

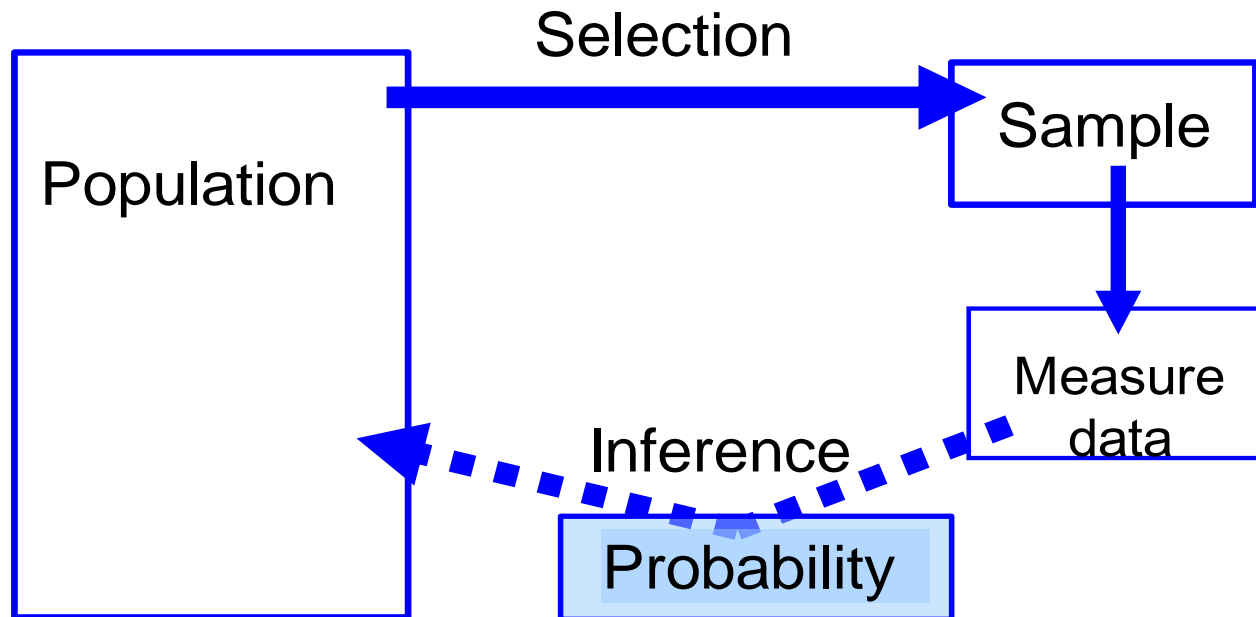


# STATISTICAL INFERENCE

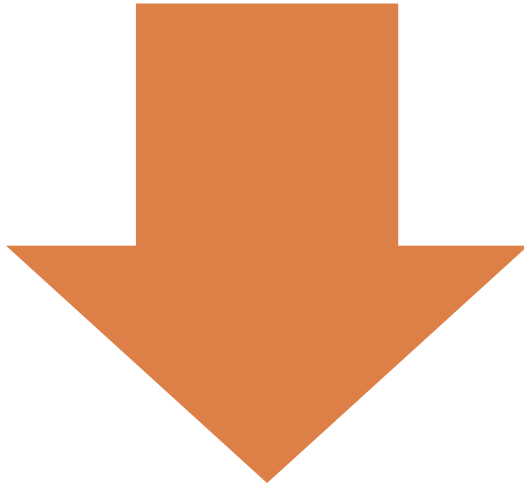
ESTIMATION

HYPOTHESIS TESTING

# Chain of Reasoning for Inferential Statistics



# INFERENCE STATISTICS



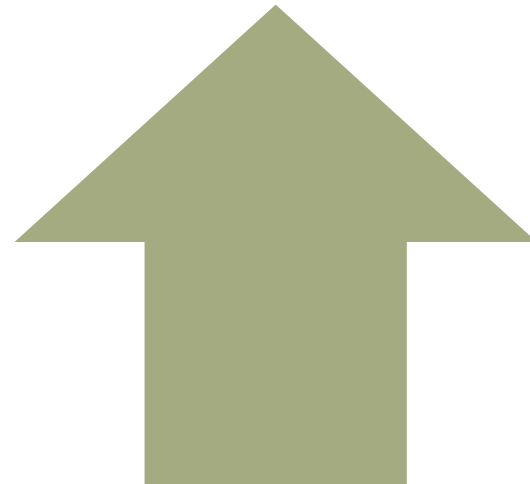
## ESTIMATION

- “guessing” the value of the parameter
- provide the measure of the quality (reliability) of the guess



