Normal distribution

Science, Statistics

Normal Distribution Why is the normal distribution so important, particularly with respect to problems dealing with statistical inference? The reason why normal distribution is so important is because of the Central Limit Theorem. This theorem states that as long as there are a significantly large number of variables involved in the sampling process, then the average, or mean, will approximately located at the center of the curve. This is important when interpreting data because the best results can be attained through a sampling process that has very few outliers. Ideally, the number of participants, or variables, in a study should be totally random in order to attain validity. If a study has a limited number of variables, then it is likely that there will be a wider margin of error, and thus the results cannot be truly trusted. For a group of statistics to be reliable, they need to have a normal or bell-shaped curve. In order to help achieve this outcome, the mean will need to be as close to the center as possible. Also, the standard deviation will likely be small in order to help the curve remain tall and narrow.

2. Computationally, why is normal distribution so useful?

The reason why normal distribution is so useful computationally is that it allows us to make conclusions from the data at hand. Let’s say, for instance, that a graph was skewed either to the left of the right. It would be very hard to come to any conclusive theories based off the data. This is because the data would not be reliable and could not be used to make any assumptions. Normal distribution is symmetrical, so it is easy to make conclusions just by looking at it with the naked eye. Other forms of data may require detailed calculations in order to make an inference or hypothesis. If a curve displays

normal distribution, then the mean will be exactly the same figure as the median and the mode. Also, the standard deviations will be within a set range of numbers. If a curve is bell-shaped, then more than two thirds of the distribution (68%) will lie between one standard deviation of the mean. More than this, 95% of the data will be located within two standard deviations of the mean. All these figures are useful when working out probability. It is possible to say that there is a certain percent of a variable falling within a certain range on the graph.

3. In tests like Goodness of Fit, Independence, and ANOVA, why are other distributions introduced and used instead of the normal distribution?

The tests such as Goodness of Fit, Independence, and ANOVA all use other distributions because the data is not symmetrically aligned. It could be that the distribution is skewed, either to the left or the right, and thus it is not possible to come to a conclusion simply by looking at the graph. The occurrence of outliers is also likely in such a case. One such method is the Chi-square test, which uses an equation to figure out the patterns of the data.