

STAT 608 Homework 4

Jack Cunningham

1)

a)

For the model $y_i = \alpha + e_i$ we have the design matrix X which has dimensions $n \times 1$:

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

So the least squares estimate for $\hat{\alpha}$ is:

$$\hat{\alpha} = (X^T X)^{-1} X^T y = \frac{1}{n} \sum_{i=1}^n y_i$$

b)

To create constant variance we use the below weights:

$$w_i = \begin{cases} 1 & 1 \leq i \leq n_1 \\ 2 & n_1 < i \leq n \end{cases}$$

Then weight matrix W is a n by n matrix with diagonal entries w_i . The weighted least squares estimator for α is $\hat{\alpha}_{\text{WLS}} = (X^T W X)^{-1} X^T W y$:

$$\hat{\alpha}_{\text{WLS}} = \frac{1}{2n - n_1} \left(\sum_{i=1}^{n_1} y_i + 2 \sum_{i=n_1+1}^n y_i \right)$$

c)

The expected value of $\hat{\alpha}$ is:

$$E[\hat{\alpha}] = (X^T X)^{-1} X^T E[y]$$

$$E[y] = X\alpha$$

$$E[\hat{\alpha}] = (X^T X)^{-1} X^T X\alpha = \alpha$$

The expected value of $\hat{\alpha}_{\text{wls}}$ is:

$$E[\hat{\alpha}_{\text{WLS}}] = (X^T W X^{-1}) X^T W E[y]$$

$$E[y] = X\alpha$$

$$E[\hat{\alpha}_{\text{WLS}}] = (X^T W X^{-1}) X^T W X\alpha = \alpha$$

Before computing variances for both estimates it is helpful to find $Var(y)$:

$$Var(y) = Var(X\alpha + e) = Var(e)$$

$$Var(e) = \text{diagonal } n \times n \text{ matrix with } \begin{cases} \Sigma_{i,i} = \sigma^2 & i = 1, 2, \dots, n_1 \\ \Sigma_{i,i} = \sigma^2/2 & i = n_1 + 1, n_1 + 2, \dots, n \end{cases}$$

Variance for $\hat{\alpha}$ is:

$$Var(\hat{\alpha}) = (X^T X)^{-1} X^T Var(y) X (X^T X)^{-1}$$

$$X^T X = n$$

$$Var(\hat{\alpha}) = \frac{1}{n^2} \left(\sum_{i=1}^{n_1} \sigma^2 + \sum_{i=n_1+1}^n \sigma^2/2 \right) = \frac{3}{4} \left(\frac{1}{n} \sigma^2 \right)$$

Variance for $\hat{\alpha}_{\text{WLS}}$ is:

$$Var(\hat{\alpha}_{WLS}) = (X^T W X)^{-1} X^T W Var(y) W^T X (X^T W X)^{-1}$$

$$X^T W X = \frac{3}{2}n$$

$$Var(\hat{\alpha}_{WLS}) = (\frac{2}{3})^2 \frac{1}{n^2} X^T W Var(y) W^T X = \frac{2}{3}(\frac{1}{n}\sigma^2)$$

To assess which estimator is better we can use theoretical MSE:

$$MSE(\hat{\alpha}) = Bias^2(\hat{\alpha}) + Var(\hat{\alpha})$$

The definition of Bias is $Bias = E[\hat{\theta} - \theta]$, so since both $E[\hat{\alpha}] = \alpha$, $E[\hat{\alpha}_{WLS}] = \alpha$ $Bias^2 = 0$ for both $\hat{\alpha}$ and $\hat{\alpha}_{WLS}$.

Then:

$$MSE(\hat{\alpha}) = Bias^2(\hat{\alpha}) + Var(\hat{\alpha}) = \frac{3}{4}(\frac{1}{n}\sigma^2)$$

$$MSE(\hat{\alpha}_{WLS}) = Bias^2(\hat{\alpha}_{WLS}) + Var(\hat{\alpha}_{WLS}) = \frac{2}{3}(\frac{1}{n}\sigma^2)$$

Since the theoretical MSE of $\hat{\alpha}_{WLS}$ is smaller than that of $\hat{\alpha}$ we would say that $\hat{\alpha}_{WLS}$ is a better estimator.

2)

We have the below regression model with $Var(e_i|x_i) = x_i^2\sigma^2$:

$$Y_i = \beta x_i + e_i$$

To find the weighted least squares estimate of β we use the weighted version of the residual sum of squares where the weights $w_i = 1/x_i^2$:

$$WRSS = \sum_{i=1}^n \frac{1}{x_i^2} (y_i - \beta x_i)^2$$

To find $\hat{\beta}$ we take the derivative with respect to β and set equal to zero:

$$\frac{d}{d\beta} \left(\sum_{i=1}^n \frac{1}{x_i^2} (y_i - \beta x_i)^2 \right) = \sum_{i=1}^n 2(-x_i) \left(\frac{1}{x_i^2} \right) (y_i - \beta x_i)$$

$$-2 \sum_{i=1}^n \frac{1}{x_i} (y_i - \beta x_i) = 0$$

$$-2 \sum_{i=1}^n \frac{y_i}{x_i} + 2 \sum_{i=1}^n \beta = 0$$

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{x_i}$$

3)

a)

The response variable Y_i is the 2006 median price per square foot. Since Y_i is measure of a median of 1922 subdivisions we use weighted least squares to deal with non-constant variance. The weights $w_i = n_i$ is appropriate since the variance of $Var(Y_i)$ is proportional to n, since $Var(Y_i) = \frac{1}{n} \sum_{i=1}^n Var(u_i)$ where u_i the variance of each group of houses with the same square foot.

b)

There is evidence of non-constant variance in the plot of $\sqrt{\text{Std. Residuals}}$ by Fitted values. As Fitted Values increases the variance of residuals increases on average.

c)

Since weighted least squares did not create a constant variance structure I would instead fit a simple linear regression model and then explore transformations to remedy the non-constant variance.

