SMM634 Exercises - Multiple Linear Regression

1. The data available from the `faraway` library in R are averages over 1960-1970 (to remove business cycle or other short-term fluctuations). `dpi` is per-capita disposable income in U.S. dollars; `ddpi` is the percent rate of change in per capita disposable income; `sr` is aggregate personal saving divided by disposable income. The percentage population under 15 (`pop15`) and over 75 (`pop75`) are also recorded. The data are in dataframe called `savings` and come from Belsley, Kuh, and Welsch (1980).

   (a) Using R, fit a linear model where `sr` is the response variable and all the rest are predictors. Write down the fitted model.

   (b) Test the null hypothesis whether any of the predictors have significance in the model.

   (c) Test the null hypothesis that the coefficient associated with `pop15` is not significant in the full model. What conclusion do you reach?

   (d) Test the same null hypothesis as in (c) using an F-testing approach.

   (e) Can you relate the result of the F-test with that based on the t-test? Justify your answer.

   (f) Construct a stepwise approach first using the model fitted in (a) and then with a model containing only the intercept. What conclusions can be drawn?

   (g) Compare the results of the model chosen in (f) with those of model in (a). What changed?

   (h) Using the scatterplots below, what do you deduce? Are you happy with the model in (g)? Would you fit another model? If yes, which one would you fit? Are conclusions changing?