Show **all** of your work on this assignment and answer each question fully in the given context.

Please staple your assignment!

1. Chapter 4, Section 1, Exercise 3 (unless directed otherwise you may use JMP; include plots as requested) (page 140) [5 pts each, 25 pts total]

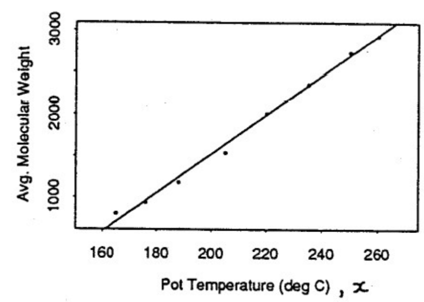
1/3

P140. 3. (a) $R^2 = .994$.

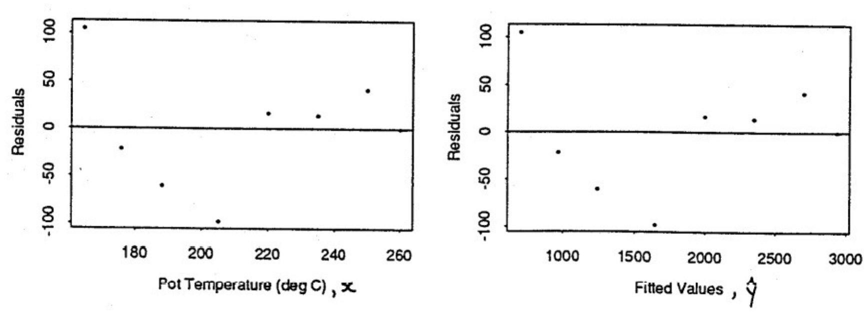
(b) The least squares equation is

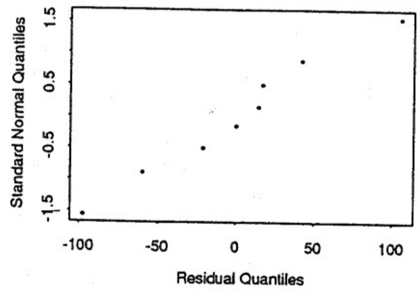
$$\hat{y} = -3174.6 + 23.5x.$$

β_1 represents the "true" average change in molecular weight that accompanies a 1°C increase in pot temperature (assuming that a straight-line model is correct). $b_1 = 23.5$ is a data-based approximation of this value.



(c) The residuals are: 105.36, −21.13, −60.11, −97.58, 16.95, 14.48, 42.00, and .02.





It is difficult to evaluate the appropriateness of the fitted equation based on so little data. The plots of residuals versus x and residuals versus $\hat{y}$ do not contain any obvious patterns, and thus provide no evidence that the equation is inappropriate. The normal plot of residuals is fairly linear, providing no evidence that the residuals are not bell-shaped.

(d) There is no replication (multiple experimental runs at a particular pot temperature). Replication would validate any conclusions drawn from the experiment, and provide more information to confirm the appropriateness of the fitted equation.

(e) For $x = 188°$C,

$$\hat{y} = -3174.6 + 23.5(188) = 1243.1.$$

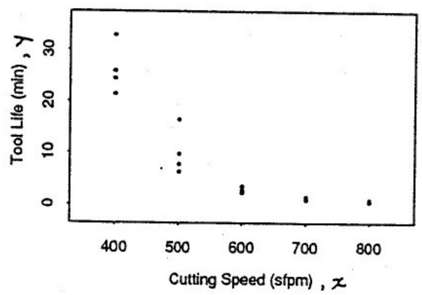
For $x = 200°$C,

$$\hat{y} = -3174.6 + 23.5(200) = 1525.1.$$

It would not be wise to make a similar prediction at $x = 70°$C because there is no evidence that the fitted relationship is correct for pot temperatures as low as $x = 70°$C. This would be an extrapolation. Some data should be obtained around $x = 70°$C before making such a prediction.

