

- 1) A (True)
- 2) B (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 refers to a probability distribution where the values of a random variable are distributed symmetrically. These values are equally distributed on the left and the right side of the central tendency. Thus, a bell-shaped curve is formed.

11) We can handle missing data through different methods

a) Mean or median imputation

A common technique is to use mean or median of non- missing observation. This can be useful in cases where the number of missing observations is low. For large number of missing values, using mean or median can result in loss of variation in data, and it is better to use imputations.

b) Multivariate imputation by chained equation (MICE)

To set up the data for MICE it is important to note that the algorithm uses all the variables in the data for predictions. In this cases variable that may not useful for predictions like the id variable should be removed before implementing this algorithm

c) Random forest

It is a non-parametric imputation method applicable to various variable type that works well with both data missing at random and not missing at random. Random forests uses multiple decision tree to estimate missing values and outputs OOB (OUT OF BAG) imputation error estimates

12) A/B testing is a type of split testing and is commonly used to drive improvements to specific variables or elements by measuring user or audience engagement. The approach is commonly used to optimize marketing campaigns or digital assets like websites. In A/B testing a specific variable is altered such as a title, image, or element layout. A sample of the audience is shown the control version and the altered version in a 50/50 split. Half traffic will interact with the original version, the other half will interact with the newer version. Engagement or the completion of a defined goal is the metric that is compared between the versions after a set period.

13) Mean imputation of missing data not acceptable practice because Any statistic that uses the imputed data will have a standard error that's too low. In other words, we get the same mean from mean-imputed data that we would have gotten without the imputations. And there are circumstances where that mean is unbiased. Even so, the standard error of that mean will be too small. Because the imputations are themselves estimates, there is some error associated with it. But statistical software doesn't know that. It treats it as real data. Ultimately, because standard errors are too low, so are p-values. Now we're making Type I errors without realizing it.

14) Linear regression models the relationships between at least one explanatory variable and an outcome variable. These variables are known as the independent and dependent variables, respectively. When there is one independent variable, the procedure is known as simple linear regression

15) The two main branches of statistics are descriptive statistics and inferential statistics.

1)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

2)inferential statistics 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.