STATISTICAL ANALYSIS - LECTURE 08

Dr. Mahmoud Mounir

mahmoud.mounir@cis.asu.edu.eg

Sampling Distribution of the Difference between Two Sample Means

Sampling Distribution of Means

Result:

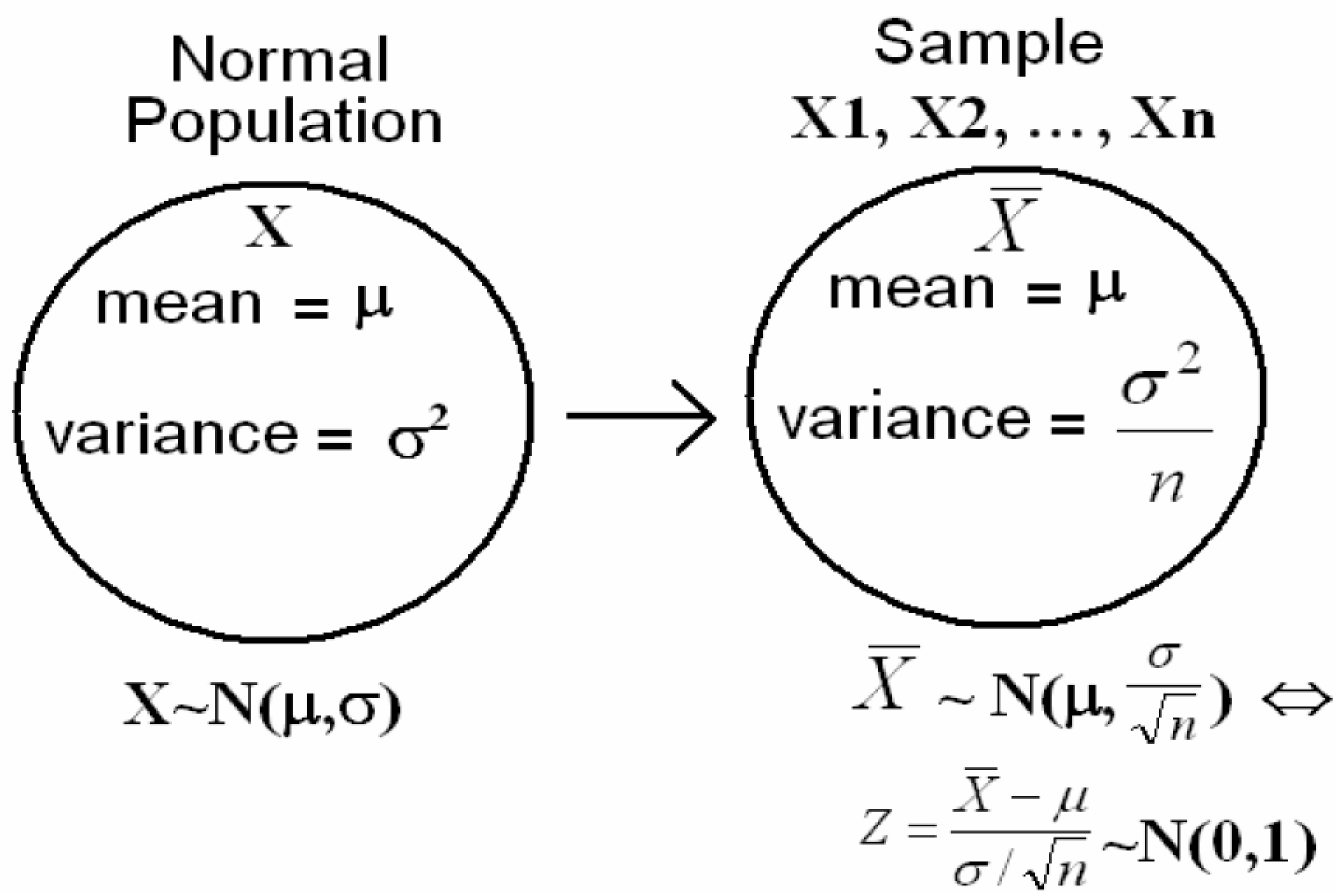
If $X_1, X_2, \ldots, X_n$ is a random sample of size $n$taken from a normal distribution with mean μ and variance σ^2, i.e. $N(\mu, \sigma^2)$, then the sample mean $\bar{X}$has a normal distribution with mean

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

and variance

$$Var(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

- If $X_1, X_2, \ldots, X_n$ is a random sample of size n from $N(\mu,\sigma)$, then $\bar{X} \sim N(\mu_{\bar{X}}, \sigma_{\bar{X}})$ or $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$.
- $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \Leftrightarrow Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$



Theorem: (Central Limit Theorem)

If X_1, X_2, …, X_n is a random sample of size n from any distribution (population) with mean μ and finite variance σ^2, then, if the sample size n is large, the random variable

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

is approximately standard normal random variable, i.e.,

$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$ approximately.

