**Univariate data analysis methods:**

In univariate data analysis, researcher analyzed only single variable without looking its relation with other variables during the data analysis

**For qualitative/categorical/non-metric variables:**

The researcher analyzed qualitative variable by finding the frequency and percent distribution for each category of qualitative variables.

**Presentation of finding of qualitative variables:**

Researcher can use any option to present the finding of qualitative variables such as table, pie chart, bar-char, line diagram, and side by side bar chart as per the nature of findings.

**For quantitative variables:**

The researcher can analyzed the quantitative variables by finding descriptive statistics such as mean, mode, median, and standards deviation. The most widely used tools are mean and standard deviation to analyze the quantitative/numerical/ metric variable.

**Presentation of finding of quantitative variables:**

Researcher can use any option to present the finding of qualitative variables such as table, pie chart, bar-char, line diagram, side by side bar chart, histogram, box-and whisker plot as per the nature of findings.

**Measures of central tendency:**

This method is used to find the central values of the variable which is used to summarize the data of the variables.

**Mean:** it is a mathematical average and it is calculated by dividing the sum of data by number of data. The mean value indicate that the data are close to mean value and it is used as a summary measure of the data of the variables. The following formula is used to calculate the mean from the raw data.

**Median:** it is a positional average and it divides the data into two equal parts i.e. 50% data are smaller than median value and rest 50% are larger than median value. It the value of middle item of ordered data. The following formula is used to calculate the median value from the raw data

Median = value of [(n+1)/2]th term of ordered data

**Mode**: it is the frequency based average. The value which has the largest frequency is called mode value i.e. the most repeated data is called mode. The variable may have unique mode or more than one mode. If no data of the observed variable is repeated then mode does not exist. The variable with one mode is called uni-modal, the variable with two modes is called bimodal, and with three modes it is called tri-modal and so on.

**Measures of dispersion:**

The variation in response to variable is the inherent characteristics. It is very important for the researcher to understand the variation of data of the observed variables. The researcher will be interested to find the other factors which may create variation in some particular variables. There are various statistical tools and methods in statistics to measure the relationship between two or more variables.

Some statistical tools which are used to measure variation of responses within the variables are range, quartile deviation, mean deviation, standard deviation, variance and coefficient of variation. The variable with less variation is considered more consistent, reliable and homogeneous than the variable with high variation. Standard deviation and range are more frequently used to measure within variation of the quantitative variables. The formula of range and standard deviation are expressed below

**Range:** The difference of largest and smallest value of the variable is called range.

Range = L – S

**Standard deviation (S):** it measures the average variation of the data from mean value. It is more used tool to measure the within variation of the quantitative variable.

**Variance (S2):** the square of standard deviation is called variance

**Coefficient of variation (C.V.):** it is a relative measure of dispersion/variation and expressed in percentage form. It is used to compare the variation of two or more quantitative variables to understand consistence of the variable as compared to others. The variable with less value of CV is considered more consistent, reliable and homogenous than the variable with higher value of CV. The formula to calculate the CV is expressed below

**Dataset 1**

The dataset 1 contains the data on five variables collected from 6 respondents. Do the analysis of the dataset 1 using appropriate tools of univariate data analysis such as mean, standard deviation, range, and frequency and percent distribution as per the nature of data under the variables.

Sex: 1= male and 2 = Female

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SN | Sex | Income | education | Age | expenditure |
| 1 | 2 | 48 | 14 | 24 | 23 |
| 2 | 1 | 50 | 10 | 40 | 25 |
| 3 | 1 | 45 | 4 | 21 | 30 |
| 4 | 1 | 29 | 17 | 23 | 20 |
| 5 | 2 | 46 | 9 | 29 | 16 |
| 6 | 1 | 30 | 11 | 32 | 14 |

**Questions:**

Q1. Determine the qualitative and quantitative variables of the dataset 1.

Q2. Which techniques are used to analyze the qualitative variable under univariate data analysis method?

Q3. Which techniques are used to analyze the quantitative variable under univariate data analysis method?

Q4. Analyze and interpret the dataset 1 by using univariate data analysis techniques.

Q5. Present the findings by creating the summary tables and interpret it.

Q6. Present the finding using suitable diagram