4)

a)

The response variable Y is the sum of n_i coin weights. Then $Var(y_i) \propto n_i$. So we have $w_i = 1/n_i$. Therefore we have the below weight matrix W:

$$W = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{bmatrix}$$

b)

Using the weighted least squares solution for $\hat{\beta}_w$ we have:

$$\hat{\beta}_w = (X^T W X)^{-1} X^T W y$$

$$X^T W = \begin{bmatrix} 1 & 0 & 1/2 & 1/2 \\ 0 & 1 & 1/2 & 1/2 \end{bmatrix}$$

$$X^T W X = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$(X^T W X)^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$X^T W y = \begin{bmatrix} y_1 + y_3/2 + y_4/2 \\ y_2 + y_3/2 + y_4/2 \end{bmatrix}$$

$$\hat{\beta}_w = (X^T W X)^{-1} X^T W y = \frac{1}{3} \begin{bmatrix} 2y_1 + (y_3 - y_2)/2 + (y_4 - y_2)/2 \\ 2y_2 + (y_3 - y_1)/2 + (y_4 - y_1)/2 \end{bmatrix}$$

5)

a)

We can rewrite $\tilde{\beta}_1$ as:

$$\tilde{\beta}_1 = \frac{1}{5} \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix}^T y$$

Bias of $\tilde{\beta}_1$ is $E[\tilde{\beta}_1 - \beta_1] = E[\tilde{\beta}_1] - \beta_1$. In order for $\tilde{\beta}_1$ to be unbiased $E[\tilde{\beta}_1]$ must equal β_1 . We calculate $E[\tilde{\beta}_1]$:

$$E[\tilde{\beta}_1] = \frac{1}{5} \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix}^T E[y|x_i]$$

$$E[y|x_i] = \beta_0 + \beta_1 x_i$$

$$E[\tilde{\beta}_1] = \frac{1}{5} \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix}^T \begin{bmatrix} \beta_0 + \beta_1 \\ \beta_0 + 2\beta_1 \\ \beta_0 + 3\beta_1 \\ \beta_0 + 4\beta_1 \end{bmatrix} = \beta_1$$

The estimator $\tilde{\beta}_1$ is unbiased.

b)

$$Var(\tilde{\beta}_1) = \left(\frac{1}{5}\right)^2 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix}^T Var(y) \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix}$$

$$Var(y) = \sigma^2 I_n$$

$$Var(\tilde{\beta}_1) = \left(\frac{1}{5}\right)^2 \sigma^2 (10) = \frac{2}{5} \sigma^2$$

c)

The least squares estimator $\hat{\beta}_1$ has the below variance:

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SXX} = \frac{1}{5} \sigma^2$$

The least squares estimator $\hat{\beta}_1$ has half the sampling variance of $\tilde{\beta}_1$.

6)

a)

I would include an interaction term. No one is eating a hot dog bun by itself, they are eating it with the hot dog. In that sense the preferred bun is the one that contributes to the tastiest combined product. An interaction term allows the effect of additional sodium to differ between the two buns and the company to explore what combination of sodium and bun creates the superior product.

b)

The linear model with an interaction term would be:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + e_i$$

With $x_1 = 1\{\text{Type A}\}$. Then the model for each type is below:

$$\text{Type A : } y_i = (\beta_0 + \beta_1) + (\beta_2 + \beta_3)x_2 + e_i$$

$$\text{Type B : } y_i = \beta_0 + \beta_2x_2 + e_i$$

The interpretation of parameters follows:

The parameter β_0 is the taste of a Bun B hot dog with no sodium content on average.

The parameter β_1 is the difference between the taste of a Bun A hot dog and a Bun B hot dog with no sodium content on average.

The parameter β_2 is the effect an additional unit of sodium has on the taste of a Bun B hot dog on average.

The parameter β_3 is the difference in taste an additional unit of sodium has on Bun A as opposed to Bun B on average.

The hypothesis we would test for a difference in groups would be $H_0 : \beta_1 = \beta_3 = 0, H_A : \text{One of } \beta_1, \beta_2 \neq 0$.

c)

We interpreted β_3 in the previous question. The interaction term allows us to examine how sodium affects Bun A hot dog taste compared to how it affects Bun B hot dog taste.

7)

A one-way ANOVA model with $k = 3$ groups and 4 observations per group without an intercept has the below model:

$$y_i = \beta_1x_{1i} + \beta_2x_{2i} + \beta_3x_{3i} + e_i$$

$$x_{1i} = 1\{\text{Category 1}\}, x_{2i} = 1\{\text{Category 2}\}, x_{3i} = 1\{\text{Category 3}\}$$

The design matrix X is:

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Then the least squares estimate $\hat{\beta}$ is:

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{bmatrix} \frac{1}{4} \sum_{i=1}^4 y_i \\ \frac{1}{4} \sum_{i=5}^8 y_i \\ \frac{1}{4} \sum_{i=9}^{12} y_i \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

a)

The hypothesis we are testing is $H_0 : (\mu_1 + \mu_2)/2 = \mu_3$. Since $E[\beta_i] = \mu_i$ we can rewrite the null hypothesis as:

$$H_0 : \begin{bmatrix} 1/2 & 1/2 & -1 \end{bmatrix} \beta = 0, H_a = \text{Not equal to zero}$$

