Project 5.1

Jared Barber

3/27/2022

Exercise 11.6

```
x = Ex11.06$Stress
y = Ex11.06$Shear.resistence
df = data.frame(x,y)
fit = lm(y~x)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.42633 -0.92139 -0.04785  0.89367  2.30506
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  42.5818     6.5065   6.544 6.52e-05 ***
## x            -0.6861     0.2499  -2.745   0.0206 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.64 on 10 degrees of freedom
## Multiple R-squared:  0.4298, Adjusted R-squared:  0.3727
## F-statistic: 7.537 on 1 and 10 DF,  p-value: 0.02064
```

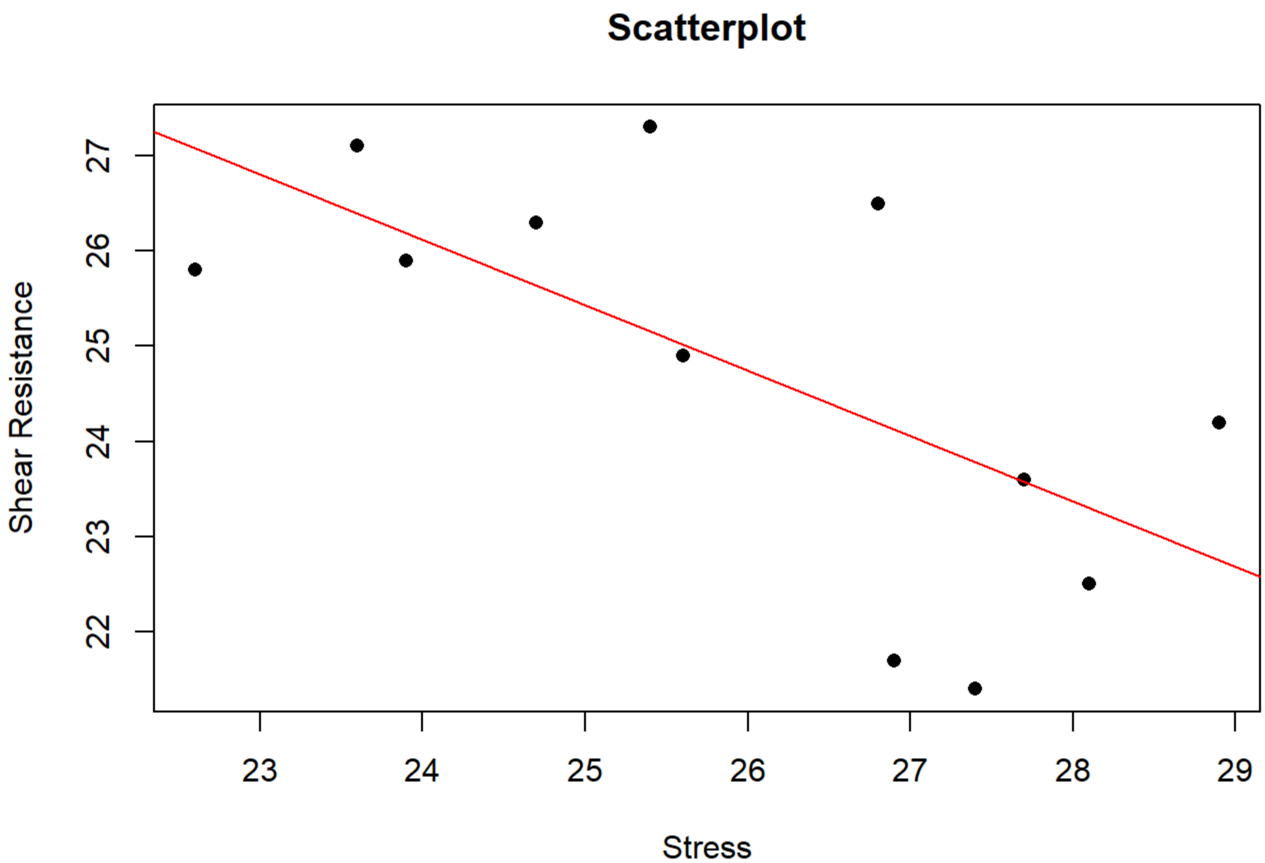
a. The regression line is

$$u_{Y|x} = +B_0 + B_1 x$$
$$y = 42.5818 - 0.6861x$$

b. When the normal stress is 24.5, estimated shear stress is 25.1549. y = 48.5818 - (0.6861)(25.4) = 25.1549

