

SEIS 632 Data Analytics and Visualization
Assignment 2
Due: midnight 9/30/21

Predictive Modeling Using Regression

- Return to the Organics diagram. Attach the StatExplore tool to the **ORGANICS** data source and run it.
- In preparation for regression, you will see there are missing values that need to be imputed.

Add an **Impute** node to the diagram and connect it to the **Data Partition** node. Set the node to impute the letter **U** for unknown class variable values (**Default Input Method** ⇒ **Default Constant Value**, and type **U** for the Default Character Value) and the overall mean for unknown interval variable values. Create imputation indicators for all imputed inputs.

- Add a **Regression** node to the diagram and connect it to the **Impute** node.
- Choose **Stepwise** as the **Selection Model** and the **Validation Error** as the **Selection Criterion**.
- Run the Regression node and view the results.

Question 1: Which variables are included in the final model?

Question 2: What is the validation ASE?

- Explore the data and have a look at the skewness of different variables.
- Disconnect the **Impute** node from the **Data Partition** node.
- Add a **Transform Variables** node to the diagram and connect it to the **Data Partition** node.
- Connect the **Transform Variables** node to the **Impute** node.
- Apply a log transformation to the **DemAffl** and **PromTime** inputs.
- Run the **Transform Variables** node. Explore the exported training data. Did the transformations result in less skewed distributions?
- Rerun the **Regression** node.

Question 3: Do the selected variables change?

Question 4: How about the validation ASE?

- Create a full second-degree polynomial model.

Question 5: How does the validation average squared error for the polynomial model compare to the original model?

Submission:

1) Answers to the questions asked above.

2) In the diagram, as a last node, add a Reporter node from the Utility tab. Change the Nodes property of the Reporter node to All. Now right click on the Reporter node and select Run. This will generate a pdf.

You should submit the above two files on canvas.