Stat 850 Final Exam

Dani Chu

Contents

Chapter 1

Question 1

1.1 Part A

Prove that the quadratic form $y^T Ay$, $y^T By$ are independent iff $B\Sigma A = 0$.

We know that if a symmetric matrix A is of full rank then it can be written $A = LL^T$ where L has full column rank

$$B\Sigma A = 0 \tag{1.1}$$
$$L_B L_B^T \Sigma L_A L_A^T = 0 \tag{1.2}$$

1.2 Part B

Let's start by rearranging R.

$$\begin{aligned}
R =& (K^T\hat{\beta} - m)^T (K^T (X^T X)^{-1} K)^{-1} (K^T \hat{\beta} - m) && (1.3)\\
=& (K^T (X^T X)^{-1} X^T y - m)^T (K^T (X^T X)^{-1} K)^{-1} (K^T (X^T X)^{-1} X^T y - m) && (1.4)\\
=& [(K^T (X^T X)^{-1} X^T)(y - XK(K^T K)^{-1} m)]^T [(K^T (X^T X)^{-1} K)^{-1}] && (1.5)\\
& [(K^T (X^T X)^{-1} X^T (y - XK(K^T K)^{-1} m)] && (1.6)\\
=& [(y - XK(K^T K)^{-1} m)]^T [X^T (X^T X)^{-1} K][(K^T (X^T X)^{-1} K)^{-1}] && (1.7)\\
& [(K^T (X^T X)^{-1} X^T][(y - XK(K^T K)^{-1} m)] && (1.8)\\
=& [y_{\text{new}}]^T A [y_{\text{new}}] && (1.9)
\end{aligned}$$

where

$$y_{\text{new}} = (y - XK(K^T K)^{-1} m) \tag{1.10}$$

and

$$A = [X^T (X^T X)^{-1} K][(K^T (X^T X)^{-1} K)^{-1}][(K^T (X^T X)^{-1} X^T] \tag{1.11}$$

Now to use the therorem we need to have $SSE = y^T(I-H)y$ to be equal to $y^T_{\text{new}}(I-H)y_{\text{new}}$.

$$y^T_{\text{new}}(I-H)y_{\text{new}} = (y - XK(K^TK)^{-1}m)^T(I-H)(y - XK(K^TK)^{-1}m) \tag{1.12}$$
$$y^T_{\text{new}}(I-H)y_{\text{new}} = (y^T - m^T(K^TK)^{-1}K^Tx^T)(I-H)(y - XK(K^TK)^{-1}m) \tag{1.13}$$
$$y^T_{\text{new}}(I-H)y_{\text{new}} = y^T(I-H)y + \tag{1.14}$$
$$y^T(I-H)(-XK(K^TK)^{-1}m)) + \tag{1.15}$$
$$(-m^T(K^TK)^{-1}K^Tx^T)(I-H)(y) + \tag{1.16}$$
$$(-m^T(K^TK)^{-1}K^Tx^T)(I-H)(-XK(K^TK)^{-1}m)) \tag{1.17}$$
$$y^T_{\text{new}}(I-H)y_{\text{new}} = y^T(I-H)y + 0 + 0 + 0 \tag{1.18}$$
$$y^T_{\text{new}}(I-H)y_{\text{new}} = y^T(I-H)y \tag{1.19}$$

So let

$$B = (I-H) \tag{1.20}$$

Finally let $\Sigma = \sigma^2 I$ then

$$B\Sigma A = (I-H)\sigma^2 I[X^T(X^TX)^{-1}K][(K^T(X^TX)^{-1}K)^{-1}][(K^T(X^TX)^{-1}X^T] \tag{1.21}$$
$$B\Sigma A = 0 \tag{1.22}$$

and we have set up the problem to use the theororem from part a. Therefore SSE and R under the genreal linear hypthesis are independent.

1.3 Part C

At the end of lecture 4 we have

$$\frac{R/\sigma^2 s}{SSE/(\sigma^2(n-(p+1)))} \sim F \tag{1.23}$$

We also know that the quotient of independently distributed chi-squared variables that have been divided by their degrees of freedom follows an F distribution. Both R and SSE have a chi-square distribution since they are sums of quadratic normals.

1.4 Part D

Chapter 2

Question 2

2.1 Part A

I fit a multiple linear regression model to the data given in the file "question2.txt". The estimated least squares regression coefficients are presented in table 2.1.

2.2 Part B

For the model fit in part A, the estimate for σ is $\hat{\sigma} = 1.29$. For comparison, if just the mean of the response was used to predict the response variable we would have a root mean squared error of 8.37. If we take the ratio of the square of these values ($\frac{SSE}{SST}$) we get the R^2 for this model which is 0.98. We can see with these 2 comparisons that the value of σ is small given the variability of the data. The response variable ranges from -23.1 to 24.8, so 1.29 is very small with respect to the range of the response.

2.3 Part C

The model assumptions are

1. Outliers do not impact the analysis in any obvious way.
2. That the errors have equal variance, are normally distributed, and are independent.
3. That the multiple linear regression is in fact the correct model to use. $E(y) = X\beta$.

Table 2.1: OLS regression coefficients

Term	Estimate
Intercept	0.02
Regressor 1	0.81
Regressor 2	1.89
Regressor 3	-3.01
Regressor 4	-4.21
Regressor 5	-4.71

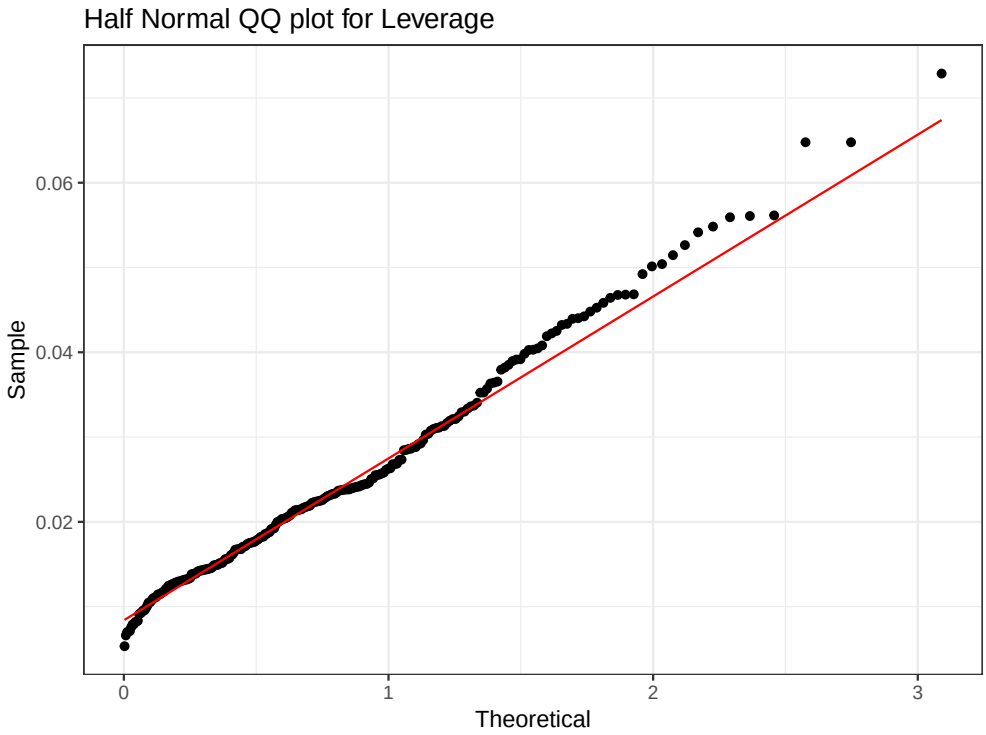

Figure 2.1: Searching for Outliers with Large Influence

2.3.1 Outliers

To asses whether there is any effect on the model due to outliers I made half normal qq plots for leverage and cooks distance. We can see in Figure 2.1 that there does not seem to be any outliers that are having a large influence on the model fit.

However in Figure 2.2 we can see that we have one very very large outlier. This value does not appear to be from the same model as the rest of the data.

We can see this clearly when we plot our fitted values vs the response in Figure 2.3.

If this was a real world situation I would investigate further about the data collection process and circumstances that led to this measurement. I would check to see if there was perhaps a data entry error or wheteher this is a realistic outcome.

Since I do not have that capacity here, I have decided to run a simulation to see how likely it was to see a residual of this size. For 10^{5} simulations, I generated 250 random values from a normal distribution with mean 0 and a standard deviation of $\hat{\sigma}$ (1.29). Out of those 10^{5} simulations, there were 0 which had a value as or more extreme as the residual value we see for our big outlier.

As a result I will remove the point as it does not appear to be from the same model as the rest of the data. In the real world I would make a note that I have removed this point and watch to see if we see anything similar pop up again.

I will refit the model with this point removed before continuing on with any of the other assumption checking. We can see in table 2.2 that the new beta estimates have not changed much. This was to be expected becasue the outlier did not have a large leverage value. The updated estimate of $\hat{\sigma}$ is 1.01.

We can now compare the half normal qq plot for cooks distance for the model with the outlier and without the outlier. The Figure 2.4 shows the plot with the outlier and Figure 2.5 shows the plot with the outlier removed.

