1. What does categorical data mean?

Categorical data means collection of information divide into groups.

Ex. Male , female, colou(red, white, yellow)

1. Formula for Probability?

Chances of an event occurring.

1. Different data measurement levels for a variable?
   1. Below are the features for a class of students.
   2. Guess the data types and measurement levels?
   3. Roll num, Name, sex, age, grades, Father’s Name, Salary

Roll Num - integer-

Name-string data

Sex-categorical data-

Age-count data

Grades-continuous data

Fathers Name- string data

Salary – continuous data

1. What is assignment operator in Python programming language?

Assignment operator is =

1. What are the various types of analytics?

Descriptive Analytics.

Diagnostic Analytics.

Predictive Analytics.

Prescriptive Analytics.

1. Inferential statistics?
2. What is the command to find out the version of R software installed?
3. What is probability distribution?

Random variable on x axis and corresponding probability on y axis is called as probability distribution.

1. What is a Sampling Frame?

A sampling frame is the list from which units are drawn for the sample.

1. Sampling techniques?

Random sampling

1. 1st moment business decision captures what?
2. What is an outlier

A outlier is an observation that lies and abnormal distance from the other sample from population.

1. Can mode be applied to continuous data?

No

1. Measures of dispersion is capturing what aspect from the dataset

The Measures of dispersion is indicates the degree of spread or distribution of data. This is only used for ordinal and interval data.

1. What is degrees of freedom
2. How to calculate standard deviation
3. Notation for Mean, variance and standard deviation

Mean=miu, variance=n, standard deviation=sigma

1. Variance for a probability distribution?

There are two types of probability distribution

1. Continuous probability distribution- if the random variables are in continuous in nature are called as continuous probability distribution.
2. Discrete probability distribution- if the random variables are discrete or categorical in nature are called as discrete probability distribution.
3. How to identify an outlier from a given data set

The most effective way is to find outliers is by using interquartile range(IQR). IQR contains the middle bulk of data set, so Outliers can be easily find once you get the IQR.

1. What are shape statistics

Shape statistics measures the shape of data , means it describe the distribution or pattern of given data set that is skewness and kurtosis.

1. If mean > median what is the shape of the distribution

Positive skewed.

1. What feature does 3rd moment business decision talk about
2. What feature/aspect does 4th moment business decision talk about
3. Summary() – output of function in Python

It gives all parameters of model such as probability, t value, p value, r-squared, adj-rsquared, A/C , B/C skew, method, date , time, f-stastics,

1. What is IQR

IQR is nothing but Interquartile range. It describe the middle 50% of data when data is order from lowest to highest. To find the IQR find the median of lower and upper half of data these value are Q1 and Q3, The IQR is difference between Q3 and Q1.

1. What % of data lies between Q3 and Q4

Bottom 75% of data lies in Q3,

Highest 25% of data lies in Q4.

1. How to identify outliers from boxplot using R
2. Formula for skewness
3. Formula for kurtosis

Z-score=X-Xbar/sigma

1. What is probability distribution?

Random variables on x axis and corresponding probability on y axis is called as probability distribution.

1. Probability distributions are classified into how many major types.

Probability distribution having two types- continuous and discrete probability distribution.

1. What is Normal distribution

If continuous distribution having perfect bell shaped curve with equal tails shaped on both side then it is called as normal distribution.

1. Properties of Normal distribution

Properties:

A continuous random variable x has a normal distribution if it have a smooth bell shaped curved at both tails.

Every normal distribution has its own mean(miu) and its own standard deviation(sigma).

The shape of normal distribution is symmetric around the mean.

The area under the curve is equal to 1.

The normal distribution are denser in center and less on both tails.

68% if the area of normal distribution is within 1 standard deviation of the mean.

95% of the data are within 2 standard deviation of the mean.

99.7% of the data are within 3 standard deviation of the mean.

1. Normal distribution is characterized by

1. Notation used to represent Normal distribution

Miu=0, sigma=1 it is also called as z distribution that is z=n(miu,sigma)

1. Properties of standard normal distribution

The mean=median=mode

The area under the curve is equal to 1.

The curve is symmetric around the center.

Exactly half value are left side of center and half value are right of center.

1. How is standard normal distribution represented

From scipy import stats

Stats.norm.cdf(x, loc=mean, scale=std)

1. What is the formula of Z calculation function in r

Z=(x-miu)/sigma

1. . What are the other applications of scale function
2. What does the line in QQ plot represent

The quantile-quantile(QQ) plot is a graphical technique for determining if two data set come from population with a common distribution.

1. What does transformation technique mean?

Transformation technique means applying some mathematical formulas or equation on given data to increase the accuracy for model, it is also called as featcher engineering.

1. What are the steps involved in EDA?
2. Import data set
3. Check for any missing value.
4. Replacing the missing value.
5. Visualization.
6. Positive correlation.
7. Negative correlation.
8. What is confidence interval formula?
9. Mean of sample means

Sample mean means the average value of sample. The sample are nothing but the randomly collected data.

1. What is the minimum sample size considered to be good enough

Minimum sample size is 100.

1. 2nd moment of sample means
2. Formula for calculating sample size
3. Can we trust a single value prediction?

No.

1. What is level of significance

The level of significance is the measurement of statistical significance. It defines whether the null hypothesis is accepted of rejected.

1. What is level of confidence
2. Z value at 5% significance level
3. What does Hypothesis testing mean?

Hypothesis testing is a statistical method that is used in making statistical decisions using experimental data. Hypothesis testing is basically an assumption that we make about the population parameter.

1. Condition of Null hypothesis

If the P < 0.05 then reject Null Hypothesis.

If the P > 0.05 then accept Null Hypothesis.

1. Condition for Alternate hypothesis

When p < 0.05,When we reject null hypothesis it means we accept Alternate hypothesis.

1. Come up with 2 \* 2 hypothesis matrix
2. Hypothesis test is done on sample, but statements are written for population:

True/False

True

1. What are the steps involved in Hypothesis testing?

1.State Null Hypothesis(p<0.05)

2.State Alternative Hypothesis.(P>0.05)

3.Collect data

4.Calculate p.

5.Compare p with 0.05

6.Draw a conclusion based on step 5.

1. How to evaluate relation between two quantitative variables