

# Homework 5 Solutions

*Jacob S Townson*

*March 26, 2018*

1)

2)

3)

First we must set up the data. It could have also been read in through a file, however this makes it easier to work from 2 different pc's on the same project.

```
set.seed(123)
x0=1:100/100
x=c(x0,10)
y0=sqrt(1+.1*x0+rnorm(100,sd=.0001))
y=c(y0,0)
hw5_3 = cbind(x,y)
outrm_hw5_3 = hw5_3[1:100,]
```

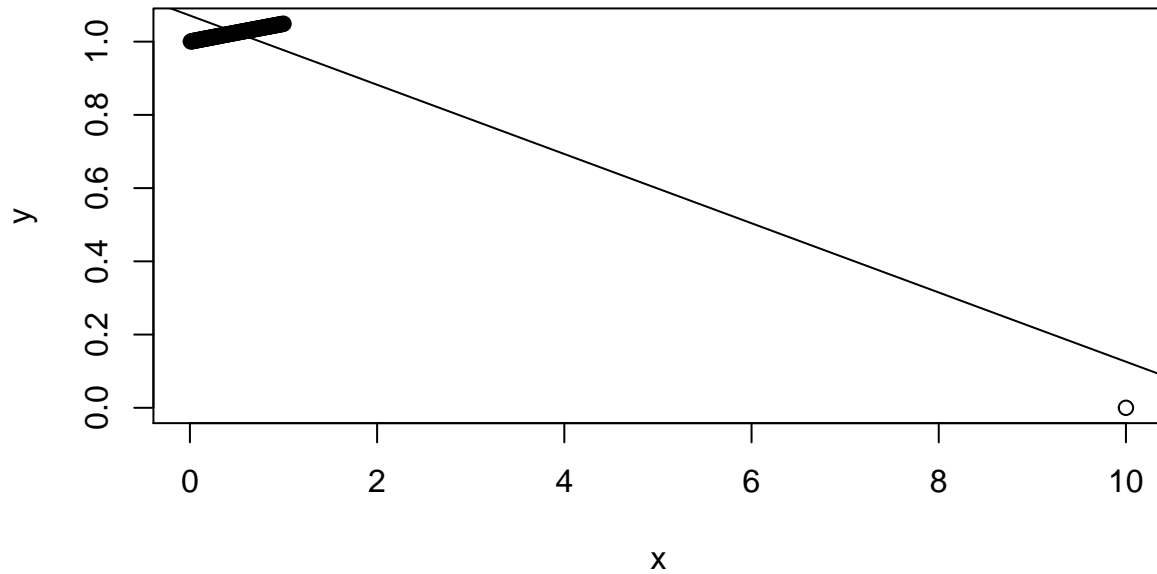
Note the last two lines of code contain the creation of two dataframes. The first of which contains the entire data set, the second of which removes the outlier.

a)

Below is a scatterplot of the data, with the line representing our model going through it. As you can see, the outlier at  $x = 10$  makes the model seemingly strange for the rest of the data. We will address this in later parts of the problem.

```
model3 = lm(y~x)
plot(x,y, main = "scatterplot for #3a")
abline(model3$coef)
```

**scatterplot for #3a**



b)

Two of the plots will be supplied below. To calculate the residuals, we use the following code:

