Bachelor of Architecture

Mathematics

Lecture 14

In this lecture we are going to see the Basic statistics. In that we will discuss the arithmetic mean, median and the mode and finally the summary.

The arithmetic mean:

Statistics deals with the methods for collection, classification and analysis of numerical data for drawing valid conclusion and making reasonable decision.

Mean:

If $x_1, x_2, x_3 \ldots\ldots x_n$ are a set of n value of variables, then the arithmetic mean (or simply mean) is given by

$$x = \frac{x_1 + x_2 + x_3 + \ldots\ldots x_n}{n}, i.e., \frac{\sum x_i}{n}$$

In a frequency distribution if $x_1, x_2, \ldots\ldots x_n$ be the mid-values of the class-intervals having frequencies $f_1, f_2, \ldots\ldots f_n$ respectively.

$$x = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots\ldots f_n x_n}{f_1 + f_2 + , \ldots\ldots + f_n} = \frac{\sum f_i x_i}{\sum f_i}$$

Calculation of mean:

Direct method of computing especially when applied to grouped data involves heavy calculations and in order to avoid these,

The following formula are generally used:

Short-cut method,

$$x = A + \frac{\sum f_i d_i}{\sum f_i}$$

Step-deviation method,

$$x = A + h\frac{\sum f_i u_i}{\sum f_i}$$

Where $d = x - A \& u = \frac{(x - A)}{h}$, A being an arbitrary origin and h the equal class interval. From these arrangements the following observation is taken.

Observation:

The algebraic sum of the deviations of all the variables from their mean is zero. For,

$$\sum f_i (x_i - x) = \sum f_i x_i - x\sum f_i$$

$$= \sum f_i x_i - \frac{\sum f_i x_i}{\sum f_i}.\sum f_i = 0$$

Corollary:

If x_1, x_2 be the means of two samples of size $n_1 \& n_2$ then the mean x is given by

$$x = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$$

Here two samples are considers and how to calculate the arithmetic for the two samples can be solved using this formula.

For $n_1 x_1$ is equal to the sum of all observation of the first sample and $n_2 x_2 =$sum of all observation of the second sample. Sum of the observation of the combined sample is equal to $n_1 x_1 + n_2 x_2$. Also number of observations in the combined sample is equal to $n_1 + n_2$.

Mean of the combined sample is,

$$= \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$$

Here two set of samples were considered instead of this two sets we can go for n set of samples. So far we learned the fundamental concepts in statistics.

Median:

If the values of a variable are arranged in the ascending order of magnitude the median is the middle item if the number is odd and is the mean of the two middle items if the number is even. If we are given a sample for ten member of students we are asked to find the median of the ten number of students, we need to consider the fifth one and the sixth one and take average of that two then that will be the median of the ten number of students.

Thus the median is equal to the mid-value that is the value which divides the total frequency into two equal parts. For the grouped data the median will be,

$$= L + \frac{\left(\frac{1}{2}N - c\right)}{f} h$$

Where,

L is the lower limit of the median class.

N is the total frequency.

F is the frequency of the median class.

H is the width of the median class.

C is the cumulative frequency upto the class preceding the median class.

Quartiles:

Quartiles are those values which divide the frequency into four equal parts, when the values are arranged in the ascending order of magnitude. The lower Quartile Q_1 is mid-way between the lower extreme and the median. The upper Quartile Q_3 is mid-way between the median and the upper extreme.

For grouped data these are calculated by the formula,

$$Q_1 = L + \frac{\left(\frac{3}{4}N - c\right)}{f}h$$

$$Q_3 = L + \frac{\left(\frac{1}{4}N - c\right)}{f}h$$

Where,

L is lower limit of the class in which Q1 and Q2 lies.

f is frequency of this class.

h is width of the class.

C is the cumulative frequency upto the class preceding the class in which Q1 or Q3 lies.

Mode:

The mode is defined as that the values of the variable which occurs most frequently that is the values of the maximum frequency. For a grouped distribution it is given by the formula,

$$\text{mod}\, e = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} h$$

Where,

L is the lower limit of the class containing the medium.

Δ_1 is the excess of modal frequency over frequency of preceding class.

Δ_2 is the excess of modal frequency over following class.

h is the size of modal class.

Next let us go to the mode in a frequency curve. For a frequency curve the abscissa of the highest ordinate determine the values of the mode. There may be one or more mode in a frequency curve. Curves having a single mode are termed as uni-modal, those having two modes are di-modal and those having more than two modes are multi-modal.

Obs:

In a symmetrical distribution the mean median and mode coincide. For other distribution they are different and are known to be connected by the empirical relationship.

Mean-mode=3(mean - median)

Now we can apply these concepts in a problem.

Example 1:

Calculate median and the lower and upper quartiles from the following distribution of number of marks obtained by 49 students in a class. Find also the semi-inter quartile range and the mode.

Solution:

Here the mark obtained by the number of 45 students is given in a data. Now first find the median using the formula,

$$= L + \frac{\left(\frac{1}{2}N - c\right)}{f}h$$

Now substituting the values in the above formula we get,

$$= 15 + \frac{\left(\frac{49}{2} - 11\right)}{15} \times 5$$

$$= 15 + \frac{13.5}{3} = 19.5 marks$$

Then lower Quartile Q_1, will be

$$Q_1 = L + \frac{\left(\frac{3}{4}N - c\right)}{f}h$$

$$Q_1 = 15 + \frac{\left(\frac{49}{4} - 11\right)}{15} \times 5$$

$$= 15 + \frac{12.5}{3} = 15.4 marks$$

Like this the upper Quartile will be,

$$Q_3 = L + \frac{\left(\frac{1}{4}N - c\right)}{f}h$$

$$Q_3 = 15 + \frac{\left(\frac{49}{4} - 11\right)}{15} \times 5$$

$$= 25.75 marks$$

Then the semi-inter Quartile will be,

$$= 5.175 marks$$

This semi-inter Quartile is obtained by taking the average of lower and the upper Quartiles.

Mode:

Then the mode value falls in the class 15-20. Employing the formula for the grouped distribution we get,

$$= 15 + \frac{15-6}{(15-6)(15-10)} 5 = 18.2 marks$$

Geometric mean:

If $x_1, x_2 ..... x_n$ are a set of n observation then geometric mean is given by

$$GM = (x_1, x_2 ..... x_n)^{1/n}$$

$$\log GM = \frac{1}{n} (\log x_1, \log x_2 ..... \log x_n) ----(1)$$

Summary:

In this lecture we have learned the Arithmetic mean of the class gives the average value of the group. It is calculated by short-cut method and step-deviation method. And the median is the middle term of array of the numbers. The mode is the value of the variable which occurs most frequently. After listening to this lecture you can answer the following questions.

Questions:

1. Define arithmetic mean of a group.
2. What is known as median?
3. Define mode of the group of sample.