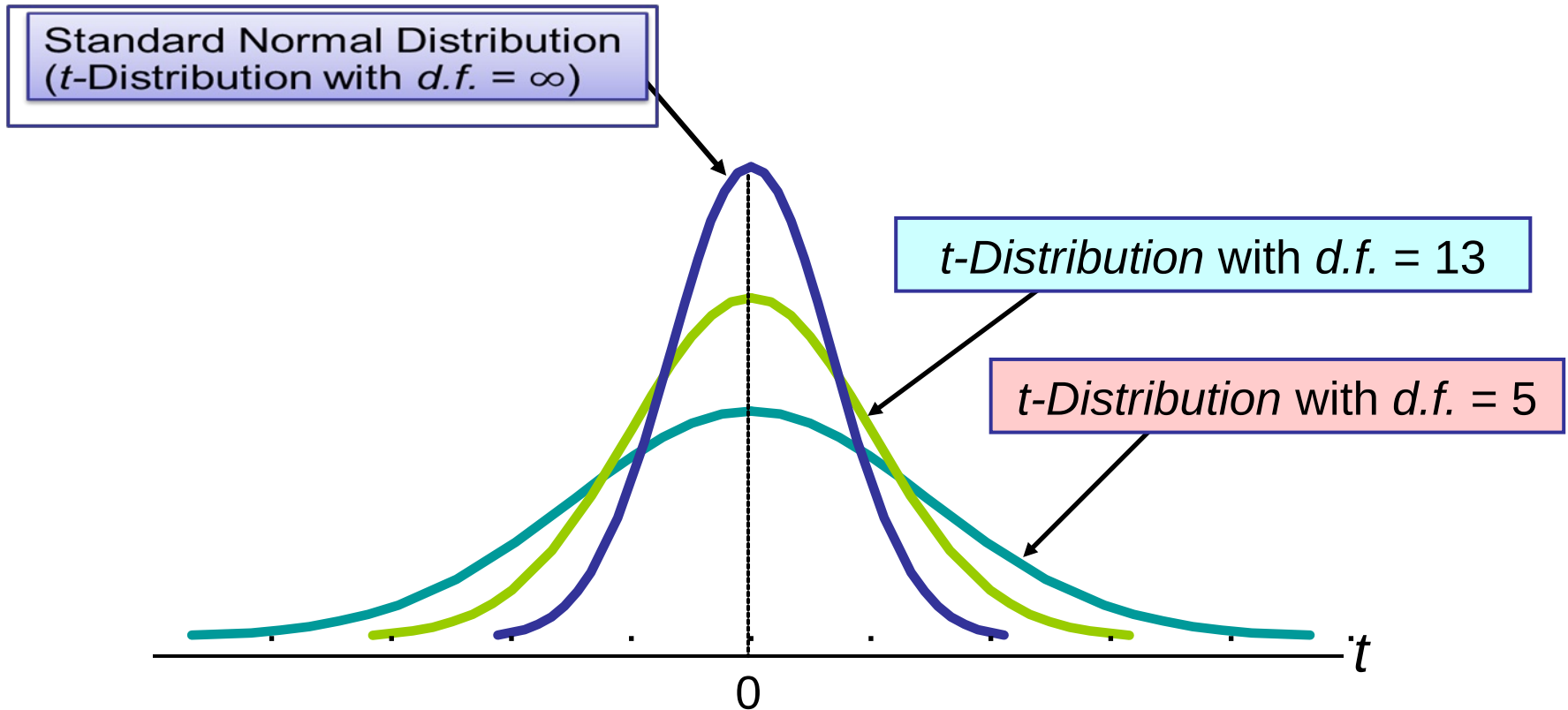
# Degrees of Freedom

- The number of independent data values available to estimate the population's standard deviation. If  $k$  parameters must be estimated before the population's standard deviation can be calculated from a sample of size  $n$ , the degrees of freedom are equal to  $n - k$
- For example:
  - The sample mean is obtained from a sample of  $n$  randomly and independently chosen data values
  - Once the sample mean has been obtained, there are only  $n - 1$  independent pieces of data information left in the sample