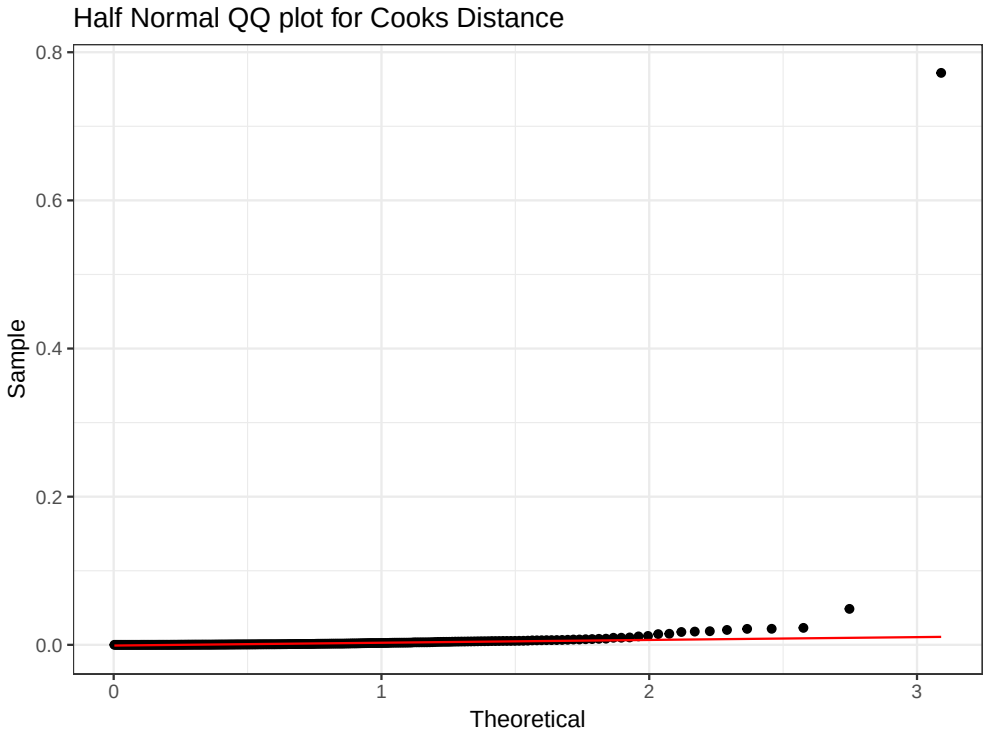

Figure 2.2: Searching for Unlikely Observations

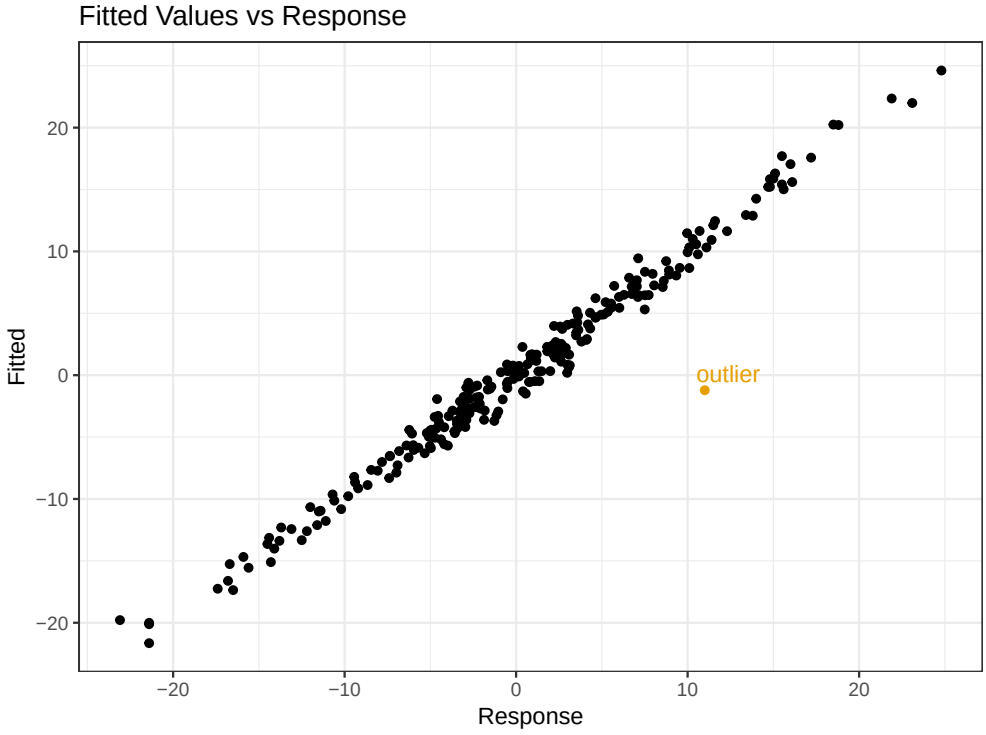

Figure 2.3: Fitted Values vs Response with one big outlier

Table 2.2: OLS regression coefficients with outlier removed

Term	Estimate
Intercept	-0.02
Regressor 1	0.96
Regressor 2	1.87
Regressor 3	-3.01
Regressor 4	-4.06
Regressor 5	-4.88

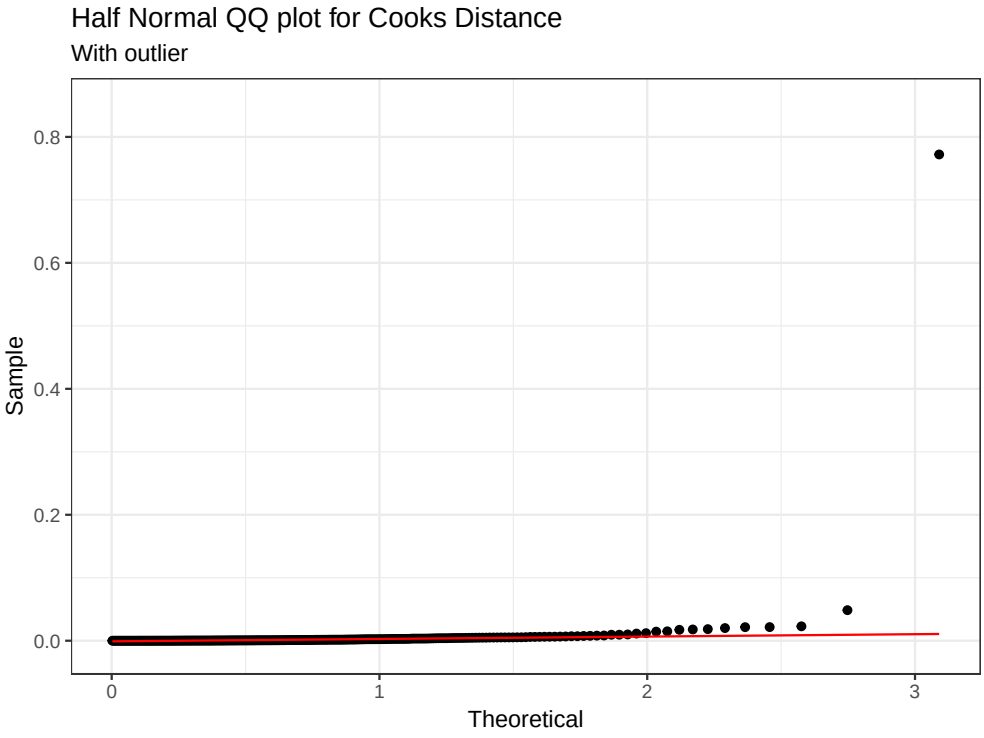

Figure 2.4: Cooks Distance Before

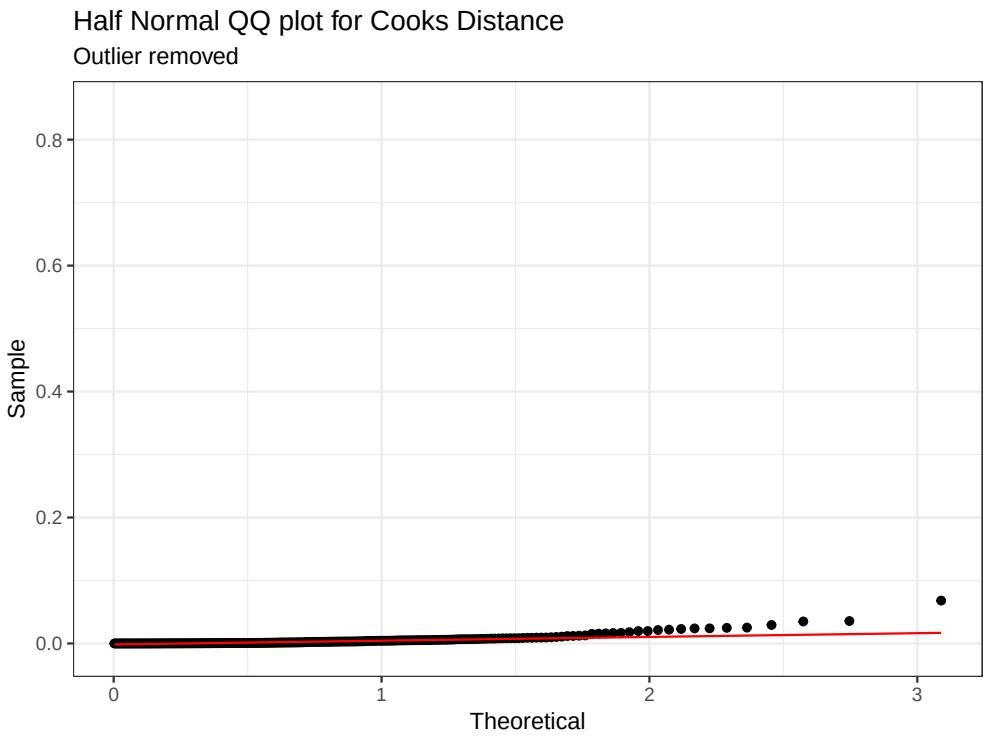

Figure 2.5: Cooks Distance After

As they are set to the same scale we can see clearly that the diagnostic has improved after the outlier is removed.

Finally, I plot the half normal qq plot of the leverage after the outlier has been removed in Figure 2.6 to show that we have not introduced any concerns by removing the observation.

We can now continue on with the rest of the model assumption checks.

2.3.2 Errors

I will now check to see if the residuals have equal variance, are normally distributed and are independent. To check for equal variance in the reisduals I plot the residuals against the fitted values in Figure 2.7. Using "Loughins Thumb Test", I do not see anything that would indicate that the constant variance assumption is violated.

Next I plotted a normal qq plot for the residuals to check to see if the data looks normally distributed. In Figure 2.8, I do not see anything that would lead me to believe that the residuals are not normal.

Finally, I checked to see if I could find a reason why the residuals would not be independent. Since there is no practical interpretations for the data the best I could do was to plot the residuals by row number. In Figure 2.9, it looks like the residuals are independent of the order in which the data was collected. Since there were no signs of autocorrelation I did not feel the need to include a Durbin Watson test.