## HYPOTHESIS TESTING

- making a “yes-no” decision regarding the parameter
- understand the chances of making incorrect decision



# STATISTICAL INFERENCE

## ESTIMATION

- estimation for single population parameter
- estimation for two population parameter
  - ▣ Independent sample
  - ▣ Paired Sample

## HYPOTHESIS TESTING

- Single population mean
- Single population proportion
- Two population Means:
  - ▣ Independent Sample
  - ▣ Paired Sample
- Two population proportion

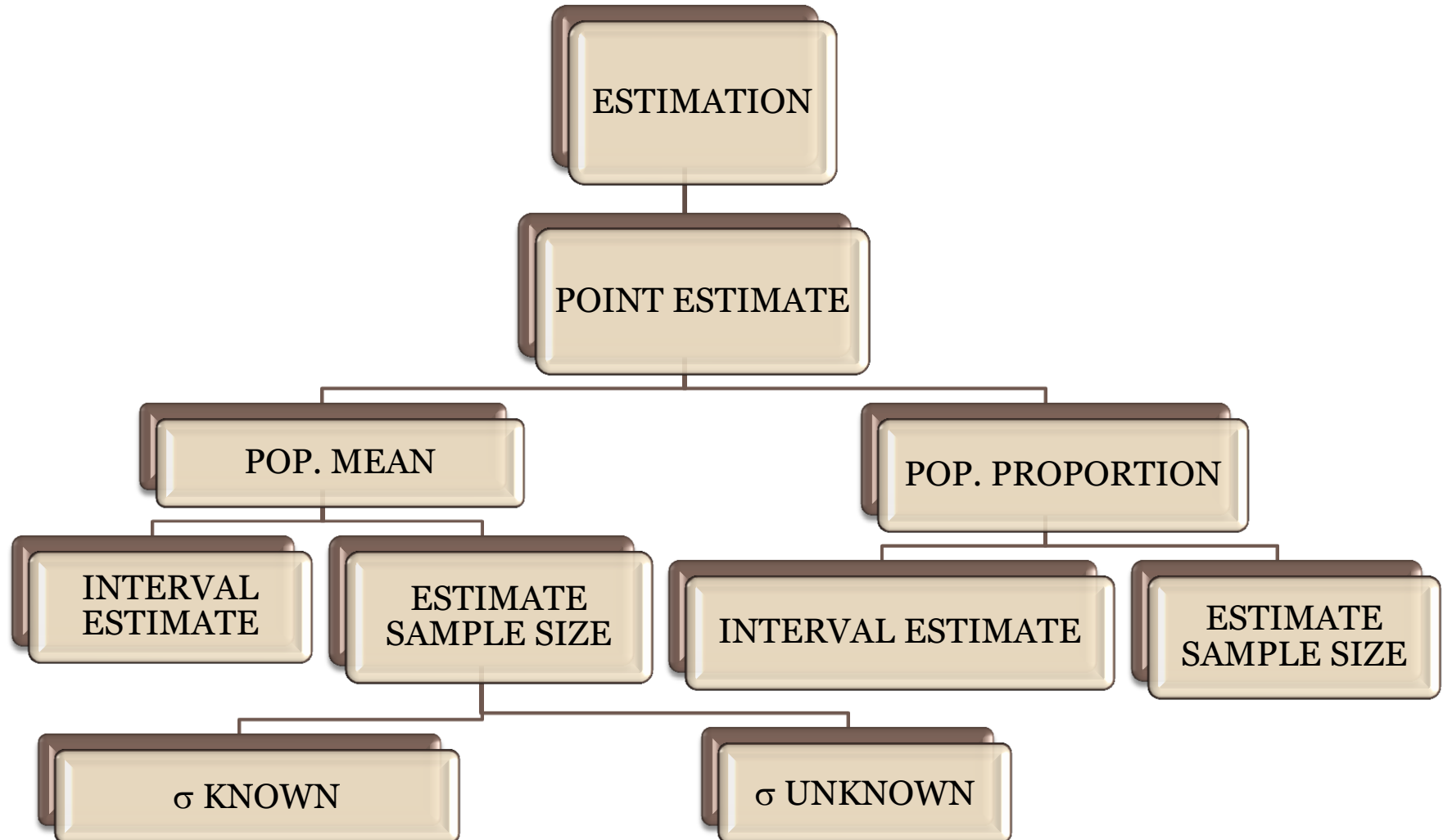
# STATISTICAL INFERENCE

## ESTIMATION WITH SINGLE POPULATION PARAMETER

Estimation?



