STATISTICAL ANALYSIS - LECTURE 07

Dr. Mahmoud Mounir

mahmoud.mounir@cis.asu.edu.eg

Sampling Distributions

Note: This PowerPoint is only a summary and your main source should be the book.

Population and Sample

- **Population:** It is the complete set of items of interest.
- **Sample:** It is a part or subset of the population used to represent the population.
- **Parameter:** A numerical value that describes some aspects of the population.
- **Statistic:** A numerical value that describes some aspects of the sample.
- **Statistics:** It is used to make inferences about population parameters.

Note: This PowerPoint is only a summary and your main source should be the book.

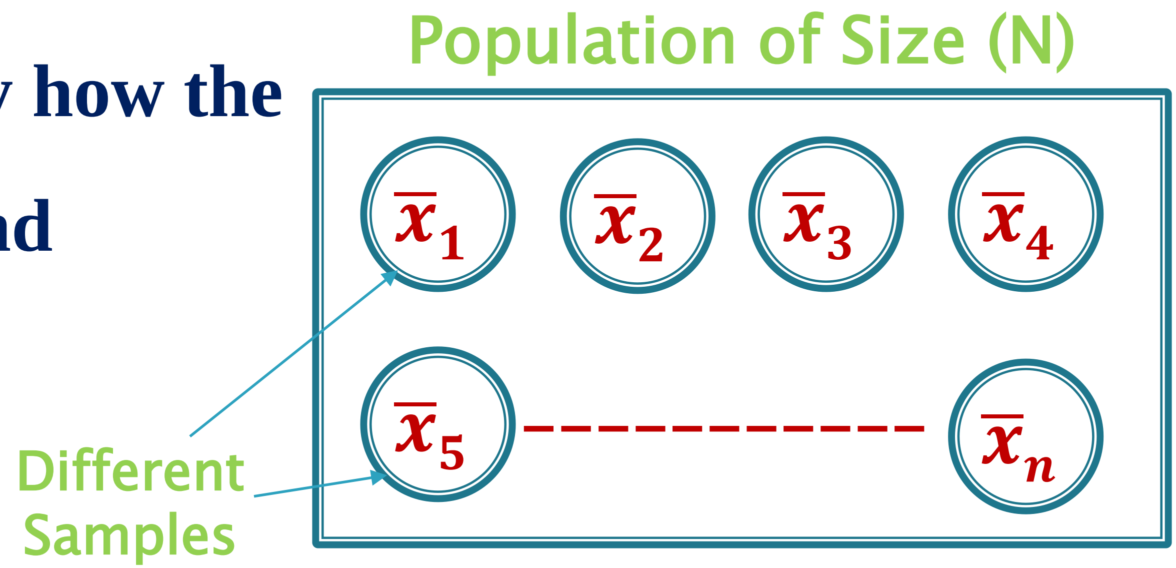
Population and Sample

	Parameter	Statistic
Size	N	n
Mean	μ	$\bar{x}$
Variance	σ^2	S^2
Standard Deviation	σ	S

- **Parameter:** It is always fixed for the same population.
- **Statistic:** It is varying depending on the type of samples taken from the population.

Sampling Distributions

- **Suppose we have a population of size (N) with Mean (μ) and Standard Deviation (σ).**
- **If we choose random samples from this population.**
- **If the size of a specific sample is (n), so the mean of this sample is $\overline{x} = \frac{\sum X}{n}$**
- **We have to study how the Sampling *Means* and *Proportion* vary.**



The Central Limit Theorem

Note: This PowerPoint is only a summary and your main source should be the book.

Distribution of Sample Means

- **A sampling distribution of sample means:** is a distribution using the means computed from all possible random samples of a specific size taken from a population.

- **Sampling error:** is the difference between the sample measure and corresponding population measure due to the fact that the sample is not a perfect representation of the population.

Note: This PowerPoint is only a summary and your main source should be the book.

Properties of the distribution of sample means:

1- The mean of the sample means will be **the same** as the population mean.

$$\mu_{\overline{X}} = \mu$$

2- The standard deviation of the sample means will be **smaller than** the standard deviation of the population, and it will be equal to the population standard deviation divided by the square root of the sample size.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Standard error of the mean

Example (1) :

If the sample size is 40 and the standard deviation of the sample mean is 3.95,Then the standard deviation of the population is ????

