

Answer 1).

a). True, Bernauli Random variable takes values 1 and 0.

Answer 2).

a). Central Limit Theorem, states that sample mean distribution becomes like normal distributed if sample size is 30 or more.

Answer 3).

b). Modelling bounded count data

Answer 4).

d). All of the mentioned

Answer 5).

c). Poisson

Answer 6).

b). False

Answer 7).

b). Hypothesis, on the basis of statistics we need to approve or reject the null hypothesis.

Answer 8).

a). 0

Answer 9).

c). Outliers cannot conform to the regression relationship

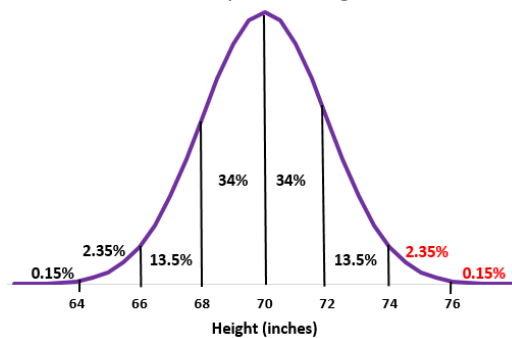
Answer 10)

Normal distribution is the statistical representation of the data which is found everywhere around us.

This distribution is bell shaped where 2 corners show the least occurred data because there are only few people with too short or too tall height.

Highest point bell is centred at average of the data because most of people have the average height.

Let's take an example of height of 100 students in a college



Answer 11)

We can use various method to fill the missing data i.e. mean, medium, mode, observation carried forward or backward, linear interpolation, MICE, Misforest.

A data scientist may consider using KNN method to impute the missing data.

Answer 12)

AB Testing is basically a segregation of entire population in 2 different group in order to compare a metric in controlled environment and test environment.

Answer 13)

No, because it doesn't take feature correlation in to consideration

No, because mean reduces the variance of data hence bring the bias-ness in the model.

Answer 14)

Linear regression is a method for predictive analysis. Here the impact of the feature is generally shown on the label/result. It is presented in the given formula

$$Y = mx + c$$

c = constant, b = coefficient, and x = data

We can use linear regression analysis for 1) determining the strength of predictors, (2) forecasting an effect, and (3) trend forecasting.

Answer 15)

Descriptive statistics

Inferential statistics