Name: - Nidhi Charde

Batch No:-Internship 19

STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0.

a) True

b) False

Answer: – a) True

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

a) Central Limit Theorem

b) Central Mean Theorem

c) Centroid Limit Theorem

d) All of the mentioned

Answer:-a) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?

a) Modeling event/time data

b) Modeling bounded count data

c) Modeling contingency tables

d) All of the mentioned

Answer:-b)Modeling bounded count data

4. Point out the correct statement.

a) The exponent of a normally distributed random variables follows what is called the log- normal distribution

b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent

c) The square of a standard normal random variable follows what is called chi-squared distribution

d) All of the mentioned

Answer:–d) All of the mentioned

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Answer:–c) Poisson

6. Usually replacing the standard error by its estimated value does change the CLT.

a) True

b) False

Answer:–b) False

7. Which of the following testing is concerned with making decisions using data?

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

Answer:-b) Hypothesis

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.

a) 0

b) 5

c) 1

d) 10

Answer:–a) 0

9. Which of the following statement is incorrect with respect to outliers?

a) Outliers can have varying degrees of influence

b) Outliers can be the result of spurious or real processes

c) Outliers cannot conform to the regression relationship

d) None of the mentioned

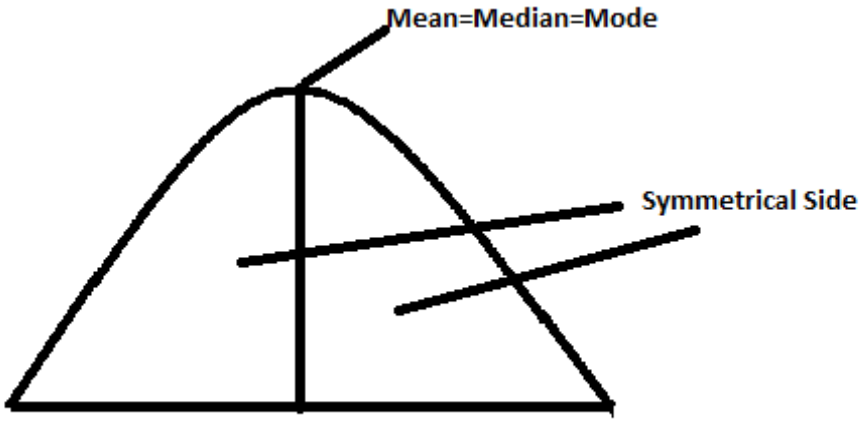
Answer:-c) Outliers cannot conform to the regression relationship

1. What do you understand by the term Normal Distribution?

Answer:- Normal distribution describe how values of random variables are distributed. It is also called Gaussian distribution or bell curve distribution. It is occurs naturally in many situations. The bell curve is symmetrical. Half of the data will fall to left of the mean and half will fall to the right.

Properties of normal distribution-

1. The mean, mode and median are all equal.
2. The curve is symmetric at centre.
3. The total area under curve is 1.



Normal Disribution Curve

2. How do you handle missing data? What imputation techniques do you recommend?

Answer:- When the no value stored in variable during observation its called missing value/data.

There are two way to Handel this missing data

1. Removing Data
2. Imputation

When the missing data is at random, related data can be deleted to reduce the bias, but deleting the data is not good practice if there are not enough observation at that time result in analysis may be affect so that we have to fill this missing data For that we have to find out the Mean or Median of the features from the existing data and replace missing values with this Mean or Median.

3. What is A/B testing?

Answer:- A/B testing is also known as bucket testing or split-run testing. It is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing and two sample hypothesis.

A/B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants are more effective.

4. Is mean imputation of missing data acceptable practice?

Answer: – In general mean imputation is bad practice while handling missing data because of following two reasons -

Mean imputation does not preserve the among variable.

Mean imputation lead to underestimate of standard error/deviation.

5. What is linear regression in statistics?

Answer:- In statistics, Linear regression is a linear approach to modelling the relationship between a scalar response and one or more variables (also known as dependent and independent variables),

The case of one variable is called simple linear regression and for more than one, the process is called multiple linear regression.

It simply uses traditional slope intercept form.

$Y = mx+b$

Where,

Y= Dependent variable

B=intercept of line

M=Linear Regression Co-efficient

X= independent variable

6. What are the various branches of statistics?

Answer:-The two main branches of statistics are –

1. Dispersive statistic: Descriptive statistics is the first part of statistics that deals with the collection of data. It having two parts.
    - Central tendency measures
    - Variability measures

2. Inferential statistic: The inference statistics are techniques that enable statisticians to use the information collected from the sample to conclude, bring decisions, or predict a defined population. It include following methods,

    - Regression analysis
    - Analysis of variance (ANOVA)
    - Analysis of covariance (ANCOVA)
    - Statistical significance (t-test)
    - Correlation analysis