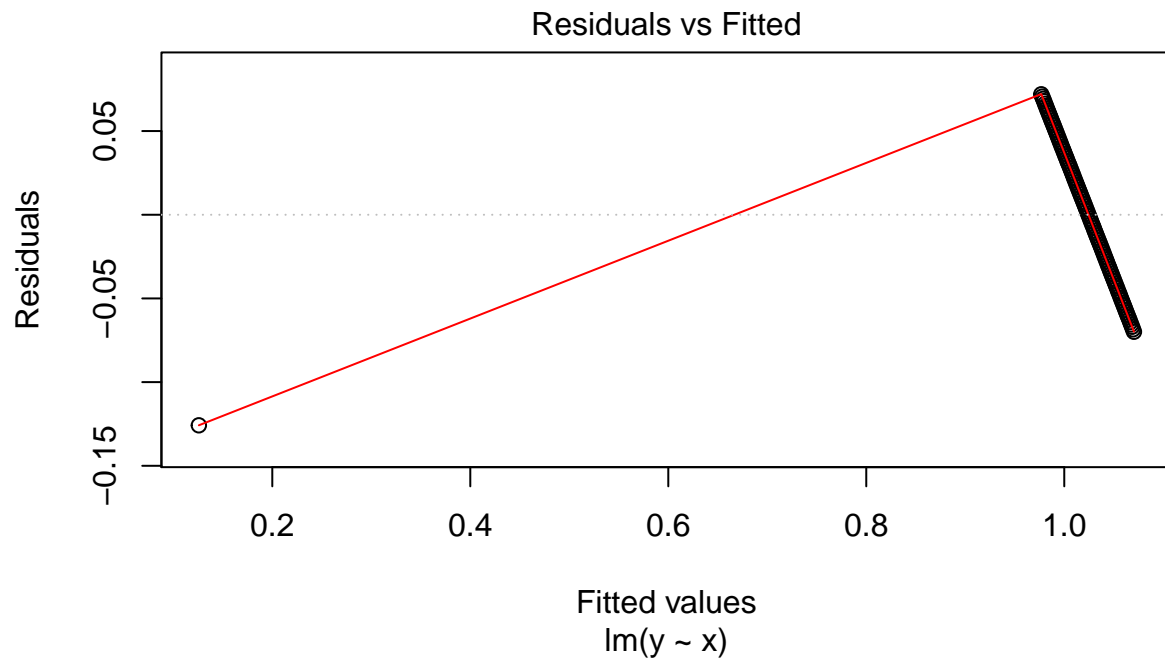
```
epsilon.hat=residuals(model3)
r=rstandard(model3)
t=rstudent(model3)
resids=cbind(epsilon.hat,r,t)
kable(resids[101,])
```

	x
epsilon.hat	-0.1258052
r	-9.9498514
t	-4602.2968772

This is a table containing the values for the residuals of the outlier in the data. As we can see, these values lead us to believe that the outlier should most likely not be contained, as it throws off the entire model.