# Degrees of Freedom - Example

- Suppose that sample size  $n = 3$  and sample mean is  $12$
- It implies that the sum of the data values is  $36$
- If  $x_1 = 10$  and  $x_2 = 9$  than  $x_3$  should be  $17$
- You are free to choose any two of the three data values before the remaining data value should be estimated ( $k = 1$ )
- Degrees of freedom =  $3 - 1 = 2$

# Degrees of Freedom and $t$ -Distribution



# Confidence Interval Estimate for the Population Mean, $\sigma$ Unknown

*t*-Value for  $\bar{x}$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Confidence Interval for  $\mu$

$$\bar{x} \pm t \frac{s}{\sqrt{n}}$$

$\bar{x}$  - Sample mean

$\mu$  - Population mean

$t$  - Critical value from the  $t$ -distribution with  $n - 1$  degrees of freedom for the desire confidence level

$s$  - Sample standard deviation

$n$  - Sample size

# t-Distribution Table Example

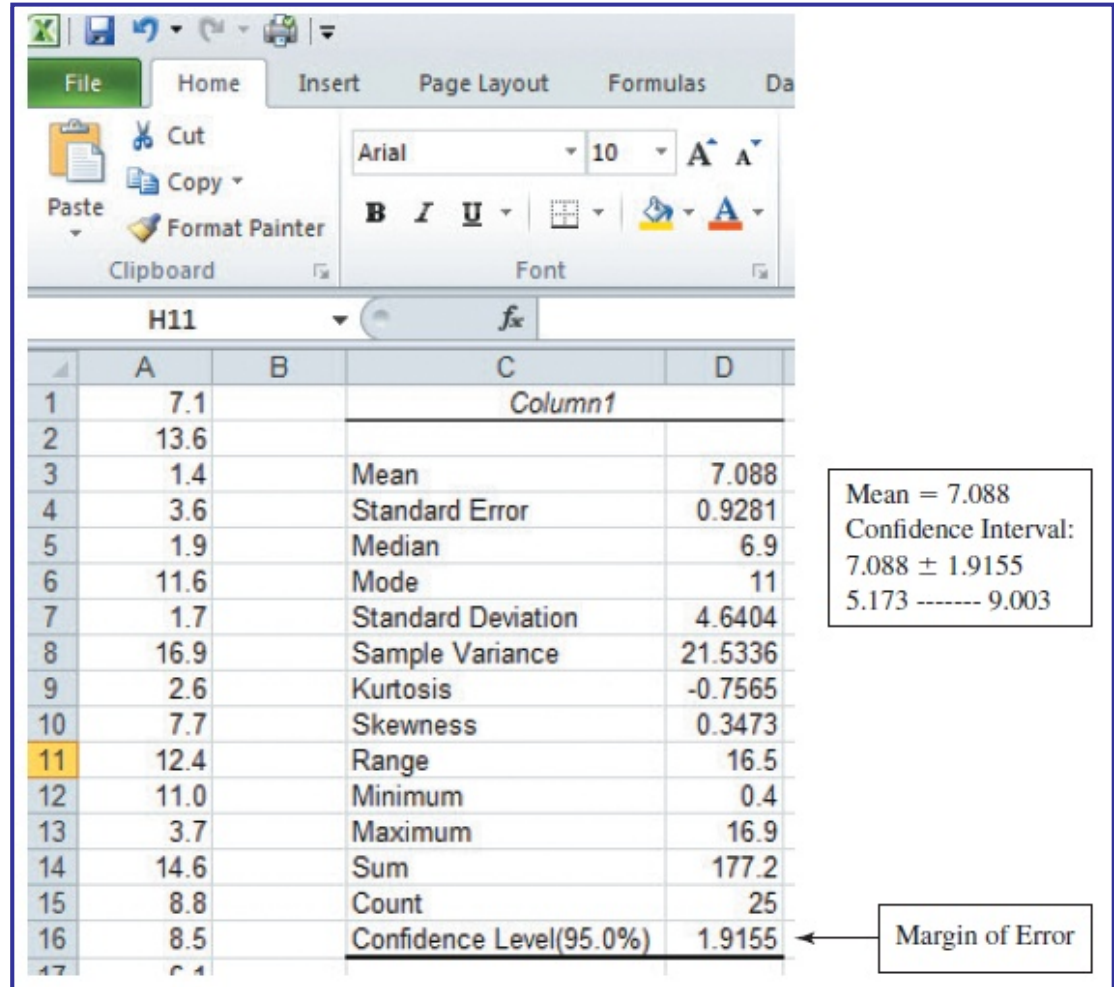
$n = 10$   
 $\bar{x} = 8.088$   
 $s = 4.64$   
 Confidence level = 95%  
 $d.f. = n - 1 = 9$

