**Statistics Assignment**

**Descriptive Statistics**

Descriptive Statistics are methods used to summarize and describe the main features of a dataset.

Here are some descriptive statistics to summarize our dataset.

**Iris-Setosa**

The general measurements of the Iris-Setosa in our dataset were determined by calculating the mean values for sepal length (5.006), sepal width (3.418), petal length (1.464), and petal width (0.244). These mean values provide an understanding of the average size and proportions of the Iris-Setosa in our analysis."

The median of the Iris-Setosa are sepal length (5), sepal width (3.4), petal length (1.5), and petal width (0.2). These median values represent the middle value of the data.

The Mode of the Iris-Setosa are sepal length (5.1), sepal width (3.4), petal length (1.5), and petal width (0.2). Mode represent the most repeated value in the data set.

The Standard Deviation of the Iris- Setosa are sepal length (0.352489687), sepal width (0.381024398), petal length (0.173511159), and petal width (0.107209503). These Standard Deviation values tells us how much the data points vary from the average. The standard deviation in dataset is lower than the mean, it suggests that the data points are relatively close to the average. This indicates less variability or spread in the dataset, and the values tend to cluster around the mean.

The Variance of the Iris- Setosa sepal length (0.12424898), sepal width (0.145179592), petal length (0.030106122), and petal width (0.011493878). It represents the average squared deviation of each data points from the mean. The variance values in dataset are less than the mean, it means that the data points are relatively close to the average. This suggests that there is less variability or spread in the dataset.

**Iris-versicolor**

The general measurements of the Iris-Versicolor in our dataset were determined by calculating the mean values for sepal length (5.843333333), sepal width (3.094), petal length (2.862), and petal width (0.785). These mean values provide an understanding of the average size and proportions of the Iris-Versicolor in our analysis."

The median of the Iris-Versicolor are sepal length (5.8), sepal width (3.05), petal length (2.45), and petal width (0.8). These median values represent the middle value of the data.

The Mode of the Iris-Versicolor are sepal length (5), sepal width (3), petal length (1.5), and petal width (0.2). Mode represent the most repeated value in the data set.

The Standard Deviation of the Iris- Versicolor are sepal length (0.828066128), sepal width (0.476057038), petal length (1.448564598), and petal width (0.566287752). These Standard Deviation values tells us how much the data points vary from the average. The standard deviation in dataset is lower than the mean, it suggests that the data points are relatively close to the average. This indicates less variability or spread in the dataset, and the values tend to cluster around the mean

The Variance of the Iris-Versicolor sepal length (0.685693512), sepal width (0.226630303), petal length (2.098339394), and petal width (0.320681818). It represents the average squared deviation of each data points from the mean. The variance values of sepal length, sepal width and petal within dataset are less than the mean, it means that the data points are relatively close to the average. This suggests that there is less variability or spread in the dataset. But the variance of petal length is greater than mean the greater variance indicates a wider range of petal lengths in dataset. This variability could signify the presence of different subgroups or distinct patterns within the data

**Iris-virginica**

The general measurements of the Iris-Virginica in our dataset were determined by calculating the mean values for sepal length (5.843333333), sepal width (3.054), petal length (3.758666667), and petal width (1.198666667). These mean values provide an understanding of the average size and proportions of the Iris-Virginica in our analysis."

The median of the Iris-Virginica are sepal length (5.8), sepal width (3), petal length (4.35), and petal width (1.3). These median values represent the middle value of the data.

The Mode of the Iris-Virginica are sepal length (5), sepal width (3), petal length (1.5), and petal width (0.2). Mode represent the most repeated value in the data set.

The Standard Deviation of the Iris- Virginica are sepal length (0.828066128), sepal width (0.433594311), petal length (1.76442042), and petal width (0.763160742). These Standard Deviation values tells us how much the data points vary from the average. The standard deviation in dataset is lower than the mean, it suggests that the data points are relatively close to the average. This indicates less variability or spread in the dataset, and the values tend to cluster around the mean

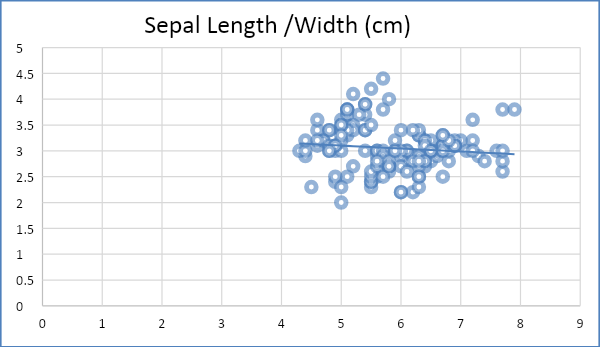
The Variance of the Iris-Virginica sepal length (0.685693512), sepal width (0.188004027), petal length (3.113179418), and petal width (0.582414318). It represents the average squared deviation of each data points from the mean. The variance values in dataset are less than the mean, it means that the data points are relatively close to the average. This suggests that there is less variability or spread in the dataset.