2.3.3 Verify Model Choice

While it is not a perfect way to check the validity of the linear relationship we have assumed on the data because for any fixed value of a variable the other variables are changing, in Figure 2.10 we can see that

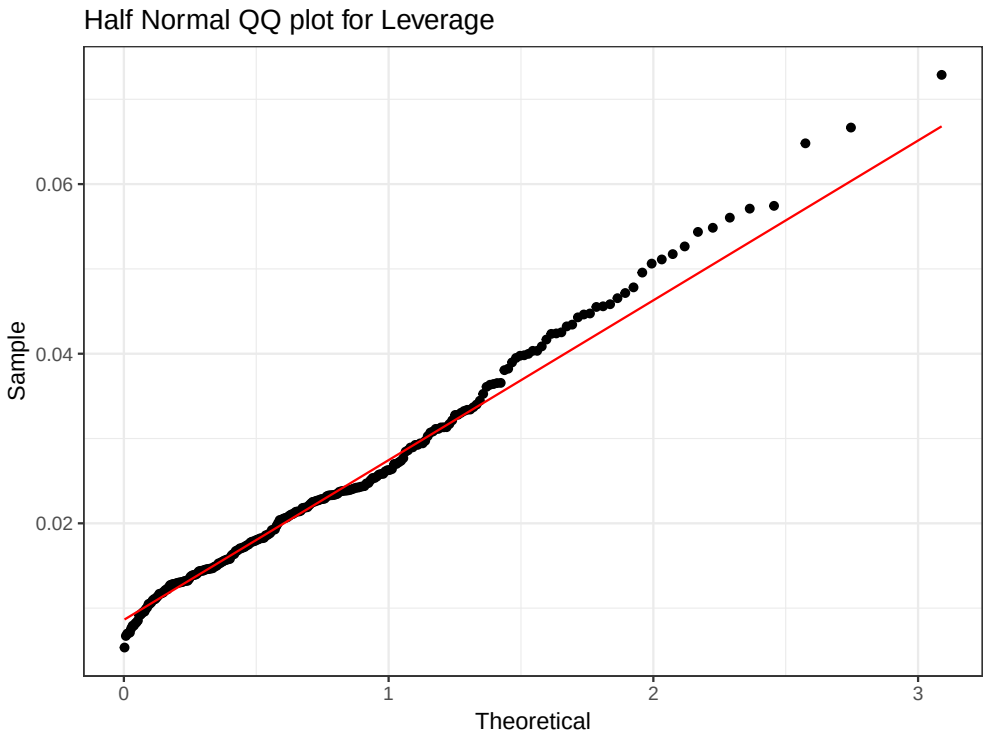

Figure 2.6: Leverage After Outlier Removed

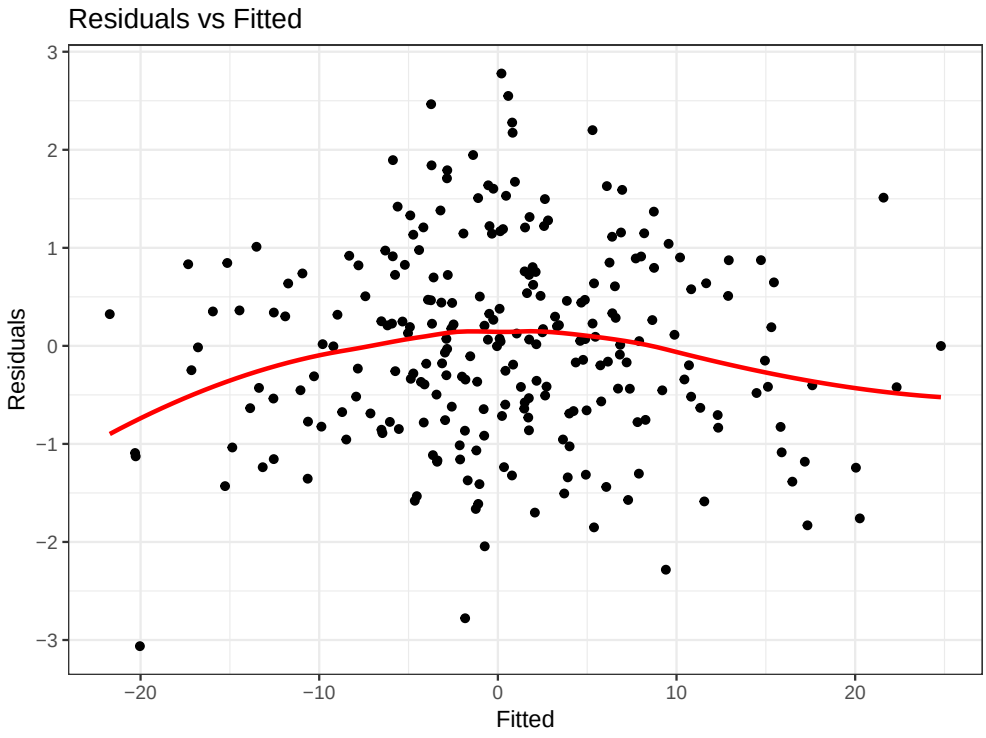

Figure 2.7: Checking for constant variance

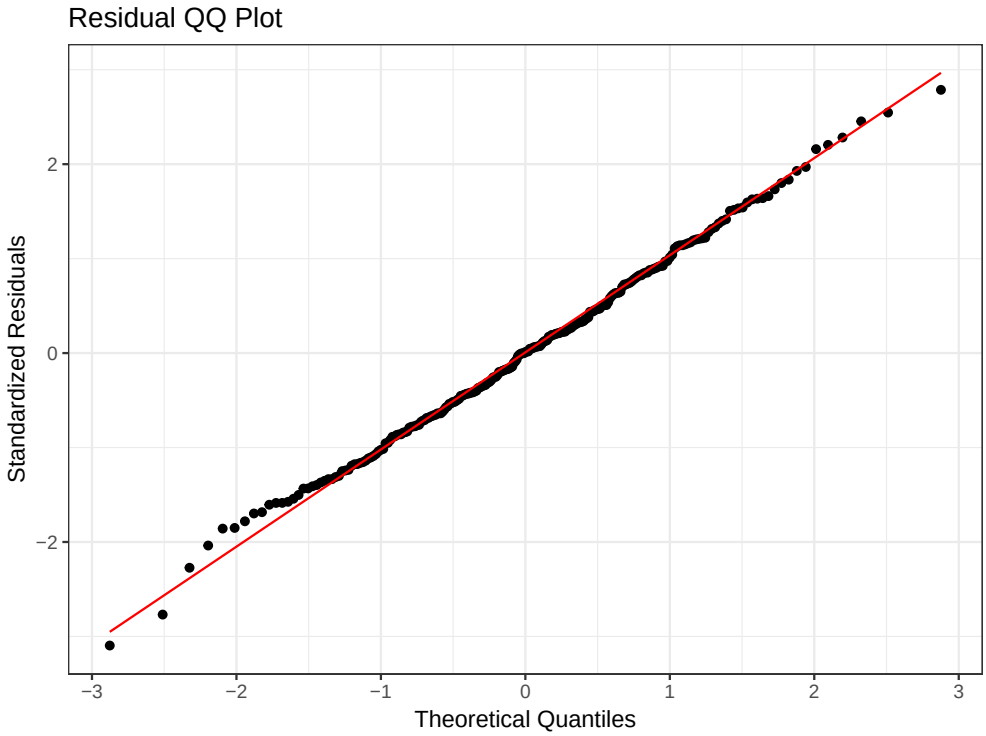

Figure 2.8: Checking for normality

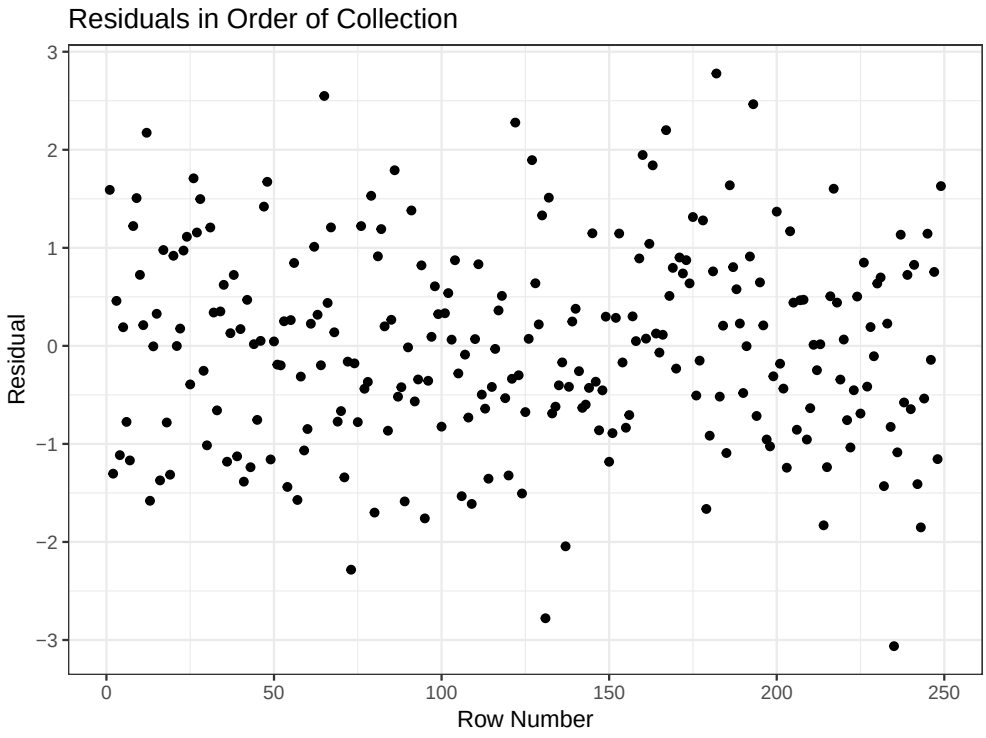

Figure 2.9: Checking for Independence of Order of Collection

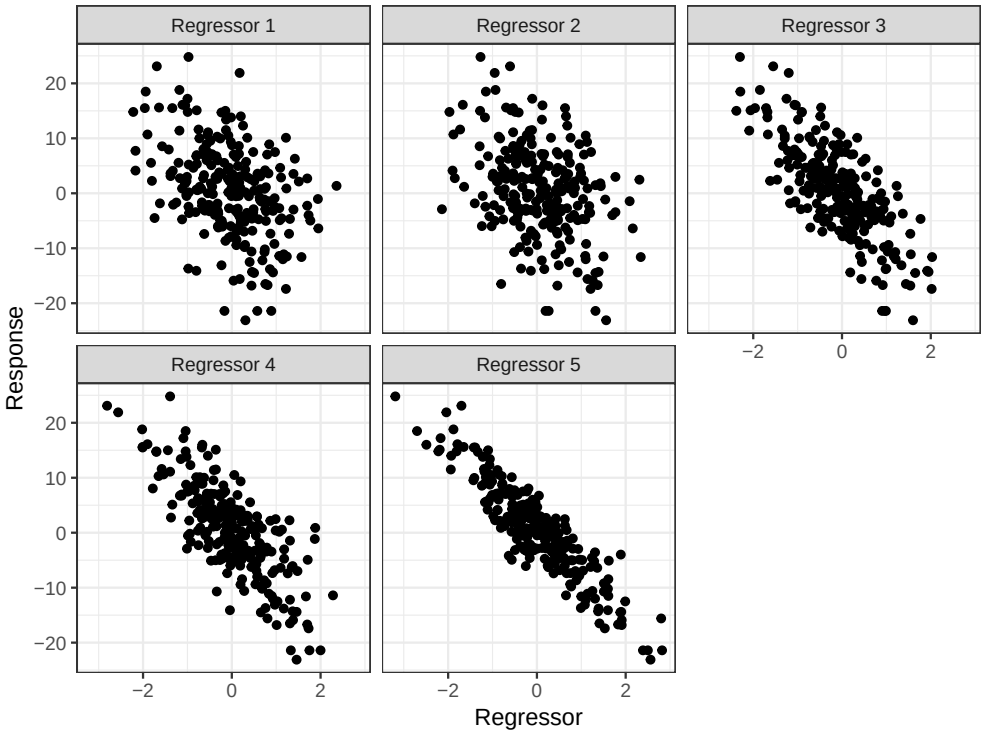

Figure 2.10: Checking for linear relationships

Table 2.3: 95 percent Confidence Intervals for model coefficients

Term	Estimate	Conf Low	Conf High
Intercept	-0.02	-0.15	0.11
Regressor 1	0.96	0.77	1.15
Regressor 2	1.87	1.68	2.07
Regressor 3	-3.01	-3.22	-2.81
Regressor 4	-4.06	-4.25	-3.86
Regressor 5	-4.88	-5.06	-4.70

all of the regressors look like they have a linear relationship with the response. There is an argument to be made that Regressor 1 and 2 are unhelpful in explaing the reponse but again we are not sure how they change with respect to the other regressors.

