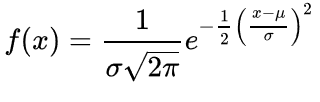
Assignment 3

1. Write the Gaussian Distribution empirical formula.

Ans. The general form of its probability density function is;

[](https://res.cloudinary.com/practicaldev/image/fetch/s--EqritkI4--/c_limit%2Cf_auto%2Cfl_progressive%2Cq_auto%2Cw_880/https:/dev-to-uploads.s3.amazonaws.com/i/depvdkp5i0r8h25gucr2.png)

Here the parameter µ is the mean or expectation of the distribution and σ is the standard deviation.

• The first empirical formula is given as

**P(µ-σ≤X≤µ+σ)≈68%**.

i.e.,68% of the total distribution will be present in the first standard deviation range.

• The second empirical formula is given as

**P(µ-2σ≤X≤µ+2σ)≈95%**.

This means that 95% of the total distribution will be present in the second standard deviation range.

• The third empirical formula is given as

**P(µ-3σ≤X≤µ+3σ)≈99.7%**.

This means that 99.7% of the total distribution will be present in the third standard deviation range.

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

Ans. The standard **score** (more commonly referred to as a z-score) is a very useful statistic because it (a) allows us to calculate the probability of a score occurring within our normal distribution and (b) enables us to compare two scores that are from different normal distributions.

3. What is an outlier, exactly?

Ans. An outlier is an observation that lies an abnormal distance from other values in a random sample from a population.

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

Ans. There are four ways to identify outliers:

* Sorting method
* Data visualization method
* Statistical tests (z scores)
* Interquartile range method

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

Ans. The formula to calculate population variance is:

σ2 = Σ (xi – μ)2 / N

where:

Σ: A symbol that means “sum”

μ: Population mean

xi: The ith element from the population

N: Population size

The formula to calculate sample variance is:

s2 = Σ (xi – x)2 / (n-1)

where:

x: Sample mean

xi: The ith element from the sample

n: Sample size

Bessel Correction

**Bessel'**s Correction is a correction applied while calculating the sample variance and sample standard deviation where the denominator is (N-1) instead of N, where N is the sample size or the number of observations in the sample.