# WHAT WE WILL DISCUSS IN ESTIMATION?



# ESTIMATION

- Estimation from sample are only guesses (of the parameter)
- Every estimate has a standard error, and it is measure of the variation in the estimates.
  - **Standard error** – A value that measures the spread of the sample means ( $\bar{x}$ ) around the population mean ( $\mu$ ). The standard error is reduced when the sample size is increased.
- Regardless of the population being estimated, we always expect sampling error:
  - **Sampling error** -the difference between a measure (statistics) computed from a sample and the corresponding measure (parameter) computed from the population.

# ESTIMATION

An estimation of a population parameter may be expressed in:

## POINT ESTIMATE

- **Definition:** a single statistics, determined from a sample that is used to estimate the corresponding population parameter.
- For example, the sample mean  $\bar{x}$  is a point estimate of the population mean,  $\mu$
- Similarly, the sample proportion  $p$  is a point estimate of the population proportion  $P$ .
- It may subject to sampling error

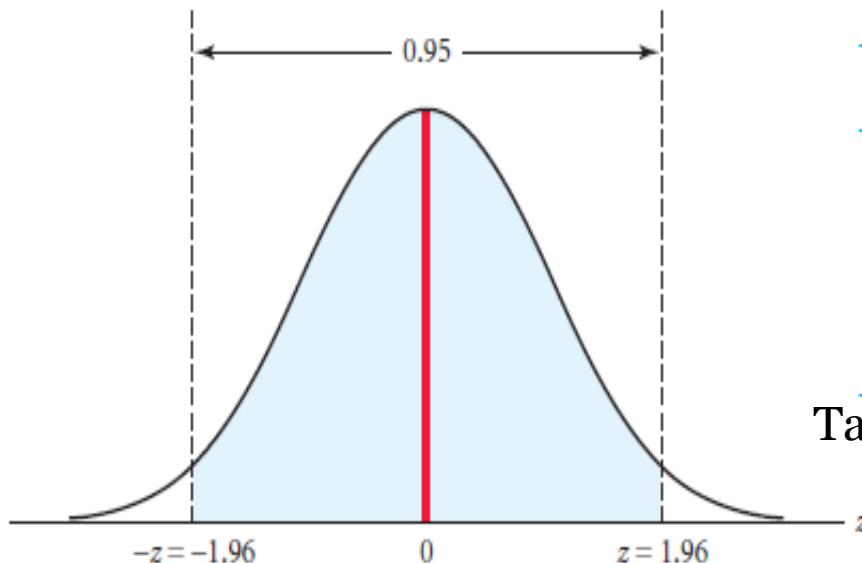
## INTERVAL ESTIMATE

- **Definition:** Is defined by two numbers, between which parameter is said to lie
- For example,  $a < x < b$  is an interval estimate of the population mean  $\mu$ . It indicates that the population mean is greater than  $a$  but less than  $b$
- A common procedure to calculate an interval estimate is know as **confidence interval**.



# INTERVAL ESTIMATE

- Confidence Interval– The percentage of all possible confidence intervals that will contain the true population parameter.



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

Table: Most commonly use confidence level

Figure: critical value for a 95% confidence interval

National Recycling operates a garbage hauling company in a southern Marine city, Australia. Each year, the 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 that goes into contract development is an estimate of the mean pounds of recycled material submitted by each customer in the city on a quarterly basis. The city has asked for both 99% and 90% confidence interval estimates for the mean. The simple random sample of  $n = 100$  customers is selected for the analysis. The sample mean,  $\bar{x} = 40.78$  pounds and the population standard deviation,  $\sigma_{\bar{x}} = 1.26$  pounds. If, after the contract has been signed, the actual mean pounds deviates from the estimate over time, an adjustment will be made (up or down) in the amount National Recycling receives. Identify the margin error if both confidence interval estimates for the mean is applied.



## **Margin Error:**

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

### Margin of Error for Estimating $\mu$ , $\sigma$ Known

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

where:

$e$  = Margin of error

$z$  = Critical value

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



# IMPACT OF CHANGING THE CONFIDENCE LEVEL?

- Values that affect the **margin error to be smaller**:
  - The **confidence level**: the margin error will be smaller when the **confidence level is smaller**.
  - **Population standard deviation ( $\sigma$ )**: the **more  $\sigma$  can be reduced**, the smaller the margin error will be.
    - However, reducing the population standard deviation ( $\sigma$ ) is not possible
  - **Sample size**: An **increase sample size** reduced the standard error of the sampling distribution. Thus as the direct way of reducing margin error

# STATISTICAL INFERENCE

ESTIMATION – Determining the Required Sample Size for Estimating a Population Mean,  $\mu$

Estimation?



