WORKSHEET

STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

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

a) True

b) False

--> a

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

--> a

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

--> b

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

--> d

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

--> c

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

a) True

b) False

--> b

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

--> d

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

--> a

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

--> c

WORKSHEET

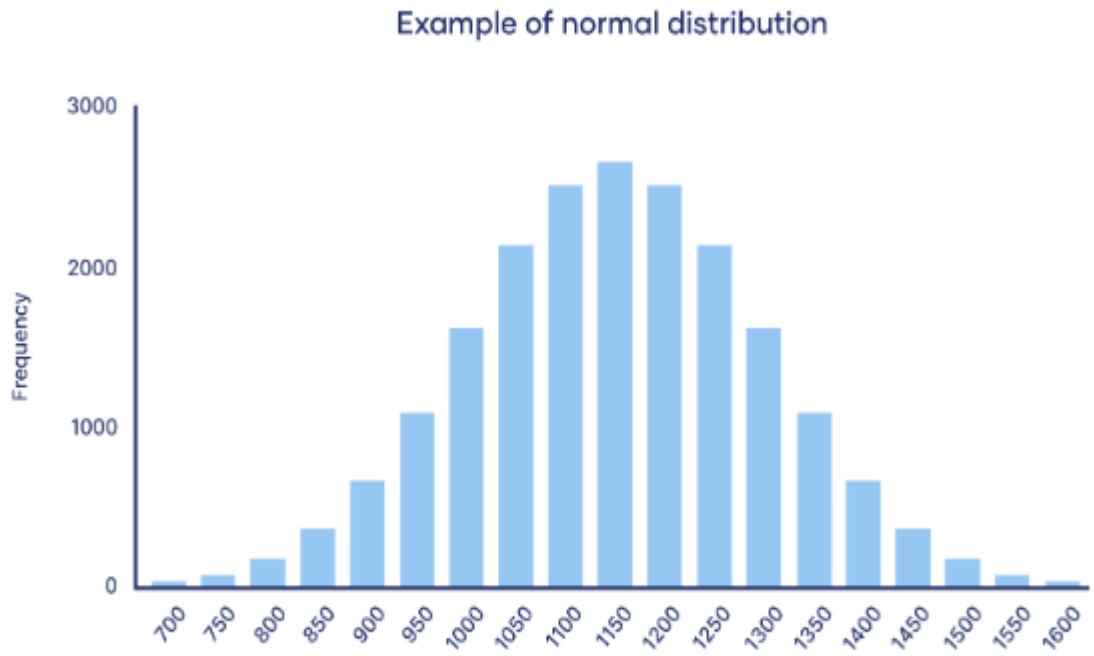
Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

--> Data is symmetrically distributed with no skew in Normal Distribution.

When plotted on a graph, the data follows a bell shape, with most values clustering around a central region and tapering off as they go further away from the center.

Normal distributions are also called Gaussian distributions or bell curves because of their shape.



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

--> Here are the most common ways of handling missing data

Zero Replacement: Here, you replace the missing value with zero irrespective of everything.

Min or Max Replacement: Replace the missing value with the minimum or maximum value of a feature.

Mean/ Median/ Mode Replacement: Replace missing value with mean or median or most frequent feature value.

Also, one can replace the value of the missing cell with the previous cell's value. This kind of technique is popular while inputting time series data. For example, if the price of an instrument is missing on the i-th day, it makes sense to replace it with the (i-1)-th day's price.

12. What is A/B testing?

--> A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to

determine which one performs better.

13. Is mean imputation of missing data acceptable practice?

--> Mean imputation (MI) is one such method in which the mean of the observed values for each variable is computed and the missing values for that variable are imputed by this mean.

This method can lead into severely biased estimates even if data are MCAR

14. What is linear regression in statistics?

--> Linear regression is the simplest and most extensively used statistical technique for predictive modelling analysis.

It is a way to explain the relationship between a dependent variable (target) and one or more explanatory variables(predictors) using a straight line.

There are two types of linear regression - Simple and Multiple.

15. What are the various branches of statistics?

--> The two major areas of statistics are known as descriptive statistics, which describes the properties of sample and population data,

and inferential statistics,

which uses those properties to test hypotheses and draw conclusions.