CHE384 Data to Decisions
Chris Mack, University of Texas at Austin

Homework #3 – Moment Testing and Outliers

Notes:
- Please name the file using this format: HW3_yourname.xlsx
- Please email the finished spreadsheet to: chris@lithoguru.com

1. Using the Newcomb data set from Data_Sets_1.xlsx, perform skewness and kurtosis tests on the data. What can you conclude?

2. Using the Newcomb data set from Data_Sets_1.xlsx, perform IQR outlier labeling, then perform the Dixon Q-test and the Grubbs' T test on the most extreme data point. What can you conclude?

3. Using the Newcomb data set from Data_Sets_1.xlsx, perform the Grubbs' test on the two most extreme data points. What can you conclude?

4. Remove the data points from the Newcomb data set that you believe are outliers, based on the above tests. Repeat the skewness and kurtosis tests on the reduced data. What can you conclude?

5. Using the Pi Digits data set from Data_Sets_1.xlsx, perform skewness and kurtosis tests on the data. What can you conclude?

6. The Grubbs T-statistic (T) studentizes the suspected outlier using the sample mean ($\bar{x}$) and the sample standard deviation (s) of the entire data set, including the suspected outlier. An alternative statstic (T') studentizes the outlier using the sample mean ($\bar{x}'$) and the sample standard deviation (s') that exclude the suspected outlier data point.

$$T = \frac{|x_{outlier} - \bar{x}|}{s}, \qquad T' = \frac{|x_{outlier} - \bar{x}'|}{s'}$$

a) Show that these two statistics are related by

$$\frac{(n-1)^2}{nT^2} - \frac{n(n-2)}{(n-1)T'^2} = 1$$

b) What are the implications of this result? Are there any advantages of using one statistic over the other?