STAT 3008
Exercises 2

Problems refer to the problem sets in the textbook: Applied Linear Regression, 3rd edition by Weisberg.

1. Problem 2.1. For 2.1.3, no need to do t-tests.
   Height and weight data
   The table below and in the data file htwt.txt gives Ht = height in centimeters and Wt = weight in kilograms for a sample of n = 10 18-year-old girls. The data are taken from a larger study described in Problem 3.1. Interest is in predicting weight from height.

Ht	*Wt*
169.6	71.2
166.8	58.2
157.1	56.0
181.1	64.5
158.4	53.0
165.6	52.4
166.7	56.8
156.5	49.2
168.1	55.6
165.3	77.8

2.1.1. Draw a scatterplot of Wt on the vertical axis versus Ht on the horizontal axis. On the basis of this plot, does a simple linear regression model make sense for these data? Why or why not?

2.1.2. Show that $\bar{x} = 165.52$, $\bar{y} = 59.47$, SXX $= 472.076$, SYY $= 731.961$, and SXY $= 274.786$. Compute estimates of the slope and the intercept for the regression of Y on X. Draw the fitted line on your scatterplot.

2.1.3. Obtain the estimate of $\hat{\sigma}^2$ and find the estimated standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$. Also find the estimated covariance between $\hat{\beta}_0$ and $\hat{\beta}_1$. Compute the t -tests for the hypotheses that $\beta_0 = 0$ and that $\beta_1 = 0$ and find the appropriate p-values using two-sided tests.

2.1.4. Obtain the analysis of variance table and F-test for regression. Show numerically that $F = t^2$ where t was computed in Problem 2.1.3 for testing $\beta_1 = 0$.

2. Problem 2.3.
Deviations from the mean
Sometimes it is convenient to write the simple linear regression model in a different form that is a little easier to manipulate. Taking equation (2.1), and adding $\beta_1\bar{x} - \beta_1\bar{x}$, which equals zero, to the right-hand side, and combining terms, we can write

$$\begin{aligned} y_i &= \beta_0 + \beta_1\overline{x} + \beta_1 x_i - \beta_1\overline{x} + e_i \\ &= (\beta_0 + \beta_1\overline{x}) + \beta_1(x_i - \overline{x}) + e_i \\ &= \alpha + \beta_1(x_i - \overline{x}) + e_i \qquad (2.29) \end{aligned}$$

where we have defined $\alpha = \beta_0 + \beta_1\bar{x}$. This is called the deviations from the sample average form for simple regression.
2.3.1. What is the meaning of the parameter α?

2.3.2. Show that the least squares estimates are

$$\hat{\alpha} = \overline{y}, \qquad \hat{\beta}_1 \text{ as given by (2.5)}$$

2.3.3. Find expressions for the variances of the estimates and the covariance between them.

3. Problem 2.4.1.
2.4. Heights of mothers and daughters
2.4.1. For the heights data in the file heights.txt, compute the regression of Dheight on Mheight, and report the estimates, their standard errors, the value of the coefficient of determination, and the estimate of variance. Give the analysis of variance table that tests the

hypothesis that E(Dheight$\|$Mheight) $= \beta_0$ versus the alternative that E(Dheight$\|$Mheight) $= \beta_0 + \beta_1$ Mheight, and write a sentence or two that summarizes the results of these computations.

4. This problem shows the unbiasedness of the error estimate, $\hat{\sigma}^2$.

   a) Show that

   $$y_i - \bar{y} = \beta_1(x_i - \bar{x}) + e_i - \bar{e}\,,$$

   where $\bar{e} = \sum_{i=1}^n e_i/n$.

   b) Using a), show that

   $$\mathrm{E}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \beta_1^2\sum_{i=1}^{n}(x_i - \bar{x})^2 + (n-1)\sigma^2\,.$$

   c) Show that

   $$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 - \hat{\beta_1}^2\sum_{i=1}^{n}(x_i - \bar{x})^2\,.$$

   d) Using b),c) and the formula for $Var(\hat{\beta}_1)$, show that

   $$E(\hat{\sigma}^2) := E\left(\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}\right) = \sigma^2\,.$$