Sampling Distribution of the Difference between Two Means

Suppose that we have two populations:

• 1-st population with mean μ_1 and variance $\sigma^2{}_1$

• 2-nd population with mean μ_2 and variance $\sigma^2{}_2$

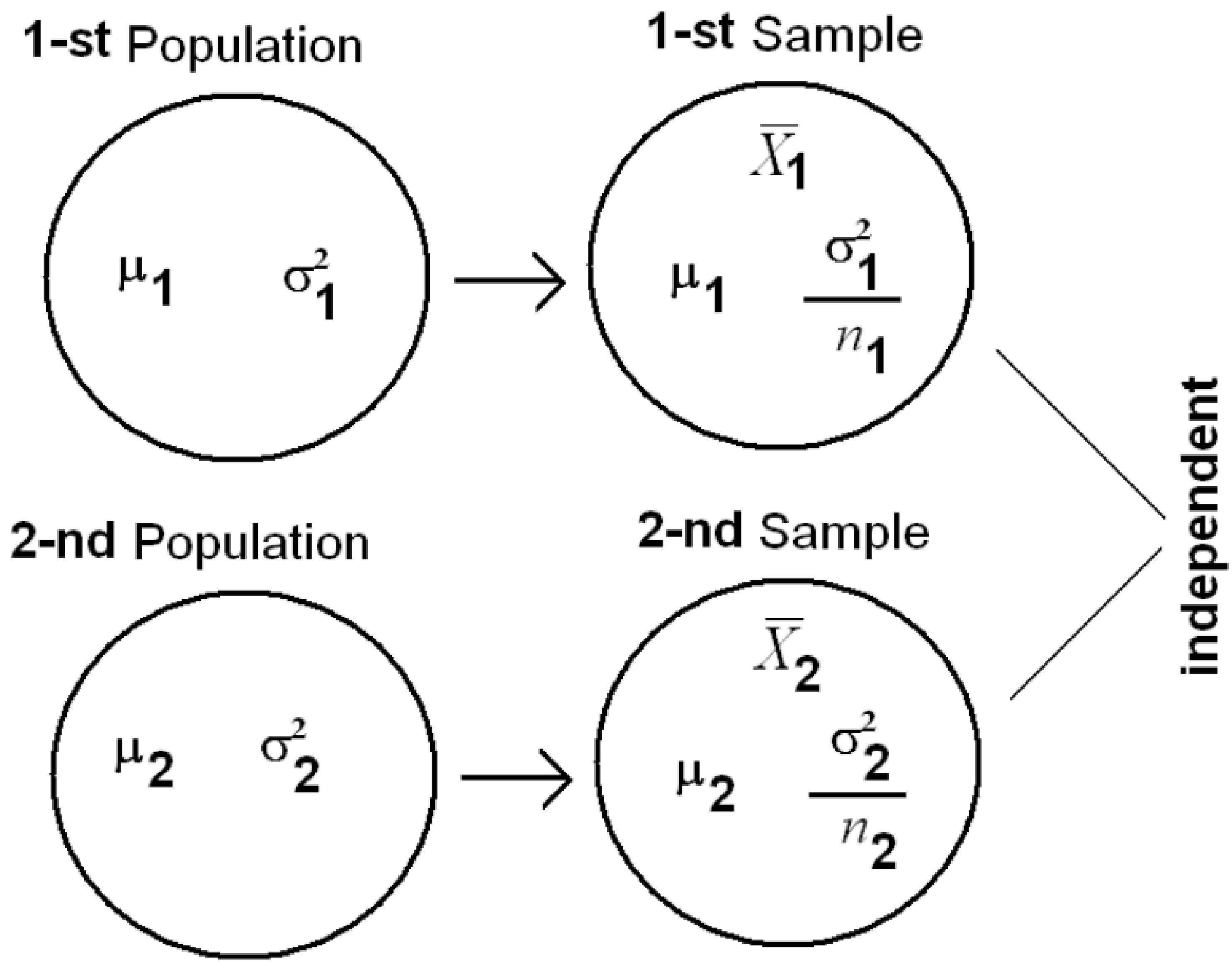
• We are interested in comparing μ_1 and μ_2, or equivalently, making inferences about $\mu_1 - \mu_2$.

