1. (30) Regression through the origin: We will consider a special case of simple linear regression where the intercept is assumed to be zero from the outset. Let

$$Y_i = \beta x_i + \epsilon_i,$$

where $\mathrm{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = \sigma^2$

(a) Define $Q(\beta) = \sum_{i=1}^{n}(Y_i - \beta x_i)^2$. Show that the minimizer of $Q(\beta)$ is $\hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2}$.

(b) Show $\mathrm{E}(\hat{\beta}) = \beta$.

(c) Show $\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_i^2}$

(d) Write the model as $Y = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, defining each matrix/vector.

(e) Verify that $\hat{\beta} = (X'X)^{-1}X'Y$ is equivalent to the minimizer in part a.

(f) Show that $\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$ is equivalent to the scalar form in part c.

2. (35) Burple, Stephens, and Gloopshire (2014) report on a study in the Journal of Questionable Research. Data were collected on the number of minutes Y_i it took n = 237 Glippers to learn how to drive a small, motorized car. Two predictors of interest are the estimated age of the glipper in months x_{i1} and the Glippers Maladaptive Score (GMS) x_{i2}, a number from 50 to 100 that summarizes how poor the glipper's vision is. Consider the following multiple regression output from R.

Coefficients

	Estimate	Std. Error
(Intercept)	-182.57923	7.41169
Age	8.56069	0.31150
GMS	0.28066	0.04621

Analysis of Variance Table

	Df	Sum Sq	Mean Sq
Regression			
Error		23565	
Total	236	104682	

(a) Complete the ANOVA table.

(b) Calculate the F-statistic from the ANOVA table and use it to test

$$H_0 : \beta_1 = \beta_2 = 0.$$

What does this imply about β_1 and β_2?

(c) Report each of $\hat{\beta}_1$ and $\hat{\beta}_2$. Construct t-tests $H_0 : \beta_1 = 0$ and $H_0 : \beta_2 = 0$ individually. Can either predictor be dropped in the presence of the other?

(d) Interpret both estimated coefficients.

(e) Report R^2; how is it interpreted here?

3. (35) Consider the "mtcars" data in R. You can load and view the dataset by typing "mtcars" in R console. Consider the response variable mileage per hour (Y=mpg), and two predictors, horsepower and weight ($X_1 = hp$, $X_2 = wt$). Ignore other variables for now.

(a) Obtain and report the scatterplot matrix; what does it tell you about the relationship between mpg and each of the predictors, horsepower and weight?

(b) Fit the regression model $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$. Report the ANOVA table and the table of regression coefficients.

(c) Comment on the significance of the overall model.

(d) Comment of the significance of each predictor. Can either predictor be dropped in the presence of the other?

(e) Obtain the normal probability plot and a histogram of the residuals. What do these plots tell you?

(f) Obtain $SSR(x_1)$, $SSR(x_2|x_1)$, and verify $SSR(x_1, x_2) = SSR(x_1) + SSR(x_2|x_1)$.

(g) Obtain and interpret an 95% interval estimate of $\mathrm{E}(Y_h)$ when $x_{h1} = 100$ and $x_{h2} = 4$.