Problem Set 1 **AQM**
Linear Models

Problem 1 - Optimization

Consider the **advertising.csv** dataset discussed. There is one outcome variable (sales) and three predictor variables (TV, radio, newspaper) and a constant term corresponding to the intercept. Write an optimization routine (gradient descent, Newton-Raphson or any other optimization approach of your choice) in R that minimizes the negative of the log-likelihood for the following linear model with respect to the β's:

$$\text{Sales}_i = \beta_0 + \beta_1 \text{TV}_i + \beta_2 \text{Radio}_i + \beta_3 \text{Newspaper}_i + \varepsilon_i$$

Compare results with a blackbox optimization routine as well as directly employing the linear regression function.

Problem 2 - Interpretation

Load the **advertising.csv**, which provides the advertising data sales (in thousands of units) for a particular product's advertising budget (in thousands of dollars) for TV, radio, and newspaper media.

1. Run a simple linear regression (you can use the in-built function) on sales and newspaper advertising. Plot Sales vs Newspaper and overlay the predicted relationship. Is there a significant relationship?
2. Run a multiple linear regression on sales and all three advertising media. What about the relationship between sales and newspaper advertising now?
3. . Why would a significant predictor with a simple regression disappear with multiple predictors?

Problem 3 - Diagnostics

Address the following questions for a few different specifications (i.e., vary the variables used as covariates in your linear regression model, incorporate interaction terms and so forth).

1. In general, What are the assumptions made in the linear regression model with respect to the errors?
2. The assumptions can be tested by plotting the residuals e_i versus the predictor x_i. Ideally, how should the residual plot look like?
3. The normality of errors can be verified by plotting the quantile-quantile (Q-Q) plot. Obtain the Q-Q plot and the residual plots for the above linear regression models.

Problem 4 - Weighted Regression

One of the assumptions of the method of OLS assumes that there is constant variance in the errors, which is known as homoscedasticity. When this assumption is violated and the errors do not have constant variance (known as heteroscedasticity), the method of weighted least squares can be used, which fortunately we can solve by the same kind of linear algebra we used to solve the ordinary linear least squares problem. Again, consider the following model:

$$y = X\beta + \varepsilon.$$

We now assume ε to be a multivariate normal distribution with a mean vector of zero and a non-constant covariance matrix:

$$\varepsilon \sim \mathcal{N}\left(\vec{\mu}, \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}\right).$$

If we write W for the matrix with elements w_i on the diagonal and zeroes everywhere else, which correspond to the reciprocal of each variance, i.e., $w_i = \sigma_i^{-2}$, then:

$$W = \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_n^2} \end{pmatrix}.$$

We can now minimize, as we did in question 1, the weighted RSS:

$$WRSS = n^{-1}(y - Xb)^T W (y - Xb).$$

1. Please solve the minimization and derive the weighted least square estimators analytically.
2. Use this solution to fit a weighted least squares to the advertising data. How does the fit differ from your original OLS?
3. Why would we weight by the inverse of the variance $\frac{1}{\sigma_i^2}$?