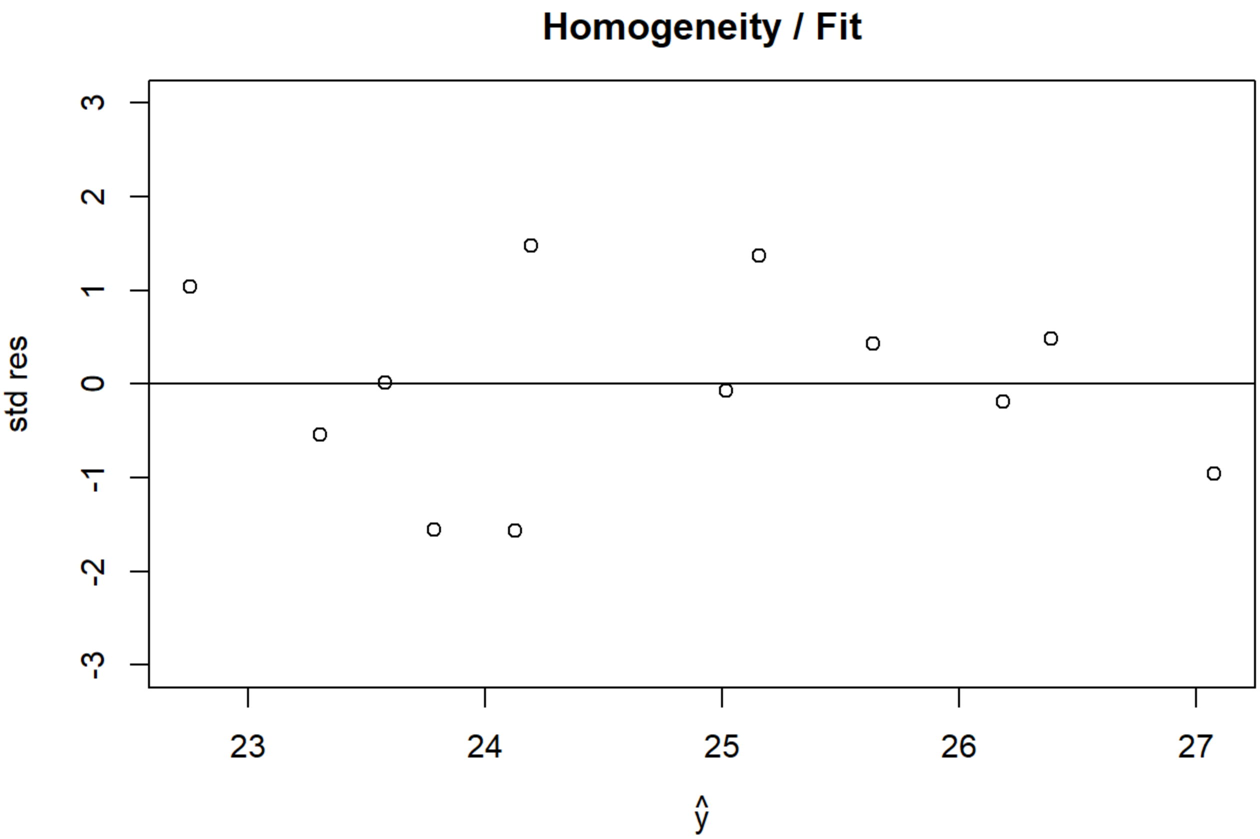
Scatterplot

```
plot(x,y,pch=16,xlab="Stress",ylab="Shear Resistance",main="Scatterplot")
# overlaid with the regression line (in red) from part (a).
abline(fit,col="red")
```



