**Qus1: Write the Gaussian distribution empirical formula.**

Empirical formula says that a random variable x belong to Gaussian distribution with some mean and std.

(Random Variable) X ≈ GD(mean, std)

Pr ( population mean – 1 population std <= X <= population mean + 1population std ) ≈ 68%

Pr ( population mean – 2 population std <= X <= population mean + 2population std ) ≈ 95%

Pr ( population mean – 3 population std <= X <= population mean + 3population std ) ≈ 99.7%

**Qus2 : What is the Z-score, and why is it important?**

z-score = [mean-x\_i]/std

a)Z-sore is use to remove the outlier from data

b)Z-score is use to compare the two result

**Qus3: What is an outlier, exactly?**

Outlier means data that’s completely different from other data.

Exp- Suppose you have a Age column where most of data is in the range of 18 to 54 but only 3 points are above 100 then it will be treated as outlier.

But it does not mean that we always remove the outlier because some time outlier is more important (detecting the fraud)

**Qus4: What are our options for dealing with outliers in our dataset?**

As discussed in 3rd question first we need to understand our data and before removing the outlier we always make sure that dataset is not imbalance.

Step 1: We should use Box plot to identify the outlier

Step 2: Apply 5 number summary

Step 3: Found the lower fence and upper fence

Step 4: Take only those values which are in the range of lower and upper fence.

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

Variance:

Population Variance: σ²= Σ(x- µ)²/N

Sample Variance: v= (Σ(x- µ)²) / (n-1)

As we take the sample randomly so sometime we take sample from only one area and then it will become bias then in order to make it unbias we divide it by n-1.