FALL 2023 MATH 484/564 HOMEWORK #4

- Due: October 30, Monday, 11:59PM
- How to submit: via Blackboard. If you have multiple files, upload a zipped file
- Submission link will disappear after 48 hours
- Homework solution is not required to be typed, but must be legible.

Problem 1 Exercise 4.8 from the TEXT.

Problem 2 For the body fat dataset, we wish to examine whether there are outliers. Use r_i, the internally studentized residual, and t_i, the externally studentized residual, what conclusion can you draw, respectively?

Problem 3 In the copier maintenance dataset (Copier.txt), X_1 (the second column) is the number of copiers serviced and Y (the first column) is the total number of minutes spent by the service person. X_2 (in Model.txt) is the binary predictor variable that indicates whether the copier model is small or large. It is coded as $X_2 = 1$ if small model is used and $X_2 = 0$ if large model is used.

1) Fit the regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$ and provide the estimated regression function. This is the pooled model.

2) Fit the regression model $y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$ and provide the estimated regression function for two groups with $X_2 = 1$ and $X_2 = 0$, respectively.

3) Show the scattered plot along with the regression lines for the small model ($X_2 = 1$) and the large model ($X_2 = 0$) by using the pooled model in 1); then show the scatter plot along with the regression lines using the separate model in 2). What conclusion can you draw from the two plots?

4) Obtain the residuals from 1) and plot them agains $x_1 x_2$. Is there any indication that an interaction term in the regression model would be helpful?

5) Fit the regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \epsilon_i$ and provide the estimated regression function.

6) Based on 5), test whether the interaction term can be dropped from the model: let the significance level $\alpha = 10\%$, state the null, alternative hypotheses, the decision rule and your conclusion. What is the P-value of the test? If the interaction term cannot be dropped from the model, describe the nature of the interaction effect.

Problem 4 Consider the data in Table 5.7 of the textbook. Use TEST as X, RACE as Z, and JPERF as response variable Y. X is quantitative predictor variable, Z is a binary indicator variable. Instead of using all 20 data points, you would use 18 data points including the first 9 data points from each group.

1) In order to answer if the minority group ($Z = 1$) has the same coefficients as the non-minority group ($Z = 0$), first provide the pooled model, and separate model for each group, and then perform a F-test to draw your conclusion? Can you use a t-test to draw the conclusion?

2) If we know as a fact that the two groups have the same intercept, we are interested in testing if the slope of X is also the same. What would you do? Would you use a t-test or F-test?

Problem 5 Derive the variance-covariance matrix $Var(\hat{\beta}_{WLS})$ for the weighted least squares estimators when the true variance of the observation y_i is unknown but only the relative magnitude is known. Let $w_i = k\frac{1}{\sigma_i^2}$. The weight matrix W used in the weighted least square estimate is a diagonal matrix with the i^{th} element on the diagonal being w_i.