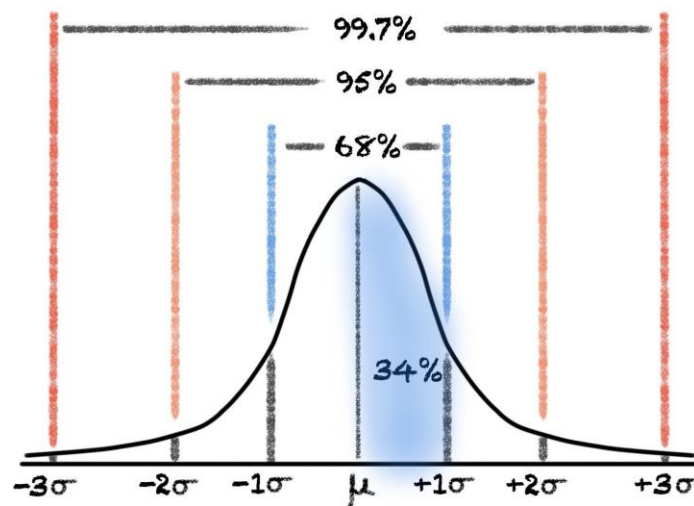


STATISTICS WORKSHEET-1

- (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) Normal distribution refers to 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. We can see it as a bell curve on a graph.



In this chart, the μ is in middle of the x-axis, that's the mean of the data set. The data point in our dataset are distributed in a bell shaped curved that is centred at the mean, as we can see that our data are symmetrically distributed or mirrored each side of y-axis, we call this a normal distribution.

As we can notice at the bottom along the x-axis the positive numbers are moving to the right and negative numbers moving to the left. The distance between each number represents one standard deviation. According to empirical rule, it says when you have symmetrical distribution of the data, we can expect 68% of all our data points to be within one standard deviation so, 68% of all the data points are within negative one and the positive one on this graph.

- (11) There are a few strategies to deal with missing information:
- a. Erase the lines
 - b. Replace with the most incessant qualities.

- c. Apply classifier calculations to anticipate.
- d. Apply solo AI.

In the first place, we need to examine and comprehend the idea of missing information, since it is basic in figuring out which strategy or treatment, we need to apply to credit the missing information.

Information can be absent in different ways.

- e. Missing totally indiscriminately: - when missing qualities are haphazardly circulated across all perceptions, then, at that point, we believe the information to be missing totally arbitrary.
- f. Missing aimlessly: - the critical contrast between missing totally indiscriminately and
- g. Missing indiscriminately (blemish) is that under blemish the information isn't missing arbitrarily across all perceptions however is missing haphazardly just with in sub-examples of information.
- h. Not missing aimlessly: - when the missing information has a design to it, we can't regard it as absent indiscriminately.

In initial two cases, it is protected to eliminate the information with missing qualities relying on their events, while in the last one eliminating perceptions with missing qualities can deliver a predisposition in the model. Thus, we must be truly cautious prior to eliminating perceptions.

Note: - imputation doesn't really give better outcomes.

(12) A/B testing, otherwise called split testing, alludes to a randomized experimentation measure wherein at least two forms of a variable (site page, page component, and so on) are displayed to various fragments of site guests simultaneously to figure out which rendition leaves the most extreme effect and drive business measurements

(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 endeavors to show the connection between two factors by fitting a straight condition to noticed information. One variable is viewed as an informative variable, and the other is viewed as a reliant variable.

(15) The two primary parts of statistics are:

1.Descriptive statistics : it is a strategy for getting sorted out, summing up, and introducing information instructively.

2. Inferential statistics : it is a strategy which is utilized in deciding something about a populace based on an example.

Populace: - the whole arrangement of people or objects of interest or the estimations got from all people or objects of interest.

Test: - A piece or part of the number of inhabitants in interest