4th exercises for SIM'2021

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 the data with the estimated linear model.
e) Compute observed residuals $\boldsymbol{e}$ and residual variance s^2.

Ex. 2

Weighted linear model. Download data `weighted-linear-model-data.dat` from the course webpage. The data comes from a linear dependency between x (first column) and y (second column), but with non-constant variance. Your task is to guess the correct weighting scheme, and fit using these weights.
a) Form data matrix $\mathbf{X}$ for linear model $y_i = \beta_0 + \beta_1 x_i$. Estimate coefficients $\boldsymbol{b}$ using regular, non-weighted least squares. Plot the data with the estimated linear model. Compute residuals $e_i = y_i - (b_0 + b_1 x_i)$. Plot residuals e_i against x_i.
b) From a), it looks as the residuals are not homoscedastic, but their variance is rather of form $\mathrm{var}(e_i) = \sigma^2/w_i$. By multiplying the residuals with correct weights, $e_i^* = e_i\sqrt{w_i}$, the residual variance would become $\mathrm{var}(e_i^*) = w_i \mathrm{var}(e_i) = w_i \sigma^2/w_i = \sigma^2$. Try to guess the correct weights w_i. They are a simple function of x, so $w_i := \mathrm{f}(x_i)$.
With your guess for weights, estimate weighted model coefficients $\boldsymbol{b}_w$ using Eq. (3.25) from lecture notes. Plot the data with the estimated linear model. Compute residuals, multiply with square root of your proposed weights, and plot against x_i. Check if the scaled residuals are now homoscedastic.