Assignment-1

Name: Renuka Rathod

1]What is data analysis ?

Ans.: Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.

2]What is data visualization? Explain all charts of visualization

Ans.: Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

1. Line charts

A line chart connects distinct data points through straight lines. Its best use case is to illuminate trends, patterns, and variable changes.

2. Bar charts

A bar chart visually represents data using rectangular bars or columns. Here, the length of each bar corresponds proportionally to its value. You can present these bars horizontally or vertically. A horizontal bar chart is best to use when the text on the x-axis of a vertical bar chart is lengthy, meaning it would have to be presented diagonally—or even worse, cut off—to fit within the visualization.

3. Scatter plots

Scatter plots are types of visualization that show a collection of data points ‘scattered’ around the graph. The data points can be evenly or unevenly distributed.

4. Pie charts

A common but limited type of visualization is the pie chart. It is a circular, statistical graphic that divides data into slices. Each slice represents a percentage or proportion of the whole.

5. Column charts

Column charts are the simplest, most versatile type of visualization used in data analytics. The horizontal chart displays your data in bars proportional to the values they represent.

6. Treemap charts

Treemaps are hierarchical charts that allow you to visualize data as nested rectangles. These rectangles or branches convey the structure and distribution of data, making treemaps useful for visualizing categorical and hierarchical relationships.

7. Heatmap charts

Heatmap charts are a type of map data visualization that uses a system of color coding to represent value. Each cell in the matrix is assigned a color based on the value it holds.

8. Pareto charts

A Pareto chart combines a bar chart and a line graph. The rectangular bars correspond to individual values in descending order, while the line graph displays the cumulative percentage total. This type of chart follows the famous Pareto principle that emphasizes that 20 percent of causes result in 80 percent of problems.

9. Geo charts

Geo charts are a type of visualization that represent data on a map. They show spatial information, such as the distribution of values across different regions, countries, or states.

10. Waterfall charts

A waterfall chart is like a visual story that helps you see how different things add up to a final result. It explains how an initial value is affected by a series of intermediate positive and negative values. The waterfall chart receives its name due to its shape as it shows cascading effects.

Q.3]Write formula of mean

Ans. Mean = (Sum of all the observations/Total number of observations)

Q.4]what is advantage and limitation of mean

Ans. Advantages:

1. It is easy to understand and calculate.
2. It is based on all observations.
3. It is useful for further mathematical treatment.

Limitation of mean:

1. It is much affected by extreme values
2. Not suitable for qualitative data analysis.
3. It can not be located graphically.

Q.5]what is advantage and limitation of median

Ans. Advantages of median:

1. It is easy to calculate and understand.
2. Not affected by extreme values .
3. Can be determined graphically using CUMMULATIVE CURVES.

Limitations of median:

1. It is not bases on all observations
2. Not suitable for further mathematical treatment.

Q.6]what is advantage and limitation of mode

Ans. Advantage of mode:

1. It is easy to calculate and understand.
2. Not affected by extreme values .
3. Can be determined graphically using histogram

Limitation of mode:

1. It is not bases on all observations
2. Not suitable for further mathematical treatment

Q.7] what is continuous variable, what is descrete variable explain with 5 examples

Ans. Continuous variables are generally measured on scales such weight, height, length, time, and temperature.With the help of continuous variables, one can measure mean, median, variance, or standard deviation. Continuous variables are the ones which in between any two numeric values have an infinite number of values.A discrete variable only allows a particular set of values, and in-between values are not included

Ex. Of descrete variables:

the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery, the number of defective light bulbs in a box of ten.

Q.8]what is difference between sample and population explain with 5 examples

Ans. Population:

* Population refers to the total membership or “universe” of a defined class of people, organizations, objects , entities or events.
* A statistical population is a set of entities, which engages the researcher’s interest as a set. E.g. the set of all petrol retailers in Pune City.

Sample:

* Statisticians use the word sample to describe a portion chosen from the statistical population so as to seek data or information, opinions or facts from the selected units.
* Thus, a finite subset of a statistical population , chosen in a specific fashion, is called a sample.

Q.9]Finite population and infinite population.

Ans. Finite Population

The finite population is also known as a countable population in which the population can be counted. In other words, it is defined as the population of all the individuals or objects that are finite. For statistical analysis, the finite population is more advantageous than the infinite population. Examples of finite populations are employees of a company, potential consumer in a market.

Infinite Population

The infinite population is also known as an uncountable population in which the counting of units in the population is not possible. Example of an infinite population is the number of germs in the patient’s body is uncountable.

Q.10]What is descriptive,diagnostic predictive, prescriptive data analytics?

Ans.descriptive analysis:what is already happened or what is happening?

Diagnostic analysis: why is it happening? Ability to drill down to root cause.

Predictive:what is likely to happen?

Prescriptive:what do I need to do?

Q.11]Find mean, median,mode of following-

1. 89,6,90,34,65,234,8,3000,7,567,6,2,45,20

|  |  |
| --- | --- |
| mean | 297.8571 |
| median | 39.5 |
| mode | 6 |

2. 5677,60,22, 34,6500,23, 869,67,900,1,2,6,1,70,1

|  |  |
| --- | --- |
| mean | 948.8667 |
| median | 34 |
| mode | 1 |

Q.12]What are data analysis steps?

Ans.Decide on the objective.

Set measurement priorities.

Data collection.

Data cleansing.

Analysis of data.

Interpretation the results.

Q13.]Explain following terms with example?

1)Nominal Scale

• When numbers assigned to objects serve as labels for identification or categorization, then such numbers are in nominal scale. Such numbers have no quantitative meaning.

• For e.g.

• Male = 1

• Female = 2

• The only permissible operation on such numbers is counting. %, mode, chi-square, and binomial tests can be performed on such data.

2)Ordinal Scale

• When assigned numbers to indicate the relation between entities in terms of greater than, equal, or less than but do not state how much greater than or less than, then the scale is called ordinal scale.

• For e.g. Ranks

• Rank following brands of TV on sound quality:

Rank

• Akai ------------

• Videocon -----------

• Samsung -----------

• This means we cannot conclude about the difference between values of two objects. We can calculate median, quartiles, deciles, percentiles & rank order correlation.

3)Interval Scale

• When assigned numbers are such that the difference in numbers is valid but not ratios, then the scale is called interval scale.

• In this scale there is no true zero indicating absence of characteristic. For e.g

• Temperature

• – To what extent do you like the sound quality of SONY TV?

Liked very much 5

Somewhat liked 4

Neither liked nor disliked 3

Somewhat not liked 2

Not liked at all 1

**Arithmetic mean, standard deviation, product-moment correlations can be applied to interval scale data.**

4)Ratio Scale

* When a scale contains absolute zero, it is called ratio scale
* All mathematical operations (+,-,\*,/) are valid on this data
* All statistical techniques can be applied to ratio scale data.

Q.14]What is skewness and kurtosis? How to measure skewness and kurtosis?

Ans. Skewness and kurtosis are statistics that describe the shape and symmetry of the distribution. These statistics are displayed with their standard errors.

Shape of data is measured by Skewness ,Kurtosis

The histogram is an effective graphical technique for showing both the skewness and kurtosis of data set.

Q.15] Find mean deviation and standard deviation , skewness, kurtosis of data

**34,56,12,34,32,89,70,65,45,678,90,890,760,55**

**Ans.** **x<-c(34,56,12,34,32,89,70,65,45,678,90,890,760,55)**

**library(moments)**

**skewness(x)**

**o/p:** skewness(x)

[1] 1.442033