With that we have checked all the model assumptions. The only change we made was removing the large outlier from the dataset.

2.4 Part D

In Table 2.3 there are the 95% confidence intervals for the model coefficients.

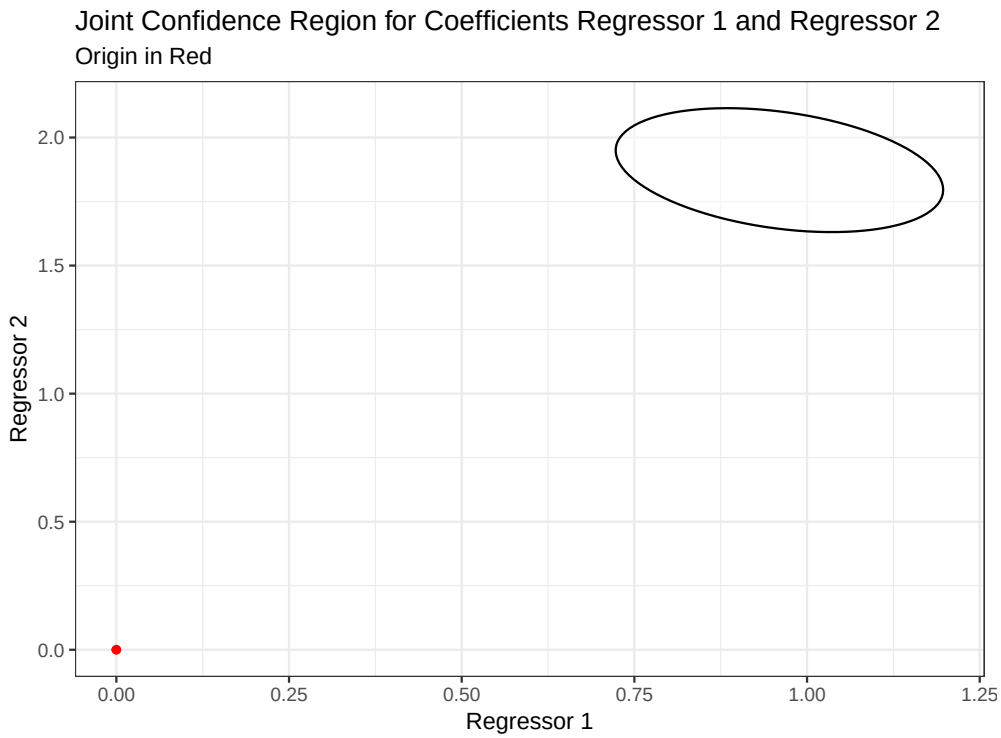

Figure 2.11: Joint 95 percent Confidence Interval for Regressor 1 and Regressor 2

2.5 Part E

In Figure 2.11 I plotted a joint 95% confidence region for β_1 and β_2. This region can be used to test the null hypothesis

$$\beta_1 = \beta_2 = 0 \tag{2.1}$$

Against the alternative hypothesis that

$$\beta_1 \neq 0 \mid \beta_2 \neq 0 \tag{2.2}$$

Since the region does not overlap with the original we can reject the null hypothesis in favour of the alternative hypothesis that both coefficients are not equal to 0 at the $\alpha = 0.05$ level.

2.6 Part F

Using the data with the outlier removed I fit ridge regression models for λ values between 0 and 50. As a result I have plotted in Figure 2.12 the coefficients of the various regressors at each λ value. Since the best lambda, the one with the smallesti cross validation mean squared error, is 0.01, the best model is actually the one without penalization. This is telling us that the model we should use is the one with the OLS coefficient estimates. These coefficient estimates are displayed in Table 2.2.

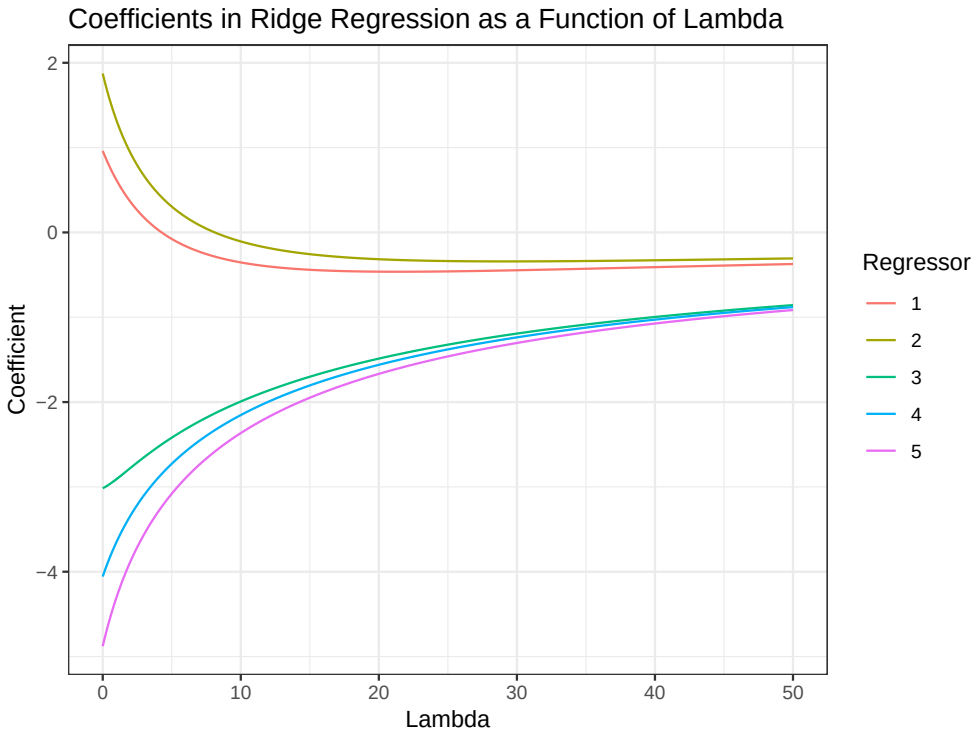

Figure 2.12: Ridge Regression Coefficients as Lambda Varies

Table 2.4: OLS regression coefficients with outlier removed rounded to nearest integer

Term	Estimate
Intercept	0
Regressor 1	1
Regressor 2	2
Regressor 3	-3
Regressor 4	-4
Regressor 5	-5

2.7 Part G

I performed forward selection using AIC as my evaluation metric and an implementation provided by the MASS package. Again, the full model is suggested as the best one. The coefficient estimates are in Table 2.2.

2.8 Part H

Given that both model selection techniques returned the OLS model, if I knew that each coefficient was originally an integer I would simply round the coefficients to the nearest integer to get the values in Table 2.4

Chapter 3

Question 3

3.1 Part A

I simulated 1000 different responses from the true model and estimated the model parameters using a OLS regression. The mean of the estimated model parameters over all the simulations are displayed in Table 3.1. They are quite close to the true value of the model parameters (as a reminder $\beta_0 = 1,\ \beta_1 = 2,\ \beta_2 = 3$). Additionally, in Table 3.2 we have the theoretical covariance matrix of the model parameters so that we can compare the mean estimated covariance matrix of the estimated model parameters from the 1000 simulations that is displayed in table 3.3

3.2 Part B

For the same 1000 simulated data sets I took a Bayesian approach to estimate the model parameters.

I used the following priors,

$$\beta \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, 5 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \tag{3.1}$$

$$\sigma^2 \sim IG(2, 0.5) \tag{3.2}$$

The marginal posterior distribution for the vector β is a multivariate t distribution which has mean μ_{new} where

$$\mu_{\text{new}} = (\Sigma^{-1} + X^T X)^{-1}(\Sigma^{-1}\mu + X^T y) \tag{3.3}$$

Table 3.1: Mean of model coefficients from OLS model over 1000 simulations

Term	Mean
Intercept	1
Regressor 1	2
Regressor 2	3

Table 3.2: Theoretical covariance matrix of the model parameters

Term	Intercept	Regressor 1	Regressor 2
Intercept	0.032600	0.000661	-0.052200
Regressor 1	0.000661	0.005310	-0.000863
Regressor 2	-0.052200	-0.000863	0.103000

Table 3.3: Mean of the estimated covariance matrix of the OLS model parameters over 1000 simulations

Term	Intercept	Regressor 1	Regressor 2
Intercept	0.032400	0.000657	-0.051900
Regressor 1	0.000657	0.005270	-0.000857
Regressor 2	-0.051900	-0.000857	0.103000

and the marginal posterior distribution for σ^2 is a Inverse Gamma distribution with parameters a_{new} and b_{new} which are defined to be,

$$a_{\text{new}} = a + n/2 \tag{3.4}$$

$$b_{\text{new}} = b + \frac{1}{2}[\mu^T \Sigma^{-1} \mu + y^T y - \mu_{\text{new}}^T \Sigma_{\text{new}}^{-1} \mu_{\text{new}}] \tag{3.5}$$

and

$$\Sigma_{\text{new}} = (\Sigma^{-1} + X^T X)^{-1}$$

the mean of an Inverse Gamma distribution is $\frac{b}{a-1}$

The mean of the estimated model parameters over all the simulations using the posterior means are displayed in Table 3.4. They are a little farther away from the true parameters than the OLS mean estimates. Regressor 1 is close to the true value.

We can compare the mean estimated covariance matrix of the estimated model parameters from the 1000 simulations that is displayed in table 3.5 to the theoretical covariance matrix of the model parameters that I have reprinted in table 3.6 for convenience. We are much farther away from the theoretical covariance matrix then in the linear model case.

3.3 Part C

I changed the priors according to part c to be,

Table 3.4: Mean of model coefficients from Bayesian model in part b over 1000 simulations

Regressor	Mean
Intercept	1.31
Regressor 1	1.96
Regressor 2	2.32

Table 3.5: Estimated covariance matrix of the model parameters for bayesian model in part b

Term	Intercept	Regressor 1	Regressor 2
Intercept	0.06980	0.00142	-0.11200
Regressor 1	0.00142	0.01140	-0.00185
Regressor 2	-0.11200	-0.00185	0.22100

Table 3.6: Theoretical covariance matrix of the model parameters

Term	Intercept	Regressor 1	Regressor 2
Intercept	0.032600	0.000661	-0.052200
Regressor 1	0.000661	0.005310	-0.000863
Regressor 2	-0.052200	-0.000863	0.103000

$$\beta \sim N\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, 1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \right) \tag{3.6}$$

$$\sigma^2 \sim IG(2, 0.5) \tag{3.7}$$

I then used the same update rule to find the means of the posterior distributions for the vector β and for σ^2.