Solution:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \longrightarrow \sigma = \sqrt{n} \times \sigma_{\bar{x}}$$

$$\sigma = \sqrt{40} \times 3.95 = 24.9 \approx 25$$

Example (2) :

The variance of a variable is 16 .If a sample of 80 individuals is selected, Compute the standard error of the mean ?????

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \longrightarrow \sigma_{\bar{x}} = \frac{\sqrt{\sigma^2}}{\sqrt{n}} = \frac{\sqrt{16}}{\sqrt{80}} = \frac{4}{8.94} = 0.447 \approx 0.45$$

Note: This PowerPoint is only a summary and your main source should be the book.

The Central Limit Theorem:

As the sample size *n* increase without limit, the shape of the distribution of the sample means taken with replacement from a population with mean μ and standard deviation σ will approach a normal distribution. This distribution will have a mean μ and a standard deviation $\sigma/\sqrt{n}$

The formula for z values is:

$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \Rightarrow z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

Note: This PowerPoint is only a summary and your main source should be the book.

❑ Two important things when we use Central limit theorem

1- when the original variable is normally distributed , the distribution of the sample means will be normally distributed , for any sample size n.

2- when the original variable might not be normally distributed, a sample size must be ***30 or more*** to approximate the distribution of sample means to normal distribution .

❑ To solve the question about sample means we must use three steps.

First Step: find the z , using the formula.
Second Step: draw the figure and Shading the area.
Third Step : find the areas under ,using table E.

Note: This PowerPoint is only a summary and your main source should be the book.

Example (3): Hours That Children Watch Television

A. C. Neilsen reported that children between the ages of 2 and 5 watch an average of 25 hours of television per week. Assume the variable is normally distributed and the standard deviation is 3 hours. If 20 children between the ages 2 and 5 are randomly selected, find the probability that **the mean** of the number of hours they watch television will be greater than 26.3 hours?

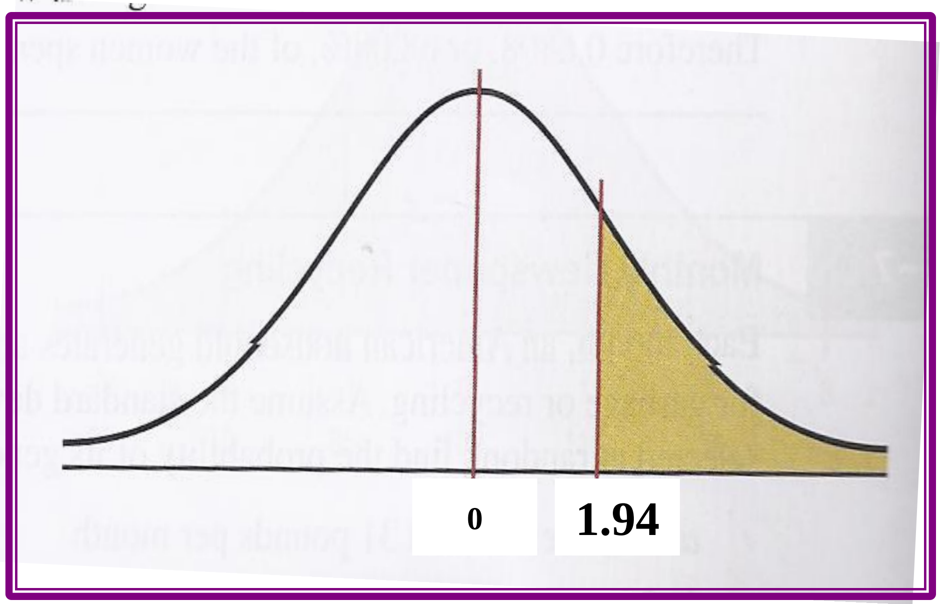
Solution:

$$\mu = 25, \sigma = 3, n = 20, \overline{X} = 26.3$$

Step 1 : Find the z value .

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{26.3 - 25}{\frac{3}{\sqrt{20}}} = \frac{1.3}{0.671} = 1.94$$

Step 2 :Draw the figure



Step 3 :Find the area ,using table E.

$P(z > 1.94) = 1.000 - 0.9738 = 0.0262$. That is, the probability of obtaining a sample mean greater than 26.3 is 0.0262 = 2.62%.

Example (4):

The average age of a vehicle registered in the United States is 8 years, or 96 months. Assume the standard deviation is 16 months. If a random sample of 36 vehicles is selected, find the probability that **the mean** of their age is between 90 and 100 months.

