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.
   1. True
   2. False

**Answer:**

1. True

Bernauli random variables are discrete random variable and takes only 0 and 1 .

It takes the value 1 when happening of event or “success” and 0 when not happening of event or “failure”.

1. 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?
   1. Central Limit Theorem
   2. Central Mean Theorem
   3. Centroid Limit Theorem
   4. All of the mentioned

**Answer:**

1. Central Limit Theorem
2. Which of the following is incorrect with respect to use of Poisson distribution?
   1. Modeling event/time data
   2. Modeling bounded count data
   3. Modeling contingency tables
   4. All of the mentioned

**Answer:**

1. Modeling bounded count data

The Poison Distribution is actually used for unbounded count data, and count of events are ranged from 0 to infinity .

1. Point out the correct statement.
   1. The exponent of a normally distributed random variables follows what is called the log- normal distribution
   2. Sums of normally distributed random variables are again normally distributed even if the variables are dependent
   3. The square of a standard normal random variable follows what is called chi-squared distribution
   4. All of the mentioned

**Answer:**

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

This is true and it is mainly used statistical interference. -hypothesis testing and constructing confidence intervals for variance.

1. random variables are used to model rates.
   1. Empirical
   2. Binomial
   3. Poisson
   4. All of the mentioned

**Answer:**

Poison

1. 10. Usually replacing the standard error by its estimated value does change the CLT.
   1. True
   2. False

**Answer:**

FALSE

Replacing the standard error by its estimated value does not change the central limit Theorem , it describes the behavior of sample means as the sample size increases and it remains valid even when the standard error is estimated from the sample..

1. 1. Which of the following testing is concerned with making decisions using data?
   1. Probability
   2. Hypothesis
   3. Causal
   4. None of the mentioned

**Answer:**

Hypothesis

Hypothesis testing is concerned with making decisions using data . It involves formulating null and alternative hypothesis , collecting data and using statistical methods to determine whether there is enough evidence to reject null hypothesis in favor of the alternative hypothesis.

1. 4. Normalized data are centered at and have units equal to standard deviations of the original data.
   1. 0
   2. 5
   3. 1
   4. 10

**Answer:**

0 (zero)

Normaliesd data are always centred at 0 , and have units equal to standard deviations of the original data.

1. Which of the following statement is incorrect with respect to outliers?
   1. Outliers can have varying degrees of influence
   2. Outliers can be the result of spurious or real processes
   3. Outliers cannot conform to the regression relationship
   4. None of the mentioned

**Answer:**

Outliers cannot confirm to the regression relationship .

This is incorrect as outliers can have significant impact on regression relationship.

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

1. What do you understand by the term Normal Distribution?
2. How do you handle missing data? What imputation techniques do you recommend?
3. What is A/B testing?
4. Is mean imputation of missing data acceptable practice?
5. What is linear regression in statistics?
6. What are the various branches of statistics?

Q10 : What do you understand by the term Normal Distribution

**Answer:**

Normal Distribution : We cans say it is kind of continues probability distribution in which most of the data points are clustered in middle of the range, and the rest are taper of symentrically towards either extremes .

This middle range is known as mean of the distribution.

It is also known as Gaussian distribution or probability bell curve . It is symmetric about the mean indicates the values near mean occur more frequently than the values that are father away from the mean.

Q11: How do you handle missing data? What imputation technique do you recommend?

**Answer:**

Generally we should use imputation technique to handle with missing data.

Imputation process is the process of substituting an estimate for missing values and analysing the entire data set as the imputed values are the true values.

Major imputation techniques we use are – Mean imputation , substitution etc We normally use Mean and Mode method to fill missing or Null values in dataset.

Q12: What is A/B testing ?

**Answer:**

A/B testing is a way to compare 2 or more versions of variables to determine which one will perform better.It is commonly used in various fields – marketing , web development to improve strategies.

Q13: Is mean imputation of missing data is acceptable practice??

**Answer:**

Acceptance is actually depends upon the nature of the data , context and any specific analysis we are doing.

Since it works in the range of min and maximum so it will not effect the overall result of the dataset but if there is any actually varience in data it is artificially reduced by Mean method .

Q14: What is linear regression statistics?

**Answer:**  The linear regression is a statistical method which gives the relationship between dependent variable and independent variable (can be one or more) using a straight line .

This is of two types –

Simple linear regression which is used when we have only one independent variable .

Multiple Linear regression which is used when we have multiple independent variables.

It is generally used for 4 assumptions – Linearity , Independence , homoscedasticity , and normality .

Q15: What are various branches of Statistics?

**Answer:**  Statistics is the branch of Mathematics which deals with data. It involves collection , description , analysis and inference of conclusion from the quantitative data .

There are 2 main branches of Statistics –

Descriptive Statistics : This describes the properties of sample and population data. This included mean, variance, skewness, and kurtosis.

Inferential Statistics : This uses above properties to test hypothesis and draw conclusion . This includes regression analysis, analysis of variance (ANOVA) , logit/probit model, and Null Hypothesis testing.

