PYTHON – WORKSHEET 1 ANSWERS

Answer 1 – (C)%

Answer 2 - (B) 0

Answer 3- (C) 24

Answer 4 - (A) 2

Answer 5 - (D) 6

Answer 6 - (B) It encloses the lines of code which will be executed if any error occurs while executing the lines of code in the try block.

Answer 7 - (A) It is used to raise an exception.

Answer 8 - (C) in defining a generator.

Answer 9 - (A) , (B), (C)

Answer 10 - (A) ,(B)

STATISTICS WORKSHEET-1 Answers

Answer 1 - (A) True

Answer 2 - (A) Central limit theoram

Answer 3 - (B) Modeling bounded count data

Answer 4 - (D) All of above

Answer 5 - (C) Poisson

Answer 6 - (A) True

Answer 7 - (B) Hypothesis

ANswer 8 - (A) 0

Answer 9 - (B) Outliers can be the result of spurious or real processes

Answer 10 – Normal distribution in statistics also called gaussian distribution. It is the most significant continuous probability distribution, It has Bell shaped curve. A large number of random variables are either nearly or exactly represent the data. The probability density function of normal distribution is given by:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

- Where x is a variable , μ is the mean , σ is the standard deviation

Answer 11 -

We often encounter missing value while we are try to analyze or understand a data. There will be missing values because the data might be corrupted or some other error. Missing values can cause bias and can affect the efficiency of how the data perform. Mean, Median , Mode is one of the most common methods of imputing value when dealing with missing data. There are may ways in which we can handle missing data like.

Imputation VS removing data , Deletion , Imputation, Time series specific methods, K- nearset neighbors.

Answer 12 -

A/B testing (also known as split testing) is a process of showing two variants of the same web page to different segments of website visitors at the same time and comparing which variant drives more conversions.

A/B testing is an efficient and effective way to “gauge your audience's response to a design or content idea” because it doesn't disturb your users' experience or send out disruptive feedback surveys. Just try something new and let the results speak for themselves. **g**

Answer 13 -

It is acceptable when the missing value proportion is not large enough. But, when the missing values are large enough and you impute them with the mean, the standard errors will be lesser than what they actually would have been.

Answer 14-

In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

Answer 15-

The types of statistics are categorised based on these features: Descriptive and inferential statistics. Based on the representation of data such as using pie charts, bar graphs, or tables, we analyse and interpret it.

Basically, the statistical analysis is meant to collect and study the information available in large quantities. Statistics is a branch of mathematics, where computation is done over a bulk of data using charts, tables, graphs, etc.

The data collected for analysis here is called measurements. Now, if we have to measure the data based on a scenario, a sample is taken out of a population. Then the analysis or calculation is done for the following measurement.

MACHINE LEARNING Worksheet –1 Answers

Answers 1 - (A) Least square method

Answer 2 - (A) Linear regression is sensitive to outliers

Answer 3 - (B) Negative

Answer 4 - (A) Regression

Answer 5 - (C) Low bias and high variance

Answer 6 - (B) Predictive model

Answer 7 - (D) Regularization

Answer 8 - (D) SMOTE

Answer 9 - (C) Sensitivity and Specificity

Answer 10 - (A) True

Answer 11 - (A) Construction bag of words from an email

Answer 12- (A) , (B) , (C)

Answer 13 -

Regularization is a technique which is used to solve the overfitting problem of the machine learning models. Overfitting is a phenomenon which occurs when a model learns the detail and noise in the training data to an extent that it negatively impacts the performance of the model on new data.

Answer 14 -

Ridge Regression and LASSO Regression

Answer 15 -