

## 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

**Ans 1 .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

**Ans 2. 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

**Ans 3. 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

**Ans 4.d) All of the mentioned**

5. \_\_\_\_\_ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

**Ans 5. c) Poisson**

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

a) True

b) False

**Ans 6.a) True**

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

**Ans 7.b) Hypothesis**

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

**Ans 8.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

**Ans 9. c) Outliers cannot conform to the regression relationship**

#### WORKSHEET

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Ans 10. Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve

In normally distributed data, there is a constant proportion of data points lying under the curve between the mean and a specific number of standard deviations from the mean. Thus, for a normal distribution, almost all values lie within 3 standard deviations of the mean

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

ans 11. Use deletion methods to eliminate missing data. The deletion methods only work for certain datasets where participants have missing fields. ...

Use regression analysis to systematically eliminate data. ...

Data scientists can use data imputation techniques.

Mean imputation. Simply calculate the mean of the observed values for that variable for all individuals who are non-missing. ...

Substitution. ...

Hot deck imputation. ...

Cold deck imputation. ...

Regression imputation. ...

Stochastic regression imputation. ...

Interpolation and extrapolation.

12. What is A/B testing?

A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

13. Is mean imputation of missing data acceptable practice?

Outliers data points will have a significant impact on the mean and hence, in such cases, it is not recommended to use the mean for replacing the missing values. Using mean values for replacing missing values may not create a great model and hence gets ruled out

14. What is linear regression in statistics?

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they—indicated by the magnitude and sign of the beta estimates—impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula  $y = c + b \cdot x$ , where  $y$  = estimated dependent variable score,  $c$  = constant,  $b$  = regression coefficient, and  $x$  = score on the independent variable.

Naming the Variables. There are many names for a regression's dependent variable. It may be called an outcome variable, criterion variable, endogenous variable, or regressand. The independent variables can be called exogenous variables, predictor variables, or regressors.

Three major uses for regression analysis are (1) determining the strength of predictors, (2) forecasting an effect, and (3) trend forecasting.

15. What are the various branches of statistics?

The two divisions of statistics are Descriptive statistics and Inferential statistics. There are two types of main division of statistics. They are Descriptive and inferential.