

**Course Title:** Experimental Design

**Course and Code:** Stat 337

**Department:** Statistics and Operations Research

**College:** Sciences

**Institution:** King Saud University

**Academic Year:** 1447      **Semester:** Second

**Course Instructor:** Dr Mahmoud Alderiny

**E.mail:** [malderiny@ksu.edu.sa](mailto:malderiny@ksu.edu.sa)

# *Lecture 1*

## *Definitions And Concepts Related to Experimental Design*

# The Contents

- 1. Types of Data.**
- 2. Symbols used in the experimental design.**
- 3. Measures of central tendency and dispersion.**
- 4. Concepts and foundations of design experience.**

# 1- Types of Data

**Qualitative  
Data**

**Nominal**

**Ordinal**

**Quantitative  
Data**

**Interval**

**Ratio**

# Qualitative Data

**Is non-numeric data, it includes mutually exclusive groups represent the levels of phenomenon under study**

## 1- Nominal scale

**Levels of phenomenon cannot be differentiated, examples:**

Gender data (Male, Female) , Methods of fertilization, Nationality (Saudi, non-Saudi).

## 2- Ordinal scale

**Levels of phenomenon can be differentiated , examples:**

Educational level (primary, middle, secondary and university or higher), The concentration of sodium acetate (0%,0.5% ,0.10%), Sets of family income thousand riyals in month (< 5 , 5 - 10, 10 - 15, 15 - 20 , > 20) .

# Quantitative Data

**Is numeric data, it represent the actual value of the phenomenon**

## 1- Interval scale

**The zero value means that the phenomenon is exist, examples:**

Temperatures , Humidity, Student test score

## 2- Ratio scale

**The value 0 indicates lack of phenomenon., examples:**

Productivity per acre tons / ha., Area by Donum that cultivated by fodder, Qquantity milk produced by the cow in the day, Number of times the farmer to use a specific type of fertilizer during season.

## 2-Mathematical symbols used in the analysis

<i>symbol</i>	<i>Name</i>
$\alpha$	<i>Alpha</i>
$\beta$	<i>Beta</i>
$\varepsilon$	<i>Epsilon</i>
$\gamma$	<i>Gamma</i>
$\lambda$	<i>Lambda</i>
$\mu$	<i>Mu</i>
$\rho$	<i>Rho</i>

<i>symbols</i>	<i>Name</i>
$\sigma$	<i>Sigma</i>
$\sigma^2$	<i>Sigma square</i>
$\sqrt{\quad}$	<i>Square Root</i>
$\Sigma$	<i>Summation</i>
$\tau$	<i>Tau</i>
$\theta$	<i>Theta</i>

- **Sum of  $n$  observations,**  
 $(x_1, x_2, \dots, x_n)$  **is written as:**

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

-**And sum of squares**  
**is written as:**

$$\sum_{i=1}^n x_i^2 = x_1^2 + x_2^2 + \dots + x_n^2$$

## 3- Measure of central tendency and

### Definitions

#### **measure of central tendency:**

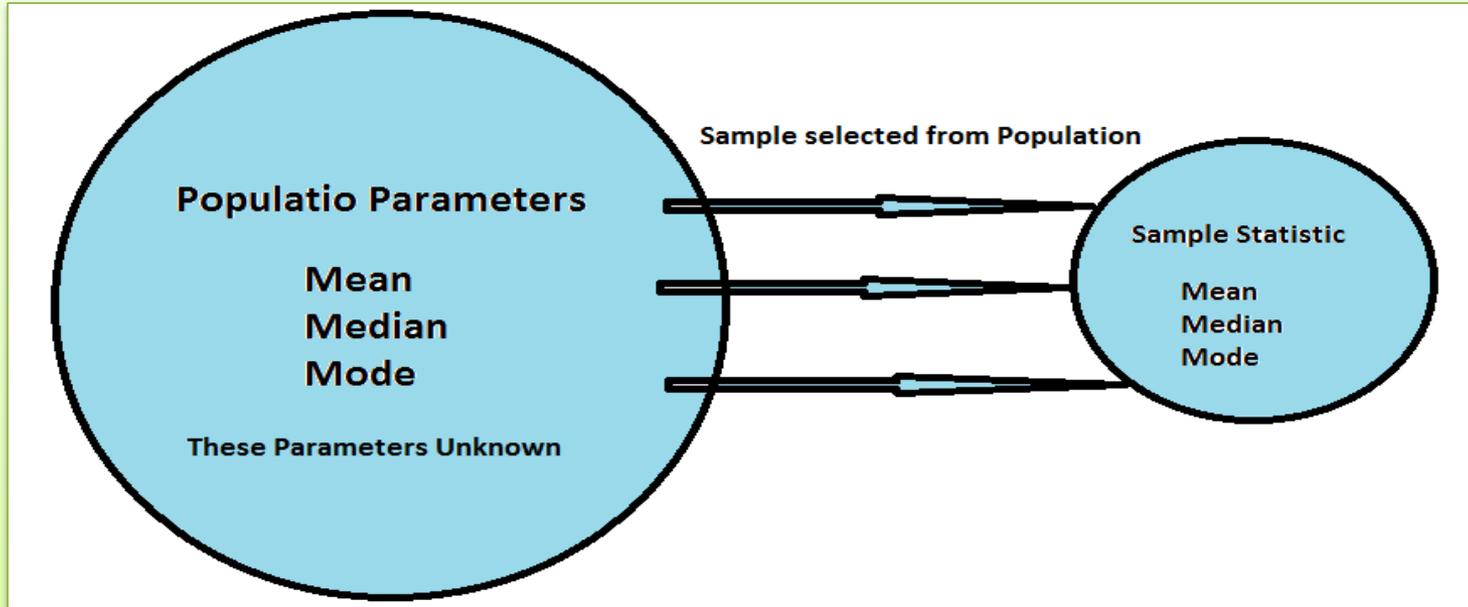
Is the measure that show where the data centered.