We test this using the F statistic:

$$F = \frac{(A\hat{\beta} - h)^T (A(X^T X)^{-1} A^T) (A\hat{\beta} - h) / r}{\hat{e}^T \hat{e} / (n - p - 1)}$$

We can simplify this statistic:

$$(A\hat{\beta} - h) = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3$$

$$(X^T X)^{-1} = \text{diag}(1/4, 1/4, 1/4)$$

$$(A(X^T X)^{-1} A^T)^{-1} = 16/6$$

$$r = 1$$

$$\hat{e}^T \hat{e} / (n - p - 1) = SS_{\text{Error Full}} / df_{\text{Error Full}}$$

$$df_{\text{error Full}} = 12 - 3 = 9 \text{ (No intercept term here)}$$

So F equals:

$$F = \frac{16}{6} \frac{(\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3)^2}{SS_{\text{Error Full}/9}}$$

b)

```
mu_1 <- 5.6
mu_2 <- 7.9
mu_3 <- 6.1
SSE <- 12.8
F <- 16/6*(1/2*mu_1+1/2*mu_2-mu_3)^2/(SSE/9)
F
```

```
[1] 0.7921875
```

The F statistic in this case is $f = 0.7921875$.

```
alpha <- 0.05
p_value <- 1 - pf(F, 1, 9)
```

The computed p-value is $P(F_{1,9} > f) = 0.3966074$. We cannot reject the null hypothesis that $(\mu_1 + \mu_2)/2 = \mu_3$.

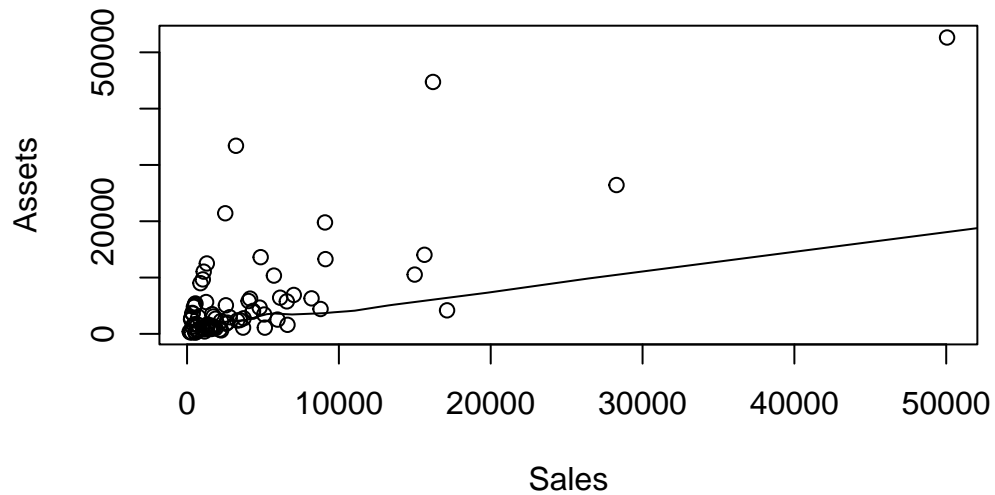
8)

Loading Company Data

```
company <- read.csv("company.csv")
attach(company)
```

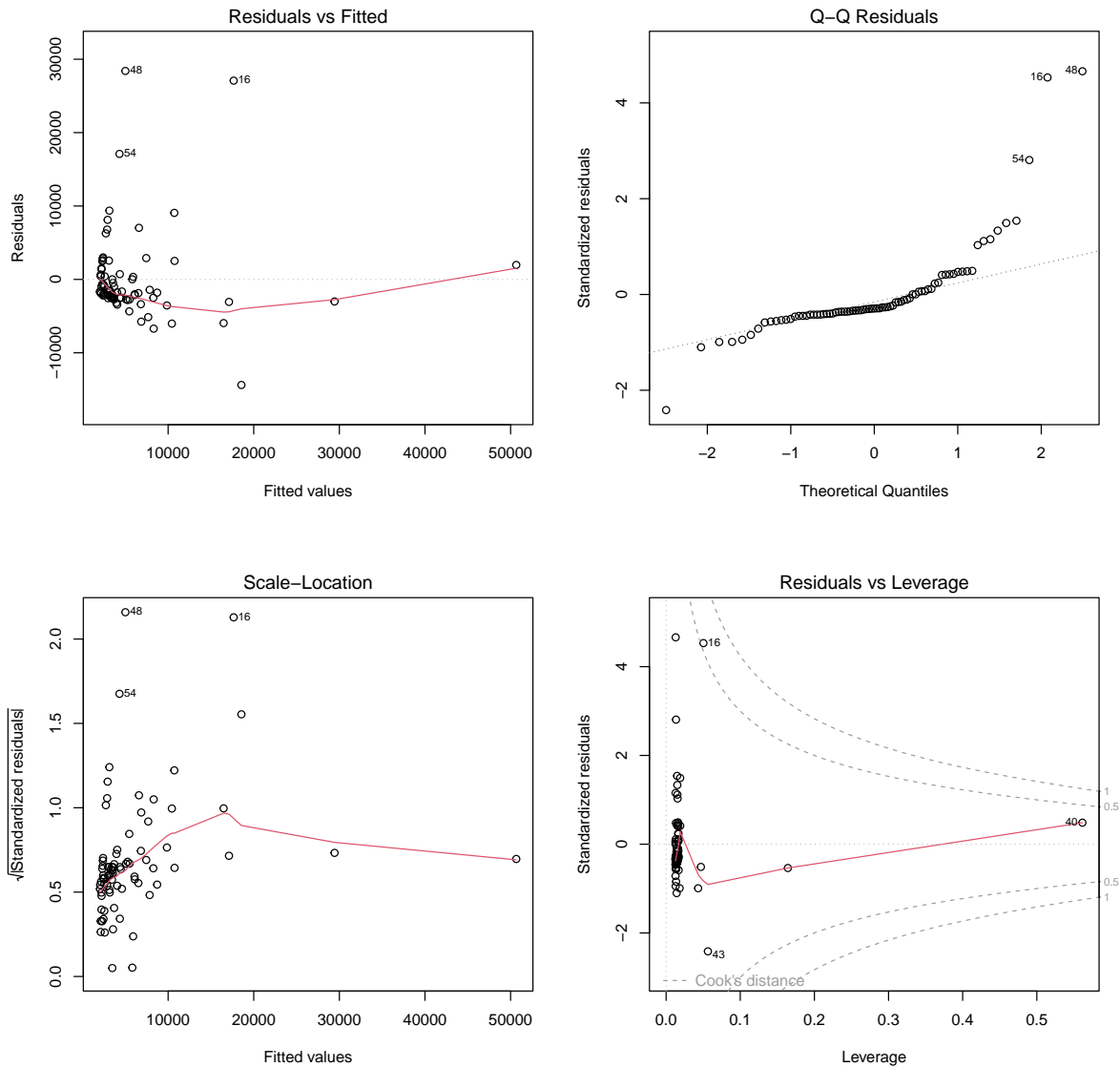
a)

```
plot(x = Sales, y = Assets)
lines(lowess(Assets, Sales))
```



