

Stat 424: Homework 1

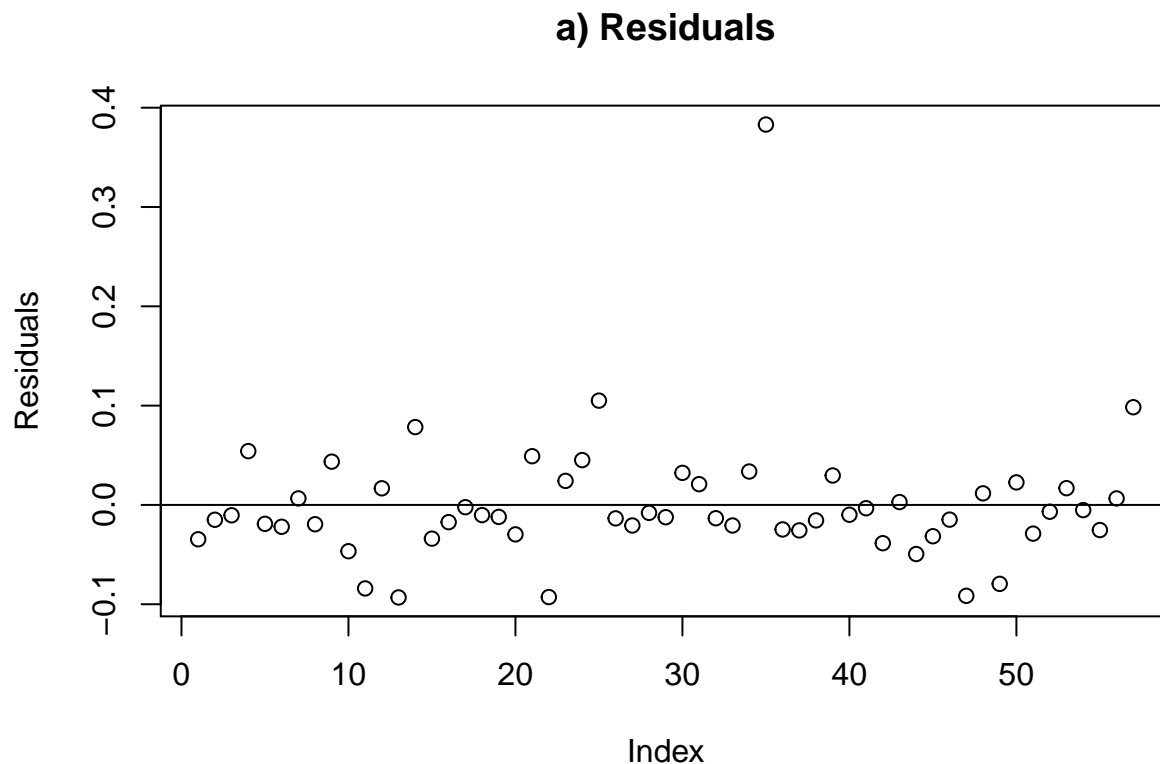
Evangeline Lim

9/23/2019

Question 9

```
myFile=read.table('rainfall.txt', header = T)

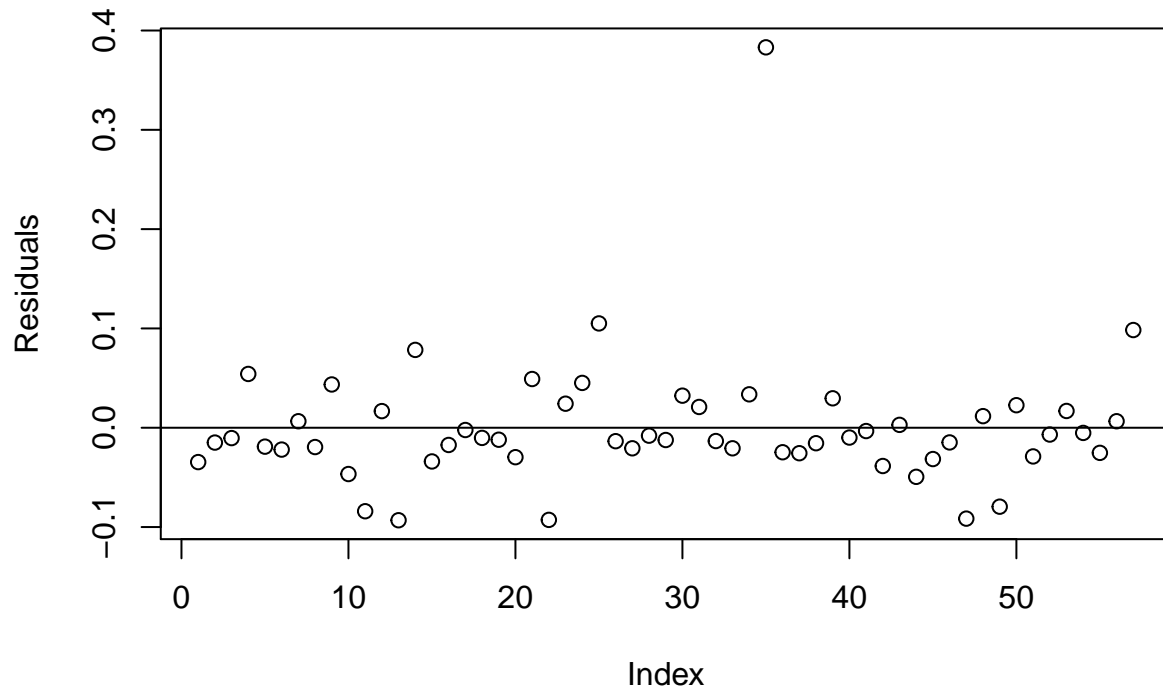
x_data = 0.44*myFile$x
rainfall=lm(myFile$y~x_data)
plot(residuals(rainfall), ylab = "Residuals", main = "a) Residuals")
abline(0,0)
```