Below is the residual plot.

```
plot(model3, which = 1, labels.id = '')
```

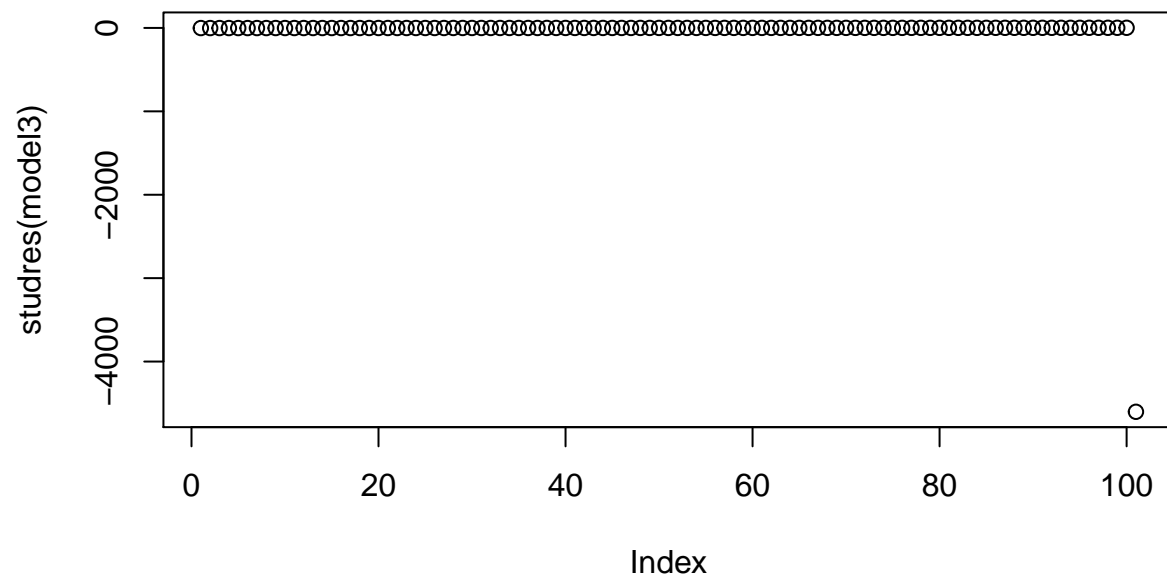


Below is the studentized residual plot with the outlier.

```
require(MASS)
```

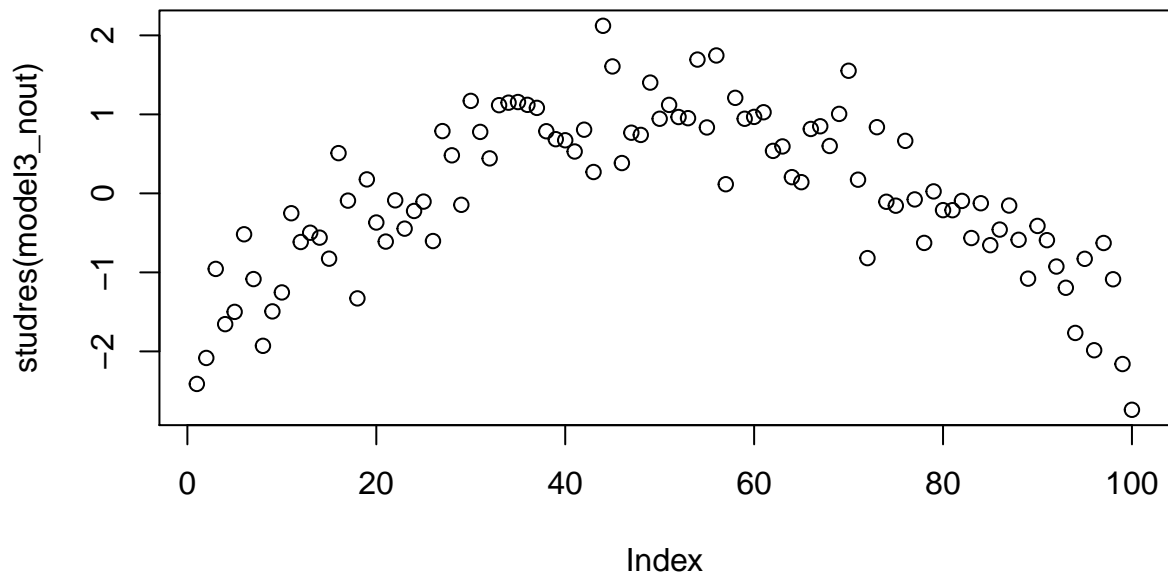
```
## Loading required package: MASS
```

```
plot(studres(model3))
```



Below is the studentized residual plot without the outlier.

```
x2=outrm_hw5_3[,1]
y2=outrm_hw5_3[,2]
model3_nout = lm(y2~x2)
plot(studres(model3_nout))
```