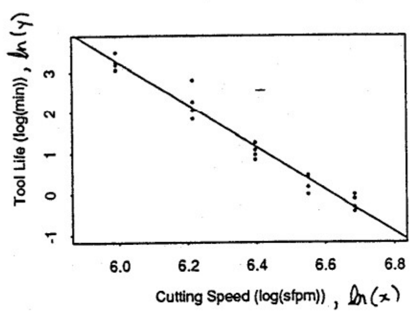
2. Chapter 4, Section 1, Exercise 4 (unless directed otherwise you may use JMP; include plots as requested) (page 140) [5 pts each, 15 pts total]

P140. 4. (a)



The scatterplot is not linear, so the given straight-line relationship does not seem appropriate. The least squares line is $\hat{y} = 44.075 - .059650x$. The corresponding R^2 is .723.

(b)



This scatterplot is much more linear, and a straight-line relationship seems appropriate for the transformed variables. The least squares line is $\widehat{\ln y} = 34.344 - 5.1857 \ln x$. The corresponding R^2 is .965.

(c) The least squares line is given in part (b). For $x = 550$,

$$\widehat{\ln y} = 34.344 - 5.1857 \ln(550) = 1.6229 \ln(\text{minutes}),$$

so $\hat{y} = e^{1.6229} = 5.07$ minutes. The implied relationship between x and y is

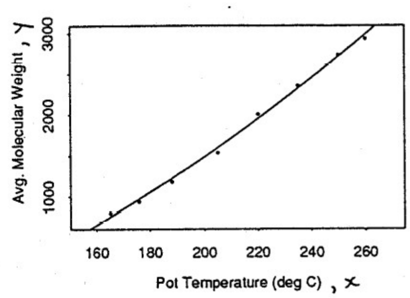
$$\begin{aligned} y &\approx e^{\beta_0 + \beta_1 \ln x} \\ y &\approx e^{\beta_0} e^{\ln x^{\beta_1}} \\ y &\approx e^{\beta_0} x^{\beta_1}. \end{aligned}$$

With slight rearrangement, this is the same as Taylor's equation for tool life.

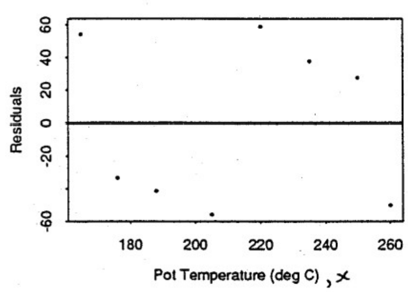
3. Chapter 4, Section 2, Exercise 1 (unless directed otherwise you may use JMP; include plots as requested) (page 161) [10 pts]

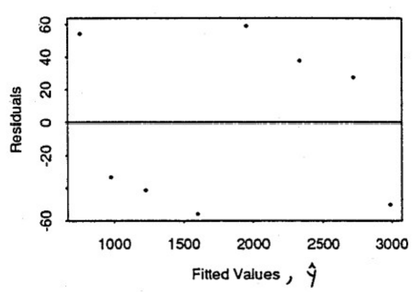
P161. 1 The least squares equation is

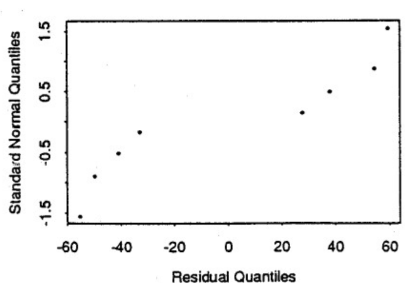
$$\hat{y} = -1315 + 5.59x + .04212x^2.$$



$R^2_Q = .996$, compared with $R^2_L = .994$. This is a very small improvement, at the cost of using a more complex equation.







The residuals here are smaller, as they will always be for a more complex model. There is no noticeable improvement in the residual plots, compared to those from the straight-line model. In fact, the residual plots for the quadratic model look more patterned. The scatterplot of y versus x also indicates that the quadratic model would be "overfitting" the data. The simpler straight-line relationship seems to be adequate.