# DETERMINING REQUIRED SAMPLE SIZE FOR ESTIMATING $\mu$

- To determine the sample size, the decision maker need to specify their confidence level and their desired margin error,  $e$
- 2 Scenario consider:
  - 1) When population standard deviation,  $\sigma$  **known**
  - 2) When population standard deviation,  $\sigma$  **unknown**
- 3 approaches to estimate unknown  $\sigma$ :
  - to use a value for  $s$  that is considered to be at least as large as the true  $s$ .
  - to select a **pilot sample**
  - to use the range of the population to estimate the population's standard deviation. ( $\sigma \approx \frac{R}{6}$ )

# DETERMINING REQUIRED SAMPLE SIZE FOR ESTIMATING $\mu$

- 1) When population standard deviation,  $\sigma$   
**known**

Sample Size Requirement for Estimating  $\mu$ ,  $\sigma$  Known

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

where:

$z$  = Critical value for the specified confidence level

$e$  = Desired margin of error

$\sigma$  = Population standard deviation



# When population standard deviation, $\sigma$ known

## **Problem:**

Consider the Mission Valley Power Company (MVP) in northwest Michigan, which has more than 6,000 residential customers. In response to a request by the Michigan Public Utility Commission, MVP needs to estimate the average kilowatts of electricity used by customers on February 1. The only way to get this number is to select a random sample of customers and take a meter reading after 5:00 P.M. on January 31 and again after 5:00 P.M. on February 1. The commission has specified that any estimate presented in the utility's report must be based on a 95% confidence level. Further, the margin of error must not exceed 30 kilowatts. Given these requirements, what size sample is needed?

## **Solution:**

Equation:  $n = \frac{z^2 \sigma^2}{e^2}$ , the z value for 95% is 1.96 (refer to Standard Normal Distribution table). For known  $\sigma$  (MPV has conduct in previous study), the value is 200.

Therefore,  $n = \frac{1.96^2 200^2}{30^2} = 170.73 \sim 171$  customers

What if the confidence level changes? What is the new sample size?

# DETERMINING REQUIRED SAMPLE SIZE FOR ESTIMATING $\mu$

- 2) When population standard deviation,  $\sigma$  **unknown** (with the 3 option as in slide #14, now  $\sigma$  is known)

## Sample Size Requirement for Estimating $\mu$ , $\sigma$ Known

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

where:

$z$  = Critical value for the specified confidence level

$e$  = Desired margin of error

$\sigma$  = Population standard deviation

# When population standard deviation, $\sigma$ unknown

## **Problem:**

Consider a situation in which the regional manager for XYZ Fuel Station in Penang wishes to know the average gallons of petrol purchased by customers each time they fill up their car. Not only does he not know  $\mu$ , he also does not know the population standard deviation,  $\sigma$ . He wants a 90% confidence level and is willing to have a margin of error of 0.50 gallons in estimating the true mean gallons purchased.



## **Steps to solution:**

### **1. Specify the desired margin of error.**

The manager wants the estimate to be within  $\pm 0.50$  gallons of the true mean.

Thus,  $e = 0.50$

### **2. Determine an estimate for the population standard deviation.**

The manager will select a pilot sample of  $n = 20$  fill-ups and record the number of gallons for each. These values are:

18.9	22.4	24.6	25.7	26.3	28.4	21.7	31.0	19.0	31.7
17.4	25.5	20.1	34.3	25.9	20.3	21.6	25.8	31.6	28.8

The estimate for the population standard deviation is the sample standard deviation for the pilot sample. This is computed using,

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{(18.9-25.05)^2 + (22.4-25.05)^2 + (24.6-25.05)^2 + \dots + (28.8-25.05)^2}{20-1}} = 4.85$$

Therefore the  $\sigma \approx 4.85$

**3. Determine the critical value for the desired level of confidence.**

The critical value will be a z-value from the standard normal table. The 90% confidence level gives  $z = 1.645$

**4. Calculate the required sample size using pilot sample's standard deviation.**

$$n = \frac{z^2 \sigma^2}{e^2} = \frac{(1.645^2)(4.85^2)}{0.50^2} = 254.61 \approx 255$$

**Summary:**

The required sample size is 255 fill-ups, but we can use the pilot sample as part of this total. Thus, the net required sample size in this case is  $255 - 20 = 235$

# STATISTICAL INFERENCE

## ESTIMATION – Estimating a Population Proportion

Estimation?



