3rd exercises for SIM'2016

Ex. 1

Linear model. Download data `linear-model-data-1.dat` from the course webpage. Complete the following task by writing yourself the functions/procedures needed. At this point, do not use existing regression analysis packages that might be available in your computing platform.
a) Import data and plot. First column is x and second is y.
b) Form data matrix $\mathbf{X}$ for linear model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$.
c) Estimate coefficient vector $\hat{\boldsymbol{\beta}} = \boldsymbol{b}$. Use either Eq. (3.14) or (3.15).
d) Plot data with the estimated linear model.
e) Compute observed residuals $\boldsymbol{e}$ and residual variance s^2.
f) Compute standardized residuals (from Eq. (3.27)) and plot against x.
g) Form diagnostic parameter table from model variables as in page 3-13 in the lecture material. What can be deducted regarding the model variables from the parameter table?

Ex. 2

We have two sets of observations, `linear-model-data-2a.dat` and `linear-model-data-2b.dat` (on course webpage). We know that both should have dependency $y = \beta_0 + \beta_1 x^2$, but we do not know if the constant β_0 is the same for both groups. Study this and make decision based on joined linear model for both sets together, with categorical variable included to separate the sets, i.e. $y_i = \beta_0 + \beta_c g_i + \beta_1 x_i^2$, where $g_i = 0/1$ is the categorical variable marking the group, and β_c is it's coefficient.

Ex. 3

Load observations from `linear-model-data-3.dat`. First two columns are explanatory variables x_1, x_2, third is dependent variable y. Find suitable model between y and x_i. You can test functions of x_i's and interaction. Test at least two different models and choose best according to model selection criterion of your choice. Check if parameters in the model are significant and that residuals against predicted $\hat{y}$ seem unbiased and homoscedastic. You can use ready-made regression package or write your own procedures.