Solution:

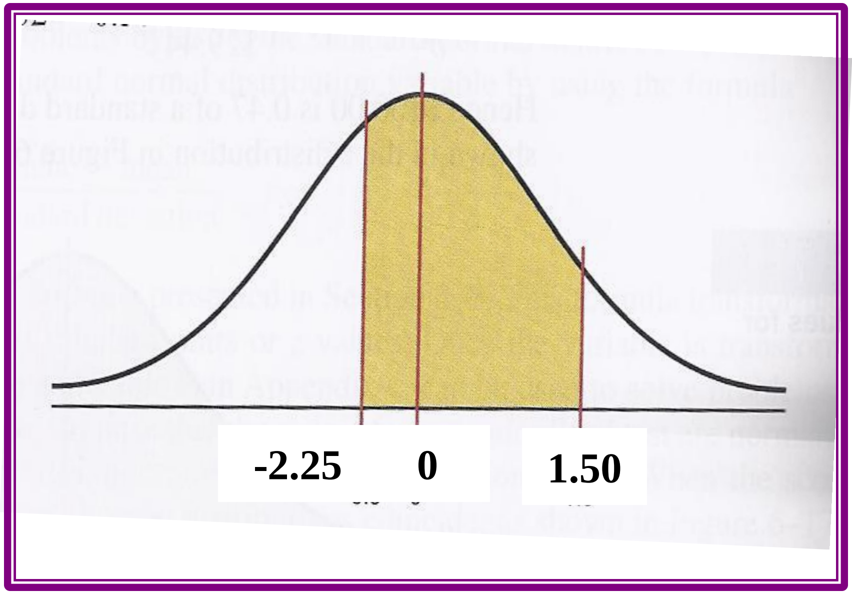
$$\mu = 96, \sigma = 16, n = 36, \bar{X}_1 = 90, \bar{X}_2 = 100$$

Step 1 : Find the two z values .

$$z_1 = \frac{\bar{x}_1 - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{90 - 96}{\frac{16}{\sqrt{36}}} = -2.25$$

$$z_2 = \frac{\bar{x}_2 - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{100 - 96}{\frac{16}{\sqrt{36}}} = 1.50$$

Step 2 :Draw the figure



Step 3 :Find the area ,using table E.

$$P(-2.25<Z<1.50)=P(Z<1.50)-P(Z<-2.25) =0.9332-0.0122$$
$$=0.9210 \text{ or } 92.1\ \%$$

The probability of obtaining a sample mean between 90 and 100 months is 92.1%; that is ,$P(90 <X<100)= 92.1\%$

Note: This PowerPoint is only a summary and your main source should be the book.

Example (5): Meat Consumption

The average number of pounds of meat that a person consumes per year is 218.4 pounds. Assume that the standard deviation is 25 pounds and the distribution is approximately normal.

a) Find the probability that a person selected at random consumes less than 224 pounds per year.

b) If a sample of 40 individuals is selected, find the probability that **the mean** of the sample will be less than 224 pounds per year.

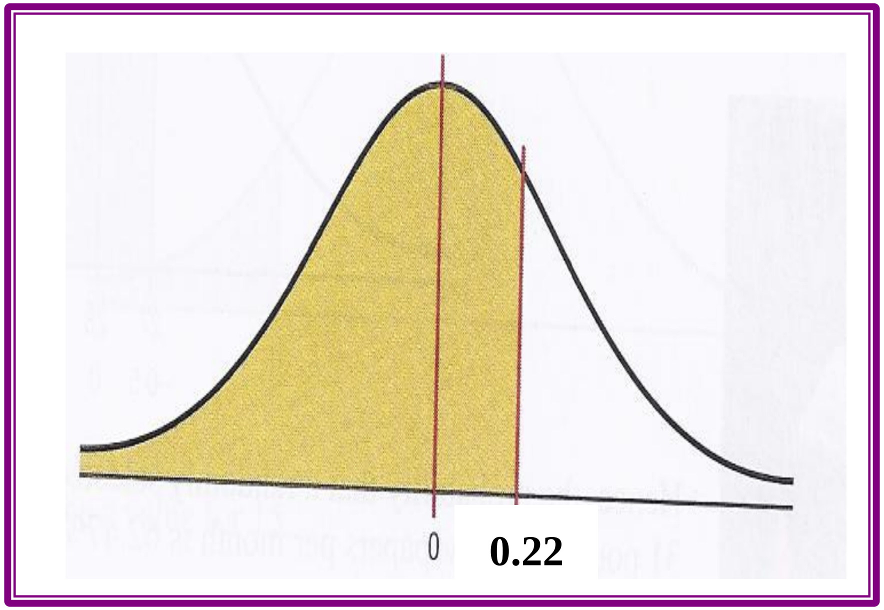
Solution(a):

$$\mu = 218.4 \ , \sigma = 25$$

Step 1 : Find the z value .