Homogeneity Plot

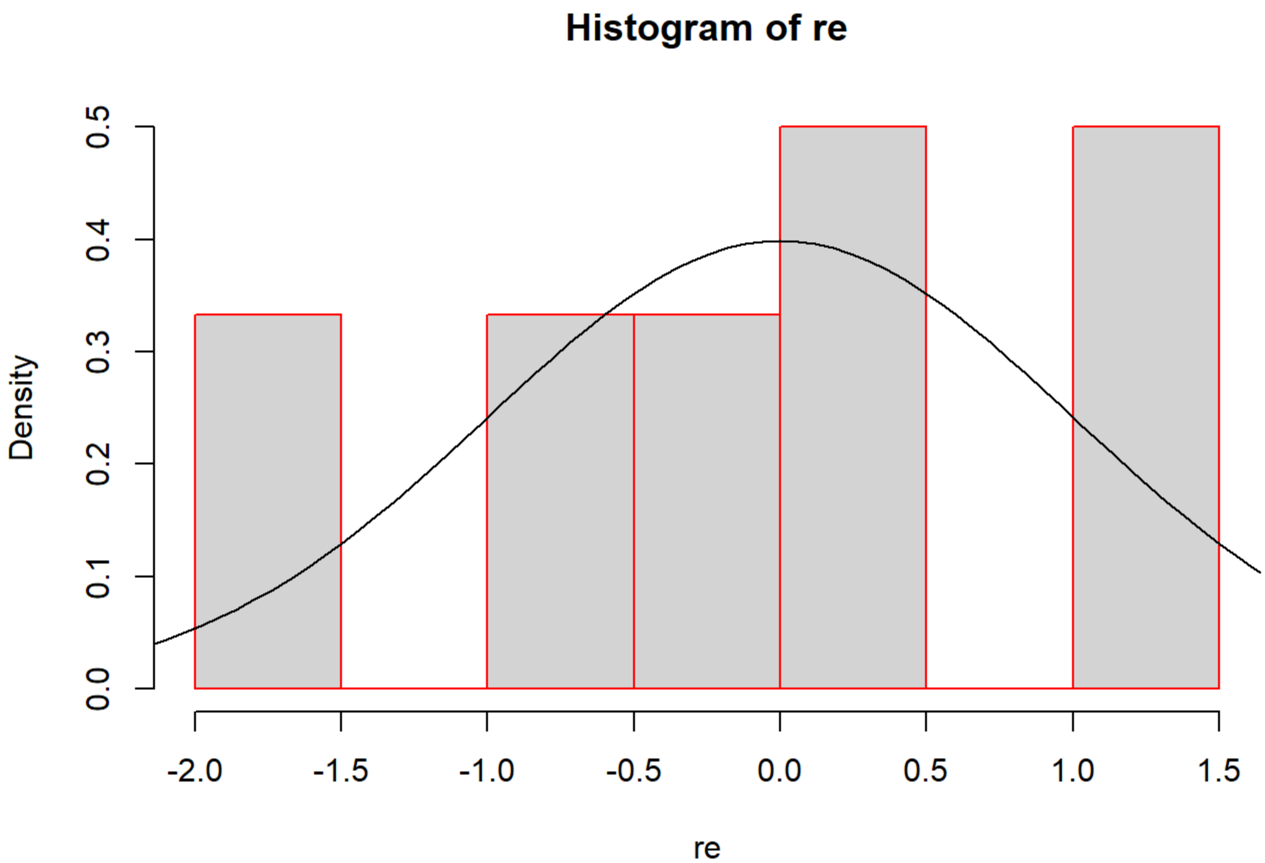
```
re = rstandard(fit)
plot(re~fitted.values(fit),ylim=c(-3,3),xlab=expression(hat(y)),ylab="std res",main="
Homogeneity / Fit")
abline(h=0)
```



The regression model assumes the errors have constant variance. If the variance of the data used (EX11.06) is constant, there should be a near constant spread on the homogeneity graph. This does not appear to be violated. The homogeneity plot implies the regression model is a good fit.

Histogram of the standardized residuals overlaid with a normal curve.

