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

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

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

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

Answer: A normal distribution is the proper term of a probability bell curve. In normal distribution the mean is zero and the standard deviation is 1.
Examples: height, birth weight, reading ability

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

Answer: Missing records may be treated in a variety of ways. I believe the most commonplace reaction is to disregard it. Choosing to make no decision, however, indicates that your statistical programme will make the choice for you.Your software will dispose of things in a listwise series most of the time. Depending on why and how much facts is long gone, listwise deletion may or may not be a good concept.
Another commonplace approach amongst folks who pay attention is imputation. Imputation is the technique of substituting an estimate for missing values and analysing the whole records set as if the imputed values were the true discovered values.

Types of Imputation:

1. Mean imputation

Calculate the mean of the observed values for that variable for all non-missing human beings.It has the gain of retaining the same suggest and sample size, but it also has a slew of drawbacks. Almost all of the methods described beneath are advanced to mean imputation.

2. Hot deck imputation

A fee picked at random from a pattern member who has comparable values on other variables.To positioned it another manner, choose all of the sample participants who are similar on different elements, then pick one of their missing variable values at random.

3. Cold deck imputation

A fee picked deliberately from an individual with comparable values on different variables.In maximum elements, that is corresponding to Hot Deck, however with out the random variance. As an instance, below the same experimental condition and block, you could constantly pick out the 0.33 man or woman.

4. Regression imputation

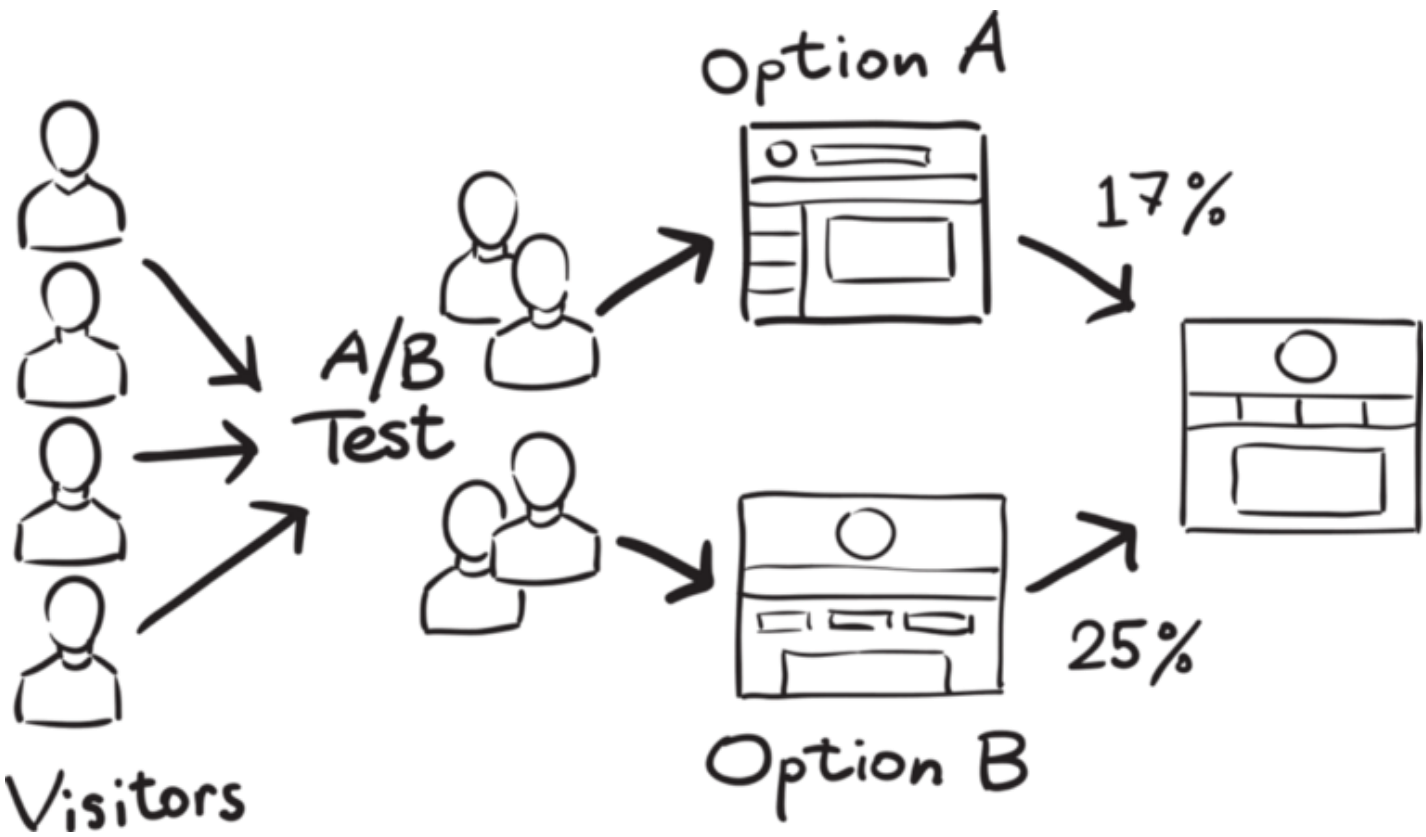
The end result of regressing the missing variable on other factors to get a anticipated price.As a end result, instead of utilizing the suggest, you are relying on the expected value, that is motivated by other elements. This maintains the associations among the variables within the imputation version, but now not the range around the predicted values.

