Statistics Assignment 3

1. Write the Gaussian Distribu**μ**tion empirical formula.

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

3. What is an outlier, exactly?

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

5. Write the sample and population variances equations and explain Bessel Correction.

    1. The empirical rule, also referred to as the 3-sigma rule or 68-95-99.7 rule is a statistical rule which states that for a normal distribution, almost all observed data will fall within 3 standard deviations(denoted by **σ**) of the mean or average(denoted by **μ**). In particular, the empirical rule predicts that 68% of the observations falls within the 1st standard deviation(***μ***±***σ***), 95% falls within the 2nd standard deviation(***μ***±2**σ**), and 99.7% within the 3rd standard deviation(***μ***±3**σ**).

2. The score (more commonly referred to as a Z-score) a very useful statistic because it allows us to calculate the probability of a score occurring within our normal distribution and enables us to compare two scores that are from different normal distribution.
3. An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. Examination of the data for unusual observations that are far removed from the mass of data. These points are often referred to as outliers.
4. Set up a filter in your testing tool. Even though this has a little cost, filtering out outliers is worth it.

   Remove or change outliers during post test analysis.

   Change the value of outliers.

   Consider the underlying distribution.

   Consider the value of mild outliers.

5. Population variance-

   $\boldsymbol{\sigma}^2=1/N\{\sum(X-\boldsymbol{\mu})^2\}$

Sample variance-

$S^2=1/n-1\{\sum(x-\bar{x})^2\}$

Bessel ‘s correction-

In estimating the population variance from a sample when the population mean is unknown, the uncorrected sample variance is the mean of the squares of deviation of sample values from the sample mean(i.e. Using a multiplicative factor 1/n) in this case the sample variance is a biased estimator of the population variance. Multiplying the uncorrected sample variance by the factor n/n-1 gives an unbiased estimator of the population variance. The above factor is called bessel’s correction.