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

2. Which of the following theorems 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

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

b) Modeling bounded count data

4. Point out the correct statement.

d) All of the mentioned

5. _____ random variables are used to model rates.

c) Poisson

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

b) False

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

b) Hypothesis

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

a) 0

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

c) Outliers cannot conform to the regression relationship

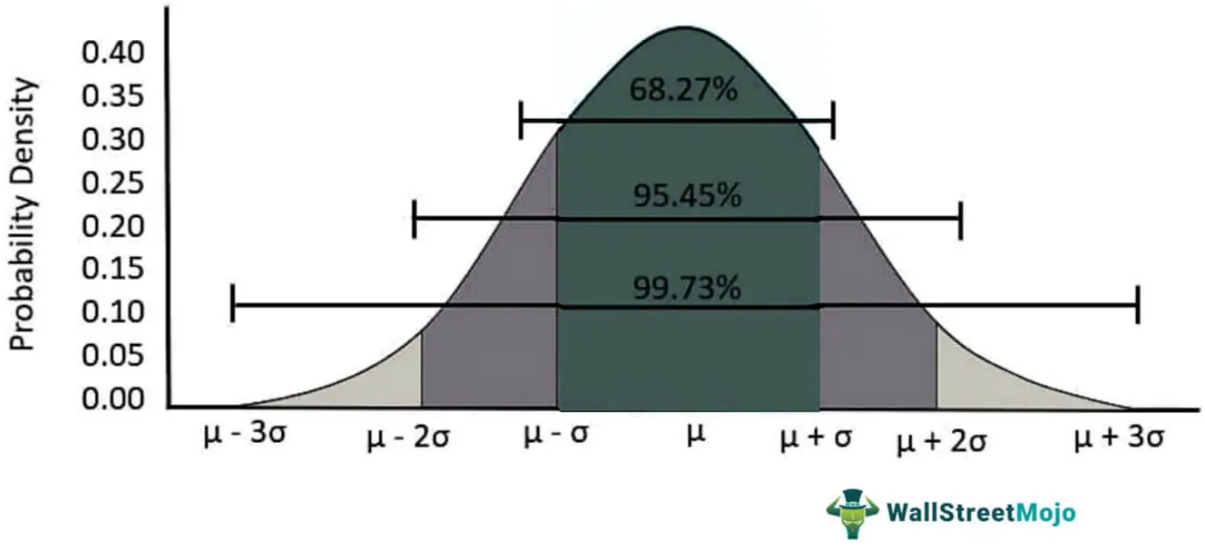
Q10and Q15 are subjective answer type questions. Answer them in your own words briefly.

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

1) Normal Distribution

A normal distribution is the proper term for a probability bell curve. In a normal distribution the mean is zero and the standard deviation is

A normal distribution is a statistical phenomenon representing a symmetric bell-shaped curve. Most values are located near the mean; also, only a few appear at the left and right tails



---

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

1) LET'S CONSIDER OUR DATA SET IS TO LARGE AND MISSING VALUE IS TOO LESS WILL DELETE IT ONLY IF IT IS NOT GOING TO AFFECT ON OUR DATASET
2) TO FILL NAN/MISSING DATA WE WILL USE A IMPUTATION TECHNIQUES
   It is the advanced technique to fill nan/null value

   KNN imputer

   Knn imputer will try to find the relation with other columns and impute the data according to the relation with other columns.

   Iterative Imputer

---

12. What is A/B testing?

A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.For instance, let’s say you own a company and want to increase the sales of your product.

---

13. Is mean imputation of missing data acceptable practice?

Mean imputation is **typically considered terrible practice**

---

14. What is linear regression in statistics?

It is the process of **predicting label** (or dependent variable/or output) based on the features (independent variable/input data)

Let's suppose we have to predict **emp_salary** based on **experience** so here ,salary is dependent on experience. = so it dependent variable/label and experience is independent variable/input data/features based data we have we are going to predict salary

---

15. What are the various branches of statistics?
1) Descriptive :- Example, statistics are used to **describe** the basic features of the data in a study. They provide simple summaries about the sample and the measures. and
2) Inferential:- Example, **voting poll** results
We can understand this topic by taking an example of election. Ever wondered how the media came up with the exit polls?
Assume an Indian state. Karnataka ,now the media will take a few samples from all the cities in karnataka.
Collecting the data from a few populations in order to form an exit poll is one of the examples of inferential statistics and with this example population and samples are also included.
After taking samples from the few populations,we will now start making predictions.
Another example can be of the electronics,in order to check the quality of the television ,not all televisions have to go through the quality control process. A few items are selected randomly and through the test,if they pass the test,all the television sets are labeled as approved.