Answers

1.

A)True

2.

a)Central Limit theorem

3

B) Modeling Bounded count data

4.

d)All of the mentioned

5.

C)Poisson

6.

B)False

7.

B)Hypothesis

8.

A)0

9.

C)Outliers cannot conform to the regression Relationship

10.

A **normal distribution** is a [probability distribution](#) used to model phenomena that have a default behaviour and cumulative possible deviations from that behaviour. For instance, a proficient [archer's](#) arrows are expected to land around the [bull's eye](#) of the target; however, due to aggregating imperfections in the archer's technique, most arrows will miss the bull's eye by some distance.

11.

Real-world data is messy and usually holds a lot of missing values. Missing data can skew anything for data scientists and, A data scientist doesn't want to design biased estimates that point to invalid results. Behind, any analysis is only as great as the data. **Missing data appear when no value is available in one or more variables of an individual.**

7 Ways to Handle Missing Values in Machine Learning

1. Deleting Rows with missing values.
2. Impute missing values for continuous variable.
3. Impute missing values for categorical variable.
4. Other Imputation Methods.
5. Using Algorithms that support missing values.
6. Prediction of missing values.

12.

A/B 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 drive business metrics.

13.

True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the **mean remains unbiased**. ... Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14.

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.

15.

The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.

Descriptive Statistics

Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.
Inferential Statistics

Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics.

In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.