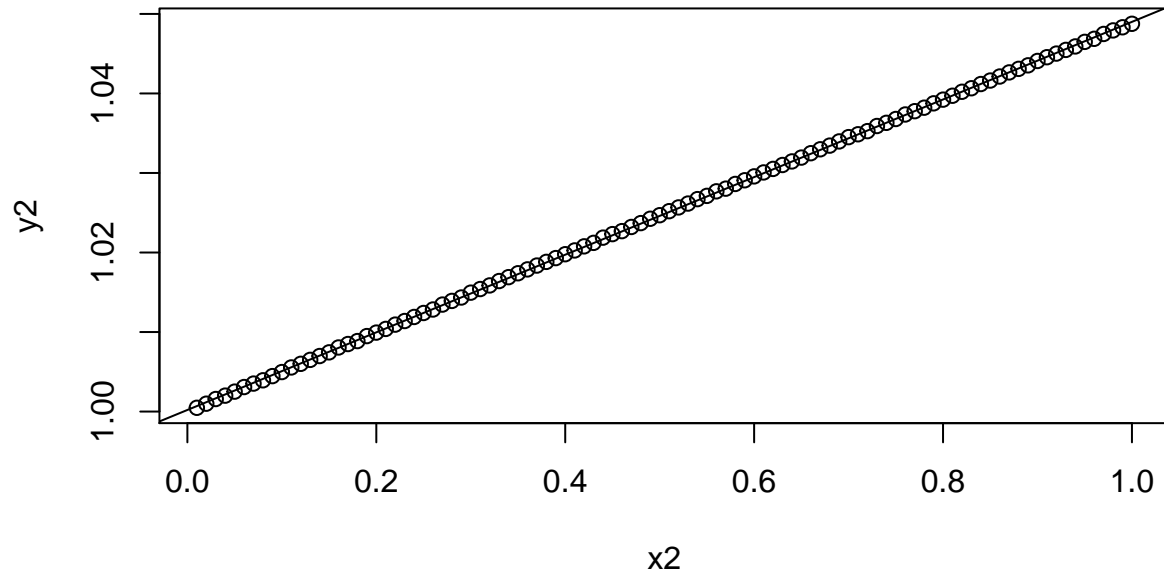


c)

Below is the scatterplot for the data, excluding the outlier. As we can see, when the outlier is removed, the model follows the data extremely well.

```
plot(x2,y2, main = "scatterplot for #3c")  
abline(model3_nout$coef)
```