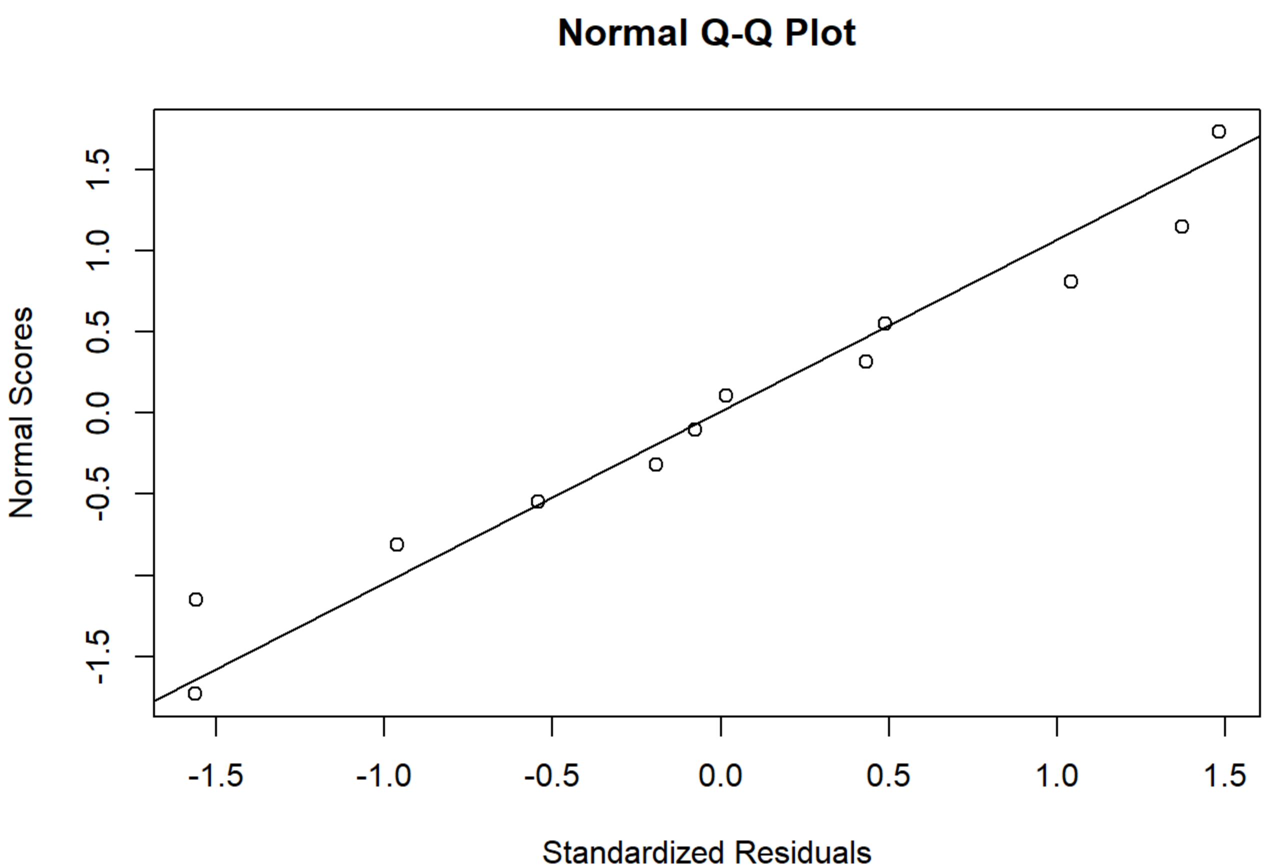
```
re = rstandard(fit)
hist(re,freq=FALSE,border='red')
plot(function(t)dnorm(t),from=-3,to=3,add=TRUE)
```



The histogram shows that the residuals are evenly split between positive and negative values, but there is not enough data to determine if this is a normal distribution.

QQ Plot

```
qqnorm(rstandard(fit), ylab="Standardized Residuals", xlab="Normal Scores", datax=TRUE)
qqline(re,datax=TRUE)
```



The qq-plot shows that the data is approximately normal, with values deviating slightly at the ends of the plot.

Exercise 11.18

```
# b and c
confint(fit,level=0.99,type="wald")
```

```
##                  0.5 %     99.5 %
## (Intercept) 21.960818 63.2027877
## x           -1.478106  0.1059517
```

a. S^2 is 1.64^2 = 2.6896
b. A 99% confidence interval for B0 is (21.960818 63.2027877). We can say with 99% confidence that the true value of B0 (the y-intercept of the regression line) is between 21.96 and 63.20.
c. A 99% confidence interval for B1 is (-1.478106 0.1059517). We can say with 99% confidence that the true value of B1 (the slope of the regression line) is between -1.48 and 0.11.