#### **Parameter of Population:**

Is a measure computed by using all population data.

#### **Statistics:**

Is a measure computed by using sample data.



Since it is difficult to measure a population parameter, a sample of size  $n$  whose values  $(x_1, x_2, \dots, x_n)$  is drawn from this population and computed the sample statistics as estimate for population parameter.

# Central tendency Measures

**Sample mean:**

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Where  $n$  is sample size

**Example (1)**

Sample  
weights  
of size 8  
person

Person	Weights $X$
1	63
2	59
3	66
4	71
5	71
6	74
7	54
8	64

$$\begin{aligned}\bar{X} &= \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \\ &= \frac{63 + 59 + 66 + 71 + 71 + 74 + 54 + 64}{8} \\ &= \frac{522}{8} = 65.25 \text{ kg}\end{aligned}$$

Sample mean is 65.25 kg

# Dispersion Measures

**Range :**

*Range* = Largest value - Smallest value

**Sample Variance:**

$$S^2 = \frac{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2 / n}{n-1}$$

**Standard Deviation:**

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2 / n}{n-1}}$$

**Coefficient of Variation:**

$$C.V = (S / \bar{X}) \times 100$$

## Example (2):

From weight data in example (1), compute the four statistical measures above.

### Solution

Range:

$$\begin{aligned} \text{Range} &= \text{Largest value} - \text{Smallest value} \\ &= 74 - 54 = 20 \end{aligned}$$

Variance:

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2 / n}{n-1} = \frac{34376 - (522)^2 / 8}{7} \\ &= \frac{315.5}{7} = 45.071 \end{aligned}$$

Standard Deviation:

$$S = \sqrt{S^2} = \sqrt{45.071} = 6.714$$

Coefficient of Variation:

$$C.V = (S / \bar{X}) \times 100 = (6.714 / 65.25) \times 100 = 10.29$$

*Weight data largely homogeneous*

## **4- Concepts and foundations of design experience**

### **Design and analysis of experiments**

Is a branch of applied statistics that applies the statistical methods in the field of agricultural research and experiments.

### **Experimental Design**

Is a series of stages to reach the optimal using of available possibilities thought application the most appropriate experimental design, in which collect data, organize, and analyze on the basis of sound science to reach decisions with high accuracy under uncertainty.

## **Experiment**

Is the foundation of the scientific method, that used to test hypotheses and explore the relationships and concepts that relate to the certain problems.

## **Factors**

Factor is categorical variable includes disjoint groups or levels called treatments, which are placed under the control of the researcher in order to allocate its on experimental units according to the design model that used.

## **Treatment**

Is the method that measured its impact on the experimental unit, may represent levels of factor (Factor), or combinations of levels of different factors.

## **Experimental unit**

Is the smallest piece of experimental material, and treatment are interviewed, they are receiving unit treatment, experimental unit may be a piece of land, the amount of bacteria, animals, people, tree, a leaf of the tree and partial of the human body.

## **Randomization**

It means the distribution of treatments on the experimental units randomly.

## **Sources of experimental Errors**

- The heterogeneity of the experimental units.
- An error in the way of implementation of the experiment.
- Technical mistakes: due to the presence of errors in the measurement processes and appreciation.

## **Factors which must be taken into account when you perform experiment according to the design**

- Number of factors.
- Fixed and random effects.
- Identify the experimental units and their characteristics.
- Duplicates replicated
- Randomization.
- The form of a mathematical model (fixed - random - Mixed).

## **Guidelines for Designing an Experiment**

1. Recognition of and statement of the problem.
2. Selection of the response variable.
3. Choice of factors, levels, and ranges.
4. Choice of experimental design.

## HW1.1 (problems 1.7 page 24)

**1.1.** Suppose that you want to design an experiment to study the proportion of unpopped kernels of popcorn. Complete steps 1–3 of the guidelines for designing experiments in Section 1.4. Are there any major sources of variation that would be difficult to control?

**1.3.** Suppose that you want to compare the growth of garden flowers with different conditions of sunlight, water, fertilizer, and soil conditions. Complete steps 1–3 of the guidelines for designing experiments in Section 1.4.

**1.8.** What is replication? Why do we need replication in an experiment? Present an example that illustrates the difference between replication and repeated measurements.

**1.9.** Why is randomization important in an experiment?