The mean of the estimated model parameters over all the simulations using the posterior means are displayed in Table 3.7. They are a closer to the true values than the first prior was (which brought the values closer to 0) but still a little farther away from the true parameters than the OLS mean estimates.

We are closer in table 3.8 to the theoretical covariance matrix of the model parameters in table 3.9 than the first bayesian approach was.

3.4 Part D

Finally, in part d I used the following priors,

$$\beta \sim N\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, 0.1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \tag{3.8}$$

$$\sigma^2 \sim IG(2, 0.5) \tag{3.9}$$

Table 3.7: Mean of model coefficients from Bayesian model in part c over 1000 simulations

Regressor	Mean
Intercept	1.11
Regressor 1	1.99
Regressor 2	2.78

Table 3.8: Estimated covariance matrix of the model parameters for bayesian model in part c

Term	Intercept	Regressor 1	Regressor 2
Intercept	0.037100	0.000751	-0.05930
Regressor 1	0.000751	0.006020	-0.00098
Regressor 2	-0.059300	-0.000980	0.11700

Table 3.9: Theoretical covariance matrix of the model parameters

Term	Intercept	Regressor 1	Regressor 2
Intercept	0.032600	0.000661	-0.052200
Regressor 1	0.000661	0.005310	-0.000863
Regressor 2	-0.052200	-0.000863	0.103000

The mean of the estimated model parameters over all the simulations using the posterior means are displayed in Table 3.10. They are a little closer to the true parameters than the ones from part b but not by much. They are farther away from the true parameters than those in part c.

The estimated covariance matrix of the estimated model parameters in table 3.11 are father away from the theoretical ones in table 3.12 than the matrix from part b but again not as close as the matrix in part c.

3.5 Part E

Knowing the true parameter values, the Bayesian model in part c (with prior pulling β s to 1 and 1, 2 and 3 down the diagonal of the Σ matrix) was the most accurate over the 1000 simulations. We can see that the model in part b pulled the estimate for Regressor 2 too far towards 0. It is interesting to see that the estimate for the Intercept was actually an overestimate.

3.6 Part F

If I did not know the true parameter values and I was handed a dataset where I had no knowledge of the situation and no one to ilicit a prior belief from I think I would use the OLS frequentist approach. It is clear that in this case that the frequentist approach is better. However, I think if there was more information about the problem it would be better to incorporate it into the model and take a Bayesian approach.

Table 3.10: Mean of model coefficients from Bayesian model in part d over 1000 simulations

Regressor	Mean
Intercept	1.27
Regressor 1	1.98
Regressor 2	2.46

Table 3.11: Estimated covariance matrix of the model parameters for bayesian model in part d

Term	Intercept	Regressor 1	Regressor 2
Intercept	0.044100	0.000892	-0.07050
Regressor 1	0.000892	0.007160	-0.00116
Regressor 2	-0.070500	-0.001160	0.14000

Table 3.12: Theoretical covariance matrix of the model parameters

Term	Intercept	Regressor 1	Regressor 2
Intercept	0.032600	0.000661	-0.052200
Regressor 1	0.000661	0.005310	-0.000863
Regressor 2	-0.052200	-0.000863	0.103000

Chapter 4

Question 4

4.1 Part A

In Figure 4.1 we can begin to explore the relationships of Density and Impurity with Absorbtion Rate. We can see that Impurity is Low that the Absorbtion Rate has a similar mean regardless of the density of the graphite. However, if the graphite impurity is high then there is fairly clear difference in absorbtion rate for high density and low density graphite.

4.2 Part B

A suitable model for this experiment would be

$$y = \beta_0 + \beta_1 I(\text{Impurity} = \text{Low}) + \beta_2 I(\text{Density} = \text{Low}) + \beta_3 I(\text{Impurity} = \text{Low}) I(\text{Density} = \text{Low}) + \epsilon \quad (4.1)$$

This uses the High/High Impurity/Density combination the baseline Intercept and the other effects are deviations form this baseline. $\epsilon \sim N(0, \sigma^2)$

4.3 Part C

I tested for each β_i the null hypothesis that $\beta_i = 0$ against the alternative hypothesis $\beta_i \neq 0$. The test statistic reported is a t-statistic with 77 degrees of freedom. We will perform these tests each at a $\alpha = 0.05$ level which does not adjust for the multiple comparisons that are being made. In Table 4.1 the model estimates and test results are displayed. All tests for the β_i have the null hypothesis that $\beta_i = 0$ rejected. I will now test furthur to see which levels of combinations have different means.

To test the which levels of combinations have different means I perforemd a Tukey Honest Significnat Differences test at a $\alpha = 0.05/10$ level. This is to account for the 4 tests we made in the first part of this question and the 6 tests we are doing here for the combinations of levels. The results of this test are outputted in table 4.2. There is only one comparison of combinations which is not significantly different.This was the combination we pointed out in the original plot. That is that when impurity is low the absorbtion rate is not affected by the density of the graphite.

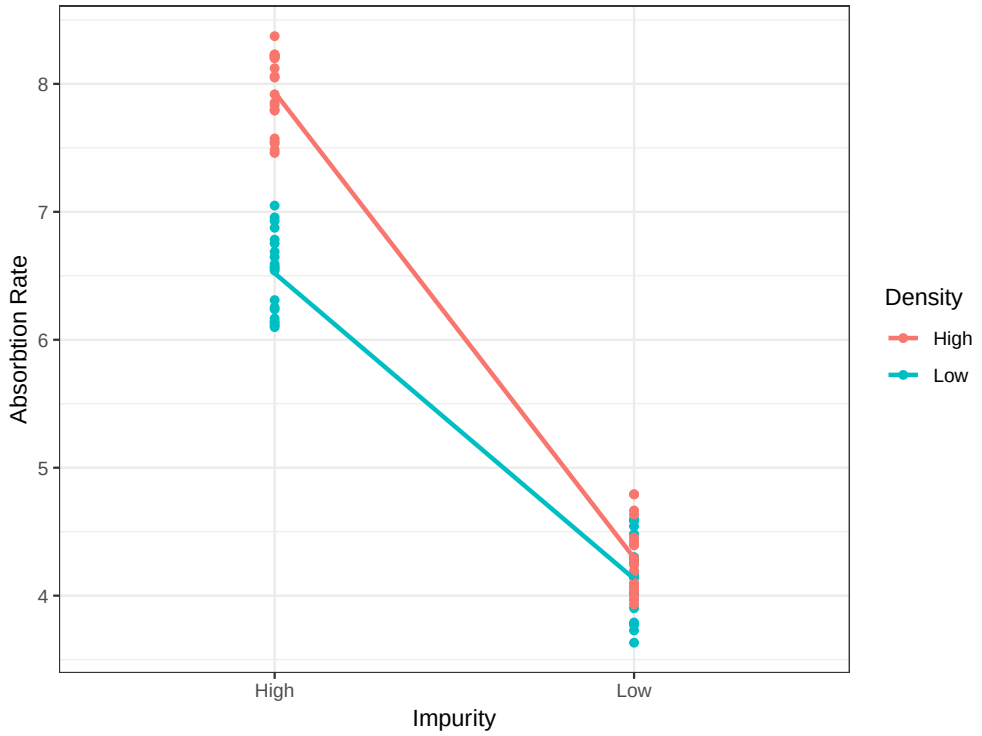

Figure 4.1: Radioactive neutrons and graphite bricks

Table 4.1: Beta estimates and test results

Term	Estimate	Std Error	Statistic	P Value
Intercept	7.9	0.066	120.0	0
Impuritylow	-3.6	0.093	-39.0	0
Densitylow	-1.4	0.093	-15.0	0
Impuritylow Densitylow	1.3	0.130	9.5	0

Table 4.2: Tukey HSD test for difference in menas in different combinations of levels

Term	Comparison	Estimate	Conf Low	Conf High	Adj P Value
impurity:density	Low:High-High:High	-3.640	-3.960	-3.320	0.000
impurity:density	High:Low-High:High	-1.420	-1.740	-1.090	0.000
impurity:density	Low:Low-High:High	-3.800	-4.120	-3.480	0.000
impurity:density	High:Low-Low:High	2.220	1.900	2.550	0.000
impurity:density	Low:Low-Low:High	-0.163	-0.486	0.159	0.306
impurity:density	Low:Low-High:Low	-2.390	-2.710	-2.070	0.000

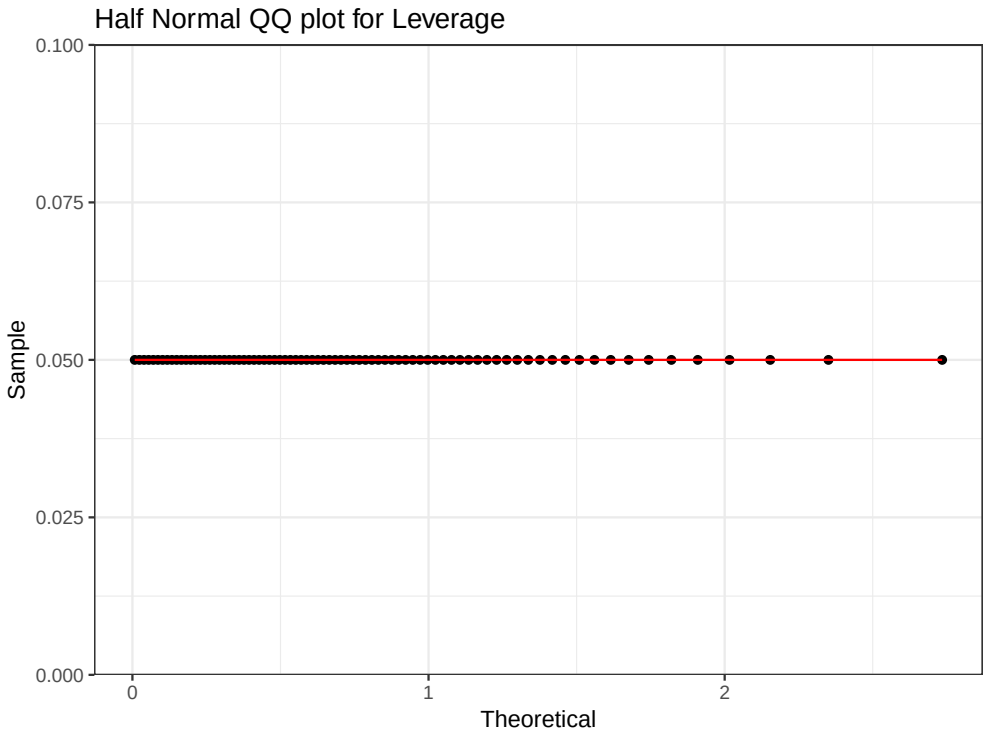

Figure 4.2: Searching for highly influential points on the model

4.4 Part D

Similarily to Question 2 I will check for influential outliers and those that seem like they not belong, I will check for equal variance, normality, and independence amongst the residuals.