Exercise 11.23

a. A 95% CI for mean shear resistance when x = 24.5 is (24.43903, 27.10679). We are 95% confident that the true mean shear resistance is at least 24.44 and at most 27.11 when x (the stress) is 24.5.

b - A 95% PI for single predicted value of shear resistance when x = 24.5 is (21.88365, 29.66217). We can be 95% confident that the next observation shear resistance will fall within this range when the stress is 24.5.

```
# a
print('95% CI for mean shear resistance when x = 24.5')
```

```
## [1] "95% CI for mean shear resistance when x = 24.5"
```

```
x<-Ex11.06$Stress
predict.lm(fit,se.fit=TRUE,newdata=data.frame(x=24.5),interval="confidence",level=0.9
5)
```

```
## $fit
##        fit      lwr      upr
## 1 25.77291 24.43903 27.10679
##
## $se.fit
## [1] 0.5986518
##
## $df
## [1] 10
##
## $residual.scale
## [1] 1.63965
```

```
# b
print('95% PI for single predicted value of shear resistance when x = 24.5')
```

```
## [1] "95% PI for single predicted value of shear resistance when x = 24.5"
```

```
predict.lm(fit,se.fit=TRUE,newdata=data.frame(x=24.5),interval="prediction",level=0.9
5)
```

```
## $fit
##        fit      lwr      upr
## 1 25.77291 21.88365 29.66217
##
## $se.fit
## [1] 0.5986518
##
## $df
## [1] 10
##
## $residual.scale
## [1] 1.63965
```

Exercise 11.9

a. Plot a scatter diagram

```
library(car)
```

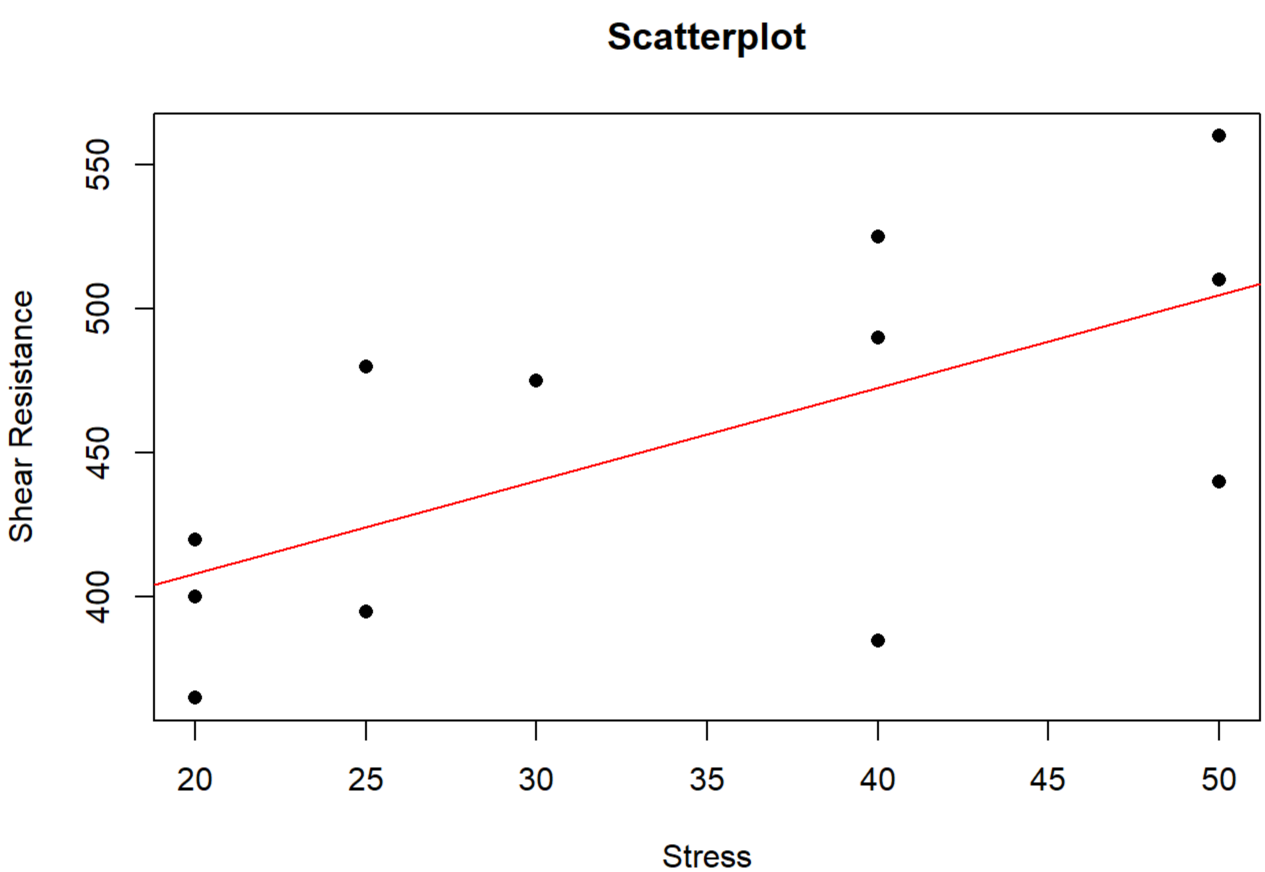
```
## Warning: package 'car' was built under R version 4.1.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.1.3
```

```
x = Ex11.09$Advertising_Costs
y = Ex11.09$Sales
fit = lm(y~x)
plot(x,y,pch=16,xlab="Stress",ylab="Shear Resistance",main="Scatterplot")

# overlaid with the regression line (in red) from part (a).
abline(fit,col="red")
```



