	Q1	Q2	Q3	Q4	Q5
50 Points	10	10	10	10	10

Question 1

Question 1. Exercise 4.18 on page 167

4.18 Coteron, Sanchez, Martinez, and Aracil ("Optimization of the Synthesis of an Analogue of Jojoba Oil Using a Fully Central Composite Design," *Canadian Journal of Chemical Engineering*, 1993) studied the relationship of reaction temperature x_1, initial amount of catalyst x_2, and pressure x_3 on the yield of a synthetic analogue to jojoba oil. The following table summarizes the experimental results.

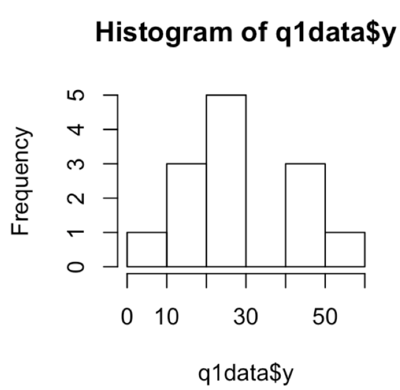
x_1	x_2	x_3	y
−1	−1	−1	17
1	−1	−1	44
−1	1	−1	19
1	1	−1	46
−1	−1	1	7
1	−1	1	55
−1	1	1	15
1	1	1	41
0	0	0	29
0	0	0	28.5
0	0	0	30
0	0	0	27
0	0	0	28

a. Fit a multiple regression of y vs. x1, x2, and x3, then perform a thorough model adequacy analysis, please include residual plots. *Note, please use lm(y ~ as.factor(x1) + as.factor(x2) + as.factor(x3)), not lm(y ~ x1 + x2 + x3) to fit your model.*
b. Perform the appropriate test for lack of fit.

Part (a)

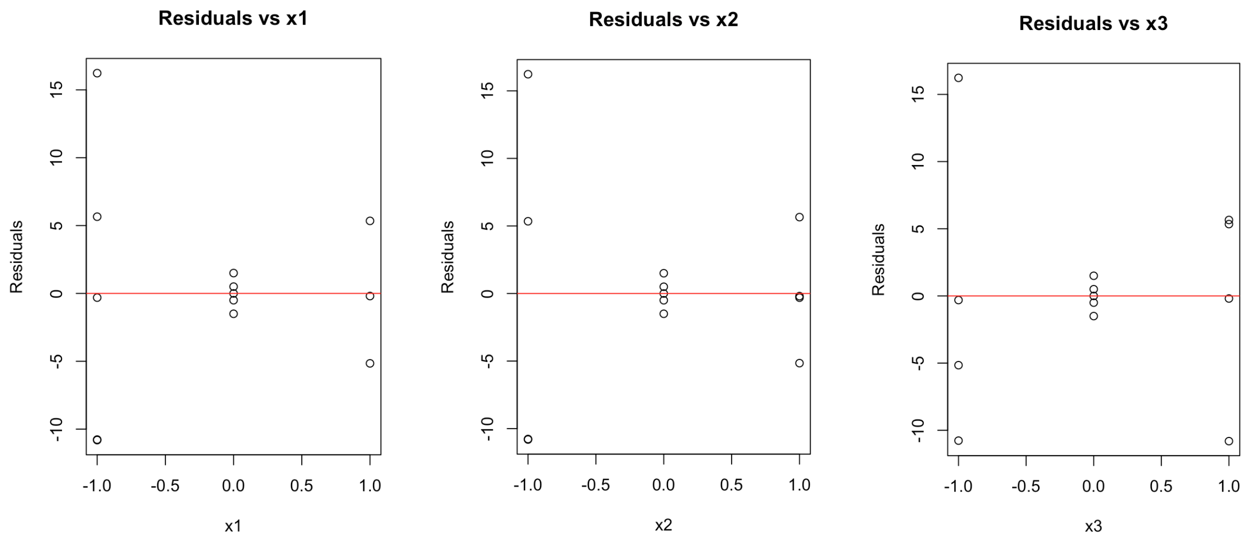
The five key assumptions:

a. Normality - Our response variable(s) (by themselves), residuals (by themselves), and residuals vs regressors, when histogrammed, look normal.