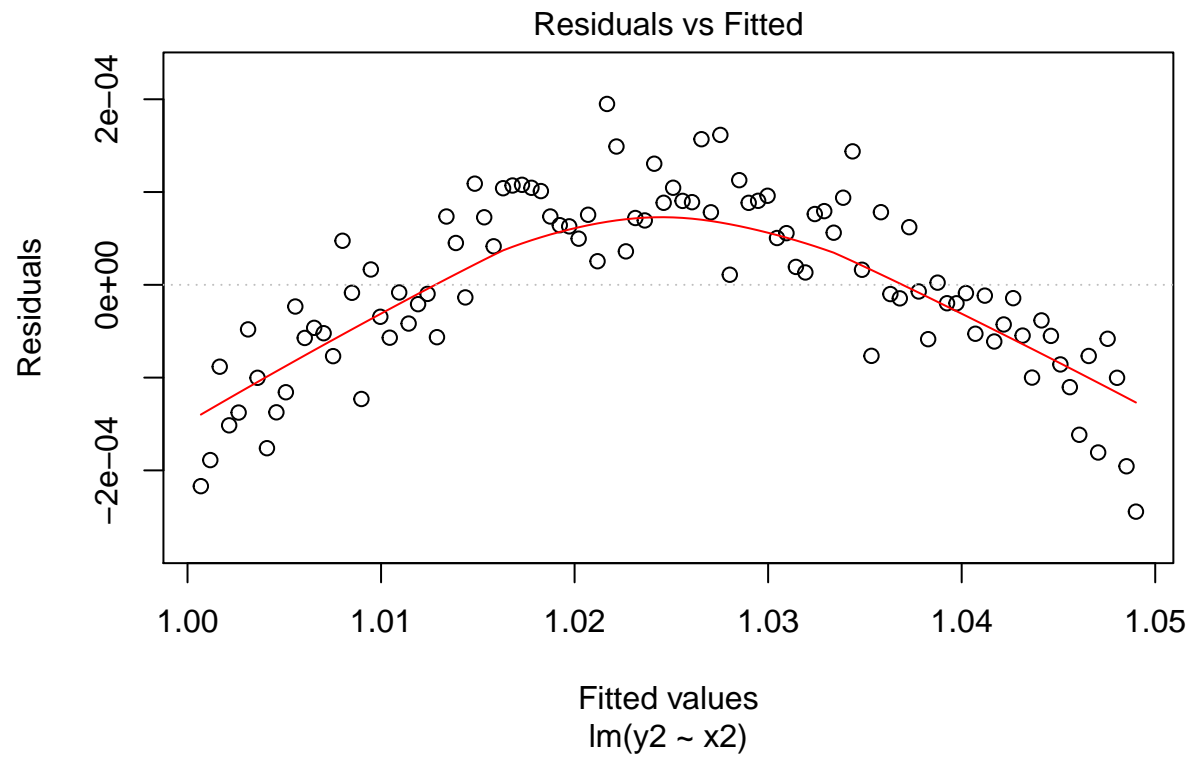
**scatterplot for #3c**



d)

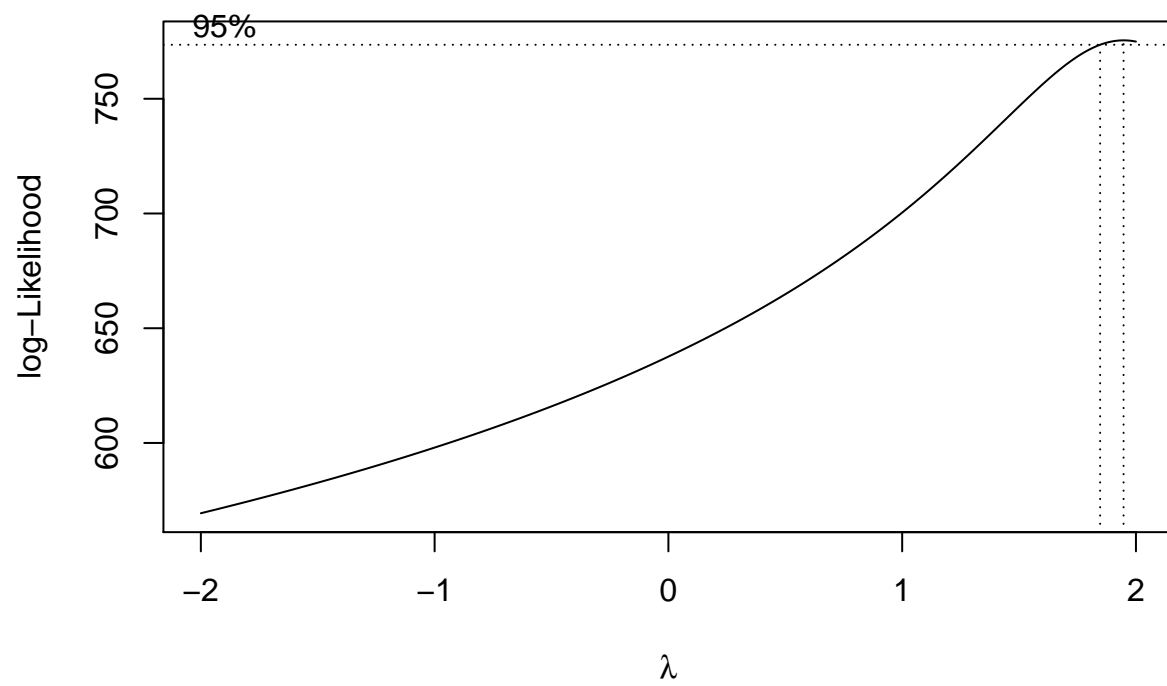
Below is the plot for the residuals without the outlier. As we can see, the model follows the line quite well, leading us to believe that the outlier was causing all of the problems in the previous model (suprise).

```
plot(model13_nout, which=1, labels.id='')
```



e)

```
bc=boxcox(y2~x2,lambda=seq(-2,2,by=.001))
```



```
bc$x[which.max(bc$y)]
```

```
## [1] 1.947
```