STAT 3008
Exercises 5

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

1. Problem 3.1.2 and 3.1.3.
   3.1. Berkeley Guidance Study The Berkeley Guidance Study enrolled children born in Berkeley, California, between January 1928 and June 1929, and then measured them periodically until age eighteen (Tuddenham and Snyder, 1954). The data we use is described in Table 3.6, and the data is given in the data files BGSgirls.txt for girls only, BGSboys.txt for boys only, and BGSall.txt for boys and girls combined. For this example, use only the data on the girls.

TABLE 3.6 Variable Definitions for the Berkeley Guidance Study in the Files `BGSgirls.txt`, `BGSboys.txt`, and `BGSall.txt`

Variable	Description
Sex	0 for males, 1 for females
WT2	Age 2 weight, kg
HT2	Age 2 height, cm
WT9	Age 9 weight, kg
HT9	Age 9 height, cm
LG9	Age 9 leg circumference, cm
ST9	Age 9 strength, kg
WT18	Age 18 weight, kg
HT18	Age 18 height, cm
LG18	Age 18 leg circumference, cm
ST18	Age 18 strength, kg
Soma	Somatotype, a scale from 1, very thin, to 7, obese, of body type

   3.1.2. Starting with the mean function $E(Soma|WT9) = \beta_0 + \beta_1 WT9$, use added-variable plots to explore adding LG9 to get the mean function $E(Soma|WT9, LG9) = \beta_0 + \beta_1 WT9 + \beta_2 LG9$. In particular, obtain the four plots equivalent to Figure 3.1, and summarize the information in the plots.
   3.1.3. Fit the multiple linear regression model with mean function $E(Soma|X) = \beta_0 + \beta_1 HT2 + \beta_2 WT2 + \beta_3 HT9 + \beta_4 WT9 + \beta_5 ST9$ Find

$\hat{\sigma}$, $R2$, the overall analysis of variance table and overall F-test. Compute the t -statistics to be used to test each of the β_j to be zero against two-sided alternatives. Explicitly state the hypotheses tested and the conclusions.
For problem 3.1.3, do also

- i A F-test for the dependence of $Soma$ on $HT9$, and compare with the t-test.
- ii A F-test for the dependence of $Soma$ on $HT9$ and $WT9$.
- iii At the point $HT2 = 50$, $WT2 = 15$, $HT9 = 100$, $WT9 = 30$, $ST9 = 10$, find the
  - a) 99% confidence interval for the fitted value of $Soma$.
  - b) 99% prediction interval for a new observation of $Soma$.
  - c) 99% confidence band for the fitted value of $Soma$.

2. Problem 3.4.
3.4. Suppose we have a regression in which we want to fit the mean function (3.1). Following the outline in Section 3.1, suppose that the two terms X1 and X2 have sample correlation equal to zero. This means that, if xij , i = 1, ... , n, and j = 1, 2 are the observed values of these two terms for the n cases in the data,$\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) = 0$
3.4.1. Give the formula for the slope of the regression for Y on X1, and for Y on X2. Give the value of the slope of the regression for X2 on X1.
3.4.2. Give formulas for the residuals for the regressions of Y on X1 and for X2 on X1. The plot of these two sets of residuals corresponds to the added-variable plot in Figure 3.1d.
3.4.3. Compute the slope of the regression corresponding to the added-variable plot for the regression of Y on X2 after X1, and show that this slope is exactly the same as the slope for the simple regression of Y on X2 ignoring X1. Also find the intercept for the added-variable plot.

3. Let $n = 100$, $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)' = (7, 1)'$, $\hat{\sigma} = 2$,

$$X'X = \begin{pmatrix} 3 & 2 \\ 2 & 5 \end{pmatrix}.$$

Find an inequality representing the 95% confidence ellipse for β. Stretch a graph for the ellipse. (Hints: Try to find some points that lie on the ellipse.)