# ESTIMATING A POPULATION PROPORTION

- Proportion - synonyms: part, portion, amount, quantity, bit, piece, percentage, fraction, section, segment, share
- Why study proportion? : some may interested in situations in which the value of interest is the proportion of items in the population that possess a particular attribute.
  - ▣ Example: You may wish to estimate the proportion of customers who are satisfied with the service provided by your company
- Notation:  $p$  = population proportion  
 $\bar{p}$  = sample proportion



# CONFIDENCE INTERVAL ESTIMATE FOR POP. PROPORTION

- Equation for estimate confidence interval for population proportion:

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

where:

$\bar{p}$  = sample proportion

$n$  = sample size

$z$  = critical value from standard normal distribution table for desired confidence interval



### **Problem:**

The Swiss Hotel, Residence and Resort is thinking of starting a new promotion. When a customer checks out of the resort after spending 5 or more days, the customer would be given a voucher that is good for 2 free nights on the next stay of 5 more or more nights at the resort. The marketing manager is interested in estimating proportion of customers who return after getting a voucher. From a simple random sample of 100 customers, 62 returned within 1 year after receiving the voucher. Help the marketing manager in estimating the confidence interval for the returning customer after getting the voucher.

### **Steps to solution:**

#### **1. Specify the desired level of confidence and the critical value.**

Assuming the marketing manager is desired for 95% confidence level, thus the critical value,  $z = 1.96$

#### **2. Compute the point estimate based on the sample data.**

$$\bar{p} = \frac{x}{n} = \frac{62}{100} = 0.62$$

where  $x$  is number of items in the sample with the attribute of interest



### 3. Compute the confidence interval.

The 95% confidence interval estimate is:

$$\bar{p} \pm z \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$$
$$0.62 \pm 1.96 \sqrt{\frac{0.62(1 - 0.62)}{100}}$$

$$0.62 \pm 0.095$$

$$[0.525 \quad 0.715]$$

# DETERMINING THE REQUIRED SAMPLE SIZE FOR ESTIMATING A POPULATION PROPORTION

□ Margin of error for estimating  $p$  
$$e = z \sqrt{\frac{p(1-p)}{n}}$$

where:

$p$  = population proportion

$n$  = sample size

$z$  = critical value from standard normal distribution table for desired confidence interval

□ Sample size for estimating  $p$  
$$n = \frac{z^2 p(1-p)}{e^2}$$

where:

$e$  = desired margin error

$p$  = population proportion

$n$  = sample size

$z$  = critical value from standard normal distribution table for desired confidence interval

### **Problem:**



The customer account manager for Neumann Research, a marketing research company located in Cincinnati, Ohio, is interested in estimating the proportion of a client's customers who like a new television commercial. She wishes to develop a 90% confidence interval estimate and would like to have the estimate be within 0.05 of the true population proportion. Can you help her to determine the required sample size?

### **Steps to solution:**

#### **1. Specify the desired level of confidence and the critical value.**

The desired confidence level is 90%, thus the critical value,  $z = 1.645$

#### **2. Determine the desired margin error.**

The account manager wishes the margin of error to be 0.05

#### **3. Obtain at a value to use for $p$**

Two options can be used to obtain a value for  $p$ :

1. Use a pilot sample and compute  $\bar{p}$ , the sample proportion. Use  $\bar{p}$  to approximate  $p$ .
2. 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.

### **Steps to solution:**

In this case, suppose the account manager has no idea what  $p$  is but wants to make sure that her sample is sufficiently large to meet her estimation requirements. Then she will use  $p = 0.50$ .

#### **4. Determine the sample size.**

$$n = \frac{z^2 p(1 - p)}{e^2} = \frac{1.645^2 (0.50)(1 - 0.50)}{0.05^2} = 270.0625 \approx 271$$

The account manager should randomly survey 271 customers.