For the quadratic model, at $x = 200°C$,

$$\hat{y} = -1315 + 5.59(200) + .04212(200)^2 = 1487.2,$$

which is relatively close to 1525.1 from 4-5 (e).

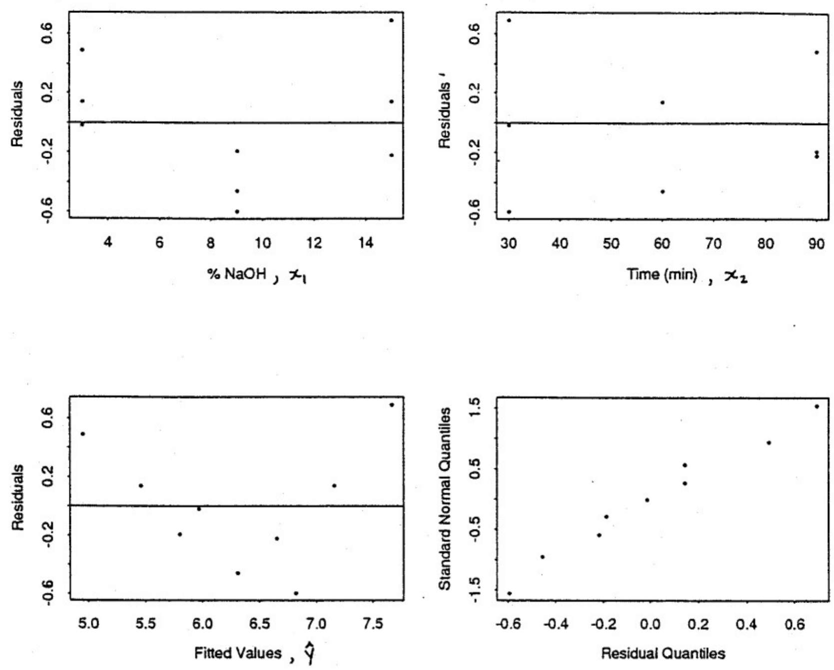
4. Chapter 4, Section 2, Exercise 2 (unless directed otherwise you may use JMP; include plots as requested; skip part h) (page 161) [5 pts each, 35 pts total]

P161. 2. (a) The least squares equation is

$$\hat{y} = 6.0483 + .14167x_1 - .016944x_2.$$

Assuming the fitted equation is appropriate, this means that as x_1 increases by 1% (holding x_2 constant), y increases by roughly .14167 cm^3/g. As x_2 increases by 1 minute (holding x_1 constant), y decreases by roughly .016944 cm^3/g. The R^2 corresponding to this equation is .807.

(b) The residuals are −.015, .143, .492, −.595, −.457, −.188, .695, .143, −.218.



Both the plots of residuals versus x_1 and residuals versus $\hat{y}$ show a positive-negative-positive pattern of residuals, indicating that the relationship between x_1 and y is not completely accounted for by the current model. These plots suggest adding an x_1^2 term. The plot of residuals versus x_2 is patternless; x_2 seems to be well represented. The normal plot of residuals is fairly linear, indicating that the residuals are bell-shaped.

(c) For $x_2 = 30$, the equation is

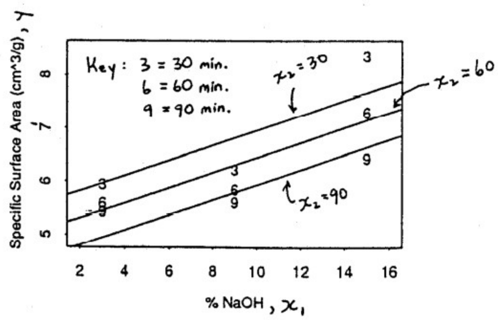
$$\begin{aligned} \hat{y} &= 6.0483 + .14167x_1 - .016944(30) \\ &= 5.53998 + .14167x_1. \end{aligned}$$

For $x_2 = 60$, the equation is

$$\begin{aligned} \hat{y} &= 6.0483 + .14167x_1 - .016944(60) \\ &= 5.03166 + .14167x_1. \end{aligned}$$

For $x_2 = 90$, the equation is

$$\begin{aligned} \hat{y} &= 6.0483 + .14167x_1 - .016944(90) \\ &= 4.52334 + .14167x_1. \end{aligned}$$



The fitted responses do not match up well, because the relationship between y and x_1 (%NaOH) is not linear for any of the x_2 values (Time).