a) The regression model shows more negative values from the line (0,0) which leads me to believe that the model is not a good model.

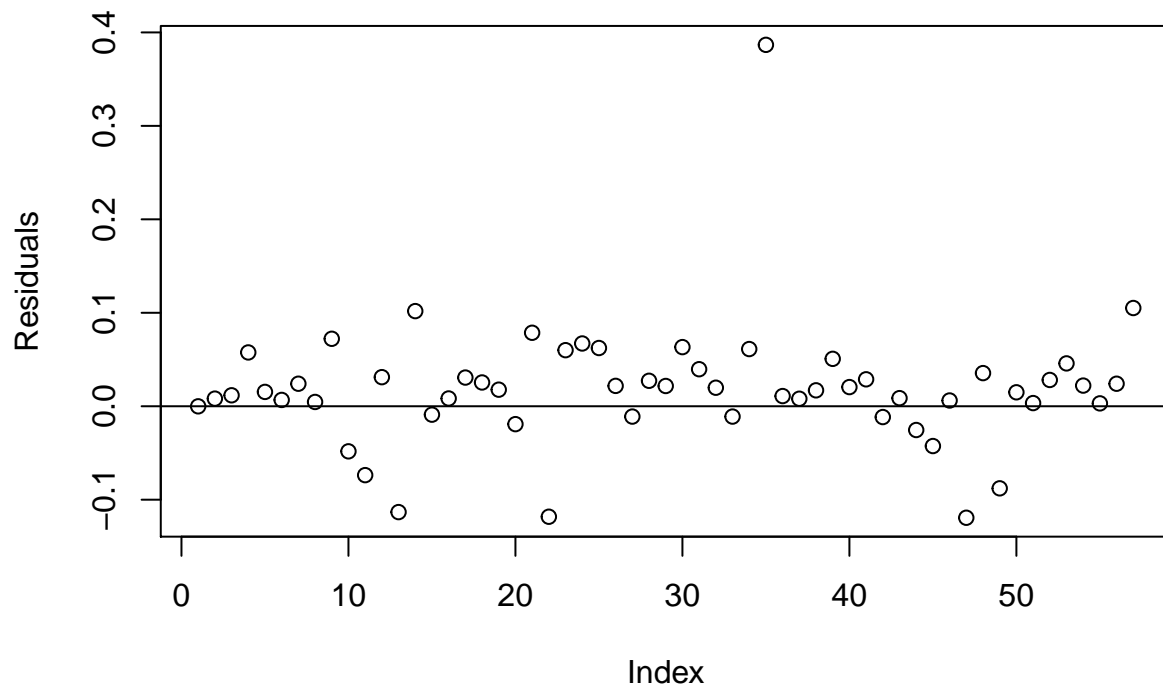
```
model1 = residuals(lm(myFile$y~myFile$x))
model2 = residuals(lm(formula=myFile$y~0+myFile$x))
plot(model1, ylab = "Residuals", main = "Residuals with intercept")
abline(0,0)
```

