

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

Homework #4 – Influence and Heteroscedasticity

Turn in your solution with the answers to the questions below. Also, email to me the supporting spreadsheet and/or R script that you used to perform the analysis. (Please name the file using this format: HW4_yourname.xlsx or HW4_yourname.R).

1. Using the Sand data set from Data_Sets_2.xlsx, generate a straight-line linear regression fit to the data. Calculate the leverage, internally studentized residual, externally studentized residual, and the Cook's Distance for each data point. Generate both the Williams graph and a plot of Cook's Distance versus x . What can you conclude?
2. Using the Heat Capacity data set from Data_Sets_2.xlsx, generate a straight-line linear regression fit to the data. Calculate the leverage, internally studentized residual, externally studentized residual, and the Cook's Distance for each data point. Generate both the Williams graph and a plot of Cook's Distance versus x . What can you conclude?
3. Using the Blood Pressure data set from Data_Sets_2.xlsx, generate a straight-line linear regression fit to the Pressure versus Age data. Calculate the leverage, internally studentized residual, externally studentized residual, and the Cook's Distance for each data point. Generate both the Williams graph and a plot of Cook's Distance versus x . What can you conclude?
4. Using the Partisan data set from Data_Sets_2.xlsx, generate a straight-line linear regression fit to the % Party Votes versus Year data. Calculate the predicted y -value and externally studentized residual for each data point. Split the data into two halves (early years and later years) and perform a Bartlett test for constant variance. What can you conclude?