The distribution of our response variable seems mostly normal - I don't know what to make of the gap in the middle, though.

b. Independence - Our samples are independent (i.e. the value of one does not affect the value of any other). This is usually the hardest one to test for - usually it's argued from a sampling side. If you plot the residuals vs the predicted values, if they're independent, you should see no pattern.



Our samples appear to be independent - there doesn't seem to be a pattern in the data (but then again, there's only 13 data points).

c. Constant Variance - The residual plots should just be bands (i.e. no funnels, cones, or any weird shape).

It does appear that we have a "bowtie" kind of pattern, so I would assume that we don't have constant variance.

d. E[ϵ] $= 0$ - This is assumed since that's the way we build our model (i.e. via least squares)

e. Linearity - The model actually fits (i.e. the data follows the shape of the model: R^2 is high)

Question 2

Question 3

Question 4

Question 5

Question 2. Exercise 5.1 on page 202

5.1 Byers and Williams (" Viscosities of Binary and Ternary Mixtures of Polyaromatic Hydrocarbons," *Journal of Chemical and Engineering Data* , **32** , 349–354, 1987) studied the impact of temperature (the regressor) on the viscosity (the response) of toluene-tetralin blends. The following table gives the data for blends with a 0.4 molar fraction of toluene.

a. Plot a scatter diagram. Does it seem likely that a straight-line model will be adequate?

b. Fit the straight-line model. Compute the summary statistics and the residual plots. What are your conclusions regarding model adequacy?

c. Basic principles of physical chemistry suggest that the viscosity is an exponential function of the temperature. Repeat part b using the appropriate transformation based on this information.

1

Temperature (°C)	Viscosity (mPa · s)
24.9	1.133
35.0	0.9772
44.9	0.8532
55.1	0.7550
65.2	0.6723
75.2	0.6021
85.2	0.5420
95.2	0.5074

Question 3. (Exercise 5.10 on page 205 and on Exercise 6.8 on page 221)

Consider the pressure drop data in Table B.9.

a. Fit a multiple regression for y and all regressors, then perform a thorough residual analysis of the above regression.
b. Identify the most appropriate transformation for these data. Fit the model and repeat the residual analysis.
c. Perform two thorough influence analyses based on the above two regression models you fit before and after the transformation. Discuss your results. (*Note, please perform the influence analysis to find some influential data points as we discussed in Chapter 6 for each of the two models separately*).

Question 4. Problem 7.17 on page 257

Chemical and mechanical engineers often need to know the vapor pressure of water at various temperatures (the "infamous" steam tables can be used for this). Below are data on the vapor pressure of water (y) at various temperatures.

Vapor.Pressure.y(mmHg)	Temperature.x(°C)
9.2	10
17.5	20
31.8	30
55.3	40
92.5	50
149.4	60

a. Fit a first-order polynomial model to the data. Overlay the fitted model on the scatterplot of y versus x. Comment on the apparent fit of the model.
b. Prepare a scatterplot of predicted y versus the observed y . What does this suggest about model fit?
c. Plot residuals versus the fitted or predicted y . Comment on model adequacy .
d. Fit a second-order model to the data. Is there evidence that the quadratic term is statistically significant?
e. Repeat parts a – c using the second-order model. Is there evidence that the second-order model provides a better fit to the vapor pressure data?

Question 5. (Problem 8.4 on page 280)

Consider the automobile gasoline mileage data in Table B.3 .

a. Build a linear regression model relating gasoline mileage y to engine displacement x_1 and the type of transmission x_{11}. Does the type of transmission significantly affect the mileage performance?
b. Modify the model developed in part a to include an interaction between engine displacement and the type of transmission. What conclusions can you draw about the effect of the type of transmission on gasoline mileage? Interpret t he parameters in this model.