HW 2

1. Investigate the bias of the regression coefficient estimators when the error variance σ^2 varies with x (the heteroscedasticity problem) for different sample sizes and different regression coefficients.

The general model will be the simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where ε_i is normally distributed with a mean of 0 and a variance equal to $e^{\gamma x_i}$, that is

$$\varepsilon_i \sim N(0, e^{\gamma x_i})$$

Set up a Monte Carlo experiment and investigate what factors affect the bias of the estimated coefficients. The factors to examine are:

Level of heterogeneity (γ = 0, 0.5, 1, 2); Regression slope (β_1 = 0, 0.5, 1, 2);
Sample sizes (n = 10, 25, 50, 100).
For each data set randomly sample the x-values from the N(0,1) distribution. These should be the steps:

a) Simulate n values of x from N(0,1)
b) Obtain the variances of the errors from the formula $e^{\gamma x_i}$
c) Simulate the ε_i from the corresponding normal distribution
d) Compute $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
e) Estimate $\hat{\beta}_0, \hat{\beta}_1$ and store them in output matrix
f) Repeat steps b)-e) R times
g) Obtain the average $\hat{\beta}_0, \hat{\beta}_1$ and compare to the true values to assess the bias
h) Repeat from part a) with different γ, β_1 and n.
i) Summarize your findings.

2. Consider the multiple regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, i = 1, \ldots, n$$

a) Generate datasets. Use n = 20, x_1 from U(0,1) and x_2 from U(0, 2) distribution. Generate the two sets of x's only once – they will remain fixed, but next we will generate many sets of y's. Now set $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 3$, and generate random errors from N(0, 1). Finally, generate the y's using the above equation. Repeat from the generation of the random errors to produce many sets of y's. Check your work with appropriate histograms.

b) Estimate the β's and σ in each dataset. Construct histograms of the distributions of $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, and $\hat{\sigma}^2$ and plot them on the same graph by splitting the screen 2x2.

c) Compute the means and standard deviations of your estimates. Are they close to what they should be per the theoretical formulas?

3. Rizzo Pr 3.11 on p. 95 (see the R code on pp. 78-79).