**Practical 3**

**Mathematical ((Numerical)) representation of data**

1. **Measurements of location ((Measurement of central tendency)).**

The most commonly used characteristic of a set of data is its centre, or the point about which many of the observations are clustered. We can measure:

a- **Arithmetic Mean** (Average)

 The most commonly used measure of central tendency. This is the sum (Σ) of all observations divided by the number of observation. We have:

 - Population Mean (µ) = Σx / N (N= No. of population)

 - Sample Mean (x) = Σx /n (n= No. of sample)

* **Simple, unique and** takes into account the magnitude of each and every observation in the data.

**EX: Suppose data are: 90, 80, 95, 85, and 65 (n=5)**

**Sample mean = 90 + 80 + 95 + 85 + 65 = 83**

 **5**

**The mean can be thought of as a “balancing point”, “center of gravity”**

****

Disadvantages: Affected by the extreme values, which in some cases so distorted that it become undesirable as a measure of central tendency.



b**- Median (50th percentile, 5th deciles).**

It is the value that divides the data into two equal parts (50% greater and 50% lesser than it)) giving that the values are arranged in o ordered of magnitude.

* If the number of values is odd, then the position of the median = n+1/ 2.
* If the number of values is even, then the median is taken to be the mean of (n/2) and (n/2 +1) observations.

**\*Advantages.** 1- Unique, simple and not affected by the extreme values.

The Median is a Better Description (than the Mean) of the Majority When the distribution is skewed (not affected by the extreme values).

**Ex:**  Data are: 14, 89, 93, 95, 96

• Skewness is reflected in the outlying low value of 14

• The sample mean is 77.4

• The median is 93



When do we use mean & median?

 If there are extreme values in a set of data we use the median, so as not to be misled by extreme value when we use the mean.

Disadvantages: it neglects all the values & takes only the median one.

c- **Mode.**

 The value which occurs most frequently. **Mode is not unique**

* If all values are different → No mode.
* If one value which occur most frequently → Uni-modal.
* If two values which occur most frequently → Bi-modal.
* If 3 values which occur most frequently → Tri-modal….etc.

**\***Advantages. 1- Simple, easy to calculate and unlike mean & median, the mode can be used for qualitative data (e.g., modal diagnosis as HT, DM, IHD …..etc.).

**EX:** Suppose data are: 90, 80, 95, 85, 90, and 65 (n=5) Mode=90

**Percentiles (*P*):** The value below which a given percentage of the cases fall. The value of the kth percentile is obtained by ranking the observations and looking at the value of the {(k/ n) X100}th observation.

 **EX1:** A teachergives a 20 point test to 10 students. The score are: 18, 15, 12, 6, 8, 2, 3, 5, 20, and 10. Find the percentile of score 12.

 **Sol:** Arrange the data in order from lowest to highest: 2, 3, 5, 6, 8, 10, 12, 15, 18, and 20.

 (k/ n) X100= (7/10)X100 = 70% thus the student whose score 12 did better than 70% of the class.

**EX2:** Find the value that corresponds to 60th percentile**.**

(60 X 10)/100=6 (position of 60th percentile), hence 10 score. Thus any one scoring 10 would have done better than 60% of the class.

**Deciles (*P*):** position in tenths a data value in the distribution → (k/ n) X10

**Quartile (*Q*);** position in fourths a data value in the distribution → (k/ n) X4

* The first quartile (Q1) is the value of the {(n + 1)÷4}th observation.
* The third quartile (Q3) is the value of the {3x(n + 1)÷4}th observation.
* Interquartile range (IQR): The middle half of the values. i.e. those lying between the first and third quartiles(Q1- Q3).

****

**Max**

**Min**

**Analyze – Descriptive Statistics ---- Frequencies ----- statistics**