From this plot we can say that it appears Sales and Assets do not have a linear relationship. The upward curving trend towards the high leverage point in the top right of the plot suggest a log transformation would be reasonable.

```
fit_1 <- lm(Assets ~ Sales)
par(mfrow = c(2,2))
plot(fit_1)
```



From the Residuals vs Fitted plot we can see that residuals do not appear to vary randomly around zero. For fitted values 0 to 20,000 residuals are below zero on average.

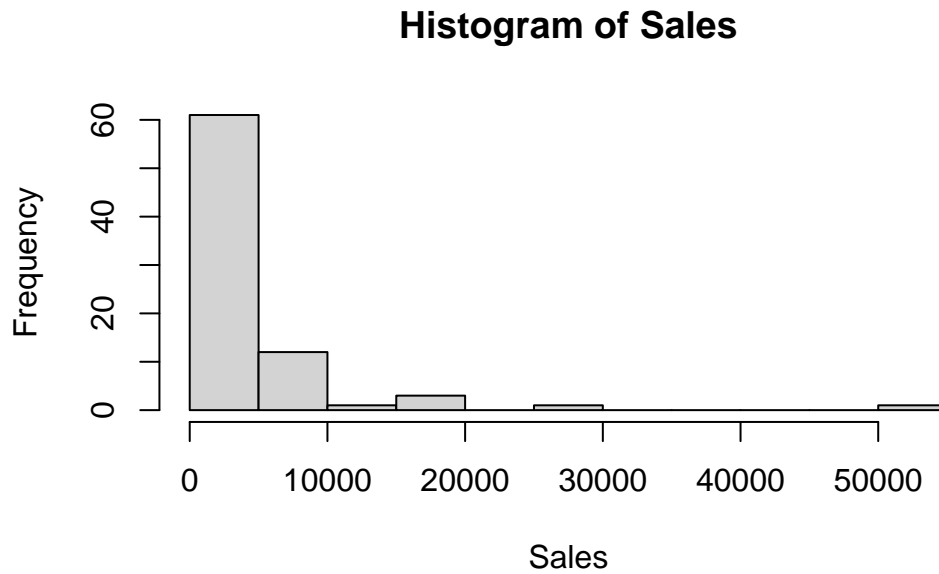
In the normal Q-Q plot we can see the standardized residuals don't appear normally distributed. Particularly the right tail is much longer and fatter than what would be expected by the theoretical normal distribution.

In the $\sqrt{|\text{std. residuals}|}$ by Fitted values plot we can see evidence of non-constant variance. The trend line has an upward slope for Fitted values 0 to 20,000 where most observations are

found, then it has a downward slope.

b)

```
hist(Sales)
```



From the histogram of Sales we can see that the distribution is highly right skewed and would be poorly fit by the normal distribution. We can use the box-cox method to transform Sales. That is the below transformation:

$$\Psi_S(X, \lambda) = \begin{cases} (X^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ \log(X) & \text{if } \lambda = 0 \end{cases}$$

Then choosing between the below options for λ :

$$\lambda : \{-1, -1/2, -1/3, -1/4, 0, 1/4, 1/3, 1/2, 1\}$$

```
library(car)
```

Warning: package 'car' was built under R version 4.3.3

Loading required package: carData

Warning: package 'carData' was built under R version 4.3.3

```
summary(powerTransform(Sales))
```

bcPower Transformation to Normality

	Est Power	Rounded Pwr	Wald Lwr Bnd	Wald Up Bnd
Sales	-0.0675	0	-0.2329	0.0979

Likelihood ratio test that transformation parameter is equal to 0
(log transformation)

	LRT	df	pval
LR test, lambda = (0)	0.6481734	1	0.42077

Likelihood ratio test that no transformation is needed

	LRT	df	pval
LR test, lambda = (1)	170.5833	1	< 2.22e-16

The likelihood ratio test with H_0 : Sales is normally distributed, H_a : Sales is not normally distributed rejects the null hypothesis with a p value extremely close to zero. We need to transform Sales.

The estimated λ found by the box-cox transformation is -0.0675. From the likelihood ratio test with H_0 : $\lambda = 0$, H_a : $\lambda \neq 0$ we see that the null hypothesis cannot be rejected. Thus, we choose the log transformation of Sales.

```
log_sales <- log(Sales)
```

c)

```
fit_2 <- lm(Assets ~ log_sales)
summary(powerTransform(fit_2))
```

bcPower Transformation to Normality

	Est Power	Rounded Pwr	Wald Lwr Bnd	Wald Up Bnd
Y1	-0.0166	0	-0.1688	0.1357

Likelihood ratio test that transformation parameter is equal to 0
(log transformation)

```

              LRT df      pval
LR test, lambda = (0) 0.04564128  1 0.83083

```

Likelihood ratio test that no transformation is needed

```

              LRT df      pval
LR test, lambda = (1) 159.6628  1 < 2.22e-16

```

The likelihood ratio test with H_0 : Residuals are normally distributed, H_a : Residuals are not normally distributed rejects the null hypothesis with a p value extremely close to zero. This is evidence that we need to transform Assets too.

The estimated λ found by the box-cox transformation is -0.0166. From the likelihood ratio test with $H_0 : \lambda = 0, H_a : \lambda \neq 0$ we see that the null hypothesis cannot be rejected. Thus, we choose the log transformation of Assets.

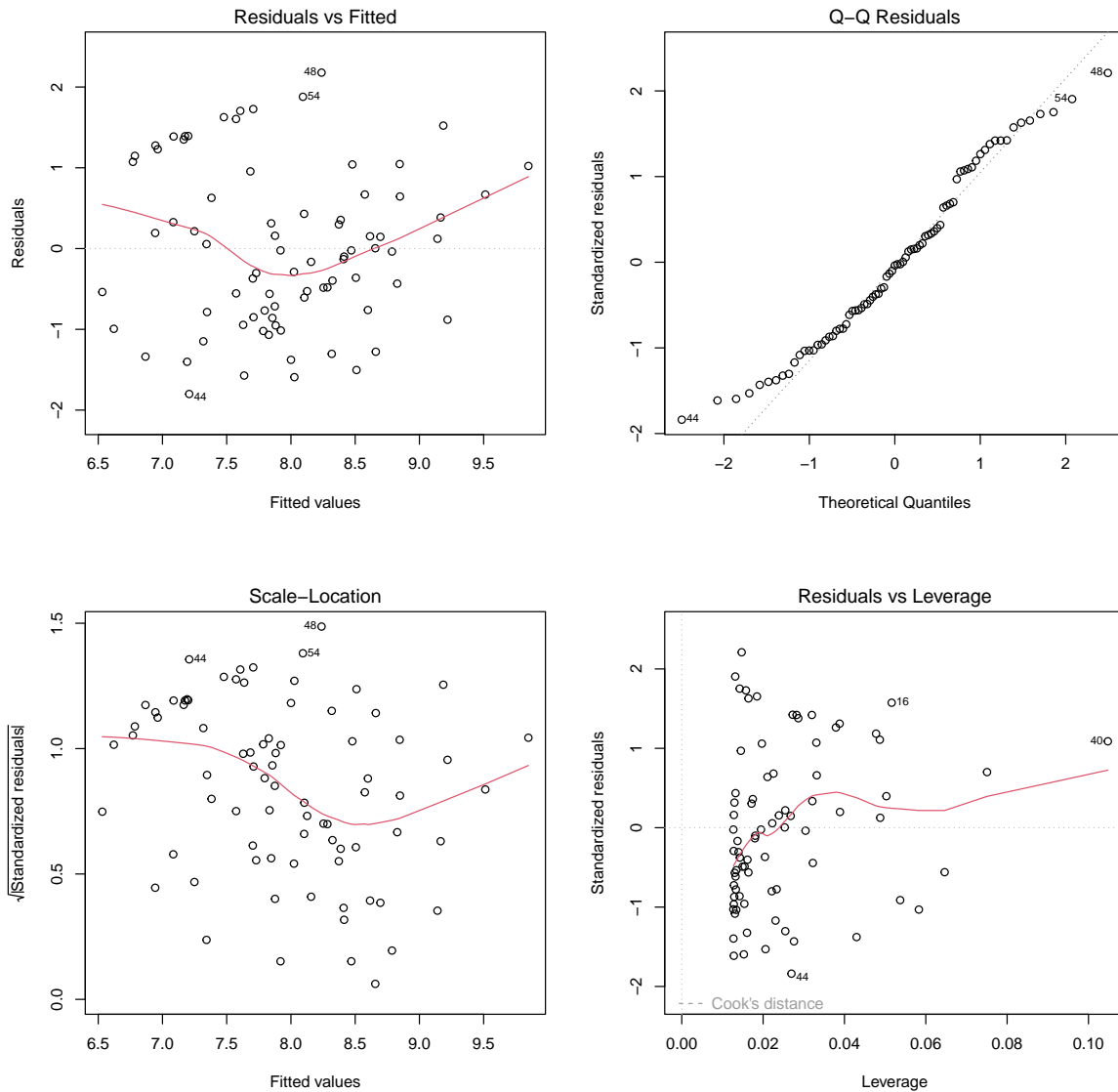
```
log_assets <- log(Assets)
```

d)

```

fit_3 <- lm(log_assets ~ log_sales)
par(mfrow = c(2,2))
plot(fit_3)

```



There are a few weaknesses in the model, first in the normal Q-Q plot of the standardized residuals there is deviation in the tails, both left and right are lighter than anticipated. There is also a trend in the scale-location plot, there is a sharp decrease in fitted values 7.5 to 8.5. This provides evidence of non-constant variance still being present.

e)