Probabilities (or Areas under t-Distribution Curve)									
Conf. Level	0.1	0.3	0.5	0.7	0.8	0.9	0.95	0.98	0.99
One Tail	0.45	0.35	0.25	0.15	0.1	0.05	0.025	0.01	0.005
Two Tails	0.9	0.7	0.5	0.3	0.2	0.1	0.05	0.02	0.01
d.f.	Values of t								
1	0.1584	0.5095	1.0000	1.9626	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.1421	0.4447	0.8165	1.3862	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.1366	0.4242	0.7649	1.2498	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.1338	0.4142	0.7407	1.1896	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.1322	0.4082	0.7267	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.1311	0.4043	0.7176	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.1303	0.4015	0.7111	1.1192	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.1297	0.3995	0.7064	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.1293	0.3979	0.7027	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.1289	0.3966	0.6998	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693

$t = 2.2622$        $8.088 \pm 2.2622 \frac{4.64}{\sqrt{10}}$        $8.088 \pm 3.319$   
 $4.769$  —————  $11.407$

# How to Do It in Excel?

1. Open file.
2. Select **Data** tab.
3. Select **Data Analysis > Descriptive Statistics** category.
4. Specify data range.
5. Define **Output Location**.
6. Check **Summary Statistics**.
7. Check **Confidence Level for Mean: 95%**.
8. Click **OK**.



	A	B	C	D
1	7.1		Column1	
2	13.6			
3	1.4		Mean	7.088
4	3.6		Standard Error	0.9281
5	1.9		Median	6.9
6	11.6		Mode	11
7	1.7		Standard Deviation	4.6404
8	16.9		Sample Variance	21.5336
9	2.6		Kurtosis	-0.7565
10	7.7		Skewness	0.3473
11	12.4		Range	16.5
12	11.0		Minimum	0.4
13	3.7		Maximum	16.9
14	14.6		Sum	177.2
15	8.8		Count	25
16	8.5		Confidence Level(95.0%)	1.9155

Mean = 7.088  
Confidence Interval:  
7.088 ± 1.9155  
5.173 ----- 9.003

Margin of Error

## 8.2 Determining the Required Sample Size

- There are three conflicting objectives:
  - High confidence level, a low margin of error, a small sample size:
    - For a given sample size, a high confidence level will tend to generate a large margin of error
    - For a given confidence level, a small sample size will result in an increased margin of error
    - Reducing the margin of error requires either reducing the confidence level or increasing the sample size, or both

# Determining the Required Sample Size, $\sigma$ Known

$$n = \left( \frac{z\sigma}{e} \right)^2 = \frac{z^2 \sigma^2}{e^2}$$

- $z$  - Critical value for the specified confidence level  
 $e$  - Desired margin of error  
 $s$  - Population standard deviation

## Example:

1. Population mean should not exceed 30
2. Population standard deviation is 200
3. Confidence level is 95% ( $z = 1.96$ )
4. Sample size = ?

## Solution:

$$n = \left( \frac{1.96(200)}{30} \right)^2 = 171$$



# Determining the Required Sample Size, $\sigma$ Known

- **Step 1:** Specify the desired margin of error
- **Step 2:** Determine the population standard deviation
- **Step 3:** Determine the critical value for the desired level of confidence
- **Step 4:** Compute the required sample size

# Determining the Required Sample Size, $\sigma$ Unknown

- Possible options:
  - to use a value for  $\sigma$  that is considered to be at least as large as the true  $\sigma$
  - select a **pilot sample**: a sample taken from the population of interest of a size smaller than the anticipated sample size used to estimate the population standard deviation
  - to use the range of the population to estimate the population's standard deviation

$$\sigma \approx \frac{R}{6}$$

# 8.3 Estimating a Population Proportion

## Sample Proportion

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

## Standard Error for $\bar{p}$

(sample size is sufficiently large  $np \geq 5$  and  $n(1 - p) \geq 5$ )

