

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

Homework #7 – Multiple Regression

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: HW7_yourname.xlsx or HW7_yourname.R).

1. Using the Body Fat data set, you wish to create a model to predict % Body Fat using one or more of the measured patient attributes as the predictor variable(s).
 - a) Using the correlation matrix, pick the first variable to add to your model (the one with the highest linear correlation coefficient). Build a model.
 - b) Pick a second variable to add as a second regressor. What variable did you pick and why? Build a two-regressor model.
 - c) Compare the two models using various appropriate measures of model goodness (likelihood ratio test, partial F-test, Bayesian information criterion). What can you conclude?
2. Using the Body Fat data set, create a model to predict % Body Fat using the following regressors to first order (no interaction terms): Chest, Abdomen, Hip, Thigh, and Knee.
 - a) Are all the model parameters statistically significant at the 95% level ($\alpha = 0.05$)?
 - b) Calculate the variance inflation factors (VIF) for each regressor in this model. Do you detect any significant multicollinearity?
 - c) What is the condition number for the correlation matrix of these variables? Does this number indicate significant multicollinearity?