b. The equation of the regression line is

$$Y|x = +B0B1xu$$
$$y = 343.706 + 3.221x$$

```
df = data.frame(x,y)
fit = lm(y~x)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -87.538 -32.700   8.566  39.118  55.774
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  343.706     44.766   7.678 1.68e-05 ***
## x              3.221      1.240   2.598   0.0266 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50.23 on 10 degrees of freedom
## Multiple R-squared:  0.403,  Adjusted R-squared:  0.3433
## F-statistic: 6.751 on 1 and 10 DF,  p-value: 0.02657
```

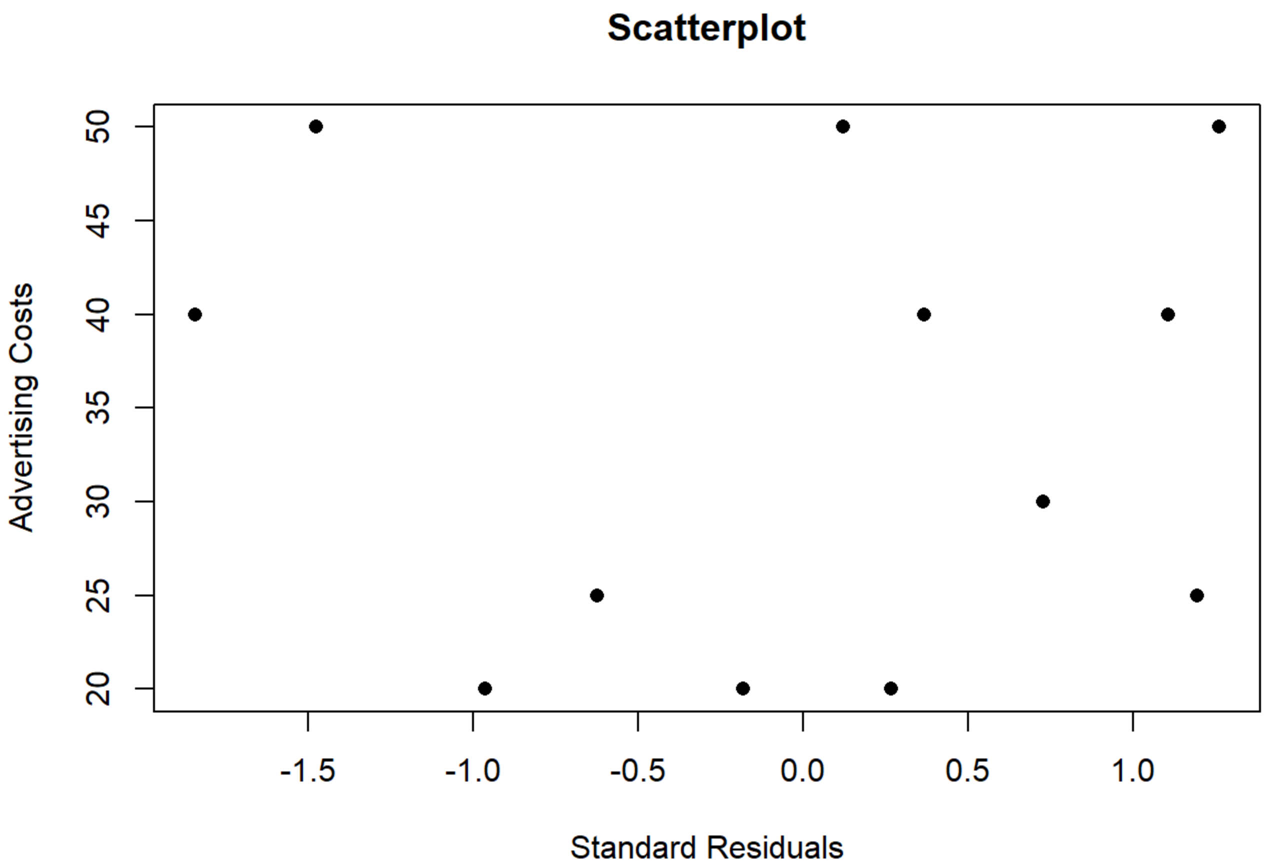
c. An estimate for the weekly sales when advertising costs are $35 is $456.44.

```
343.706 + 3.221 * (35)
```

```
## [1] 456.441
```