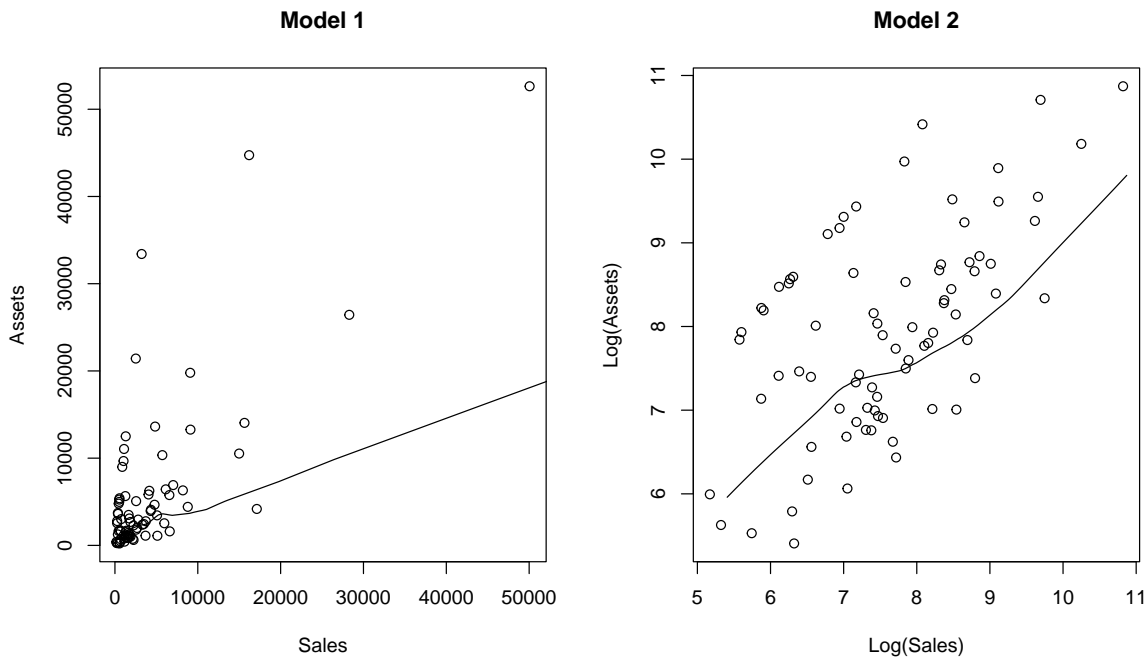
To compare the two models I will step through the assumptions we make and how reasonable they are.

Beginning with linear association in the explanatory and response variables.

```
par(mfrow = c(1,2))

plot(x = Sales, y = Assets, main = "Model 1")
lines(lowess(Assets, Sales))

plot(log_sales, log_assets, xlab = "Log(Sales)", ylab = "Log(Assets)", main = "Model 2")
lines(lowess(log_assets, log_sales))
```



In Model 1 we assume that Sales and Assets are linearly associated. In the left plot we clearly observe that this assumption is poorly met.

In Model 2 we assume that $\log(\text{Sales})$ and $\log(\text{Assets})$ are linearly associated. In the right plot we see that this assumption seems reasonable, a linear trend is evident.

The assumption of independent errors won't be a differentiator for either model as we are using the same data.

Next we look at the assumption of normality in errors.

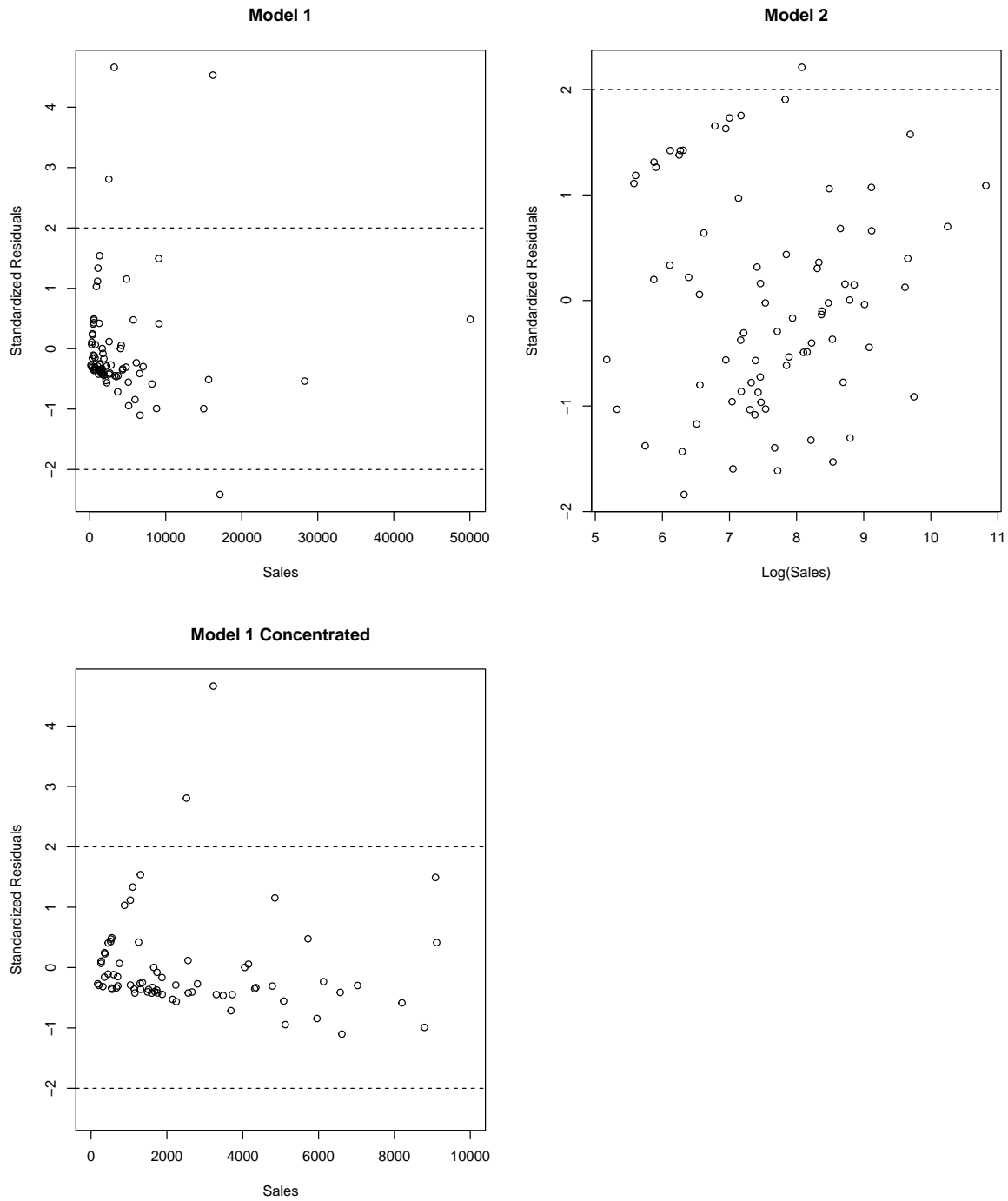

```

par(mfrow = c(2,2))
StanRes1 <- rstandard(fit_1)
plot(x = Sales, y = StanRes1, xlab = "Sales", ylab = "Standardized Residuals", main= "Model 1")
abline(h=2,lty=2)
abline(h=-2,lty=2)

StanRes2 <- rstandard(fit_3)
plot(x = log_sales, y = StanRes2, xlab = "Log(Sales)", ylab = "Standardized Residuals", main= "Model 2")
abline(h=2,lty=2)
abline(h=-2,lty=2)

plot(x = Sales, y = StanRes1, xlab = "Sales", ylab = "Standardized Residuals", main= "Model 1")
abline(h=2,lty=2)
abline(h=-2,lty=2)