12. What is A/B testing?

Answer: A/B trying out is a basic randomized control experiment. It is a way to evaluate the 2 versions of a variable to find out which plays better in a managed surroundings.

For example, allow's say you own a business enterprise and need to increase the income of your product. Here, both you could use random experiments, or you can observe clinical and statistical techniques. A/B checking out is one of the most outstanding and extensively used statistical gear.

In the above situation, you could divide the products into elements – A and B. Here A will remain unchanged even as you're making big adjustments in B's packaging. Now, on the premise of the response from client groups who used A and B respectively, you try and determine that's appearing higher.



It is a hypothetical trying out method for making selections that estimate population parameters based on sample facts. The populace refers to all the clients shopping for your product, whilst the pattern refers to the range of clients that participated in the take a look at.

13. Is mean imputation of missing data acceptable practice?

Answer: The manner of replacing null values in a records collection with the data's mean is called mean imputation.
Mean imputation is usually considered terrible practice since it ignores feature correlation. Consider the subsequent state of affairs: we've got a table with age and fitness scores, and an eight-12 months-antique has a missing health rating. If we average the health rankings of human beings between the a while of 15 and eighty, the eighty-yr-antique will appear to have a notably greater health level than he definitely does.
Second, suggest imputation decreases the variance of our data at the same time as growing bias. As a result of the reduced variance, the version is less correct and the confidence c programming language is narrower.

14. What is linear regression in statistics?

Answer: Linear regression evaluation is used to expect the price of a variable based totally at the cost of any other variable. The variable you need to are expecting is called the based variable. The variable you are the use of to expect the opposite variable's fee is referred to as the impartial variable.
This form of analysis estimates the coefficients of the linear equation, involving one or greater impartial variables that excellent are expecting the cost of the based variable. Linear regression fits a instantly line or floor that minimizes the discrepancies among anticipated and real output values. There are easy linear regression calculators that use a "least squares" method to discover the fine-fit line for a set of paired statistics. You then estimate the price of X (dependent variable) from Y (independent variable).

15. What are the various branches of statistics?

Answer: Types of statistics

Descriptive and Inferential

Descriptive Statistics: when we can describe the data we called as descriptive statistics.

- **➔ Measure of central tendency(Mean,Median,Mode)**
- **➔ Measure of dispersion(Variance and Standard deviation)**

Inferential Statistics: when we pick random data and infer that it is know as Inferential Statistics.

- **➔ Sampling data and infer the result to describe population.**