# **BAN210- Workshop 4**

| Total Mark: | 10 marks |
| --- | --- |
| Submission file(s): | * WS04.docx |

Please work in **groups** to complete this lab. This workshop is worth 5% of the total course grade. Please submit the submission file(s) through Blackboard. Only one person must submit for the group and only the last submission will be marked.

## **Part I: Linear Regression**

1. Start a new **project** named *W04\_Regression*.
2. Create a **diagram** and name it as *Regression*.
3. Select the **Sample** tab and find the **File Import** node. Drag and drop the File Import node to the diagram. In property panel, under Train, select the **Import File** item and click on the properties indicated by the three dots. Choose W04\_2D.xlsx file.
4. Right click on the File Import node and click on **Edit Variables**.

* Click on Y. Choose **Target** as the Role for Y. Then click **Explore**.
* Paste the frequency plot of Y here. Chart, histogram

  Description automatically generated
* How many rows were imported?

50

* Close the explore window. Then click on X and choose Explore. Paste the frequency plot of X here.

Chart, histogram

Description automatically generated

* Is the distribution of X suitable for a linear regression model?

No, the distribution of X is not suitable for a linear regression model. It requires transformation.

* Press OK.

1. Click on the **Modify** Tab and drag the **Transform** node into the diagram and connect to the previous node. Then in the property panel, select the ellipsis button associated with variables.
2. Select the X variable. Then under the **Method** column, select **Log** from the drop-down menu and then OK. Select Run and examine the results.

* What is the formula selected?

log(X + 1)

1. Close the output window of the Transform node and click the ellipsis button associated with Exported Data in the property panel of the Transform node. Highlight the Train line in the table and select Explore.
2. In the window that appears, select Actions > Plot > Histogram > Next, the variable LOG\_X, and Role of X. Select Finish.

* Paste the plot here.

Chart, histogram

Description automatically generated

1. Close the window, and press OK.
2. Drag the **MultiPlot** node from the Explore tab and connect to the Transform node. From the property panel, choose **scatter** plot as **Type of Charts** and Run.

* Paste the graph here.

Chart, scatter chart

Description automatically generated

1. Close the window.
2. Drag the **Regression** node from **Model** tab and connect to the **Transform** node.
3. In property panel, set the Regression Type to **Linear Regression**. Then Run and inspect the results.

* Paste the effects plot here.

Chart, bar chart

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

* What is the coefficient for Log\_X?

10.90688

* What is the intercept?

-0.59349

* Based on this model, what is the predicted Y for X= 10

Our formula for Log\_X is ln(X+1) and Y = w0 + w1 \* Log\_X

Y = -0.59349 + 10.90688 \* ln(X+1)

Y = 25.56

1. Close the results window.

## **Part II: Multivariate Regression**

1. Similar to step 3 above, select the **Sample** tab and find the **File Import** node. Drag and drop the File Import node to the diagram. In property panel, under Train, select the **Import File** item and click on the properties indicated by the three dots. This time, choose W04\_3D.xlsx file.
2. Right click on the File Import node and click on **Edit Variables**.

* Click on Y. Choose **Target** as the Role for Y.

1. Click on the **Modify** Tab and drag the **Transform** node into the diagram and connect to the previous node. Then in the property panel, select the ellipsis button associated with variables.
2. Select the X1 variable. Then under the Method column, select **Log** for X1 from the drop-down menu. Then select X2 and choose **Max. Normal** as Method for X2 and then OK. Select Run and examine the results.

* What are the formulas selected?

1. Close window. Then drag the **MultiPlot** node from the Explore tab and connect to the Transform node. Choose **scatter** plot as **Type of Charts** and Run.

* Paste the graphs (for LOG\_X1 and X2) here.

1. Close the window. Then drag the **Regression** node from **Model** tab and connect to the **Transform** node.
2. In property panel, set the Regression Type to **Linear Regression** in properties. Then Run and inspect the results.

* Paste the effects plot here.
* What is the formula for the model?
* Based on this model, what is the predicted Y for X1= 10 and X2=5000?

1. Paste the diagram here.

## **Part III: Regression assessment**

1. Compare the two regression models above. Fill in the following table by checking the output window in the results of each.

|  |  |  |
| --- | --- | --- |
|  | R squared | Average squared error |
| First regression model |  |  |
| Second regression model |  |  |

* What do the above numbers mean? Is it fair to compare them if they are using different data? Does a better fit on training data result in a better fit on validation or test data?

## **Part IV: Group work**

1. Add this declaration to your file:

We, ------------ (mention your names), declare that the attached assignment is our own work in accordance with the Seneca Academic Policy. We have not copied any part of this assignment, manually or electronically, from any other source including web sites, unless specified as references. We have not distributed our work to other students.

1. Specify what each member has done towards the completion of this work:

|  | Name | Task(s) |
| --- | --- | --- |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |