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

If you cannot submit your homework in the class, you can drop it at my office door in 3220 Snedecore Hall by Thursday at 03:30 PM.

In this homework, you CAN use JMP to plot the data or calculate coefficients of regression model whenever it is asked in the question.

Please staple your assignment and write your name !

1. 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.
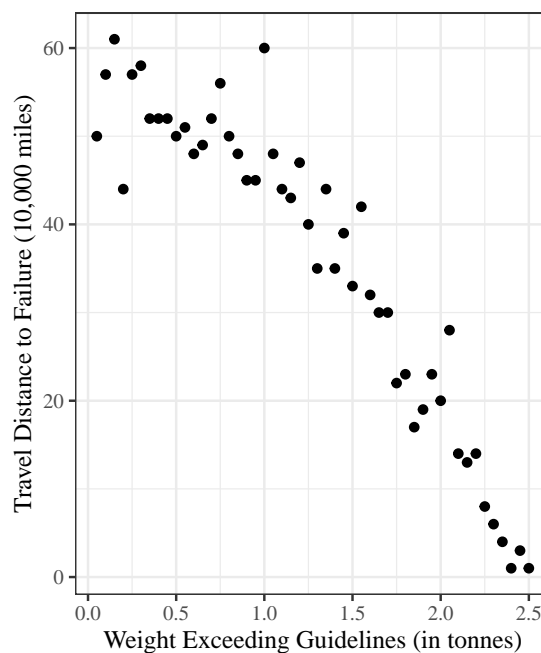
   (a) Using the summaries above, fit a linear relationship between **weight exceeding guidelines** (x) and **travel distance to failure** (y).[10 pts]

   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:

$$
\begin{aligned}
b_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \\
&= \frac{(1699) - (50)(64/50)(1795/50)}{107 - 50(64/50)^2} \\
&= -23.8676236044657
\end{aligned}
$$

and with b_1 we can find the value fo b_0:

$$
\begin{aligned}
b_0 &= \bar{y} - b_1 \bar{x} \\
&= (1996/50) - (-23.8676236044657)(64/50) \\
&= 64.5305582137161
\end{aligned}
$$

Which gives us the fitted equation of

$$\hat{y} = 64.53 - 23.86 \cdot x$$

(b) Write the equation of the fitted linear relationship. [5 pts]

$$\hat{y} = 64.53 - 23.86 \cdot x$$

(c) Find and <u>interpret</u> 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 (slides 29 and 51 of chapter 4 slides):

$$
\begin{aligned}
r &= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}} \\
&= \frac{(1982) - (50)(64/50)(1699/50)}{\sqrt{\left(107 - 50(64/50)^2\right)\left(94078 - (50)(1699/50)^2\right)}} \\
&= -0.965183339978807
\end{aligned}
$$

So $R^2 = (r)^2 = 0.931578879772645$

This means that 93.15% of our the variablity in travel distance to failure can be explained by the linear relationship with weight exceeding guidelines.

(d) 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]

$$\hat{y} = 64.53 - 23.86(3.4) = -16.594$$

(e) If the observed travel distance is 0.37 when the weight exceeding the guidelines is 3.4, what is the residual? Based on the sign of the residual, explain if we are overfitting or under fitting.[5 pts]

Hint: You have already achieved the predicted value of travel distance when the weight exceeding the guideline is 3.4 in part (d).
Using the formula for residuals on page 36-37 of chapter 4 slides,

$$e_i = y_i - \hat{y}_i = 0.37 - (-16.594) = 16.964$$

The positive residual means that we are underpredicting.

The JMP output below comes from fitting a quadratic model using x and x^2.

Response Distance to Failure

Summary of Fit

RSquare	REDACTED
RSquare Adj	REDACTED
Root Mean Square Error	5.281589
Mean of Response	0.16
Observations (or Sum Wgts)	50

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	13229.647	6614.82	237.1314
Error	47	1311.073	27.90	Prob > F
C. Total	49	14540.720		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	16.27602	2.333507	6.97	<.0001*
Weight Exceeding Limit	4.6604349	4.221593	1.10	0.2752
(Weight Exceeding Limit)^2	-10.2775	1.604983	-6.40	<.0001*

(f) Write the equation of the fitted quadratic relationship. [5 pts]

$$\hat{y} = 16.27602 + 4.6604349x - 10.2775x^2$$

(g) Find and interpret the value of R^2 for the fitted quadratic relationship.[5 pts]

$$R^2 = 1 - SSE/SSTO = 1 - (1311.073/14540.720) = 0.909834382341452$$

In other words, 90.98% of the variability in travel distance to failure can be explained by the linear relationship with weight exceeding guidelines.

(h) Using the fitted quadratic relationship, provide a predicted value of travel distance to failure when the weight exceeding the guidelines is 3.4 tonnes.[5 pts]

$$\hat{y} = 16.27602 + 4.6604349(3.4) - 10.2775(3.4)^2 = -86.68640134$$

2. **[Ch. 4.1 Exercise 3, pg. 140]** The article "Polyglycol Modified Poly (Ethylene Ether-Carbonate) Polyols by Molecular Weight Advancement" by R. Harris (*Journal of Applied Polymer Science*, 1990) contains some data on the effect of reaction temperature on the molecular weight of resulting poly polyols. The data for eight experimental runs at temperature 165°C and above are as follows (see website for 'polyols.csv'):

Pot temperature (°C)	Average molecular weight
165	808
176	940
188	1183
205	1545
220	2012
235	2362
250	2742
260	2935

Use a statistical package (JMP or 'R') to help you complete the following (plots and computation):

(a) What fraction of the observed raw variation in molecular weight of resulting poly polyols (y) is accounted for by a linear equation in reaction temperature (x)?[5 pts]

Using JMP, the R^2 value is 0.9942.
hint: The question asks for the coefficient of determination.

(b) Fit a linear relationship $y \approx \beta_0 + \beta_1 x$ to these data via least squares. Then explain how the average mulecular weight changes if pot temperature increases for a 1°C ?[5pts]
Using JMP, the fitted linear relationship is

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

(c) Compute and plot residuals from the linear relationship fit in b). Discuss what they suggest about the appropriateness of that fitted equation.[10 pts]

Note: You should provide both residual plots vs. experimental variable and normal QQ-plot vs. residual quantiles to see if the assumptions of the model are met.
Using JMP, you can save the residuals of the fit, and then plot residuals vs. experimental variable (reaction temperature) we have
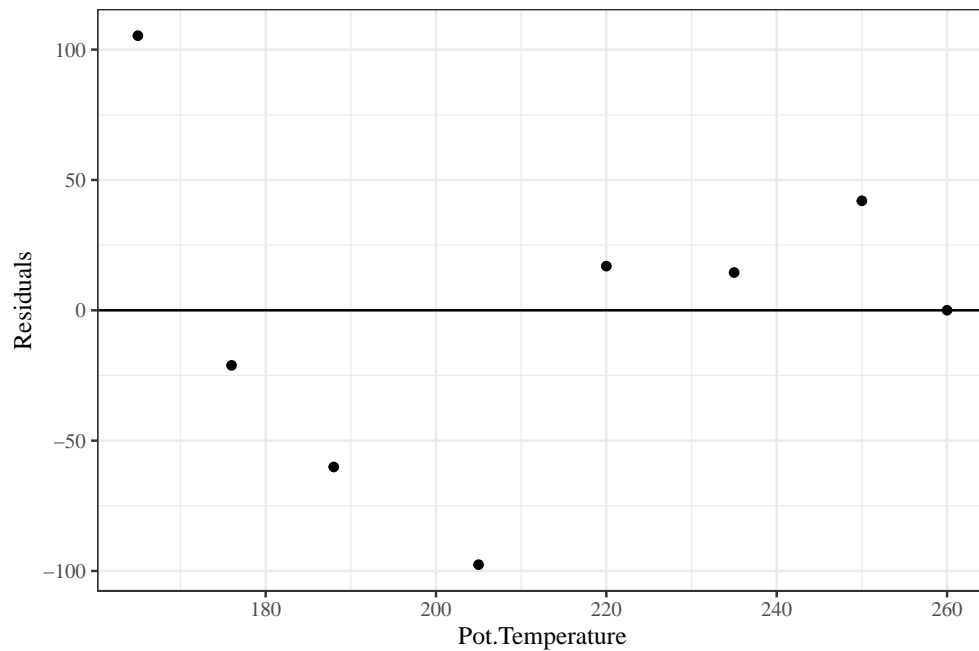
The residual vs. pot temperature plot shows patternless points centered at zero. So, the assumptions of patternless residuals and centered residual at zero seems to be met. To check if the residuals are normally distributed, we can draw a histogram of the residuals and also check normal QQ-plot of residuals.
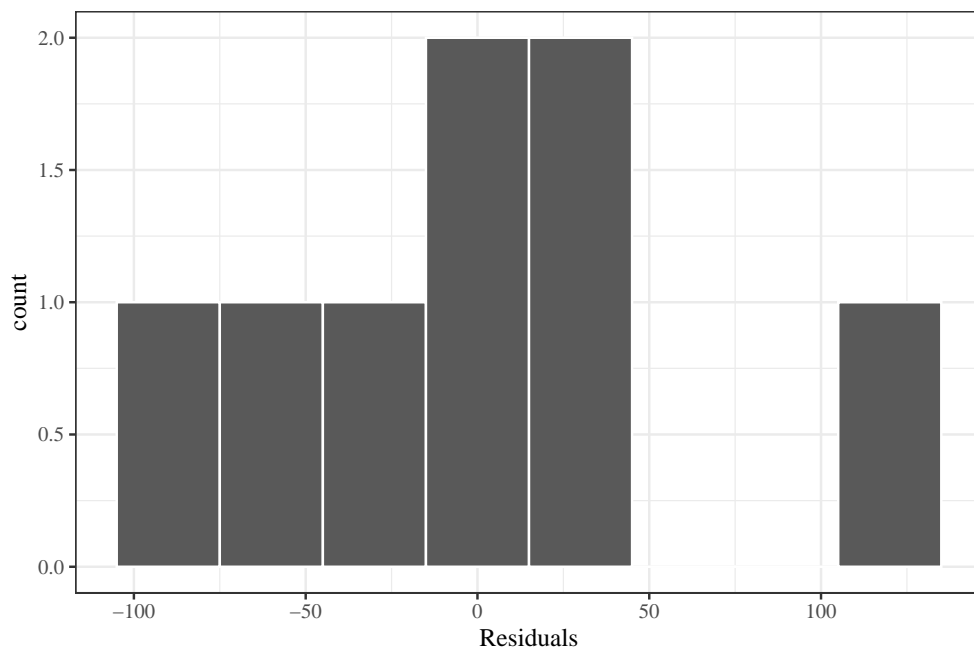
It is hard to evaluate the normality of the residuals based on the histogram of residuals. To investigate that more, a better tool is Normal QQ-plot as follow
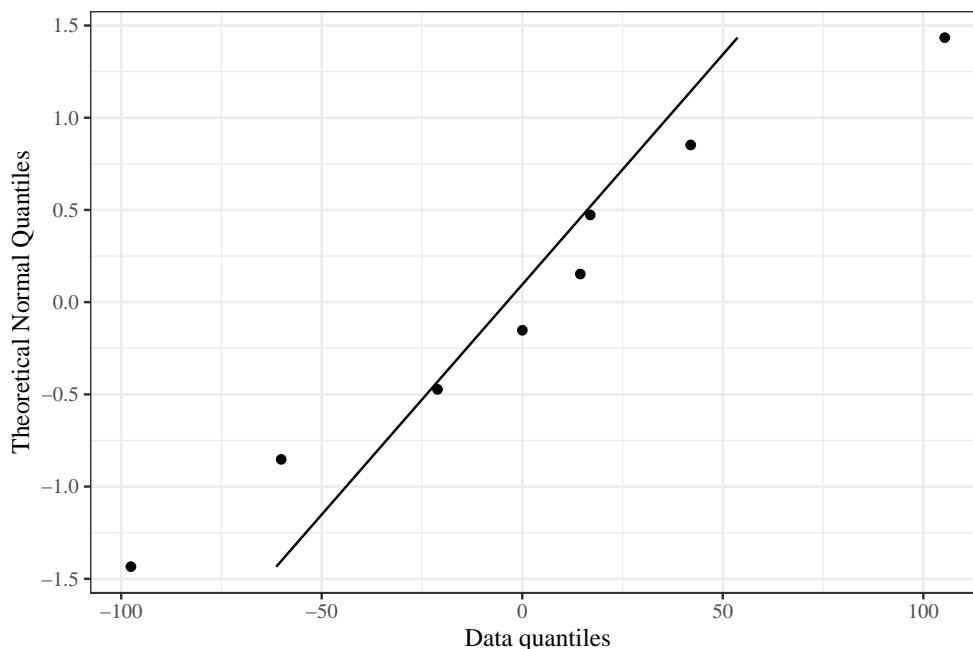
The normal qq-plot of residuals is fairly linear, providing no evidence hat the residuals are not bell shaped. **So, we can say that the residuals are roughly bell-shaped** and therefore, the assumption of normality is met.

(d) Based on your analysis of these data, what average molecular weight would you predict for an additional reaction run at 188°C? At 200°C? Why would or wouldn't you be willing to make a similar prediction of average molecular weight if the reaction is run at 70°C?[6 pts]

Hint: You may consider extrapolation and/or intrapolation.

for $x = 188$°C

$$\hat{y} = -3174.6 + 23.5 * (188) = 1243.4$$

and for $x = 200$°C,

$$\hat{y} = -3174.6 + 23.5 * (200) = 1525.4$$

It would not be wise to make a similar prediction at $x = 70$°C because there is no evidence that the fitted relationship is correect for pot temeratures as low as $x = 70$°C. In other word, that would be extrapolation. some data should be obtained around $x = 70$°C before making such prediction.

3. **[Ch. 4.2, Exercise 1, pg. 161]** Return to problem 2, (Exercise 3 of Section 4.1 on pg. 140 of the textbook).