4.4.1 Outliers

There were no points that had high leverage. In fact, we can see in Figure 4.2 that due to the design all the points have the same leverage.

Next we look for any highly irregular points used cooks distance. In Figure 4.3 we can see that there are small deviations from the half normal expected quantiles. However, the scale of the graph is very small and so there are no points that I will examine furthur.

4.4.2 Errors

Next I will check for constant variance in the residuals using the Residuals vs Fitted plot in Figure 4.4. There is nothing in the graph that leads me to believe that the variance is not consistent across the fitted values.

Now I will check the normality of the residuals. In Figure 4.5 the distribution looks very flat, possibly uniform.

In fact if we plot the Uniform(-1, 1) qq plot (like in Figure 4.6) we can see that the residuals probably do follow a uniform distribution. I think that the model assumption can still be approximately satisfied.

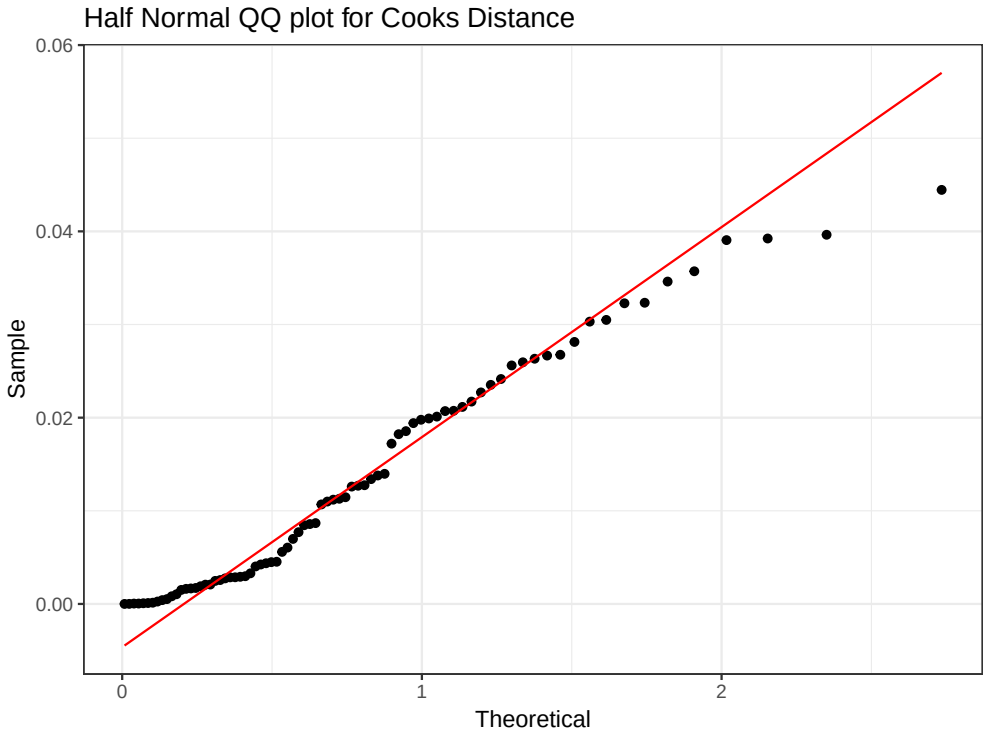

Figure 4.3: Searching for irregular points

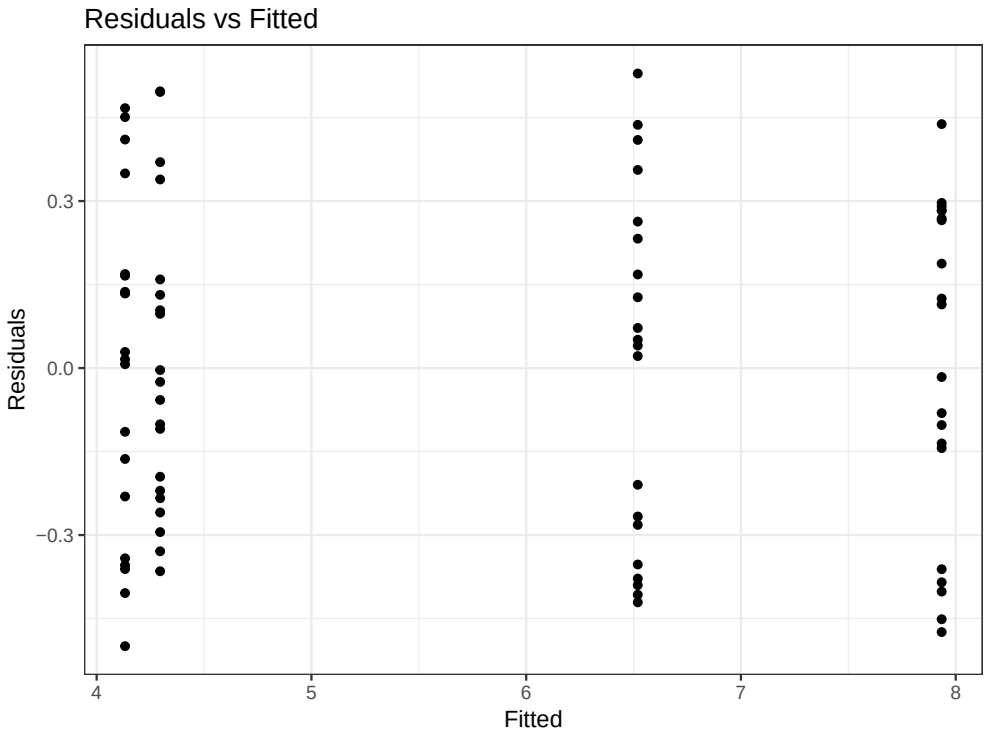

Figure 4.4: Looking for non-constant variance

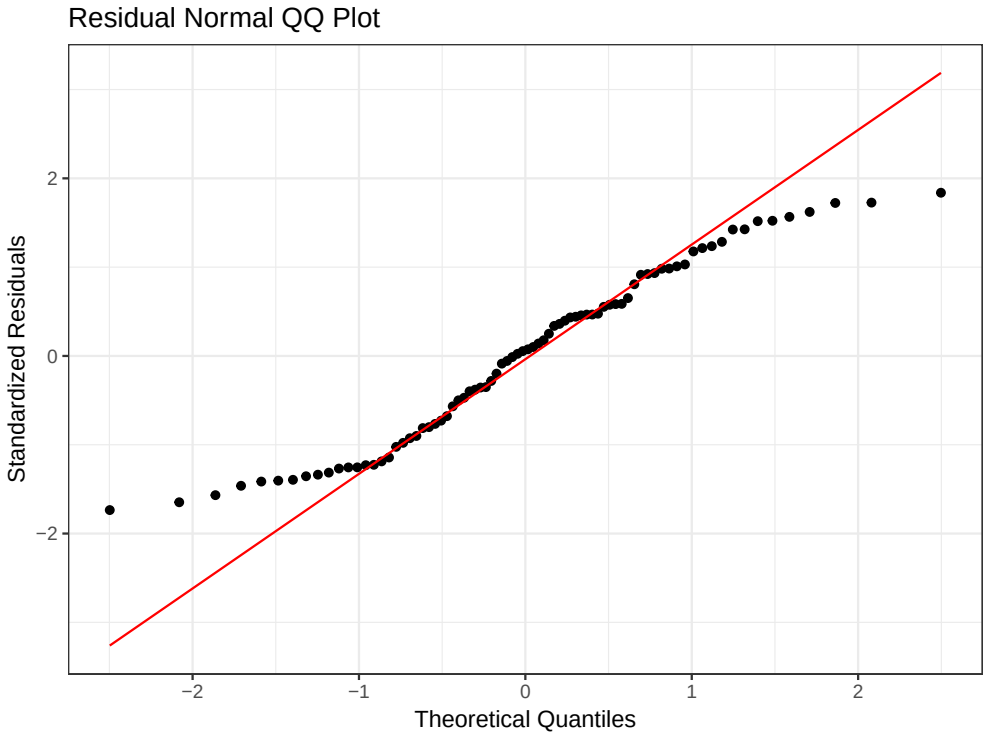

Figure 4.5: Normal QQ plot for residuals

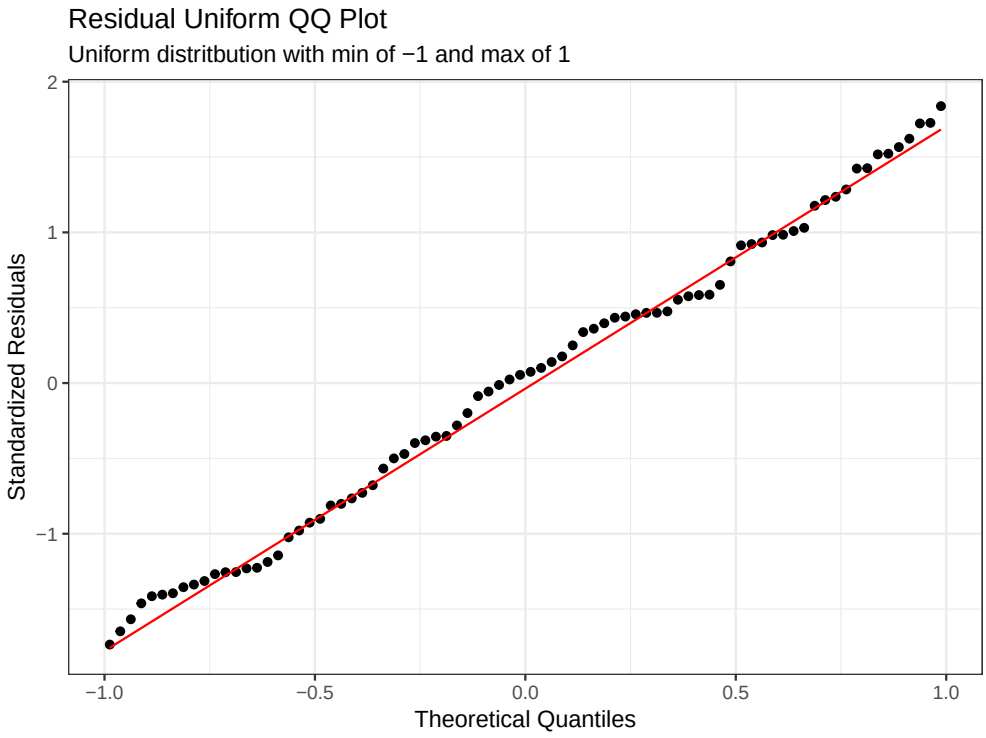

Figure 4.6: Uniform QQ plot for residuals

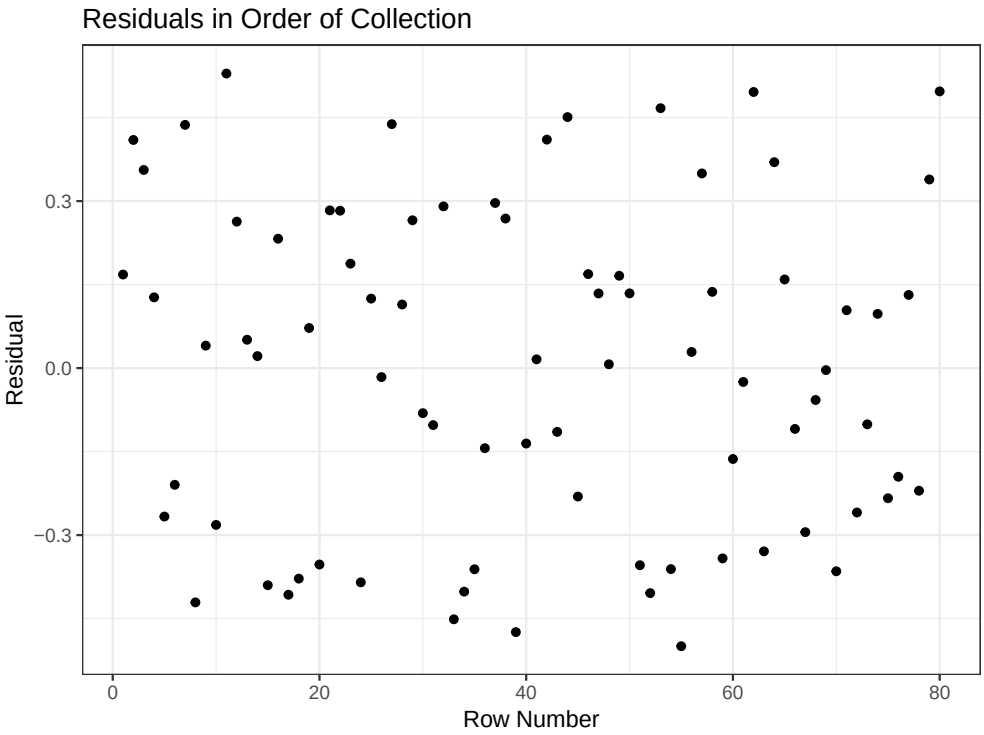

Figure 4.7: Independence of Order Reported

Finally, we do not have access to the researchers to ask them questions about the data collection process so the only check for independence amongst the residuals that I will do will be to plot the residuals in the order that they were reported in the text file to make sure there is no autocorrelation. This is done in Figure 4.7

Overall, I do not think that any of the assumptions are not at least approximately satisified. I believe it is important to note that the residuals look to be uniformly distributed but that it does not change our inferences.

4.5 Part E

If I had a balanced expereiment with no interaction between the factors and I was searching for an difference in means of 1 unit or more that can be detected 80% of the time with significance level 0.05 then I could run a power analysis to determine the sample size needed.

Since we are assuming there is no interaction between factors we can treat this like 2 independent 2 sample t tests. The power anlaysis to determine the sample size needed in a 2 sample t test depends on the alpha level (0.05), the power (0.80), the effect size (1) and the pooled standard deviation (not given in the question).

For impurity, the pooled standard deviation for the two groups is 0.5875189 and for density the pooled standard deviation for the two groups is 1.5854313. With these values we can compute the sample size needed for one group in each case independently. We can then take the larger of the two values as the sample size needed for one group. We can multiply that value by 2 to get the total sample size we need.

To calculate the sample size needed for one group in each caes I used the following formula

$$n_{\text{var}} = \left(\sigma_{\text{var}}(z_{0.975} + z_{0.80})/1\right)^2 \tag{4.2}$$

For impurity the sample size needed for one group is 6 and for density it is 40. The bigger of the two is 40 so 80 is the total sample size needed for the experiment.

Chapter 5

Question 5

5.1 Part A

This experimental design is called Latin Squares since each treatment is assigned to each block once and only once and there are two blocking variables each with the same amount of levels (4) which is also equal to the number of treatments.

5.2 Part B

I made two plots (Figure 5.1) to investigate the effect of the blocking factors and the treatment on sales. Since we have such little data it is hard to tell if there is an effect or interactions between the blocking factors and treatment. With this data there seems to be interactions between the blocking factors and the treatment.

5.3 Part C

We cannot estimate interactions in the Latin Square so an appropriate model would be

$$y_{h,i,j} = \mu + \alpha_h + \beta_i + \delta_j + \epsilon_{h,i,j} \tag{5.1}$$

where $h = 1, 2, 3, 4$ is the week number, $i = 1, 2, 3, 4$ is the store number and $j = 1, 2, 3, 4$ is the treatment number. $\epsilon \sim N(0, \sigma^2)$. We have to assume that the treatment and two blocking factors do not interact. Note that the μ is a basleline with the first level of each blocking factor and treatment absorbed in. There is an argument to be made to treat week as a random effect and a smaller case to treat store as a random effect (if there are other stores that we care about we should), however, I have not done so here.

The estimated parameters for the described model are shown in Table 5.1.

5.4 Part D

Similarily to Question 2 and Question 4 I will check for influential outliers and those that seem like they not belong, I will check for equal variance, normality, and independence amongst the residuals. We have assumed also here that there is no interactions between the regressors. We saw in the initial plot that this may not be exactly the case.

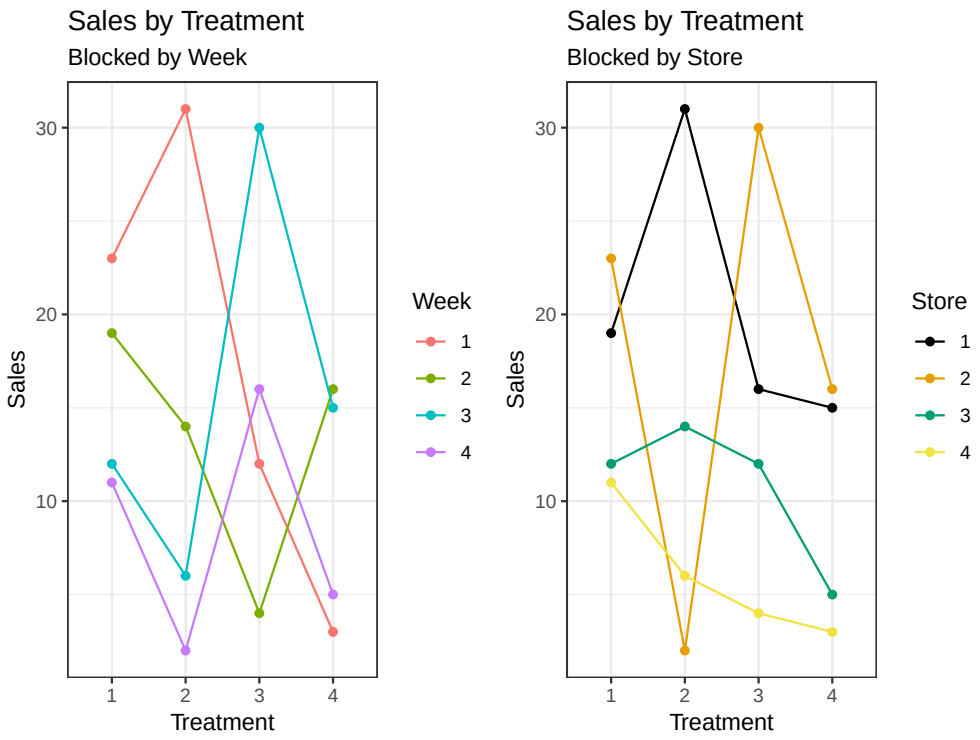

Figure 5.1: Investigating effect of blocking factors and treatment on sales

Table 5.1: Model coefficients

Term	Estimate
Intercept	26.38
Treatment2	-3.00
Treatment3	-0.75
Treatment4	-6.50
Week2	-4.00
Week3	-1.50
Week4	-8.75
Store2	-2.50
Store3	-9.50
Store4	-14.25

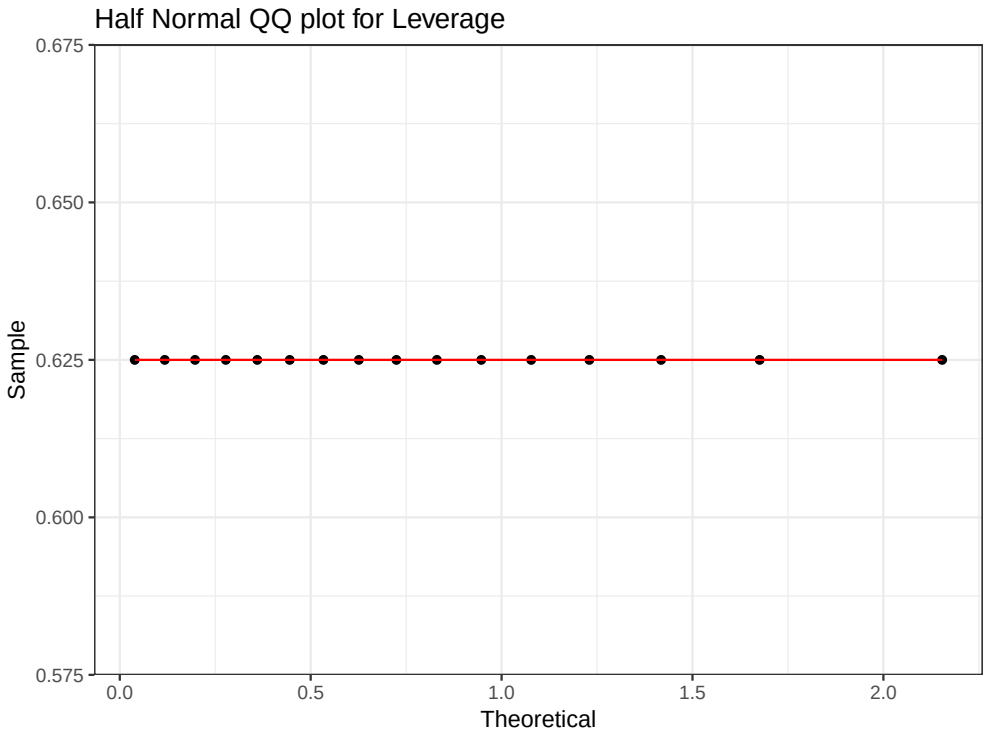

Figure 5.2: Searching for influential points on the model

5.4.1 Outliers

Like in Question 4 due to the design all the points have the same leverage. Therefore Figure 5.2 is not very interesting.

There are some very large outliers as can be seen in 5.3. This may be because we have not accounted for interactions between regressors. However, there is no real action to take since we have such little data.

5.4.2 Resiudals

Since we have such few points it is tough to see if there is any non-constant variance in the residuals. However, with what we have in Figure 5.4 I think this is a safe assumption.

With only 16 points I do not see enough evidence to believe that the residuals are not normal when looking at Figure 5.5.

