

STAT 210

Applied Statistics and Data Analysis:

Homework 8

Due on Nov. 13/2022

Question 1

For this question we will use the data set **dataB** which has a response variable **res** and five covariates.

- (i) Do a exploratory analysis of this data set, including a scatterplot matrix and a graphical representation of the correlation matrix. Comment on your results.
 - (ii) Fit a complete model for **res** including all the other variables. Produce a summary table and interpret the t tests in the table. What is the p-value for the overall significance test for the regression?
 - (iii) Starting with the model fitted in section (ii), fit a minimal model using a backwards selection procedure with a critical α of 0.15.
 - (iv) Plot the standard diagnostic graphs for the model that you selected and comment on what you observe.
 - (v) Predict the **res** value for a subject with covariates $(\text{var1}, \text{var2}, \text{var3}, \text{var4}, \text{var5}) = (65, 100, 50, 0.02, 3)$. Add a confidence interval at level 98%.
 - (vi) Print an anova table for the final model and find the estimated variance of the errors. Describe explicitly the sampling distribution for the estimated parameters.
-

Question 2

The file **dataC** has information on two variables, **yvar** and **xvar**. We want to build a regression model for **yvar** as a function of **xvar**.

- (i) Fit a simple regression model for **yvar** in terms of **xvar**. Print the summary table and comment on the results. Draw a scatterplot and add the regression line. Comment.
- (ii) State clearly the assumptions on which the model is based and, using the standard diagnostic plots and any tests that are necessary, verify if these assumptions are valid in this case.
- (iii) Use the function **residualPlots** in the package **car**. This function was introduced in problem 2 of Problem List 8. The result of applying this function is twofold. On the one hand, graphs of residuals against fitted values and regressors are plotted, including (in blue) a quadratic term, and on the other hand, a couple of tests are performed and printed in the console. The first one tests whether a quadratic term in the regressor variable would be significant. Interpret the result you obtain.
- (iv) Add a quadratic term to the initial regression model. Print the summary table, and interpret the results. Draw the diagnostic plots and comment on them.
- (v) Write an equation for the final model. Do a scatter plot and add the initial regression line and the curve for the quadratic model that you fitted in (iv).