Python Answer sheet

1. The **modulo operator**, denoted by %, is an arithmetic operator. The modulo division operator produces the remainder of an integer division.
2. The decimal form of 2/3 is **0.666**
3. 36
4. True
5. 0
6. B option
7. A option
8. C option
9. D option
10. D option

Statistics

1. A
2. C
3. B
4. D
5. C
6. A
7. B
8. A
9. C
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 a normal distribution, the mean is zero and the standard deviation is 1. It has zero skew and kurtosis of 3.

Normal distributions are symmetrical, but not all symmetrical distributions are normal.

Many naturally-occurring phenomena tend to approximate the normal distribution.

In finance, most pricing distributions are not, however, perfectly normal.

1. Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical programme will make the decision for you.

Your application will remove things in a listwise sequence most of the time. Depending on why and how much data is gone, listwise deletion may or may not be a good idea.

Another common strategy among those who pay attention is imputation. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

1. Mean imputation
2. Substitution
3. Hot deck imputation
4. Regression imputation
5. The last observation carried forward
6. Maximum likelihood
7. Expectation-Maximization
8. Multiple imputation
9. [A/B testing](https://vwo.com/testing/ab-testing/), also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

1) Solve visitor pain points

2) Get better ROI from existing traffic.

3) Reduce bounce rate.

4) Make low-risk modifications.

5) Achieve statistically significant improvements.

1. It considers to be a bad idea for imputation for the following reasons.

1) Mean imputation reduces the variance of the imputed variables.

2) Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval.

3) Mean imputation does not preserve relationships between variables such as correlations.

1. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a “least squares” method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable) from Y (independent variable).

1. **The Branches of Statistics**

Two branches, *descriptive statistics* and *inferential statistics*, comprise the field of statistics.

1. **Descriptive Statistics**

**CONCEPT** The branch of statistics that focuses on collecting, summarizing, and presenting a set of data.

**EXAMPLES** The average age of citizens who voted for the winning candidate in the last presidential election, the average length of all books about statistics, the variation in the weight of 100 boxes of cereal selected from a factory's production line.

**INTERPRETATION** You are most likely to be familiar with this branch of statistics, because many examples arise in everyday life. Descriptive statistics forms the basis for analysis and discussion in such diverse fields as securities trading, the social sciences, government, the health sciences, and professional sports. A general familiarity and widespread availability of descriptive methods in many calculating devices and business software can often make using this branch of statistics seem deceptively easy.

**Inferential Statistics**

**CONCEPT** The branch of statistics that analyzes sample data to draw conclusions about a population.

**EXAMPLE** A survey that sampled 2,001 full-or part-time workers ages 50 to 70, conducted by the American Association of Retired Persons (*AARP*), discovered that 70% of those polled planned to work past the traditional mid-60s retirement age. This statistic could be used to draw conclusions about the population of all workers ages 50 to 70.

**INTERPRETATION** When you use inferential statistics, you start with a hypothesis and look to see whether the data are consistent with that hypothesis. Inferential statistical methods can be easily misapplied or misconstrued, and many inferential methods require the use of a calculator or computer.

Machine learning

1. A
2. B
3. B
4. C
5. A
6. B
7. D
8. C
9. A
10. B
11. B
12. A,b,c.
13. Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.

Sometimes the [machine learning](https://www.javatpoint.com/machine-learning) model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when dealing with unseen data by introducing noise in the output, and hence the model is called overfitting. This problem can be dealt with the help of a regularization technique.

This technique can be used in such a way that it will allow maintaining all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.

Techniques of Regularization

There are mainly two types of regularization techniques, which are given below:

* **Ridge Regression.**
* **Lasso Regression.**

Ridge Regression

* Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.
* Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as **L2 regularization**.
* In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called **Ridge Regression penalty**. We can calculate it by multiplying with the lambda to the squared weight of each individual feature

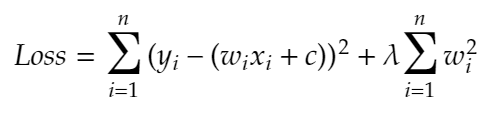
**Lasso Regression**:

* Lasso regression is another regularization technique to reduce the complexity of the model. It stands for **Least Absolute and Selection Operator.**
* It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.
* Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.
* It is also called as  **regularization.**

1. **Different Regularization algorithms**
2. Ridge Regression.
3. LASSO (Lea.st Absolute Shrinkage and Selection Operator.
4. Regression.
5. Elastic-Net Regression.

**Ridge Regression**

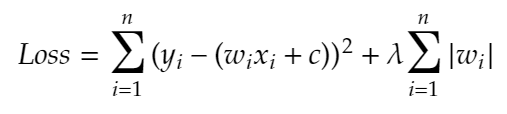
Ridge regression is a method for analyzing data that suffer from multi-collinearity.



Loss Function for Ridge Regression.

**LASSO Regression**

LASSO is a regression analysis method that performs both feature selection and regularization in order to enhance the prediction accuracy of the model.



Loss Function for LASSO Regression.