# STATISTICAL INFERENCE

ESTIMATION FOR TWO POPULATION  
MEANS:

Independent Samples

Paired Samples

# WHY NEED TO KNOW?

- Previously, we have discuss estimation involving a single population parameter.
- In many business decision-making situations, managers must decide between two or more alternatives.
  - For example, farmers must decide which of several brands and types of wheat to plant.
- There are statistical procedures that can help decision makers use sample information to **compare different populations**.
- The samples from the two populations will be either:
  - Independent Samples
  - Paired Samples

# ESTIMATION FOR TWO POPULATION MEANS USING **INDEPENDENT** SAMPLES

- Independent Samples:
  - Samples selected from two or more populations in such a way that the occurrence of values in one sample has no influence on the probability of the occurrence of values in the other sample(s).
- The estimation between the means will cover:
  1. The population standard deviation are **known** and the samples are independent
  2. The population standard deviation are **unknown** and the samples are independent

## ESTIMATING THE DIFFERENCE BETWEEN TWO POPULATION MEANS WHEN $\sigma_1$ AND $\sigma_2$ ARE KNOWN, USING INDEPENDENT SAMPLES

- In business application, sometimes it may be interested in estimating the difference between two population means.
  - For instance:
    - Estimating the difference in mean starting salaries between males and females
    - Estimating the difference in mean service times at two different fast-food businesses.
- Point estimate =  $\bar{x}_1 - \bar{x}_2$



# ESTIMATING THE DIFFERENCE BETWEEN TWO POPULATION MEANS WHEN $\sigma_1$ AND $\sigma_2$ ARE KNOWN, USING INDEPENDENT SAMPLES

## Standard Error of $\bar{x}_1 - \bar{x}_2$ When $\sigma_1$ and $\sigma_2$ Are Known

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where:

$\sigma_1^2$  = Variance of population 1

$\sigma_2^2$  = Variance of population 2

$n_1$  and  $n_2$  = Sample sizes from populations 1 and 2

## Standard Error of $\mu_1 - \mu_2$ When $\sigma_1$ and $\sigma_2$ Are Known, Independent Samples

$$(\bar{x}_1 - \bar{x}_2) \pm z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

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

Table: The z values for several of the most commonly used confidence levels

# ESTIMATING THE DIFFERENCE BETWEEN TWO POPULATION MEANS WHEN $\sigma_1$ AND $\sigma_2$ ARE KNOWN, USING INDEPENDENT SAMPLES

## **Problem:**

Healthcare Associates operate 33 medical clinics in Minnesota and Wisconsin. As part of the company's ongoing efforts to examine its customer service, Healthcare worked with one of the Minnesota universities on a project in which a team of students observed patients at the clinics to estimate the difference in mean time spent per visit for men and women patients. The student team has selected simple random samples of 100 males, the  $\bar{x}_{males} = 34.5$  minutes and 100 females, the  $\bar{x}_{females} = 42.4$  minutes at different times in different clinics owned by the company across Minnesota and Wisconsin. Previous studies indicate that the standard deviation is 11 minutes for males and 16 minutes for females. Determine the 95% confidence interval estimate for the difference in mean times for men and women patients.

## **Steps to solution:**

- 1. Define the population parameter of interest and select independent samples from the two populations.**

The healthcare association is interested in estimating the difference in mean time spent in the clinic between males and females. The measure of interest is  $\mu_1 - \mu_2$ .

### **Steps to solution:**

#### **2. Specify the desired confidence level.**

To have a 95% confidence interval estimate.

#### **3. Compute the point estimate.**

Point estimate =  $\bar{x}_{males} - \bar{x}_{females} = 34.5 - 42.4 = -7.9$  minutes

Women in the sample spent an average of 7.9 minutes longer in the clinic.

#### **4. Determine the standard error of the sampling distribution**

standard error,  $\sigma_{\bar{x}_1, \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{11^2}{100} + \frac{16^2}{100}} = 1.9416$

#### **5. Determine the critical value, z, from the standard normal table.**

The critical value z at 95% confidence level are  $z = 1.96$

#### **6. Develop the confidence interval estimate**

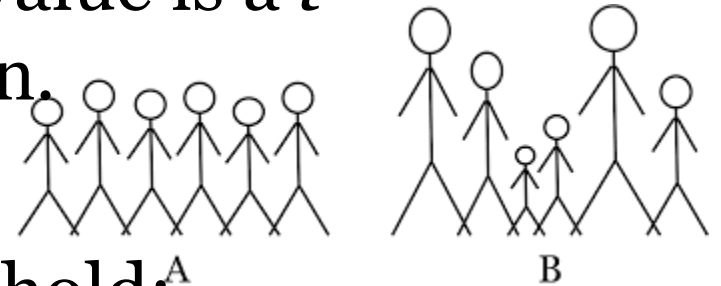
confidence interval,  $(\bar{x}_1 - \bar{x}_2) \pm z(\sigma_{\bar{x}_1, \bar{x}_2}) = (-7.9) \pm 1.96(1.9416)$   
 $= -7.9 \pm 3.8056$

#### **7. Identify the summary for the problem studied**

The 95% confidence interval estimate for the difference in mean time spent in the medical clinics between men and women is  $-11.706 \text{ minutes} \leq (\mu_1 - \mu_2) \leq -4.094$  minutes. Thus, based on the sample data and the specified confidence level, women spend on average between 4.09 and 11.71 minutes longer at the Healthcare Associates clinics.