• We independently select a random sample of size n_1 from the 1-st population and another random sample of size n_2 from the 2-nd population:

• Let $\bar{X}_1$ be the sample mean of the 1-st sample.

• Let $\bar{X}_2$ be the sample mean of the 2-nd sample.

• The sampling distribution of $\bar{X}_1 - \bar{X}_2$ is used to make inferences about $\mu_1 - \mu_2$.


1-st Population
μ_1 σ_1^2
1-st Sample
$\bar{X}_1$
μ_1 $\frac{\sigma_1^2}{n_1}$
2-nd Population
μ_2 σ_2^2
2-nd Sample
$\bar{X}_2$
μ_2 $\frac{\sigma_2^2}{n_2}$
independent

Theorem

If n_1 and n_2 are large, then the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is approximately normal with mean

$$E(\bar{X}_1 - \bar{X}_2) = \mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$

and variance

$$Var(\bar{X}_1 - \bar{X}_2) = \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

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

that is:

$$\bar{X}_1 - \bar{X}_2 \sim \mathrm{N}(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$$

$$\Leftrightarrow$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1)$$

Note:

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\sigma^2_{\bar{X}_1 - \bar{X}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \neq \sqrt{\frac{\sigma_1^2}{n_1}} + \sqrt{\frac{\sigma_2^2}{n_2}} = \frac{\sigma_1}{\sqrt{n_1}} + \frac{\sigma_2}{\sqrt{n_2}}$$

Example (1)

The television picture tubes of manufacturer ***A*** have a mean lifetime of 6.5 years and standard deviation of 0.9 year, while those of manufacturer ***B*** have a mean lifetime of 6 years and standard deviation of 0.8 year. What is the probability that a random sample of **36** tubes from manufacturer ***A*** will have a mean lifetime **that is at least 1 year more** than the mean lifetime of a random sample of **49** tubes from manufacturer ***B***?

Solution:

Population 1	Population 2
$\mu_1 = 6.5$	$\mu_2 = 6.0$
$\sigma_1 = 0.9$	$\sigma_2 = 0.8$
$n_1 = 36$	$n_2 = 49$

We need to find the probability that the mean lifetime of manufacturer A is at least 1 year more than the mean lifetime of manufacturer B which is $P(\bar{X}_1 \geq \bar{X}_2 + 1)$

The sampling distribution of $\bar{X}_1 - \bar{X}_2$ is

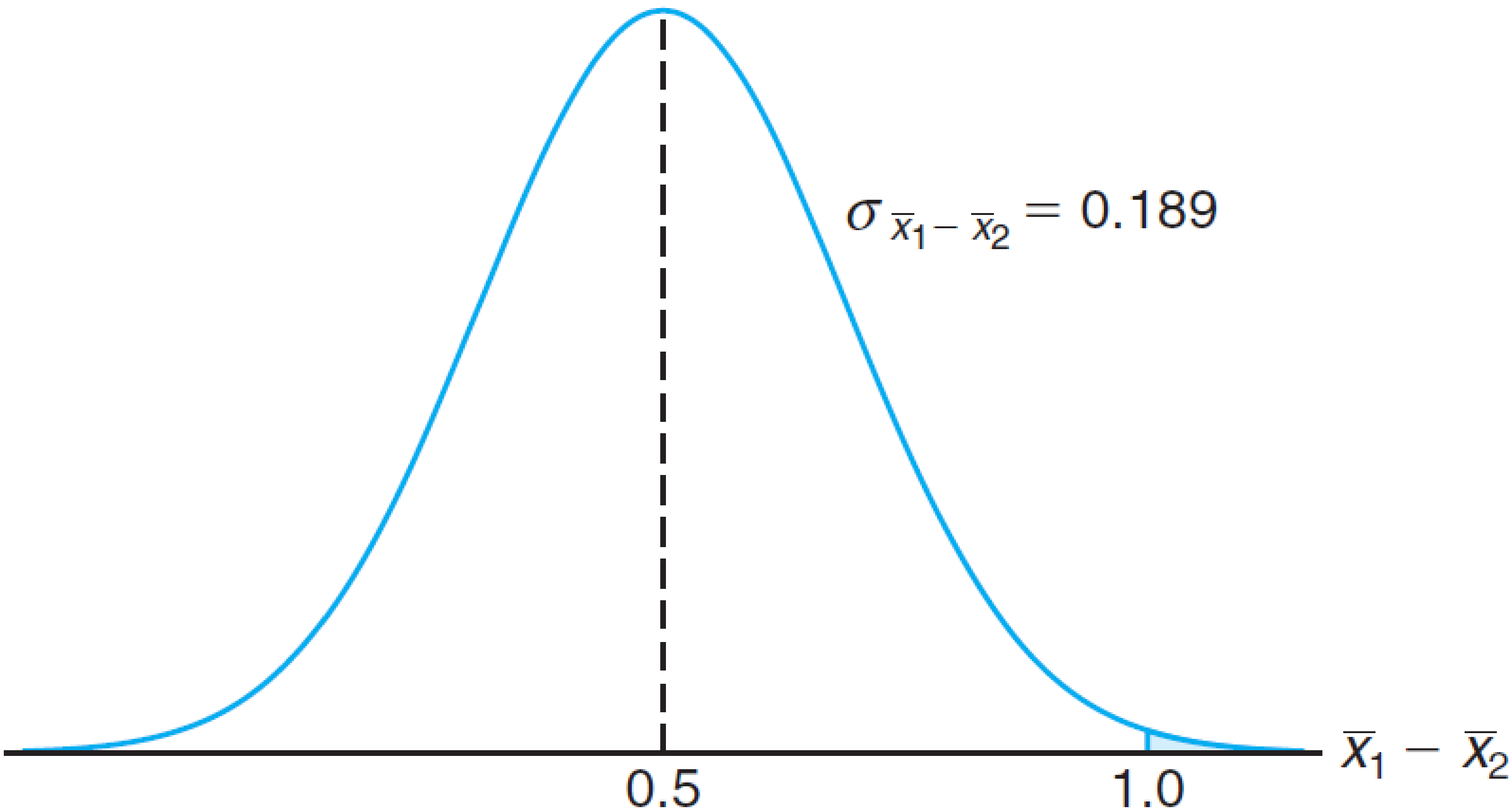
$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$$

