**Assignment(1)**

**Class:MCA-I(Sem-I)**

**Subject: Statistics and Probability**

**Topics**:Mean,How to find mean,Correlation and types of correlation

**Submitted To:**

Silky Bansal

Prof.of Comp.Sci.Department

**Ques1: What is Mean? How to Find Mean?**

**Ans.Mean** is the average of the given numbers and is calculated by dividing the sum of given numbers by the total number of numbers.

|  |
| --- |
| Mean = (Sum of all the observations/Total number of observations) |

**Example:**

What is the mean of 2, 4, 6, 8 and 10?

**Solution:**

First, add all the numbers.

2 + 4 + 6 + 8 + 10 = 30

Now divide by 5 (total number of observations).

Mean = 30/5 = 6

In the case of a discrete probability distribution of a random variable X, the mean is equal to the sum over every possible value weighted by the probability of that value; that is, it is computed by taking the product of each possible value x of X and its probability P(x) and then adding all these products together.

Mean Symbol (X Bar)

The symbol of mean is usually given by the symbol ‘x̄’. The bar above the letter x, represents the mean of x number of values.

X̄ = (Sum of values ÷ Number of values)

X̄ = (x1 + x2 + x3 +….+xn)/n

**Mean Formula**

The basic formula to calculate the mean is calculated based on the given data set. Each term in the data set is considered while evaluating the mean. The general formula for mean is given by the ratio of the sum of all the terms and the total number of terms. Hence, we can say;

**Mean = Sum of the Given Data/Total number of Data**

To calculate the arithmetic mean of a set of data we must first add up (sum) all of the data values (x) and then divide the result by the number of values (n). Since ∑ is the symbol used to indicate that values are to be summed (see Sigma Notation) we obtain the following formula for the mean (x̄):

**x̄=∑ x/n**

**How to Find Mean?**

As we know, data can be grouped data or ungrouped data so to find the mean of given data we need to check whether the given data is ungrouped. The formulas to find the mean for ungrouped data and grouped data are different. In this section, you will learn the method of finding the mean for both of these instances.

### Mean for Ungrouped Data

The example given below will help you in understanding **how to find the mean** of ungrouped data.

**Example:**

In a class there are 20 students and they have secured a percentage of 88, 82, 88, 85, 84, 80, 81, 82, 83, 85, 84, 74, 75, 76, 89, 90, 89, 80, 82, and 83.

Find the mean percentage obtained by the class.

**Solution:**

Mean = Total of percentage obtained by 20 students in class/Total number of students

= [88 + 82 + 88 + 85 + 84 + 80 + 81 + 82 + 83 + 85 + 84 + 74 + 75 + 76 + 89 + 90 + 89 + 80 + 82 + 83]/20

= 1660/20

= 83

Hence, the mean percentage of each student in the class is 83%.

## Mean of Negative Numbers

We have seen examples of finding the mean of positive numbers till now. But what if the numbers in the observation list include negative numbers. Let us understand with an instance,

**Example:**

Find the mean of 9, 6, -3, 2, -7, 1.

**Solution:**

Add all the numbers first:

Total: 9+6+(-3)+2+(-7)+1 = 9+6-3+2-7+1 = 8

Now divide the total from 6, to get the mean.

Mean = 8/6 = 1.33

**Ques2:What is Correlation and Types of Correlation?**

## Ans. Correlation in Statistics

This section shows how to calculate and interpret correlation coefficients for ordinal and interval level scales. Methods of correlation summarize the relationship between two variables in a single number called the correlation coefficient. The correlation coefficient is usually represented using the symbol r, and it ranges from -1 to +1.

A correlation coefficient quite close to 0, but either positive or negative, implies little or no relationship between the two variables. A correlation coefficient close to plus 1 means a positive relationship between the two variables, with increases in one of the variables being associated with increases in the other variable.

A correlation coefficient close to -1 indicates a negative relationship between two variables, with an increase in one of the variables being associated with a decrease in the other variable. A correlation coefficient can be produced for ordinal, interval or ratio level variables, but has little meaning for variables which are measured on a scale which is no more than nominal.

For ordinal scales, the correlation coefficient can be calculated by using Spearman’s rho. For interval or ratio level scales, the most commonly used correlation coefficient is Pearson’s r, ordinarily referred to as simply the correlation coefficient.

**What Does Correlation Measure?**

In statistics, Correlation studies and measures the direction and extent of relationship among variables, so the correlation measures co-variation, not causation. Therefore, we should never interpret correlation as implying cause and effect relation. For example, there exists a correlation between two variables X and Y, which means the value of one variable is found to change in one direction, the value of the other variable is found to change either in the same direction (i.e. positive change) or in the opposite direction (i.e. negative change). Furthermore, if the correlation exists, it is linear, i.e. we can represent the relative movement of the two variables by drawing a straight line on graph paper.

**Correlation Coefficient**

The correlation coefficient, r, is a summary measure that describes the extent of the statistical relationship between two interval or ratio level variables. The correlation coefficient is scaled so that it is always between -1 and +1. When r is close to 0 this means that there is little relationship between the variables and the farther away from 0 r is, in either the positive or negative direction, the greater the relationship between the two variables.

The two variables are often given the symbols X and Y. In order to illustrate how the two variables are related, the values of X and Y are pictured by drawing the scatter diagram, graphing combinations of the two variables. The scatter diagram is given first, and then the method of determining Pearson’s r is presented. From the following examples, relatively small sample sizes are given. Later, data from larger samples are given.

**Scatter Diagram**

A scatter diagram is a diagram that shows the values of two variables X and Y, along with the way in which these two variables relate to each other. The values of variable X are given along the horizontal axis, with the values of the variable Y given on the vertical axis.

Later, when the regression model is used, one of the variables is defined as an independent variable, and the other is defined as a dependent variable. In regression, the independent variable X is considered to have some effect or influence on the dependent variable Y. Correlation methods are symmetric with respect to the two variables, with no indication of causation or direction of influence being part of the statistical consideration. A scatter diagram is given in the following example. The same example is later used to determine the correlation coefficient.

**Types of Correlation**

The scatter plot explains the correlation between the two attributes or variables. It represents how closely the two variables are connected. There can be three such situations to see the relation between the two variables –

* Positive Correlation – when the values of the two variables move in the same direction so that an increase/decrease in the value of one variable is followed by an increase/decrease in the value of the other variable.
* Negative Correlation – when the values of the two variables move in the opposite direction so that an increase/decrease in the value of one variable is followed by decrease/increase in the value of the other variable.
* No Correlation – when there is no linear dependence or no relation between the two variables.