d. Plot of the residuals versus advertising costs

```
library(car)
x = rstandard(fit)
y = Ex11.09$Advertising_Costs
plot(x,y,pch=16,xlab="Standard Residuals",ylab="Advertising Costs",main="Scatterplo
t")
```



There is no obvious pattern, this corresponds to the regression model being a ‘good’ fit.

Exercise 11.21

The p-value obtained is 0.0244336. This is slightly less than the significance level of 0.05, so reject the null hypothesis. There is significant evidence that B1 is < 6.

$$H_0 : B_1 = 6$$
$$H_A : B_1 < 6$$

```
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -87.538 -32.700   8.566  39.118  55.774
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  343.706     44.766   7.678 1.68e-05 ***
## x              3.221      1.240   2.598   0.0266 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50.23 on 10 degrees of freedom
## Multiple R-squared:  0.403,  Adjusted R-squared:  0.3433
## F-statistic: 6.751 on 1 and 10 DF,  p-value: 0.02657
```

```
x= Ex11.09$Advertising_Costs
Sxx = sum((x-mean(x))^2)
b1 = 3.221
MSE = 50.23
TS = (b1-6)/(MSE/sqrt(Sxx))
pt(TS,10)
```

```
## [1] 0.02443361
```

Exercise 11.51

```
# a 95% CI for average weekly sales when $45 spent on ads
predict.lm(fit,se.fit=TRUE,newdata=data.frame(x=45),interval="confidence",level=0.95)
```

```
## $fit
##        fit      lwr      upr
## 1 488.6421 444.6084 532.6758
##
## $se.fit
## [1] 19.76254
##
## $df
## [1] 10
##
## $residual.scale
## [1] 50.2257
```

```
# b 95% PI for average weekly sales when $45 spent on ads
predict.lm(fit,se.fit=TRUE,newdata=data.frame(x=24.5),interval="prediction",level=0.9
5)
```

```
## $fit
##        fit      lwr      upr
## 1 422.6155 303.1152 542.1158
##
## $se.fit
## [1] 18.80976
##
## $df
## [1] 10
##
## $residual.scale
## [1] 50.2257
```

a. A 95% CI for average weekly sales when $45 spent on ads is (444.6084, 532.6758). We are 95% confident that the true mean weekly sales is at between $444.60 and $532.68 when $45 is spent on advertising.

b. A 95% PI for average weekly sales when $45 spent on ads is (303.1152, 542.1158) We can be 95% confident that the next observation of weekly sales will fall within this range when $45 is spent on advertising.

Exercise 11.55

a. The data appears to have a parabolic distribution, and the data points are not randomly scattered. It appears that an improved model could be found using a transformation.

```
x = Ex11.55$WT
y = Ex11.55$MPG
fit = lm(y~x)
summary(fit)
```

