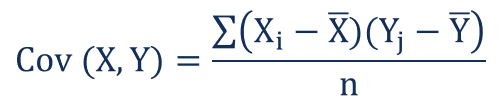
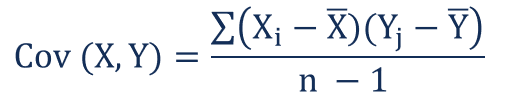
Assignment 4

1. What is the definition of covariance? Create the formula for it.

Ans. Covariance is a measure of the relationship between two random variables and to what extent, they change together. Or we can say, in other words, it defines the changes between the two variables, such that change in one variable is equal to change in another variable.



For a sample covariance, the formula is slightly adjusted:



Where:

* **Xi**– the values of the X-variable
* **Yj**– the values of the Y-variable
* **X̄**– the mean (average) of the X-variable
* **Ȳ** – the mean (average) of the Y-variable
* **n** – the number of data points

2. What makes Correlations better than Covariance?

Ans. Correlation is better than covariance for these reasons: 1 — Because correlation removes the effect of the variance of the variables, it provides a standardized, absolute measure of the strength of the relationship, bounded by -1.0 and 1.0.

3. Explain the process as well as Pearson and Spearman Correlation.

Ans. Pearson Correlation Coefficient(PCC):

Pearson Correlation is the coefficient that measures the degree of relationship between two random variables. The coefficient value ranges between +1 to -1. Pearson correlation is the normalization of covariance by the standard deviation of each random variable.



Speraman Rank Correlation Coefficient(SRCC):

SRCC is a test that is used to measure the degree of association between two variables by assigning ranks to the value of each random variable and computing PCC out of it.



4. What are the advantages of Spearman Correlation over Pearson Correlation?

Ans. Spearman’s correlation is more robust to outliers than Pearson’s correlation. If S >> P or S << P, that means the correlation is monotonic but not linear.

5. Describe the Central Limit Theorem.

Ans. Central limit theorem is a statistical theory which states that when the large sample size is having a finite variance, the samples will be normally distributed and the mean of samples will be approximately equal to the mean of the whole population.