STATISTICS WORKSHEET-1

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

a) True

b) False

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

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

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

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

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

a) True

b) False

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

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

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

original data.

a) 0

b) 5

c) 1

d) 10

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

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

- Normal Distribution is a probability distribution which is symmetric about the mean.
- In a graph form normal distribution will appear as a probability bell curve
- In ND the mean is 0 and the standard deviation is 1
- It shows the data along the mean is more frequent then the data far away from the mean.

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

- First we need to find the missing values by reading the file and getting the sum of NA
- We need to handle the columns with null data we can either impute the null values or delete.
- Replacing with the Arbitrary value (0).
- Replacing with the Mean, Mode , Median.
- Replacing with the previous value
- Replacing with the next value
- Imputing the most frequent VALUE.

12. What is A/B testing?

- It is also known as Split testing or Bucket testing
- It is a research method which compares 2 different variants with each other
- The testing is done to determine what is preferred by the targeted audience.
- It helps to identify the how the change of the variants affects the end users.
- It solves user problems ,saves money, reduces risks.

13. Is mean imputation of missing data acceptable practice?

- It’s the most common method if imputation.
- In simple words it is replacing the nul, values with the mean value
- Its not recommended when we have outliners.
- For eg if loanamount has null values we can replace the nul values with mean as follows
- Bank_df[‘loanamount’] = Bank_df[[‘loanamount’].fillna(Bank_df[‘loanamount’].mean())

14. What is linear regression in statistics?

- It is a machine learning algorithm based on supervised learning
- It performs a regression task.
- It is mostly used to find relationship between variables and forecasting
- It performs tasks to predict a dependent variable based on a given independent variable

15. What are the various branches of statistics?

- Data collection

  Its all about how the data is collected- what kind of data is collected.
- Descriptive statistics

  This deals with the data presenting- its presenting the data we have collected in the form of graphs-chats-various insights
- Inferential statistics.

  It is the aspect that deals with concluding the data- what the actual data is telling us –what information we can get though the data.