

# Statistics

## Assignment 3

**1. Write the Gaussian Distribution empirical formula?**

**Ans:** According to Gaussian Distribution Empirical Formula, 68% of the data falls within one standard deviation, 95% percent within two standard deviations, and 99.7% within three standard deviations from the mean.

**2. What is the Z-score, and why is it important?**

**Ans:** Z-score is a value's relationship to the mean of a group of values. It is expressed with the letter "z" and found by subtracting sample mean from observed value and dividing the result by sample standard deviation. Z-score is always between 0 and 1 for all distribution. So it is important because we can compare data from different distributions with z-score.

**3. What is an outlier, exactly?**

**Ans:** An outlier is an extreme value that affects the measures of central tendency and dispersion of a distribution. For example, a dataset containing: 10, 30, 15, 23, 5000, 21, 39. 12 has an outlier, which is 5000.

**4. What are our options for dealing with outliers in our dataset?**

**Ans:** First of all, we need to find the outlier by calculating the lower fence and the upper fence. IQR, first quartile and third quartile are used to find the fences. Once we find it, we can remove the outlier.

**5. Write the sample and population variances equations and explain Bessel's correction**

**Ans:**

Population Variance,  $\sigma^2 = \sum(x - \mu)^2/N$

Where, x = observed value,  $\mu$  = population mean, N = population size

Sample Varuance,  $s^2 = \sum(x - \bar{x})^2/(n - 1)$

Where, x = observed value,  $\bar{x}$  = sample mean, n = sample size

Bessel's Correction:

*There could be bias in the estimation of the population variance when we take a sample from it. Bessel's Correction is the use of  $n - 1$  instead of  $n$  in the formula for the sample variance and sample standard deviation to correct the bias.*