Practical #7: Anomaly Detection

Objective:

1. To consolidate the understanding of anomaly/outlier scores and anomaly detection methods.
2. Understand how to use the Filter node of SAS Enterprise Miner to filter out outliers.

Submission:

- *What to submit*: A document (.doc or PDF) contains the information (steps, diagram, results etc.) about the exercise you did for **all questions** (Q1, Q2, and Q3) of this practical.
- *Deadline of the submission*: 11:59PM (Adelaide Time), Tuesday of **Week 10**.
- *Submission link*: "**Submission Link of Prac #7"** in **Week 9 section** on Learnonline course site.
- *Marks*: Prac#7 (part of the ongoing assessment of the course) is worth 2% of the total marks of the course.

Answer the following questions:

The file named "**Prac#7-data.csv**" (given in Week 9 section of Learnonline course site) contains 20 data points (values).

1. Assume that the given data follows a normal distribution. Use the **3-sigma method** to detect if there are outliers in this dataset. **List** the outlier(s) and **present** the steps in your answer (e.g., how you calculated / what approach you used to detect the outlier(s)). (Hint: calculate the mean and standard deviation of the data set and decide the outlier using the individual value's z-score)

2. Within SAS Enterprise Miner, the Filter node can be used to remove outliers from a data set using the 3-Sigma method. **Import** the given dataset to SAS Enterprise Miner and learn to use the **Filter node** to identify and remove outlier(s) from the given dataset with the 3-sigma method. Please include SAS screenshots in your answer (Hint: review "Help" contents about the Filter node for the meaning of the various options available).

3. To apply the 3-Sigma method to remove outlier(s) from the given dataset using the Filter node, how would you set up the Filter node? Please include the SAS screenshot in your answer. (Hint: check the setting of the property pane of the Filter node)