(d) At $x_1 = 10\%$ and $x_2 = 70$ minutes,

$$\hat{y} = 6.0483 + .14167(10) - .016944(70) = 6.279 \text{ cm}^3/\text{g}.$$

It would not be wise to make a similar prediction at $x_1 = 10\%$ and $x_2 = 120$ minutes because there is no evidence that the fitted relationship is correct under these conditions. This would be extrapolating. Some data should be obtained around $x_1 = 10\%$ and $x_2 = 120$ minutes before making such a prediction.

(e) The least squares equation is

$$\hat{y} = 4.9833 + .260x_1 + .00081x_2 - .001972x_1x_2,$$

and the corresponding R^2 is .876. The increase in R^2 from .807 to .876 is not very large; using the more complicated equation may not be desirable (this is subjective). Residual plots for this more complicated equation should also be examined before evaluating its appropriateness.

(f) For $x_2 = 30$, the equation is

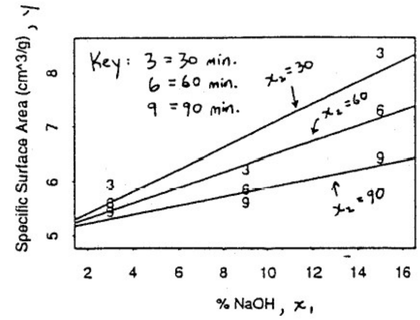
$$\begin{aligned} \hat{y} &= 4.9833 + .260x_1 + .00081(30) - .001972x_1(30) \\ &= 5.0076 + .20084x_1. \end{aligned}$$

For $x_2 = 60$, the equation is

$$\begin{aligned} \hat{y} &= 4.9833 + .260x_1 + .00081(60) - .001972x_1(60) \\ &= 5.0319 + .14168x_1. \end{aligned}$$

For $x_2 = 90$, the equation is

$$\begin{aligned} \hat{y} &= 4.9833 + .260x_1 + .00081(90) - .001972x_1(90) \\ &= 5.0562 + .08252x_1. \end{aligned}$$

