Q1. A
Q2. A
Q3. B
Q4. A
Q5. C
Q6. B
Q7. B
Q8. A
Q9. C
Q10. Normal distribution also known as the Gaussian distribution and 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 graphical form, the normal distribution appears as a "bell curve".

Q11. We can handle missing data using Imputation, in this process substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.
We can calculate the mean of the observed values for that variable for all non-missing data. It has the advantage of maintaining the same mean and sample size.
There are various imputation techniques which are as follows:
a). Substitution
b). Hot deck imputation
c). Cold deck imputation
d). Regression imputation
e). Single or Multiple imputation.

Q12. A/B testing is 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.
A/B testing is a shorthand for a simple randomized controlled experiment, in which two samples (A and B) of a single vector-variable are compared.
These values are similar except for one variation which might affect a user's behavior. A/B tests are widely considered the simplest form of controlled experiment. However, by adding more variants to the test, its complexity grows.

Q13. The process of replacing null values in a data collection with the data’s mean is known as mean imputation.
Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score.
If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.
Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

Q14. 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*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

Q15. There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.

a). Data Collection: Data collection is all about how the actual data is collected.

b). Descriptive Statistics: Descriptive statistics is the part of statistics that deals with presenting the data we have.

This can take two basic forms –presenting aspects of the data either visually (via graphs, charts, etc.) or numerically (via averages)

c). Inferential Statistics: Inferential statistics is the aspect that deals with making conclusions about the data.