IC252: Lab 7

Preamble: This assignment focuses on sample covariance and sample correlation of given data sets. For datasets, these quantities are defined as follows:

Dataset notation: $X: (x_1, x_2, \cdots, x_n)$, $Y: (y_1, y_2, \cdots, y_n)$

Sample mean: $\hat{E}(X) = \frac{1}{n} \sum_{i=1}^{n} x_n$

Sample variance: $\hat{Var}(X) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{E}(X))^2$

Sample covariance: $\hat{Cov}(X,Y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{E}(X))(y_i - \hat{E}(Y))$

Sample standard deviation: $\hat{SD}(X) = \sqrt{\hat{Var}(X)}$

Sample correlation: $\frac{\hat{Cov}(X,Y)}{\hat{SD}(X)\,\hat{SD}(Y)}$

Q1) Calculate the covariance between the variables X and Y.

X	15	17	20	21	25
Y	9	13	16	18	21

Q2) Calculate correlation between the variables X and Y.

X	15	17	20	21	25
Y	9	13	16	18	21

Q3) Broadband Network Gateway (BNG) device is the access point for subscribers, through which they connect to the broadband network. When a connection is established between BNG device and Customer Premise Equipment (CPE), the subscriber can access the broadband services provided by the Network Service Provider (NSP) or Internet Service Provider (ISP). A network management company monitors the performance of each of the BNG devices to provide better services. The company records different performance measures (PMs) for every 15 minutes. Recording of the different PMs for one of the BNG devices is in BNG-Device.csv. This dataset contain the values recorded for 4 different PMs, such as *Active-Count*, *CPU-Utilization*, *Total-Memory-Usage*, and *Average-Temperature* between 9th August 2018 and 6th September 2018. Table 1 provides a description of

different columns i.e. attributes (PMs) in this dataset. Assume that the observations are independent of the *Dates*.

Table 1: Description of attributes in the dataset.

Attribute Names	**Attribute Descriptions**
Dates	Date and time of the recordings
Active-Count	Number of active subscribers connected to the device
CPU-Utilization	% of usage of processor in the device
Total-Memory-Usage	Total % of memory used in the device
Average-Temperature	Average of temperatures recorded from the different slots in the device

Find the Pearson's correlation coefficient, *r* between (i) *Active-Count* & *CPU-Utilization*, (ii) *CPU-Utilization* & *Total-Memory-Usage*, (iii) *CPU-Utilization* & *Average-Temperature*, (iv) *Active-Count* & *Average-Temperature* and (v) *Total-Memory-Usage* & *Average-Temperature* for each of the BNG devices using the equation (1) and display the relationship based upon the Table 2.

$$r = \frac{\hat{Cov}(X,Y)}{\hat{SD}(X)\,\hat{SD}(Y)}$$

Table 2: Interpretation of values of Pearson's correlation coefficient *r* in the range [-1, +1]

r* range for a direct relationship between *X* and *Y	***r* range for a direct relationship between *X* and *Y***	**Relationship between *X* and *Y***
0.0	0.0	None
(0.0, 0.1]	(-0.0, -0.1]	Weak
(0.1, 0.3]	(-0.1, -0.3]	Moderate
(0.3, 0.5]	(-0.3, -0.5]	Strong
(0.5, 1.0)	(-0.5, -1.0)	Very Strong
1.0	-1.0	Perfect