```



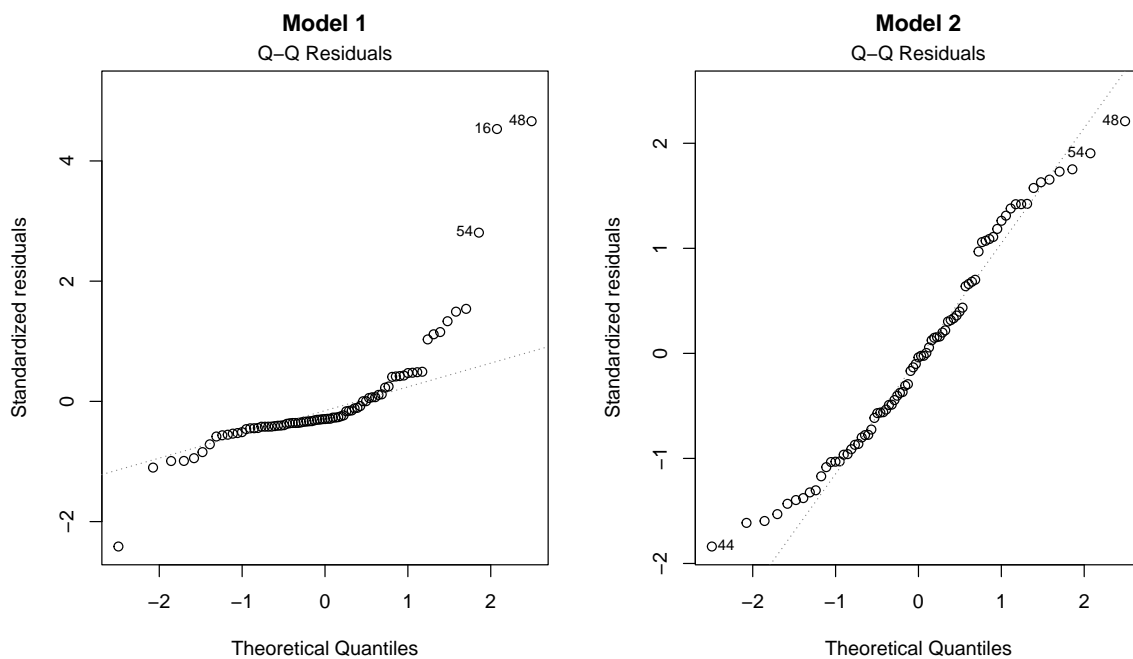
In Model 1 we can see that standardized residuals do not deviate randomly around zero how we would expect from a normal distribution. We can see this clearly if we take a closer look

at residuals for Sales 0 to 10,000. The majority of residuals in this range are beneath zero and particularly concentrated in the -.3 to -1 standardized residual range. That is not what we would expect in the normal distribution.

In Model 2 the Standardized Residuals randomly deviate around 0 as we would expect in a normal distribution.