# ESTIMATING THE DIFFERENCE BETWEEN TWO POPULATION MEANS WHEN $\sigma_1$ AND $\sigma_2$ ARE UNKNOWN, USING INDEPENDENT SAMPLES

- When estimating two population mean when the population **standard deviation is unknown** and the **sample sizes are small**, the critical value is a  $t$ -value from the  $t$ -distribution.



- The following assumptions hold:
  - ▣ The populations are normally distributed.
  - ▣ The populations have equal variances.
  - ▣ The samples are independent.

# ESTIMATING THE DIFFERENCE BETWEEN TWO POPULATION MEANS WHEN $\sigma_1$ AND $\sigma_2$ ARE UNKNOWN, USING INDEPENDENT SAMPLES

## **Problem:**

The head of research and development at Sneva Pharmaceutical Research is interested in estimating the difference between individuals age 50 and under and those over 50 years old with respect to the mean time from when a patient takes a new medication until the medication can be detected in the blood. Once she estimates the difference, if a difference does exist, the company can use this information to guide doctors in how to advise their patients to use this medication. A simple random sample of six people age 50 or younger and eight people over 50 participated in the study. The research manager wishes to have a 95% confidence interval estimate. The resulting sample means and sample standard deviations for the two groups are:

Age $\leq$ 50 years	Age $>$ 50 years
$\bar{x}_1 = 13.6$ minutes	$\bar{x}_1 = 13.6$ minutes
$S_1 = 3.1$ minutes	$S_2 = 5.0$ minutes

### **Steps to solution:**

#### **1. Define the population parameter of interest and select independent samples from the two populations.**

The objective here is to estimate the difference in mean time between the two age groups with respect to the speed at which the medication reaches the blood. The research lab has selected simple random samples of six “younger” and eight “older” people. Because the impact of the medication in one person does not influence the impact in another person, the samples are independent.

#### **2. Specify the desired confidence level.**

To have a 95% confidence interval estimate.

#### **3. Compute the point estimate.**

Point estimate =  $\bar{x}_1 - \bar{x}_2 = 13.6 - 11.2 = 2.4$  minutes

#### **4. Determine the standard error of the sampling distribution**

The pooled standard deviation is computed using:  $s_p = \sqrt{\frac{(n_1-1)s^2_1 + (n_2-1)s^2_2}{n_1+n_2-2}}$

$$s_p = \sqrt{\frac{(6-1)3.1^2 + (8-1)5.0^2}{6+8-2}} = 4.31$$

then standard error is calculated as,  $s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 4.31 \sqrt{\frac{1}{6} + \frac{1}{8}} = 2.3277$

## **Steps to solution:**

### **5. Determine the critical value, $z$ , from the standard normal table.**

Because the population standard deviations are unknown, the critical value will be a  $t$ -value from the  $t$ -distribution as long as the population variances are equal and the populations are assumed to be normally distributed.

The critical  $t$  for 95% confidence,  $t = (n_1 + n_2 - 2)$  degree of freedom, thus  $t = 6 + 8 - 2 = 12$  degrees of freedom is  $t = 2.1788$

### **6. Develop the confidence interval estimate**

$$\begin{aligned} \text{confidence interval, } (\bar{x}_1 - \bar{x}_2) \pm t(s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}) &= (2.4) \pm 2.1788(2.3277) \\ &= 2.4 \pm 5.0715 = -2.6715 \leq (\mu_1 - \mu_2) \leq 7.4715 \end{aligned}$$

### **7. Identify the summary for the problem studied**

Because the interval crosses zero, the research manager cannot conclude that a difference exists between the age groups with respect to the mean time needed for the medication to be detected in the blood. Thus, with respect to this factor, it does not seem to matter whether the patient is 50 or younger or over 50, so no advice for doctors is warranted.

