Lecture 4 **25/01/2022**

Measures of central tendency

A measure of central tendency is a summary statistic that represents the center point or typical value of a dataset. These measures indicate where most values in a distribution fall and are also referred to as the central location of a distribution. In statistics, the three most common measures of central tendency are the mean, median, and mode. Each of these measures calculates the location of the central point using a different method.

Mean

It is an essential concept in mathematics and statistics. The mean is the average or the most common value in a collection of numbers. Calculating the mean is very simple. You just add up all of the values and divide by the number of observations in your dataset. There are different ways of measuring the central tendency of a set of values. There are multiple ways to calculate the mean, here are the most popular ones:

Arithmetic mean

It is the total of the sum of all values in a collection of numbers divided by the number of numbers in a collection, see the equation below.

$$\bar{X} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Hint: The calculation of the mean incorporates all values in the data. If you change any value, the mean changes.

Example

If the heights of five people are 48, 51, 52, 54, and 56 inches, what is the mean of the people?

Solution

The average heights of the people is (48 + 51 + 52 + 54 + 56 / 5 = 52.2)

So, the heights is 52.2 inches.

Ideally, the mean indicates the region where most values in a distribution fall. However, the mean does not always find the center of the data.

If the frequency is given in a frequency distribution form, then the arithmetic mean will be calculated according to the formula:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots . + x_n f_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

Example

For the following table, find the arithmetic mean

Pile length	3	7	8	12	19	21	25	31	38	45	52	60	68	74
frequency	2	3	1	5	4	2	1	4	6	1	5	6	3	4

Solution

Pile length (x_i)	3	7	8	12	19	21	25	31	38	45	52	60	68	74
Frequency (f_i)	2	3	1	5	4	2	1	4	6	1	5	6	3	4
$x_i f_i$	6	21	8	60	76	42	25	124	228	45	260	360	204	296

$$\sum x_i f_i = 1755$$

The mean is = $\frac{1755}{47}$ =37.34

Geometric mean

The geometric mean is a type of average, usually used for growth rates, like population growth or interest rates. While the arithmetic mean adds items, the geometric mean multiplies items. Also, you can only get the geometric mean for positive numbers. In other reference, the definition is “the nth root of the product of n numbers”. It is defined by G, we follow the equation below (for raw data):

G = $\sqrt[n]{x_1 * x_2 * ... * x_n}$ ….. root equation.

Or

Log G = $\frac{1}{n} \sum_{i=1}^{n} \log x_i$ ….. logarithmetic equation.

Example

What is the geometric mean of 4, 8, 3, 9 and 17?

Solution

First, multiply the numbers together and then take the 5th root (because there are 5 numbers), so:

G= $(4 * 8 * 3 * 9 * 17)^{(1/5)}$ = 6.81

However, If the data are given in a frequency distribution form the geometrical mean (G) becomes:

G = $\sqrt[n]{x_1^{f_1} * x_2^{f_2} * x_3^{f_3} * \ldots * x_n^{f_n}}$ ….. root equation.

Or

Log G = $\frac{1}{n} \sum_{i=1}^{n} f_i \log x_i$ ….. logarithmetic equation.

Example

Find the geometric mean for the data in table below:

Load (x)	Frequency (f)
5	5
10	8
15	3
20	4

Solution

Load (x)	Frequency (f)	Log x	F * log x
5	5	0.699	3.495
10	8	1.176	9.408
15	3	1.397	4.193
20	4	1.544	6.176
	20		23.273

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

Log G = 23.273/20 = 1.16

Example

load	frequency
50 - 69	3
70 – 89	7
90 – 109	4
110 – 129	4
130 - 149	9

Solution

load	frequency	Class limit (mid point) (x)	Log x	$f_i \log x_i$
50 - 69	3	59.5	1.7745	5.3235
70 – 89	7	79.5	1.9003	13.3525
90 – 109	4	99.5	1.9978	7.9912
110 – 129	4	119.5	2.0773	8.3094
130 - 149	9	139.5	2.1445	19.3011
	27			54.2280

Log G = ???

Harmonic mean

The harmonic mean is a type of numerical average. It is calculated by dividing the number of observations by the reciprocal of each number in the series. Thus, the harmonic mean is the reciprocal of the arithmetic mean of the reciprocals.

For example, the harmonic mean of 1, 4, and 4 is?

As simple as can,

Number of sample is n, here n = 3

$$harmonic\ mean = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}}$$

So,

Harmonic mean = $\frac{3}{\frac{1}{1}+\frac{1}{4}+\frac{1}{4}}$

= 3/(6/4) ….= 12/6=2

Median

The median is the middle number in a data set. To find the median, list your data points in ascending order and then find the middle number.

Example

The set of numbers 3,4,4,5,6,8,8,8,10 has median (M) (6)

Note: If you have an even set of numbers, average the middle two to find the median.

For example, the median for the set of the numbers 23, 24, 26, 26, 28, 29, 30, 31, 33, 34

is 28.5 (28 + 29 / 2).

Example

Find the median and the location for the data below>

30, 23, 26, 29, 34, 28, 24

Solution

23, 24, 26, 28, 29, 30, 34… ascending

The number of data is odd (7)

The median (M) = X ($\frac{n+1}{2}$) = X ($\frac{7+1}{2}$) =X_4

The M = 28

The location is X_4

If the data is given in frequency distribution form, the median class is the first class that have accumulative frequency equal or greater than the half total frequency

$$M = a + \left[\frac{\frac{n}{2} - n_1}{f_m}\right] * C$$

a= the lower true limit (mid point) of median class

n = summation of the frequencies

n_1 = the cumulative frequency of the previous limit of the median

f_m = the frequency of the median class

C = the length of the median class interval

<u>Example</u>

Find the median for the following data

Load	Frequency
50 -69	3
70 – 89	7
90 -109	4
110 – 129	4
130 - 149	9

<u>Solution</u>

Load	Frequency	Cumulative f.	True class limit
50 -69	3	3	49.5 – 69.5
70 – 89	7	10	69.5 – 89.5
90 -109	4	14	89.5 – 109.5
110 – 129	4	18	109.5 – 129.5
130 - 149	9	27	129.5 – 149.5

n = 27

m = 27/2… M=13.5

So, the median class is (90-109) because its accumulative frequency is (14) and this is the first class which its frequency greater or equal 13.5.

Applying the median equation,

$$M = a + \left[\frac{\frac{n}{2} - n_1}{f_m}\right] * C$$

a = 89.5

$n_1 = 10$

C= 20

f_m = 4

So, M =107

Mode

In statistics, the mode is the value that is repeatedly occurring in a given set. We can also say that the value or number in a data set, which has a high frequency or appears more frequently, is called mode or modal value. It is one of the three measures of central tendency, apart from mean and median. In the given set of data: 2, 4, 5, 5, 6, 7, the mode of the data set is 5 since it has appeared in the set twice.

For frequency distribution data

In a class of 30 students marks obtained by students in mathematics out of 50 is tabulated as below. Calculate the mode of data given.

marks	No. of student
10 – 20	5
20 – 30	12
30 – 40	8
40 - 50	5

Solution:

The maximum class frequency is 12 and the class interval corresponding to this frequency is 20 – 30. Thus, the modal class is 20 – 30.

Lower limit of the modal class (l) = 20

Size of the class interval (h) = 10

Frequency of the modal class (f1) = 12

Frequency of the class preceding the modal class (f0) = 5

Frequency of the class succeeding the modal class (f2) = 8

$$Mode = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h = 20 + \left(\frac{12 - 5}{2 \times 12 - 5 - 8}\right) \times 10 = 26.364$$