1. What are the most important topics in statistics?

Ans.

• Probability definitions and properties

• Common discrete and continuous distributions

• Univariate and bivariate transformations

• Convergence of random variables: in probability

• Central Limit Theorem,

• Estimation: bias, variance

• Hypothesis testing: significance level and power

• Confidence Intervals: definitions, duality with hypothesis tests

2. What is exploratory data analysis?

Ans.

EDA is the initial operation to analyse and summarize the data using visual techniques. Using EDA we can discover patterns, spot anomalies, test hypothesis and check assumption.

3. What are quantitative data and qualitative data?

Ans.

- Quantitative data is the data that can be counted or measured and expressed in numbers. Mainly we are having two types of quantitative data :

i. Discrete data

ii. Continuous data

- Qualitative data describes qualities or characteristics. Which is defined in categories.

4. What is the meaning of KPI in statistics?

Ans.

Key Performance Indicators (KPIs) is a metric that is used to a measure the performance over time for a specific objective.

In simple language KPI is a tool to check whether or not we are hitting the target.

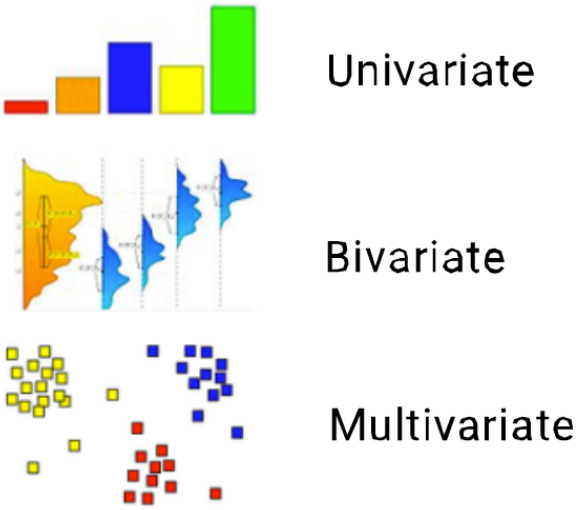
5. What Is the Difference Between Univariate, Bivariate, and Multivariate Analysis?

Ans.

Univariate Analysis is analysis of one variable at a time.

Bivariate Analysis compares two variables.

Multivariate Analysis compares more than two variables.



6. How Would You Approach a Dataset That's Missing More Than 30 Percent of Its Values?

Ans.

i. If the data set is large :

```
We can simply remove the rows with missing data.

It is the quickest way, we can use rest of the data to predict the value.
```

ii. If the dataset is small :

```
We can impute the missing values with various statistical methods, by using df.
fillna(mean).
```

7. Give an example where the median is a better measure than the mean

Ans.

It is better to use mean when the dataset is symmetric or normally distributed. If the dataset contains very small or very large outliers compared to the dataset in left or right side in this case we should use median.

8. What is the difference between Descriptive and Inferential Statistics?

Ans.

Descriptive statistics gives summary or describes about the dataset.

Inferential statistics is making conclusion about population data using sample data. We validate this conclusion using various tests.

9. What are descriptive statistics?

Ans.

Descriptive statistics describes, summarizes and show the basic features of the dataset found in the given study.

Distribution – refers to the frequencies of responses.

Central Tendency – gives a measure or the average of each response.

Variability – shows the dispersion of a data set.

10. Can you state the method of dispersion of the data in statistics?

Ans.

Dispersion represents the spread of the dataset. It is of three types:

```
i. Range:
Range is measure of difference between the largest and smallest value with a da
ta variability.

ii. Variance:
Variance shows the spread of the data. It basically shows how far the data in t
he dataset away from the other data.

iii. Standard Deviation:
Standard deviation is the square root of variance. It displays the dispersion o
f the dataset in relation to mean. Low standard deviation means dataset is clus
tered to the mean. High standard deviation means data are more spread out.
```

Thanks

github: https://github.com/saisubhasish/statistics (https://github.com/saisubhasish/statistics)

In [ ]:

```
1
```