# STATISTICAL INFERENCE

ESTIMATION FOR TWO POPULATION  
MEANS:

Paired Samples



# ESTIMATION FOR TWO POPULATION MEAN WITH PAIRED SAMPLES

- Previously, we have look into estimating the two population mean when the samples are independent.
- There are many situations in business where using paired samples should be considered.
  - What is paired samples? : Samples that are selected in such a way that values in one sample are matched with the values in the second sample for the purpose of controlling for extraneous factors.
  - Another term for paired samples is **dependent samples**.
  - Why paired samples? – to control for any variation in a sample. Example: different cars (and drivers), different painter and etc

# ESTIMATION FOR TWO POPULATION MEAN WITH PAIRED SAMPLES

## **Example of business application:**

A major oil company wanted to estimate the difference in average mileage for cars using a regular engine oil compared with cars using a synthetic-oil product. A random sample of 10 motorists (and their cars) was selected. Each car was filled with gasoline, the oil was drained, and new, regular oil was added. The car was driven 200 miles on a specified route. The car then was filled with gasoline and the miles per gallon were computed. After the 10 cars completed this process, the same steps were performed using synthetic oil. Because the same cars and drivers tested both types of oil, the miles-per-gallon measurements for synthetic oil and regular engine oil will most likely be related.



$$d = x_1 - x_2$$

where:

$d$  = Paired difference

$x_1$  and  $x_2$  = Values from samples 1 and 2, respectively

## Paired Difference

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$$

where:

$d_i$  =  $i$ th paired difference value

$n$  = Number of paired differences

## Point Estimate for the Population Mean Paired Difference, $\mu_d$

$$s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}}$$

where:

$d_i$  =  $i$ th paired difference

$\bar{d}$  = Mean paired difference

## Sample Standard Deviation for Paired Differences

## Confidence Interval Estimate for Population Mean Paired Difference, $\mu_d$

$$\bar{d} \pm t \frac{s_d}{\sqrt{n}}$$

where:

$t$  = Critical  $t$  value from  $t$ -distribution with  $n - 1$  degrees of freedom

$\bar{d}$  = Sample mean paired difference

$s_d$  = Sample standard deviation of paired differences

$n$  = Number of paired differences (sample size)

# ESTIMATION FOR TWO POPULATION MEAN WITH PAIRED SAMPLES

## **Problem:**

Technology has done more to change golf than possibly any other sport in recent years. Titanium woods, hybrid irons and new golf ball designs have impacted professional and amateur golfers alike. PGA of America is the association that only professional golfers can belong to. The association provides many services for golf professionals, including operating an equipment training center in Florida. Recently, a maker of golf balls developed a new ball technology, and PGA of America is interested in estimating the mean difference in driving distance for this new ball versus the existing best-seller. To conduct the test, the PGA of America staff selected six professional golfers and had each golfer hit each ball one time. The sample data is as follows:

Calculate the 95% confidence interval estimate for the difference in population means for paired samples.

<b>Golfer</b>	<b>Existing Ball</b>	<b>New Ball</b>
1	280	276
2	299	301
3	278	285
4	301	299
5	268	273
6	295	300

### **Steps to solution:**

#### **1. Define the population parameter of interest**

Because the same golfers hit each golf ball, the company is controlling for the variation in the golfers' ability to hit a golf ball. The samples are paired, and the population value of interest is  $\mu_d$ , the mean paired difference in distance. We assume that the population of paired differences is normally distributed.

#### **2. From the sample data collected, compute the point estimate, $\bar{d}$**

Golfer	Existing Ball	New Ball	$\bar{d}$
1	280	276	4
2	299	301	-2
3	278	285	-7
4	301	299	2
5	268	273	-5
6	295	300	-5

The Point estimate is,  $\bar{d} = \frac{\sum d}{n} = -2.17$  yards

## **Steps to solution:**

### **3. Calculate the standard deviation, $s_d$**

Standard deviation is computed using:  $s_d = \sqrt{\frac{\sum(d-\bar{d})^2}{n-1}} = 4.36 \text{ yard}$

### **4. Determine the critical value, $t$ , from the $t$ distribution table.**

The critical  $t$  for 95% confidence,  $t = (n-1)$  degree of freedom, thus  $t = 6 - 1 = 5$  degrees of freedom is  $t = 2.5706$

### **5. Develop the confidence interval estimate**

confidence interval,  $\bar{d} \pm t \left( \frac{s_d}{\sqrt{n}} \right)$

$$(-2.17) \pm 2.5706 \left( \frac{4.36}{\sqrt{6}} \right)$$

$$=-2.17 \pm 4.58 = -6.75 \text{ yards} \quad \text{_____} \quad 2.41 \text{ yards}$$

### **6. Identify the summary for the problem studied**

Based on these sample data and the confidence interval estimate, which contains zero, the PGA of America must conclude that the new ball's average distance may not be any longer than that of the existing best-seller. This may affect whether the company that developed the new ball will continue to make it.

# STATISTICAL INFERENCE

- end for ESTIMATION
- next: HYPOTHESIS TESTING