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

## Estimate for the Standard Error of $\bar{p}$

$\bar{p}$  - Sample proportion

$p$  - Population proportion

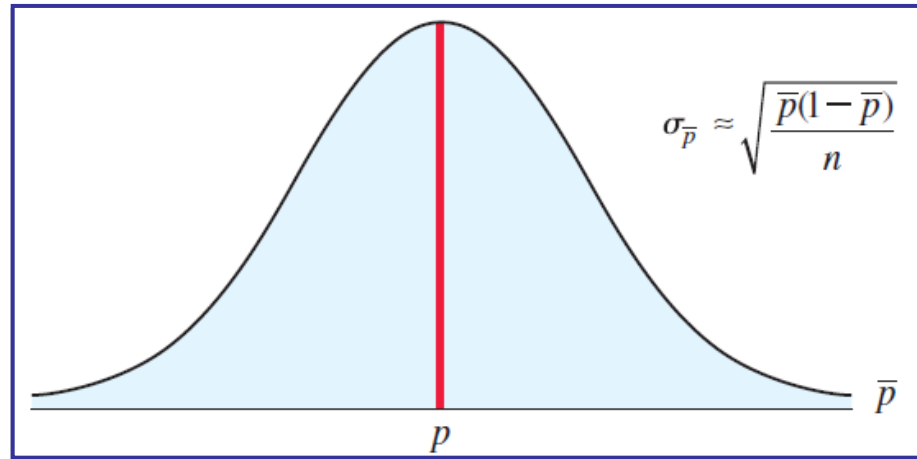
$x$  - Number of items in the sample with the attribute of interest

$n$  - Sample size

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

# Confidence Interval for the Population Proportion $p$

- Sample Distribution for  $\bar{p}$



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

$\bar{p}$  - Sample proportion

$n$  - Sample size

$z$  - Critical value from the standard normal distribution for the desired confidence level

# Confidence Interval for the Population Proportion $p$

- **Step 1:** Define the population and variable of interest for which to estimate the population proportion
- **Step 2:** Determine the sample size and select a random sample that must be large enough
- **Step 3:** Specify the level of confidence and obtain the critical value from the standard normal distribution table
- **Step 4:** Calculate the sample proportion
- **Step 5:** Construct the interval estimate

# Confidence Interval for the Population Proportion $p$ - Example

- A random sample of 100 people shows that 25 are left-handed. Define a 95% confidence interval for the true proportion of left-handers
- Sample proportion
- z-value for 95% confidence level
- Confidence interval

$$\bar{p} = \frac{25}{100} = 0.25$$

$$z = 1.96$$

$$0.25 \pm 1.96 \sqrt{\frac{0.25(1 - 0.25)}{100}} = 0.25 \pm 0.085$$

0.165 ————— 0.335

# Required Sample Size

- Changing the confidence level affects the interval width
- Changing the sample size will affect the interval width
- An increase in sample size will reduce the standard error and reduce the interval width
- A decrease in the sample size will have the opposite effect

# Required Sample Size

- Margin of Error for Estimating  $p$

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

$$n = \frac{z^2 p(1-p)}{e^2}$$

$p$  - Population proportion

$z$  - Critical value from standard normal distribution for the desired confidence level

$n$  - Sample size



# Sample Size Determination

- **Step 1:** Define the population and variable of interest
- **Step 2:** Determine the level of confidence and find the critical z-value using the standard normal distribution table
- **Step 3:** Determine the desired margin of error
- **Step 4:** Arrive at a value to use for  $p$ 
  - Use a pilot sample and compute  $\bar{p}$  to approximate  $p$
  - Select a value for  $p$  that is closer to 0.50 than you actually believe the value to be
  - If you have no idea what  $p$  might be, use  $p = 0.50$ , which will give the largest possible sample size for the stated confidence level and margin of error.
- **Step 5:** Determine the sample size

# Sample Size Determination - Example

- How large a sample would be necessary to estimate the true proportion defective in a large population **within 3%, with 95% confidence?**  
(Assume a pilot sample yields  $p = 0.12$ )
- **Sample size**

$$z = 1.96 \quad e = 0.03 \quad p = 0.12$$

$$n = \frac{1.96^2 (0.12)(1 - 0.12)}{0.03^2} = 450.74 \approx 451$$