STAT 210
Applied Statistics and Data Analysis
Problem List 9
(Due on Week 10)

Problem 1

For this exercise we use the data set `ais` in library `DAAG`. We concentrate in six variables, `bmi`, `lbm`, `ssf`, `ht`, `wcc`, and `hc`.

(i) Use the function `scatterplotMatrix` in the package `car` to obtain a graph matrix for the variables. Use the `corrplot.mixed` function in the package `corrplot` to draw a plot of the correlation coefficients for the four variables. Use also the `ggcorr` function in the package `GGally`. Comment.

(ii) Fit a multiple regression model for `bmi` as a function of the other variables. Print the summary table and discuss the results.

(iii) Using a stepwise procedure, select a minimal adequate model.

(iv) Fit also models using the adjusted R^2 and AIC as criteria. Select a minimal adequate model out of all these procedures. Justify your answer.

(v) Draw the diagnostic plots for your final model and discuss them.

Problem 2

For this question use the data set `PL925FPr2.csv`.

(i) Read the data and plot `vary` as a function of `varx`. Fit a simple linear regression for `vary` as a function of `varx` and add the regression line to the plot. Comment. Obtain a summary for the regression and draw the diagnostic plots. Comment on the results.

(ii) Now you are going to use the Box-Cox transformations to improve the model. To simplify this problem, you have to choose between two transformations of the output variable `vary`, a square root or a logarithm. Use the function `boxcox` on the package `MASS` with the argument set to the model you fitted in (i). If the confidence interval in the graph includes zero, choose a logarithmic transformation for `vary`. If the confidence interval in the graph includes 0.5 then choose a square root transformation.

(iii) Fit a new model with the transformation that you choose in (ii). Obtain a summary of the new regression and compare with the previous one. Draw the diagnostic plots and compare with the previous results.

(iv) Write down the final model in terms of the original variables. What would be the predicted `vary` for a point with `varx` $= 45$? Draw a scatterplot of `vary` against `varx` and add the regression line for the first model and the curve you obtained with the second regression.

Problem 3

The data for this question is stored in the file `PL925FPr3.csv`. It has two variables, `yvar` and `xvar`. We want to look at `yvar` as a function of `xvar`.

(a) Draw a graph. Do these variables seem to be linearly related? Calculate the correlation.
(b) Fit a linear model and comment on the results. What is the R^2?
(c) Draw residual plots using the function `residualPlots` in the `car` package. Comment on your results.
(d) Fit a quadratic model to these data and comment on the results. What are the R^2 and adjusted R^2?
(e) Draw residual plots using the `car` function `residualPlots`. Comment on your results.
(f) Fit a cubic model to these data and comment on the results. What are the R^2 and adjusted R^2? Is this a better model than the quadratic model? Give reasons for your answer.
(g) Draw a scatterplot of `yvar` as a function of `xvar` and add the quadratic regression curve that you fitted.

Problem 4

Use the data set `PL925FPr4.csv` for this problem. This set has six variables, and we are interested in building a regression model for `vary` in terms of the other variables in the set.

(a) Start by using the function `scatterplotMatrix` in the `car` library to obtain a matrix of scatterplots for the variables in the data set. Calculate and plot the correlation matrix. Comment on what you observe.

(b) Fit a complete linear regression model and find a minimal adequate model using backward elimination with a critical p-value of 0.05.

(c) Find a minimal adequate model using AIC and the function `stepAIC` in the `MASS` package. Compare with the model in (b)

(d) Find a minimal adequate model maximizing the R^2. Compare with the models in (b) and (c).

(e) How many different models do you have? What would you do if you had to select one of them?