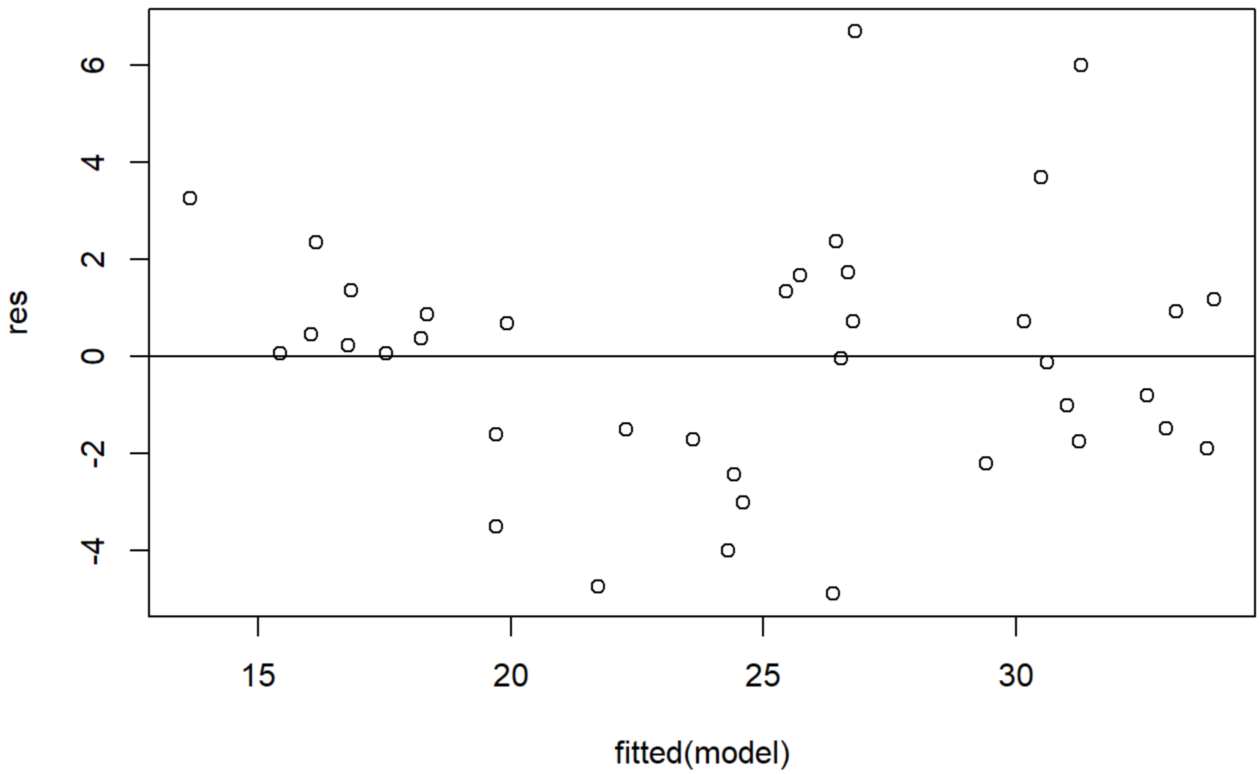
```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4370 -1.8789  0.1892  1.1712  6.4327
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  48.6793     1.9405   25.09  < 2e-16 ***
## x            -8.3624     0.6591  -12.69 7.51e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.838 on 36 degrees of freedom
## Multiple R-squared:  0.8172, Adjusted R-squared:  0.8122
## F-statistic:   161 on 1 and 36 DF,  p-value: 7.512e-15
```

b. The data has less of a parabolic shape than before, but it could still possibly be improved on. Also, the R-squared value has increased.

```
x = log(Ex11.55$WT)
y = Ex11.55$MPG
model = lm(y~x)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4370 -1.8789  0.1892  1.1712  6.4327
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  48.6793     1.9405   25.09  < 2e-16 ***
## x            -8.3624     0.6591  -12.69 7.51e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.838 on 36 degrees of freedom
## Multiple R-squared:  0.8172, Adjusted R-squared:  0.8122
## F-statistic:   161 on 1 and 36 DF,  p-value: 7.512e-15
```

```
res = resid(model)
plot(fitted(model), res)
abline(0,0)
```



c. The data appears to have a more random than either of the two previous graphs. Also, the R-squared value is now 0.8568, which higher than the two previous models. Accordingly, this is the preferred model.

```
x = Ex11.55$WT
y = c(100/Ex11.55$MPG)
model = lm(y~x)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4370 -1.8789  0.1892  1.1712  6.4327
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  48.6793     1.9405   25.09  < 2e-16 ***
## x            -8.3624     0.6591  -12.69 7.51e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.838 on 36 degrees of freedom
## Multiple R-squared:  0.8172, Adjusted R-squared:  0.8122
## F-statistic:   161 on 1 and 36 DF,  p-value: 7.512e-15
```

```
res = resid(model)
plot(fitted(model), res)
abline(0,0)
```