(a) Fit a quadratic relationship $y \approx \beta_0 + \beta_1 x + \beta_2 x^2$ to the data via least squares.[5 pts]

Using JMP, the quadratic fitted relationship is

$$\hat{y} = -1315 + 5.59 * x + 0.04212 * x^2$$

(b) Provide the value of R^2 and interpret that.[5 pts]

Using JMP, the $R^2 = 0.996$ which mneans that 99.6% of the variation in average molecular weight can be explained by **the quadratic relationship** between average molecular weight and pot temperature.

(c) Plot residuals against the experiemntal variable and normal QQ-plot and discuss if the assumptions of the model are met.[5 pts]
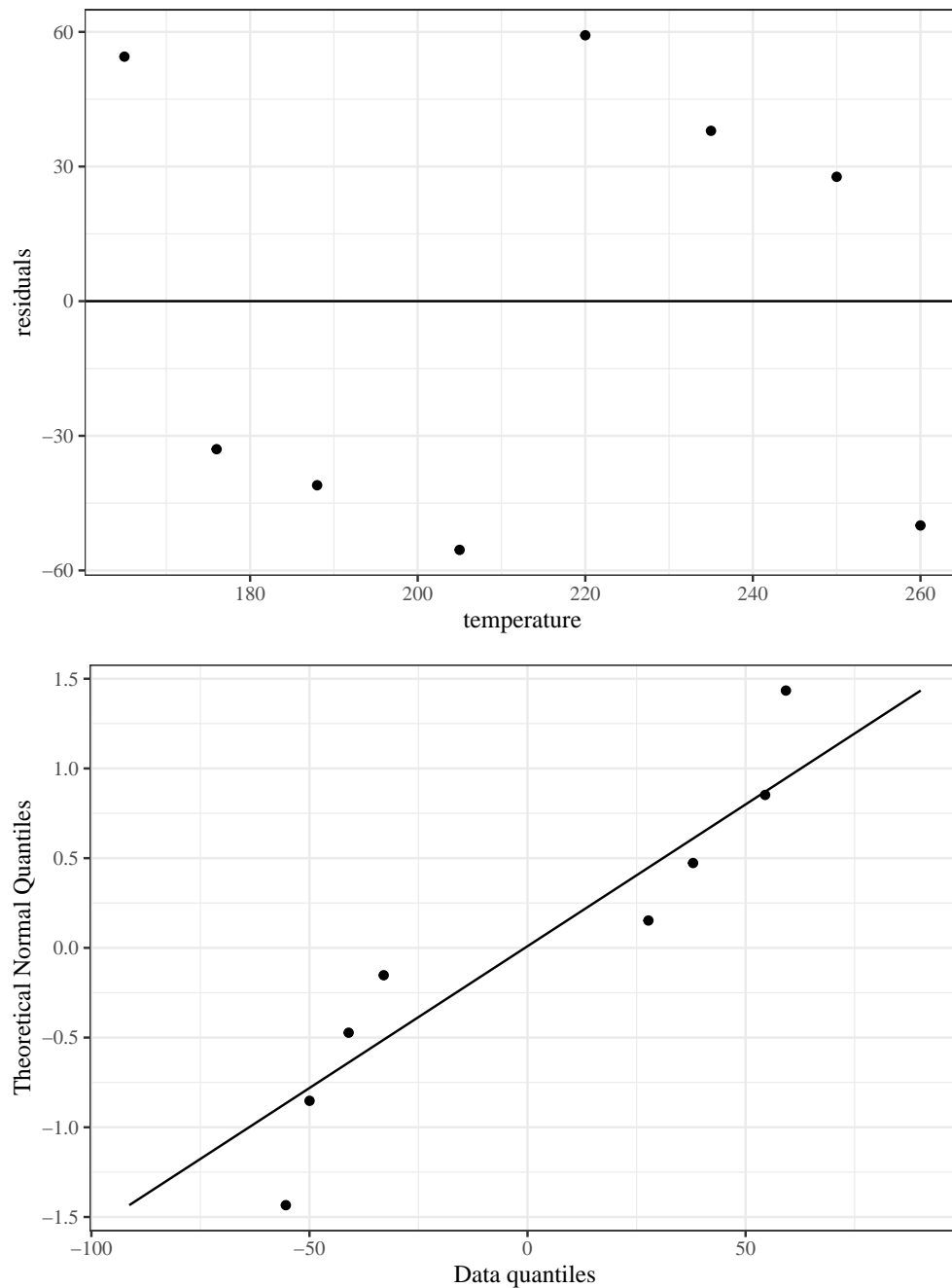
In general, the residual plots do not show normality of residuals. The residual vs. experimental plot also show some patterned data.

The residuals here are smaller, as they will always be more for a more complex model. There is no noticeable improvement in residual plot compared to previous question (fitting a linear line). In fact, the residual plot for the quadratic model looke even more ptterned.

Total: 86 pts