MATHEMATICS

Chapter 14: Statistics

Statistics

1. Three measures of central tendency are:
   i. Mean
   ii. Median
   iii. Mode

2. The **arithmetic mean**, also called the average, is the quantity obtained by adding all the observations and then dividing by the total number of observations.

3. Arithmetic mean may be computed by anyone of the following methods:
   i. Direct method
   ii. Short-cut method/ Assumed mean method
   iii. Step-deviation method

4. **Direct method** of finding mean:

   If a variant X takes values x_1, x_2, x_3.... x_n with corresponding frequencies f_1, f_2, f_3 ,... f_n respectively, then arithmetic mean of these values is given by:

   $$\overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{N} \text{ where } N = \sum_{i=1}^{n} f_1 + f_2 + f_2 \ldots \ldots + f_n$$

5. **Class mark** $= \frac{1}{2}$ (Upper class limit + Lower class limit)

6. **Short-cut method/ assumed mean method** of finding mean:

   Let x_1, x_2....,x_n be values of a variable X with corresponding frequencies f_1, f_2, f_3 ,fn respectively. Let A be the assumed mean. Then:

   $$\overline{X} = A + \frac{1}{N}\left\{\sum_{i=1}^{n} f_i d_i\right\}$$

   Note that in case of continuous frequency distribution, the values of x1, x2, x3 ... x_n, are taken as the mid-points or class-marks of the various classes.

7. **Step-deviation method** of finding mean:

   Let x1, x2....,xn be values of a variable X with corresponding frequencies f_1, f_2, f_3,.....fn respectively. Let A be the assumed mean. Then:

   $$\overline{X} = A + h\left\{\frac{1}{N}\sum_{i=1}^{n} f_i u_i\right\}$$

   Here, h is generally taken as common factor of the deviations, in case of ungrouped frequency distribution. And, in case of grouped frequency distribution, h is the class width, $u_i = \frac{x_i - A}{h} = \frac{d_i}{h}$

   Note that in case of continuous frequency distribution, the values of x1, x2, x3 ..., xn are

taken as the mid-points or class-marks of the various classes.

8. The step deviation method will be convenient to apply if all the deviations (d's) have a common factor.

9. If class mark obtained, are in decimal form, then step deviation method is preferred to calculate mean.

10. **Median** is a measure of central tendency which gives the value of the middle observation in the data, arranged in order. It is that value such that the number of observations above it is equal to the number of observations below it.

11. For finding the median of a raw data, we arrange the given data in increasing or decreasing order. If n is odd, then median is the value of $\left(\frac{n+1}{2}\right)^{th}$ observation.

If n is even, then median is the arithmetic mean of the values of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observations.

12. The **cumulative frequency** of a class is the frequency obtained by adding the frequencies of all the classes preceding the given class to the frequency of the class.

13. In case of an **ungrouped frequency distribution**, we calculate the **median** by following the steps given below:

Step 1: Find the cumulative frequencies (c.f.) and obtain $N = \sum f_1$.

Step 2: Find $\frac{n}{2}$

Step 3: Look for the cumulative frequency (c. f.) just greater than $\frac{n}{2}$ and determine the corresponding value of the variable. The value so obtained is the median.

14. In case of a **continuous frequency distribution**, we calculate the **median** by following the steps:

Step 1: Find the cumulative frequencies (c.f.) and obtain $N = \sum f_1$.

Step 2: Find $\frac{N}{2}$

Step 3: Look for the cumulative frequency (c. f.) just greater than $\frac{N}{2}$ and determine the corresponding class. This class is known as the median class. (Note that the value of the median will lie in this class)

Step 4: Use the following formula to find median:

$$\text{Mediun} = l + \left[\frac{\frac{N}{2} - cf}{f}\right] \times h$$

Here, l = lower limit of the median class

f = frequency of the median class

h = width (size) of the median class

cf = cumulative frequency of the class preceding the median class

$N = \sum f_1$.

15. **Mode** is the value of the most frequently occurring observation in the data.

16. In an ungrouped frequency distribution, mode is the value of the variable having maximum frequency.

17. In a **grouped frequency distribution**, the modal class is the one with highest frequency and the

    mode can be calculated by the following formula

    $$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

    l = lower limit of the modal class

    h = size of the class interval

    f_1 = frequency of the modal class

    f_0 = frequency of the class preceding the modal class

    f_2 = frequency of the class succeeding the modal class

18. The most frequently used measure of central tendency is the mean, because the mean is calculated by taking into account all the observations of a given data. And it lies between the smallest and the largest value of the data.

19. The biggest drawback in considering mean is that it is affected by the extreme values. One large or small number can distort the average. In that case, median is a better measure of central tendency. While, when the most repeated value or the most wanted one is required, then mode is used.

20. When all three measures of central tendency are equal, the distribution is called **symmetrical distribution**.

21. When the values of mean, median and mode are not equal, then the distribution is known as **asymmetrical or skewed**. In this case, the distribution can be positively skewed or negatively skewed.

    Negatively skewed distributions have a few extremely low scores, while positively skewed distributions have a few extremely high scores.

    i. When the data is negatively skewed, then Mean < Median < Mode

    ii. When the positively skewed, then Mean > Median > Mode