Next we look at the respective QQ-plots:

```
par(mfrow = c(1,2))
plot(fit_1, which = 2, main = "Model 1")
plot(fit_3, which = 2, main = "Model 2")
```



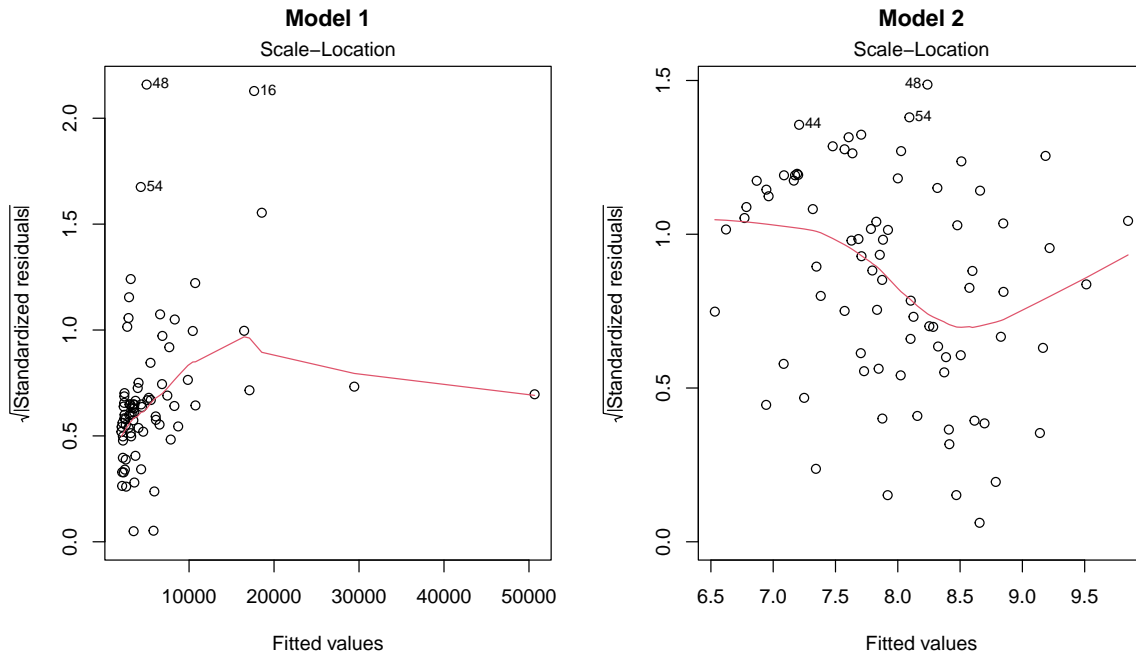
In Model 1 we see a far longer and fatter right tail than we would expect in the normal distribution.

In Model 2 we see slightly lighter tails than what we would expect in the normal distribution.

Overall the assumption of normality in errors is reasonable for Model 2 and not reasonable for Model 1.

The next assumption we check is equal variance.

```
par(mfrow = c(1,2))
plot(fit_1, which = 3, main = "Model 1")
plot(fit_3, which = 3, main = "Model 2")
```



Both models appear to have issues with this assumption as previously discussed.

Overall Model 2 is superior to Model 1, Model 2 meets the LINE assumptions well and the biggest difference between the two models is the assumption of Linear Association between the explanatory and response variables. Assets and Sales are not linearly associated which calls into question any conclusion taken away from Model 1. On the other hand $\log(\text{Assets})$ and $\log(\text{Sales})$ are linearly associated, so Model 2 is the more valid model.

f)

```
summary(fit_3)
```

Call:

```
lm(formula = log_assets ~ log_sales)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.80081	-0.77679	-0.03703	0.66919	2.17828

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.4969	0.7181	4.870	5.83e-06 ***
log_sales	0.5870	0.0934	6.284	1.82e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.993 on 77 degrees of freedom

Multiple R-squared: 0.339, Adjusted R-squared: 0.3304

F-statistic: 39.49 on 1 and 77 DF, p-value: 1.817e-08

The slope is interpreted as follows: As Sales increases by one percent we expect a .587 percent increase in Assets.

g)

Our transformation shows us that $\log(Y)$ is close to normally distribution. That means that Y is close to the log normal distribution. The mean of the log-normal distribution is:

$$E[Y] = e^{\mu + \sigma^2}$$

Then our correction to move from our transformed model back to the units of the original question we have the correction $e^{\text{endpoint} + MSE/2}$.

```
x_star_transformed <- log(6571)
sigma_hat <- sqrt(sum(fit_3$residuals^2)/(79-1))
n <- length(Sales)
x_bar_transformed <- mean(log_sales)
SXX <- sum((log_sales - x_bar_transformed)^2)
beta_0 <- as.numeric(fit_3$coefficients[1])
beta_1 <- as.numeric(fit_3$coefficients[2])
conf_int <- beta_0 + beta_1*x_star_transformed + c(-1,1)*qt(1 - .025,df = 79-1)*sigma_hat*sqrt(SXX)
conf_int
```

```
[1] 8.344006 8.969037
```

The interval we need to transform is 8.3440063, 8.9690375. We apply the previously stated correction.

```
corrected_conf_int <- exp(conf_int + sigma_hat^2/2)
corrected_conf_int
```

```
[1] 6841.006 12781.082
```

Among companies with 6571 million in sales we are 95% confident that the mean of assets will fall between 6841.006 and 12781.082 million.

9)

With $E[Y] = \mu$, $Var(Y) = \mu^2$ we want to use a transformation that provides us with constant variance.

Using the Taylor expansion, $f(Y) = f(E[Y]) + f'(E[Y])(Y - E[Y]) + \dots$, that allows us to estimate $Var(f(Y))$:

$$Var(f(Y)) = f'(E[Y])^2 Var(Y)$$

Subbing in $E[Y]$, $Var(Y)$:

$$Var(f(Y)) = f'(\mu)^2 \mu^2$$

If we use $f(y) = \log(y)$, which has derivative $f'(y) = \frac{1}{y}$ we get the following variance:

$$Var(f(Y)) = \frac{1}{\mu^2} \mu^2 = 1$$

Variance is now constant.