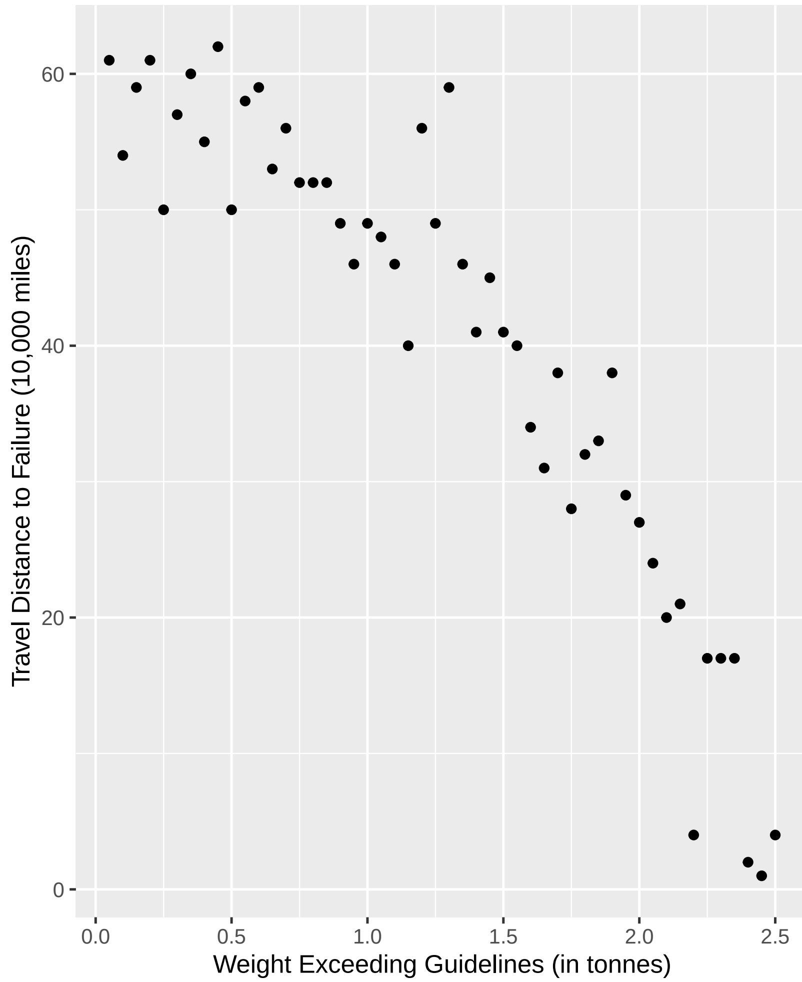


The new model allows there to be a different slope for different values of x_2, so these lines fit the data better than the lines in part (c). But they still do not account for the non-linearity between x_1 and y. An equation with an x_1^2 term would fit much better.

(g) There is no replication (multiple experimental runs at a particular NaOH-Time combination). Replication would validate any conclusions drawn from the experiment, and it would allow for better comparisons among the different possible fitted equations.

5. The major cause of axel failure in freight trucks is when shippers exceed the recommended weight limits that can be handled by the axels. Issues resulting from these failures have been becoming more frequent as shippers try to cut corners, leading members of the state's Department of Transportation to ask one of their civil engineers to look into the available data and better advise them on the relationship between excessive weight and axel failure.

   A company manufacturing axels provides the engineer with data gathered from conducting experiments loading axels with excessive weight and simulating traveling conditions. The data consists of two columns, **excessive weight (in tonnes)** is the amount of weight over the limit that was placed on the axel, and **distance to failure (in tens of thousands of miles)** is the simulated distance to the axel's failure.



Here are some summaries of the data:

6. Using the summaries above, fit a linear relationship between **weight exceeding guidelines** (x) and **travel distance to failure** (y).[10 pts]

   The fitted line equation is

$$\hat{y} = b_0 + b_1 \cdot x$$

   We can use the information above to get the value for b_1 and b_0:

$$\begin{aligned} b_1 &= \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \\ &= \frac{(1982) - (50)(64/50)(1996/50)}{107 - 50(64/50)^2} \\ &= -22.8421052631579 \end{aligned}$$

and with b_1 we can find the value fo b_0:

$$\begin{aligned} b_0 &= \bar{y} - b_1\bar{x} \\ &= (1996/50) - (-22.8421052631579)(64/50) \\ &= 69.1578947368421 \end{aligned}$$

Which gives us the fitted equation of

$$\hat{y} = 69.15 - 22.84 \cdot x$$

7. Write the equation of the fitted linear relationship. [5 pts]

$$\hat{y} = 69.15 - 22.84 \cdot x$$

8. Find and interpret the value of R^2 for the fitted linear relationship.[5 pts]

Since we are using a linear relationship, we can get R^2 from r:

$$\begin{aligned} r &= \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^n y_i^2 - n\bar{y}^2\right)}} \\ &= \frac{(1982) - (50)(64/50)(1996/50)}{\sqrt{(107 - 50(64/50)^2)\,(94078 - (50)(1996/50)^2)}} \\ &= -0.953352777638285 \end{aligned}$$

So $R^2 = (r)^2 = 0.908881518630634$

This means that 93.00% of our the variablity in travel distance to failure can be explained by the linear relationship with weight exceeding guidelines.

9. Using the fitted line, provide a predicted value of travel distance to failure when the weight exceeding the guidelines is 3.4 tonnes.[5 pts]

$$\hat{y} = 69.15 - 22.84(3.4) = -8.50599999999999$$

Total: 100 pts