**Correlation Analysis**

Correlation is a statistical measure that expresses the extent to which two variables are linearly related.

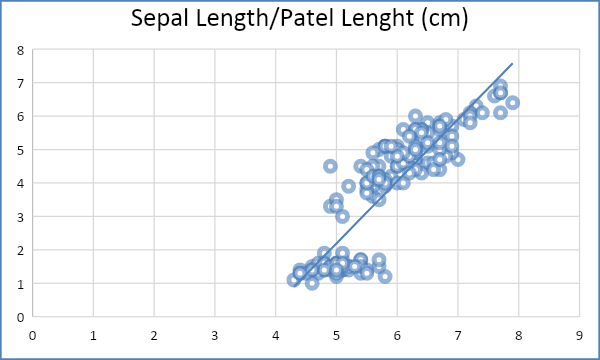
**Sepal Length Correlation with Sepal Width**

Based on the correlation value of **-0.10936925** between sepal length and sepal width, we can conclude that there is a weak negative relationship between these two variables. This means that as the sepal length increases, the sepal width tends to slightly decrease, but the relationship is not very strong.



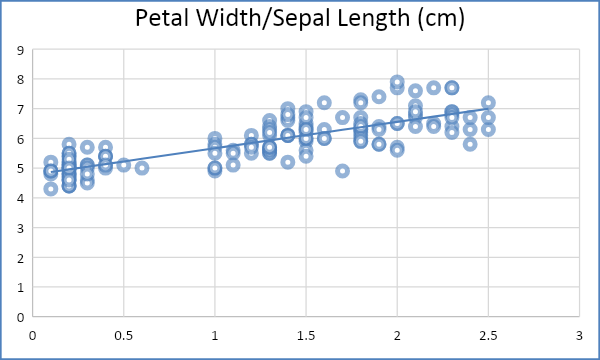
**Sepal Length Correlation with Patel Length**

The value of **0.871754157** between sepal length and Patel length, represent that there is a strong positive relationship between these two variables. This means that as the petal length increases, the sepal length tends to increase as well.



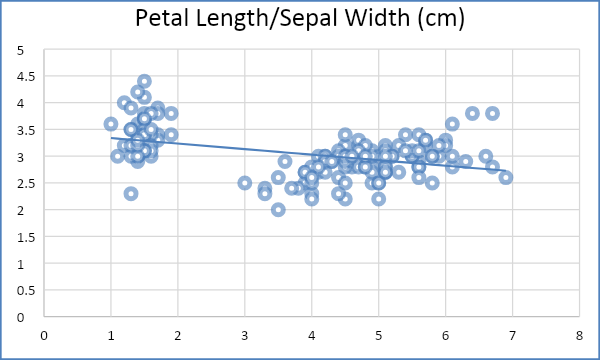
**Petal Width Correlation with Sepal Length**

The value of **0.817953633** between petal length and sepal length, represent that there is a strong positive relationship between these two variables. This means that as the petal length increases, the sepal length tends to increase as well.



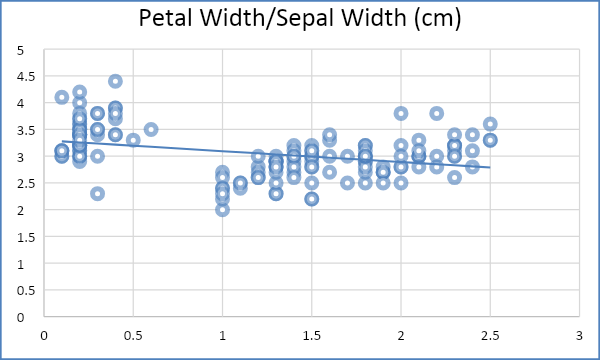
**Petal Length Correlation with Sepal Width**

The value of **-0.420516096** between petal length and sepal width suggests a moderate negative relationship between these two variables. This means that as the petal length increases, the sepal width tends to slightly decrease.



