Instructions: This homework assignment covers **Chapter 11** of the course notes. On each part, there is opportunity for partial credit, so show all of your work and explain all of your reasoning. No work/no explanation means no credit even if your answer is correct. If you use R to answer any part or to check your work, please include all code and output as attachments. Do not just write out the code you used.

1. A chemical engineer is studying the impact of temperature (x_1) and concentration (x_2) on the percentage of impurities Y for a chemical production process. The following measurements were obtained from $n = 14$ different batches:

Observation	Y	x_1	x_2
1	14.9	85.8	42.3
2	16.9	83.8	43.4
3	17.4	84.5	42.7
4	16.9	86.3	43.6
5	16.9	85.2	43.2
6	16.7	83.8	43.7
7	17.1	86.1	43.3
8	16.9	85.9	43.4
9	16.7	85.7	43.3
10	16.9	86.3	42.6
11	16.7	83.5	44.0
12	17.1	85.8	42.8
13	17.6	85.9	43.1
14	16.9	84.2	43.5

Consider the multiple linear regression model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \iff \underbrace{E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2}_{\text{a plane in } \mathbb{R}^3}$$

for the population of all batches for this process.

(a) Estimate the population model using the data above. What is your prediction equation?
(b) Find the ANOVA table for the regression fit in part (a).
(c) What two hypotheses are tested by the overall F test? Perform this test using $\alpha = 0.05$. What is your conclusion?
(d) Write 95% confidence intervals for β_1 and β_2 and interpret them.
(e) What if temperature (x_1) and concentration (x_2) interact with each other? This would occur if

- the relationship between the percentage of impurities Y and x_1 depended on x_2.
- the relationship between the percentage of impurities Y and x_2 depended on x_1.

Consider the multiple linear regression model which accounts for interaction

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon \iff \underbrace{E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2}_{\text{a curvilinear surface in } \mathbb{R}^3}$$

Estimate the interaction model. What is your prediction equation? Write a 95% confidence interval for β_3, the population regression parameter associated with the interaction. Do these data provide evidence that temperature (x_1) and concentration (x_2) interact with each other in the population of all batches?
(f) Which model would you select for these data: the no-interaction model or the interaction model? For the model you select, perform residual diagnostics to assess model assumptions. Do you detect any model violations?

2. The brake horsepower (`BHP`, Y) of an automobile engine is thought to be a function of the engine speed in revolutions per minute (`RPM`, x_1), the road octane number of the fuel (`OCT`, x_2), and the engine compression (`COM`, x_3). An experiment is run in a laboratory at twelve different times; on each run, the temperature (`TEMP`, x_4) is also recorded. The data from the experiment are below:

Y	x_1	x_2	x_3	x_4
225	2000	90	100	71.2
212	1800	94	95	70.3
229	2400	88	110	72.3
222	1900	91	96	69.9
219	1600	86	100	73.2
278	2500	96	110	70.0
246	3000	94	98	70.7
237	3200	90	100	70.8
233	2800	88	105	72.1
224	3400	86	97	71.8
223	1800	90	100	71.1
230	2500	89	104	70.6

Consider the multiple linear regression model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon \iff E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

for the population of all runs of this experiment.

(a) Estimate the population model using the data above. What is your prediction equation?
(b) Find the ANOVA table for the regression fit in part (a).
(c) Find two different sets of sequential sums of squares for the 4 independent variables and verify they each add to SS_R. How many sets of sequential sums of squares are there?
(d) Consider the following setting of the independent variables:

$$\mathbf{x}_0 = \begin{pmatrix} x_{10} \\ x_{20} \\ x_{30} \\ x_{40} \end{pmatrix} = \begin{pmatrix} 2500 \\ 90 \\ 100 \\ 72 \end{pmatrix}.$$

Write a 95% confidence interval for $E(Y)$ and a 95% prediction interval for Y^* at this setting of the independent variables. Carefully interpret each interval.
(e) Perform residual diagnostics to assess model assumptions. Do you detect any model violations?