DS064-Statistics 4

By: Eng. Esraa Madhi

Advanced Statistical Concepts for Data Scientists

Using data analysis and statistics to make conclusions about a population from a sample is called statistical inference.

The main types of statistical inference are:

- Estimation (with confidence interval)
- Hypothesis testing

https://www.w3schools.com/statistics/statistics_statistical_inference.php

Statistics - Statistical Inference •

1. Confidence interval

https://www.w3schools.com/statistics/statistics_estimation.php

Statistics - Estimation •

https://www.w3schools.com/statistics/statistics_estimation_proportion.php

Statistics - Estimating Population Proportions •

https://www.w3schools.com/statistics/statistics_estimation_mean.php

Statistics - Estimating Population Means •

https://makemeanalyst.com/inferential-statistics/confidence-intervals/

A Confidence Interval - MAKE ME ANALYST • makemeanalyst.com

2. Hypotheses testing

https://www.w3schools.com/statistics/statistics_hypothesis_testing.php

Statistics - Hypothesis Testing •

https://www.w3schools.com/statistics/statistics_hypothesis_testing_mean.php

 Statistics - Hypothesis Testing a Mean •

---

3. Central Limit Theorem

It provides insight into population data by using the mean of the samples, and if the mean value of samples is plotted, it approaches a Normal Distribution that holds irrespective of the type of distribution of population.

Central limit theorem (CLT) justifies why normal distribution can be used in such cases. According to the CLT, as we take more samples from a distribution, the sample averages will tend towards a normal distribution regardless of the population distribution.

Consider a case that we need to learn the distribution of the heights of all 20-year-old people in a country. It is almost impossible and, of course not practical, to collect this data. So, we take samples of 20-year-old people across the country and calculate the average height of the people in samples. CLT states that as we take more samples from the population, sampling distribution will get close to a normal distribution.