**Petal Width Correlation with Sepal Width**

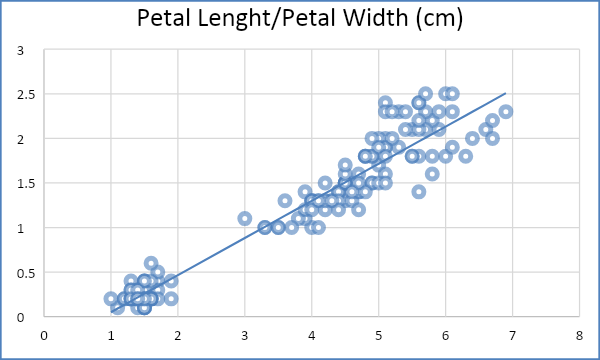
The value of -0.35654409 between petal width and sepal width suggests a moderate negative relationship between these two variables. This means that as the petal length increases, the sepal width tends to slightly decrease.



**Petal Length**

**Correlation with Petal Width**

The value of **0.962757097** between petal length and petal width, represent that there is a strong positive relationship between these two variables. This means that as the petal length increases, the petal width tends to increase as well.



**Hypothesis Testing**

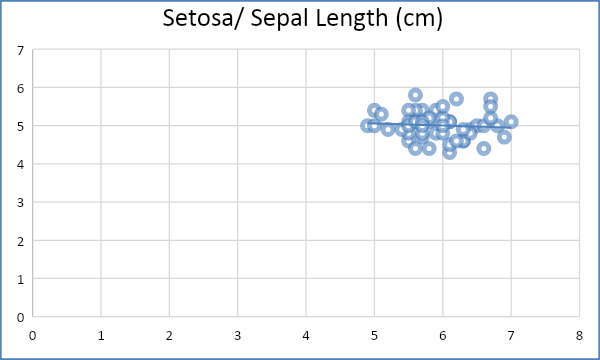
Hypothesis testing is a formal procedure for investigating our ideas and test specific predictions.

There are different statistical Tests to Test the Hypothesis. Here we are using T-Test with two tailed.

Ho = There is a significant difference in sepal length between the Iris-setosa and Iris- versicolor species.

H1 = There is no significant difference in sepal length between the Iris- setosa and Iris- versicolor species.

Based on the p-value (0.428733425), it appears that there is no significant difference in the mean sepal length between the Setosa and Versicolor species. This means that the observed difference in the means of sepal length could likely be due to random chance rather than a true difference between the two species.



Scatter-Chart represent the negative relationship.

**Regression Analysis**

Regression Analysis is a set of statistical processes for estimating the relationships between a dependent and one or more independent variable.

**Regression Equation**

**Y= a + bX**

b = Slope of the line

a = Y-intercept of the line

X = Values of the first data set

Y = Values of the second data set.

Here we are estimating the relationship between effect of sepal width on sepal length. We consider Sepal Length as an independent variable and Sepal Width as dependent variable.

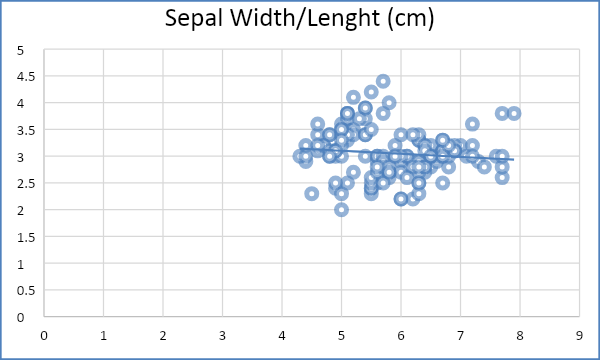
**Test the effect of sepal width on sepal length through a simple linear regression**

The R-squared value of 0.011961633 suggests that only a small portion of the variation in sepal length can be explained by the predictor variables.

The coefficient of -0.208870294 indicates that there is a negative relationship between sepal width and sepal length. However, the p-value for the coefficient (0.182765215) suggests that this relationship is not statistically significant at the conventional significance level of 0.05.

The p-value for the intercept (1.72623E-27) is very small, indicating that the intercept is statistically significant. This means that the estimated average sepal length at zero sepal width is significantly different from zero.

The significance F value of 0.182765215 suggests that the overall model is not statistically significant.



Based on these findings, it seems that sepal width does not have a significant impact on sepal length.