$$z = \frac{x - \mu}{\sigma} = \frac{224 - 218.4}{25} = 0.22$$

Step 2 :Draw the figure



Step 3 :Find the area ,using table E.

$P(Z<0.22) = 0.5871$

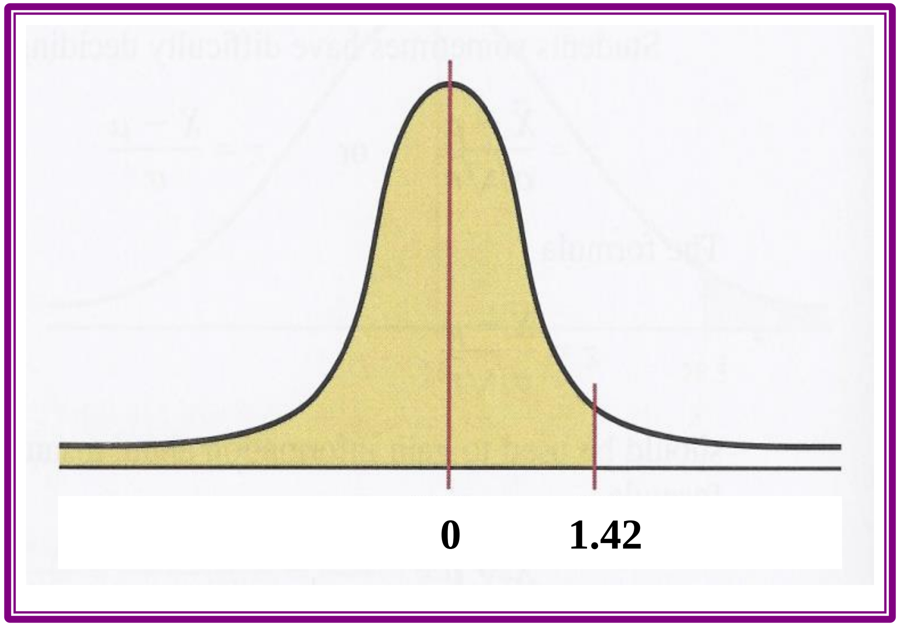
Note: This PowerPoint is only a summary and your main source should be the book.

Solution(b):

Step 1 : Find the z value .

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{224 - 218.4}{\frac{25}{\sqrt{40}}} = 1.42$$

Step 2 : Draw the figure



Note: This PowerPoint is only a summary and your main source should be the book.

Step 3 :Find the area ,using table E.

$P(z < 1.42) = 0.9222.$

The probability that mean of a sample of 40 individuals is less than 224 pounds per year is 0.9222 or 92.22%.

Note: This PowerPoint is only a summary and your main source should be the book.

Example (6):

Suppose it is known that in a certain large human population cranial الجمجمة length is approximately normally distributed with a mean of 185.6 mm and a standard deviation of 12.7 mm.

What is the probability that a random sample of size 10 from this population will have a mean greater than 190?

Solution:

$\mu = 185.6$	$\sigma = 12.7$	$n = 10$	$\bar{x} = 190$
$\mu_{\bar{x}} = \mu = 185.6$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12.7}{\sqrt{10}} = 4.0161$	$z_{\bar{x}} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{190 - 185.6}{4.0161} = 1.1$	
$P(\bar{x} > 190) = P(z_{\bar{x}} > 1.1)$			
$= 1 - P(z_{\bar{x}} \leq 1.1)$			
$= 1 - 0.8643 = \mathbf{0.1357}$			

Example (7):

If the mean and standard deviation of a serum iron الحديد في الدم values for healthy men are 120 and 15 micrograms per 100 ml, respectively,

what is the probability that a random sample of 50 normal men will yield a mean between 115 and 125 micrograms per 100 ml?

