PH245
Introduction to Multivariate Statistics
Homework Set 2

Due date: October 28, Monday

Problems:

1. Percentage of body fat, age, weight, height, and ten body circumference measurements (e.g., abdomen) were recorded for 252 men. Body fat, a measure of health, is estimated through an underwater weighing technique. Fitting body fat to the other measurements using multiple regression provides a convenient way of estimating body fat for men using only a scale and a measuring tape. The data can be found in "Data-HW2-Bodyfat.txt", along with a read-me file "Data-HW2-Bodyfat-Readme.txt". Remove the two outliers as we discussed in class.

   (a) Fit a linear regression model with *percent body fat* using Siri's equation as the response, *age, weight, height*, and the ten body circumference measurements as the predictors. Present the summary of the linear regression fit.

   (b) Interpret the coefficient associated with the predictor, *age*. If one wishes to test the null hypothesis that this coefficient equals zero, what is the p-value of this test? If the significance level is set at 0.05, what is your conclusion of this hypothesis test?

   (c) Repeat the above questions for the coefficient associated with the predictor, *abdomen*.

   (d) Draw a residual plot, with the fitted values on the x-axis, and the residuals on the y-axis. Does the plot suggest any violation of the key assumptions of the linear model? What are those key assumptions?

   (e) Compare the model you fitted with *age, weight, height*, and the ten body circumference measurements as the predictors, to the model we discussed in class with only *age, weight, height* as the predictors, in terms of adjusted R^2 and the significant predictors each model found.

   (f) Test the null hypothesis that the reduced model is preferred *versus* the alternative hypothesis that the full model is preferred, given the data. Use the significance level 0.05.

(g) Draw a scatter plot matrix of the ten body circumference measurements. Please summarize what you see from the plot.

(h) Draw a plot of the Lasso solution path for the regression on *age, weight, height*, and the ten body circumference measurements.

Policy: You must do the homework on your own. Please ask the Instructor or the GSIs if you have any question.