$$E(\bar{X}_1 - \bar{X}_2) = \mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 = 6.5 - 6.0 = 0.5$$

$$Var(\bar{X}_1 - \bar{X}_2) = \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{(0.9)^2}{36} + \frac{(0.8)^2}{49} = 0.03556$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{0.03556} = 0.189$$

$$\bar{X}_1 - \bar{X}_2 \sim N(0.5, 0.189)$$



$\sigma_{\bar{x}_1 - \bar{x}_2} = 0.189$
0.5
1.0
$\bar{x}_1 - \bar{x}_2$

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$P(\overline{X}_1 \geq \overline{X}_2 + 1) = P(\overline{X}_1 - \overline{X}_2 \geq 1)$$

$$= P\left(\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq \frac{1 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)$$

$$= P\left(Z \geq \frac{1 - 0.5}{0.189} \right)$$

$= P(Z \geq 2.65)$

$= 1 - P(Z < 2.65)$

$= 1 - 0.9960$

$= 0.0040$

Example (2)

Suppose it has been established that for a certain type of client that average length of a home visit by a public health nurse is 45 minutes with a standard deviation of 15 minutes, and that for a second type of client the average visit is 30 minuets long with a standard deviation of 20 minutes. If a nurse randomly visits **35** clients from the first and **40** from the second population, what is the probability that the average length of home visit **will differ between the two groups by 20 or more minutes**?

Solution:

$$z_{\bar{x}_1-\bar{x}_2} = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}}$$

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

Sample (1)	$\mu_1 = 45$	$\sigma_1 = 15$	$n_1 = 35$
Sample (2)	$\mu_2 = 30$	$\sigma_2 = 20$	$n_2 = 40$

$\mu_1 = 45$	$\sigma_1 = 15$	$n_1 = 35$
$\mu_2 = 30$	$\sigma_2 = 20$	$n_2 = 40$
$(\bar{x}_1-\bar{x}_2) = 20$	$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ $= \sqrt{\frac{15^2}{35}+\frac{20^2}{40}} = \sqrt{\frac{115}{7}}$ $= 4.053$	$z_{\bar{x}_1-\bar{x}_2} = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}}$ $= \frac{20-15}{4.053} = 1.23$
$P((\bar{x}_1-\bar{x}_2) \geq 20) = P(z_{\bar{x}_1-\bar{x}_2} \geq 1.23)$		
$= 1 - P(z_{\bar{x}_1-\bar{x}_2} < 1.23)$		
$= 1 - 0.8907 = \mathbf{0.1093}$		