STAT 512 HW2

Reading: Ott & Longnecker 12.1-12.6, 13.4
See Canvas calendar for the due date.
32 points Total, 2 points per question unless otherwise noted.

A study investigated body fat of n = 20 (female) subjects. The amount of body fat was measured by a cumbersome and expensive procedure requiring immersion of the person in water. For each subject, the following information was recorded:

BodyFat (Y)
Triceps (X_1) = triceps skinfold thickness
Thigh (X_2) = thigh circumference
Midarm (X_3) = midarm circumference

The data is available from Canvas as "BodyFat.csv". This data is taken from "Applied Linear Statistical Models" by Neter, Kutner, Nachtsheim and Wasserman.

1. Calculate pairwise (Pearson) correlations between the 4 variables (BodyFat and each of the predictors). You should also briefly examine the pairwise scatterplots, but you do NOT need to include them in your assignment.
2. Fit the "full" model using BodyFat as the response and including all 3 predictors. Include the parameter estimate information ("Coefficients" table) and R^2 value for the full model in your assignment.

Questions 3 through 5 are based on the **"full"** model from question 2.

3. Based on the "full" model, give 95% confidence intervals for each of the four β's (intercept and three partial regression coefficients)
4. Based on the "full" model, test the null hypothesis that all three of the partial regression coefficients are simultaneously zero. In other words, test H_0: $\beta_1 = \beta_2 = \beta_3 = 0$. Give the F-statistic and p-value and make a conclusion about the test. (**4 pts**)
5. Based on the "full" model, test the null hypothesis that the partial regression coefficients for Thigh and Midarm are <u>simultaneously</u> zero. In other words, test H_0: $\beta_2 = 0$ AND $\beta_3 = 0$. Give a test statistic, p-value and conclusion. (**4 pts**)

6. Now we will sequentially eliminate any terms from the model that are not significant at the 0.05 level. Starting from the "full" model, eliminate the least significant predictor variable (highest p-value) and rerun the regression. Continue that process until all predictor variables are significant at the 0.05 level. Include the parameter estimate information ("Coefficients" table) and R^2 value for the final model in your assignment. (**4 pts**) **We will use this "final" model for the remaining questions.**
7. In the initial inspection of the pairwise correlations and plots (question 1) it appeared that there was a relationship between BodyFat and Thigh; however, Thigh was dropped from the multiple regression because it was not significant. Speculate about why this is the case.

Questions continue on next page….

Questions 8 through 10 are based on the **"final"** model from question 6.

8. Working from the "final" model, look at the residual plots, paying particular attention to the (A) plot of residuals versus fitted values and (B) qqplot of residuals. Discuss each of these plots and whether the regression assumptions appear to be satisfied. You do not need to include the graphs in your assignment, just discuss your findings and conclusions. (**4 pts**)
9. Consider a subject with Triceps = 20 and Midarm = 25. Working from the "final" model, give (A) predicted body fat for this subject, (B) 95% confidence interval for the mean BodyFat of subjects with the same values and (C) 95% prediction interval for the predicted BodyFat for a new subject with these values. (**4 pts**)
10. Working from the "final" model, identify the largest RStudent residual and do an outlier test for that value. Give the test statistic, unadjusted p-value and Bonferonni adjusted p-value. Based on the Bonferonni adjusted p-value, can we conclude this observation is an outlier? Note: The outlierTest() function from the car package can be used for this question, but may return an NA for the Bonferonni p-value. I still want the Bonferonni adjusted p-value! (**4 pts**)