Solution:

$\mu = 120$	$\sigma = 15$	$n = 50$	$\bar{x}_1 = 115$ $\bar{x}_2 = 125$
$\mu_{\bar{x}} = \mu = 120$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{50}} = 2.12$	$z_{\bar{x}_1} = \frac{\bar{x}_1 - \mu}{\sigma_{\bar{x}}} = \frac{115 - 120}{2.12} = -2.36$	$z_{\bar{x}_1} = \frac{\bar{x}_2 - \mu}{\sigma_{\bar{x}}} = \frac{125 - 120}{2.12} = 2.36$

$$P(115 \le \bar{x} \le 125) = P(-2.36 \le z_{\bar{x}} \le 2.36)$$
$$= P(z_{\bar{x}} \le 2.36) - P(z_{\bar{x}} \le -2.36)$$
$$= 0.9909 - 0.0091 = \mathbf{0.9818}$$

Note: This PowerPoint is only a summary and your main source should be the book.

Example (7):

Researchers found the mean sodium intake إمتصاص in men and women 60 years or older to be 2940 mg with a standard deviation of 1476 mg. Use these values for the mean and standard deviation of the U.S. population.

Find the probability that a random sample of 75 people from the population will have a mean

a) less than 2450 mg

b) over 3100 mg

c) Between 2500 and 3300 mg

d) Between 2500 and 2900 mg

Solution (a):

$\mu = 2940$	$\sigma = 1476$	$n = 75$	$\bar{x} = 2450$
$\mu_{\bar{x}} = \mu = 2940$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1476}{\sqrt{75}} = 170.43$	$z_{\bar{x}} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{2450 - 2940}{170.43} = -2.88$	
$P(\bar{x} < 2450) = P(z_{\bar{x}} < -2.88)$			
$= \mathbf{0.0020}$			

Solution (b):

$\mu = 2940$	$\sigma = 1476$	$n = 75$	$\bar{x} = 3100$
$\mu_{\bar{x}} = \mu$ $= 2940$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $= \frac{1476}{\sqrt{75}}$ $= 170.43$	$z_{\bar{x}} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ $= \frac{3100 - 2940}{170.43}$ $= 0.94$	
$P(\bar{x} > 3100) = P(z_{\bar{x}} > 0.94)$			
$= 1 - P(z_{\bar{x}} \leq 0.94)$			
$= 1 - 0.8264 = \mathbf{0.1736}$			

Solution (c):

$\mu = 2940$	$\sigma = 1476$	$n = 75$	$\bar{x}_1 = 2500$ $\bar{x}_2 = 3300$
$\mu_{\bar{x}} = \mu = 120$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1476}{\sqrt{75}} = 170.43$	$z_{\bar{x}_1} = \frac{\bar{x}_1 - \mu}{\sigma_{\bar{x}}} = \frac{2500 - 2940}{170.43} = -2.58$	$z_{\bar{x}_1} = \frac{\bar{x}_2 - \mu}{\sigma_{\bar{x}}} = \frac{3300 - 2940}{170.43} = 2.11$
$P(2500 \leq \bar{x} \leq 3300) = P(-2.58 \leq z_{\bar{x}} \leq 2.11)$			
$= P(z_{\bar{x}} \leq 2.11) - P(z_{\bar{x}} \geq -2.58)$			
$= 0.9826 - 0.0049 = \mathbf{0.9777}$			

Note: This PowerPoint is only a summary and your main source should be the book.

Solution (d):

$\mu = 2940$	$\sigma = 1476$	$n = 75$	$\bar{x}_1 = 2500$ $\bar{x}_2 = 2900$
$\mu_{\bar{x}} = \mu = 120$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $= \frac{1476}{\sqrt{75}} = 170.43$	$z_{\bar{x}_1} = \frac{\bar{x}_1 - \mu}{\sigma_{\bar{x}}}$ $= \frac{2500 - 2940}{170.43}$ $= -2.58$	$z_{\bar{x}_1} = \frac{\bar{x}_2 - \mu}{\sigma_{\bar{x}}}$ $= \frac{2900 - 2940}{170.43}$ $= -0.23$
$P(2500 \leq \bar{x} \leq 2900) = P(-2.58 \leq z_{\bar{x}} \leq -0.23)$			
$= P(z_{\bar{x}} \leq -0.23) - P(z_{\bar{x}} \geq -2.58)$			
$= 0.4090 - 0.0049 = \mathbf{0.4041}$			