Measure of central tendency

مقاييس النزعة المركزية

- **Frequency التكرار:**

Frequency is a measure of the number of occurrences of a repeating event per unit time.

Example:

Given data 2, 3, 4, 2, 6, 2, 4, 3

x	2	3	4	6
F(x)	3	2	2	1

The values of x with frequencies are called frequency distribution. In the data given above $x = 2$ has frequency as 3, $x = 3$ has frequency 2 and $x = 4$ has frequency 2 and $x = 6$ has frequency 1.

There are many ways of capturing important information about a random variable, the famous types of these measures are:

1. **The mean (arithmetic mean) الوسط الحسابى**

The arithmetic mean (the mean) of a list of numbers is the sum of all the numbers of the list divided by the number of items in the list:أى أن الوسط الحسابى للبيانات يساوى مجموع البيانات مقسوما على عددها

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$

If the list is a statistical population, the mean of that population is called a population mean. If the list is a statistical sample, we call the resulting statistic a sample mean. If each number in the data has some frequency given as

x	x_1	x_2	x_3	…………	x_n
F(x)	f_1	f_2	f_3	…………	f_n

Then the mean is calculated using the formula

$$Mean = \bar{x} = \mu_1 = \frac{\sum f_i x_i}{\sum f_i}.$$

Example (1) for the sample of data 2, 3, 5, 7, 11, find the mean?

The mean is given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{2+3+5+7+11}{5} = \frac{28}{5} = 5.6.$$

Example (2) Find the mean of the data given in the following table?

The mean is given by

x	3	5	7	9
F(x)التكرار	2	1	2	4

$$The\,Mean = \mu_1 = \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{2(3)+(1)5+(2)7+4(9)}{2+1+2+4} = \frac{6+5+14+36}{9} = \frac{61}{9} = 6.77.$$

2. **The Median الوسيط**

The median is the number separating the higher half of a sample, a population, or a probability distribution, from the lower half. The median of finite list of numbers can be found by arranging all the observations from the lowest value to highest value and taking the middle one.

For example the sequence of first five odd numbers; 5, 7, 9, 11, 13 has the median 9.

If the number of the sample data is even, the mean is obtained by taking the mean of the two middle values. For example the median of data sample: 6, 9, 10, 12, 14, 16 is given as:

$$\text{Median} = \text{mean of the middle} = \frac{10+12}{2} = 11.$$

3. The Mode المنوال

The mode is the value that occurs most frequently in a data set or probability distribution.

أى أنه القيمة التى تأخذ أكبر تكرار فى الحدوث

For the example the sample data 3, 6, 8, 3, 4, 5, 3, 4, 3, 9 has the mode 3 (because 3 is repeated four times).

The mode is not necessarily unique, i.e., the data may have more than one mode for example: the sample data 2, 5, 7, 2, 6, 2, 5, 8, 5 the mode is two numbers; 2 and 5 (because 2 is repeated three times and 5 is repeated three times).

4. Moments العزوم :

The moments about the origin العزوم حول نقطة الأصل are denoted by $\mu_1', \mu_2', \mu_3', \ldots$ where;

$$\mu_1' = \bar{x} = \frac{\sum f_i x_i}{\sum f_i},$$

$$\mu_2' = \frac{\sum f_i x_i^2}{\sum f_i},$$

$$\mu_3' = \frac{\sum f_i x_i^3}{\sum f_i}.$$

Exc. Calculate the first four moments about the origin for the given data:

x	1	2	3	4	5
F(x)التكرار	1	6	13	24	30

The moments about the mean العزوم حول الوسط الحسابى are denoted by $\mu_1, \mu_2, \mu_3, \ldots$are defined as:

$$\begin{aligned}\mu_1 = 0 &= \frac{\sum f_i(x_i - \bar{x})}{\sum f_i} \\ &= \frac{\sum f_i x_i}{\sum f_i} - \bar{x}\frac{\sum f_i}{\sum f_i} \\ &= \bar{x} - \bar{x} = 0,\end{aligned}$$

$$\begin{aligned}\mu_2 &= \frac{\sum f_i(x_i - \bar{x})^2}{\sum f_i} \\ &= \frac{\sum f_i x_i^2}{\sum f_i} - 2\bar{x}\frac{\sum f_i x_i}{\sum f_i} + (\bar{x})^2 \\ &= \mu_2' - 2\bar{x}\bar{x} + (\bar{x})^2 \\ &= \mu_2' - (\bar{x})^2 \\ &= \mu_2' - (\mu_1')^2.\end{aligned}$$

Similarly we can determine μ_3.

5. **Geometric mean: الوسط الهندسى**

The geometric mean is a kind of mean or average, which indicates the central tendency or typical value of a set of numbers. It is similar to the arithmetic mean. The geometric mean only applies to positive numbers. It is also often used for a set of numbers whose values are meant to be multiplied together or are exponential in nature, such as data on the growth of the human population or interest rates of financial investment.

The geometric mean of a data set is smaller than or equal to the data set's arithmetic mean (the two means are equal if and only if all members of the data set are equal).

For the given data;

x	x_1	x_2	x_3	…………	x_n
F(x)التكرار	f_1	f_2	f_3	…………	f_n

the geometric mean is defined as:

$$G = (x_1^{f_1}\ x_2^{f_2}\ \dots\ x_n^{f_n})^{1/N}, \quad \text{where} \quad N = \sum f_i$$

By taking log for both sides of the last equation, then:

$$\log G = \frac{1}{N}\sum_{i=1}^{n} f_i \log x_i$$

Measure of variation

The variance: التباين

The formula for the variance of data sample is given as:

$$\sigma^2 = \mu_2 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i} = \mu_2' - (\mu_1')^2 ,$$

and its **standard deviation** (S.D.) الإنحراف المعيارى is given by:

$$\sigma = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}} .$$

Example (3) Find σ^2 and standard deviation for the data:

x	1	3	5
F(x)	2	4	6

Solution

$$\mu_1' = \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{2+12+30}{12} = \frac{11}{3} ,$$

$$\mu_2' = \frac{\sum f_i x_i^2}{\sum f_i} = \frac{2+36+150}{12} = \frac{47}{3} .$$

Now, the variance $\sigma^2 = \mu_2' - (\mu_1')^2 = \frac{47}{3} - (\frac{11}{3})^2 = \frac{20}{9}$,

And the standard deviation (S.D.) $\sigma = \sqrt{\frac{20}{9}} = \frac{2\sqrt{5}}{3}$.