Finally, I check in Figure 5.6 if there is an effect based on the order the data was collected. I do not see anything to indicate theres anything wrong.

5.5 Part E

I will test to see if there is an effect of treatment on sales. I will do so by taking the null hypothesis that $\text{treatment}_1 = \text{treatment}_2 = \text{treatment}_3 = \text{treatment}_4 = 0$. For the alternative that atleast one of the effects is not equal to 0. The test statisitc follows an F distribution. The results are displayed in Table 5.2. Since the p value is greater than 0.05 we will fail to reject the null hypothesis. Of course this is not terribly surprising considering the small size of the data set.

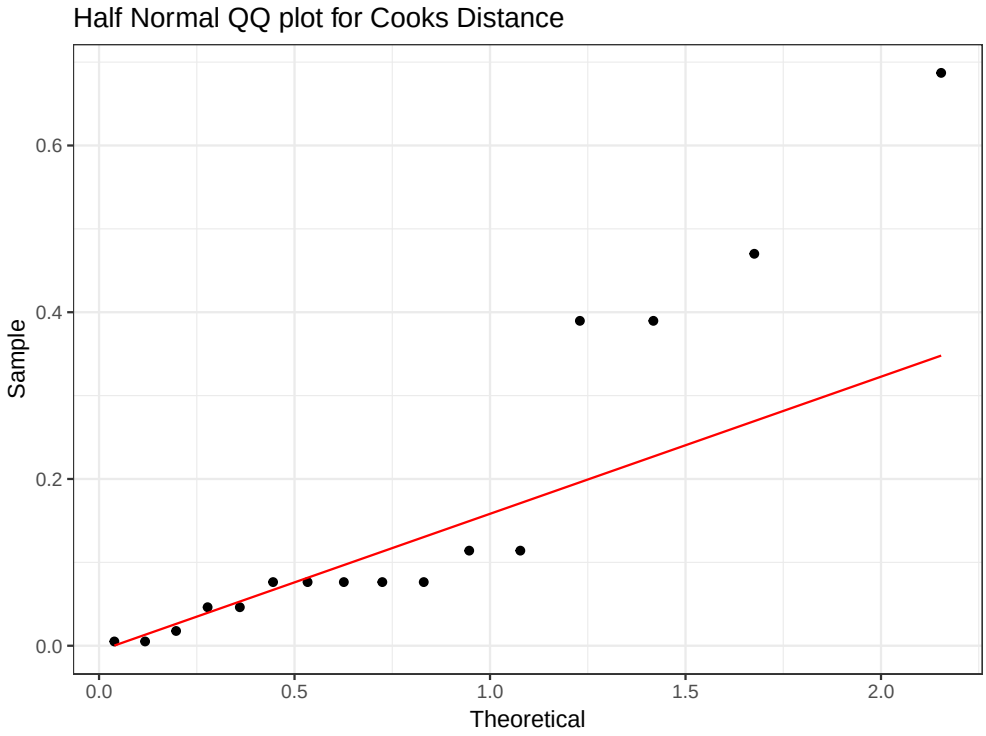

Figure 5.3: Searching for Outliers

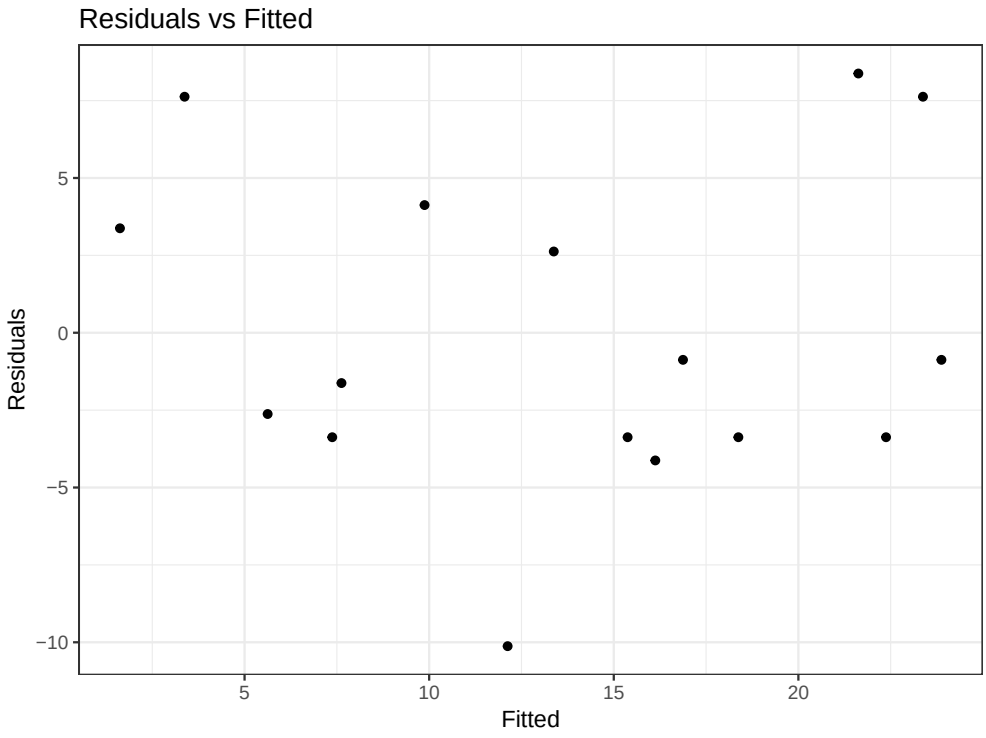

Figure 5.4: Looking for Non-Constant Variance

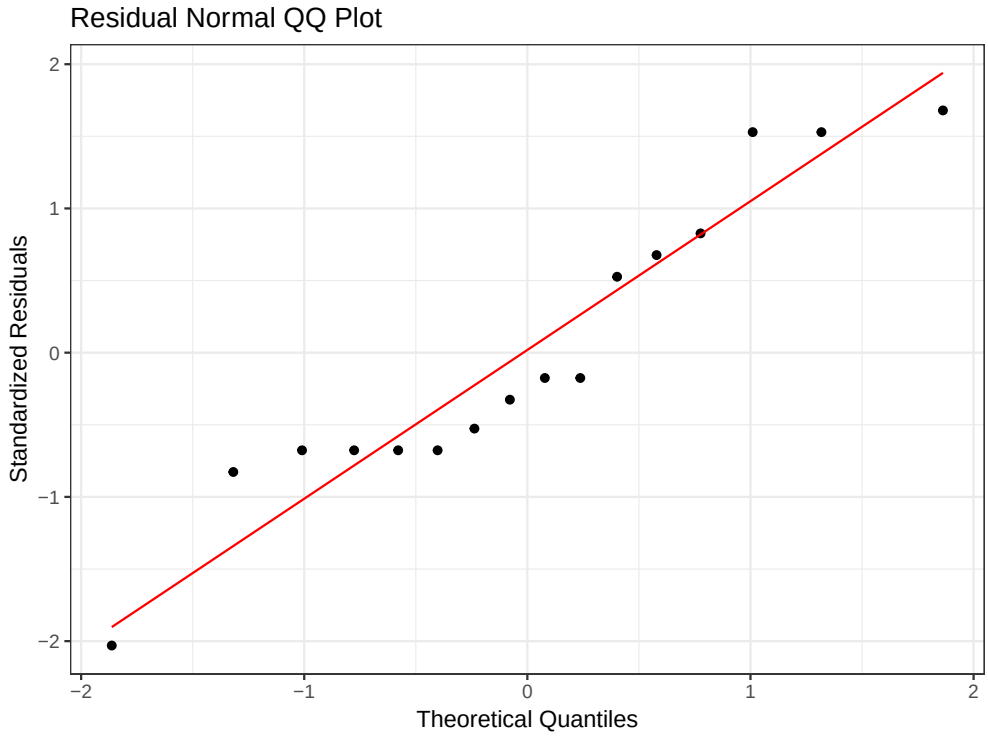

Figure 5.5: Are the residuals normal

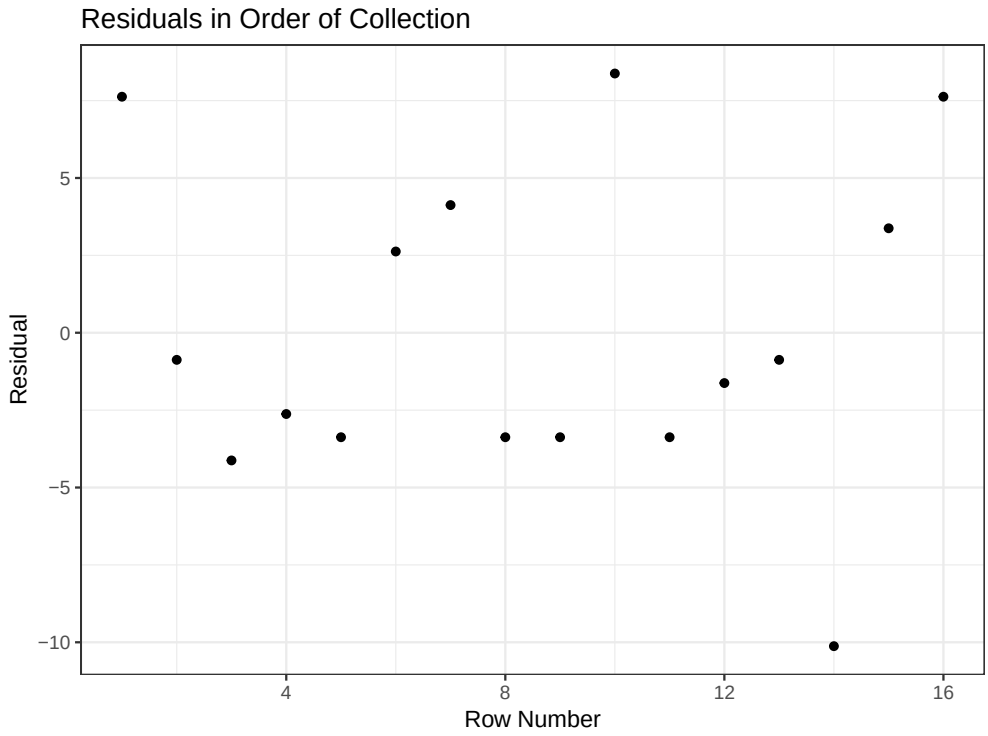

Figure 5.6: Order of Collection

Table 5.2: Test for the effect of Treatment on Sales

Term	Statistic	P Value
Treatment	0.51	0.69

5.6 Part F

The model I would propose would be,

$$y_{h,i,j} = \mu + \alpha_h + \beta_i + \delta_j + \epsilon_{h,i,j} \tag{5.2}$$

W where $h = 1, 2, 3, 4, 5, 6, 7, 8$ is the week number, $i = 1, 2, 3, 4, 5, 6, 7, 8$ is the store number and $j = 1, 2, 3, 4$ is the treatment number. $\epsilon \sim N(0, \sigma^2)$.