Statistics

1. A
2. A
3. B
4. D
5.C
6. B
7. B
8. A
9. C
10. In a normal distribution, data are symmetrically distributed with no skew. Most values cluster around a central region, with values tapering off as they go further away from the center. The measures of central tendency (mean, mode, and median) are exactly the same in a normal distribution.
A normal distribution has a probability distribution that is centered around the mean. This means that the distribution has more data around the mean. The data distribution decreases as you move away from the center. The resulting curve is symmetrical about the mean and forms a bell-shaped distribution.

11. Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. 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.

The following are some of the most prevalent methods:

1. Mean imputation
2. Substitution
3. Hot deck imputation
4. Cold deck imputation
5. Regression imputation
6. Stochastic regression imputation
7. Interpolation and extrapolation
8. Single or Multiple Imputation

12. A/B testing is the process of comparing two variations of a page element, usually by testing users' response to variant A vs. variant B and concluding which of the two variants is more effective. It is also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

13. Mean imputation is considered awful practice as it avoids feature correlation. Since most research studies are interested about the relationship among variables, mean imputation is not a effective solution.

14. Linear regression is basic and commonly used predictive analysis technique. Regression estimates are used to describe data and to explain the relationship. Linear regression is also used for modelling.
For example, it can be used to quantify the impacts of age, gender, weight and diet on height

15. There are three branches of statistics-
1. Data collection- Data collection is fairly collecting data (sometimes getting approximate numbers) to deal with it and make conclusions to derive appropriate inference.
2. Descriptive statistics - Descriptive statistics is presenting data we have visually (graphs, charts, etc.) Or numerically.
3. Inferential statistics-Where inferential statistics is making conclusions about the data.
Example: getting conclusions about finding a solution to reduce road accidents via monitoring data of average speed of cars and making respective speed limits on the road.