Residuals with intercept



```
plot(model2, ylab = "Residuals", main = "Residuals without intercept")  
abline(0,0)
```

Residuals without intercept

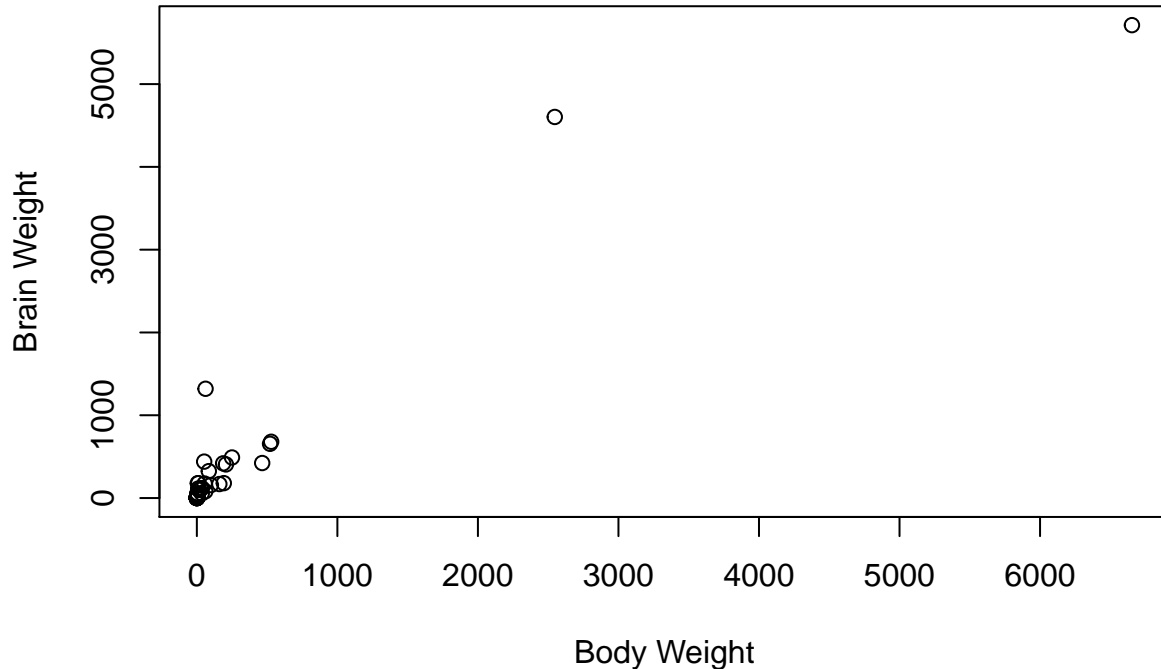


b) The residual plot without the intercept gives more positive values from the line (0,0) which leads me to believe that the model without the intercept does not fit the data very well

compared to the residual model with the intercept. However, the model with the intercept has significantly more variation and hence, the intercept is significant.

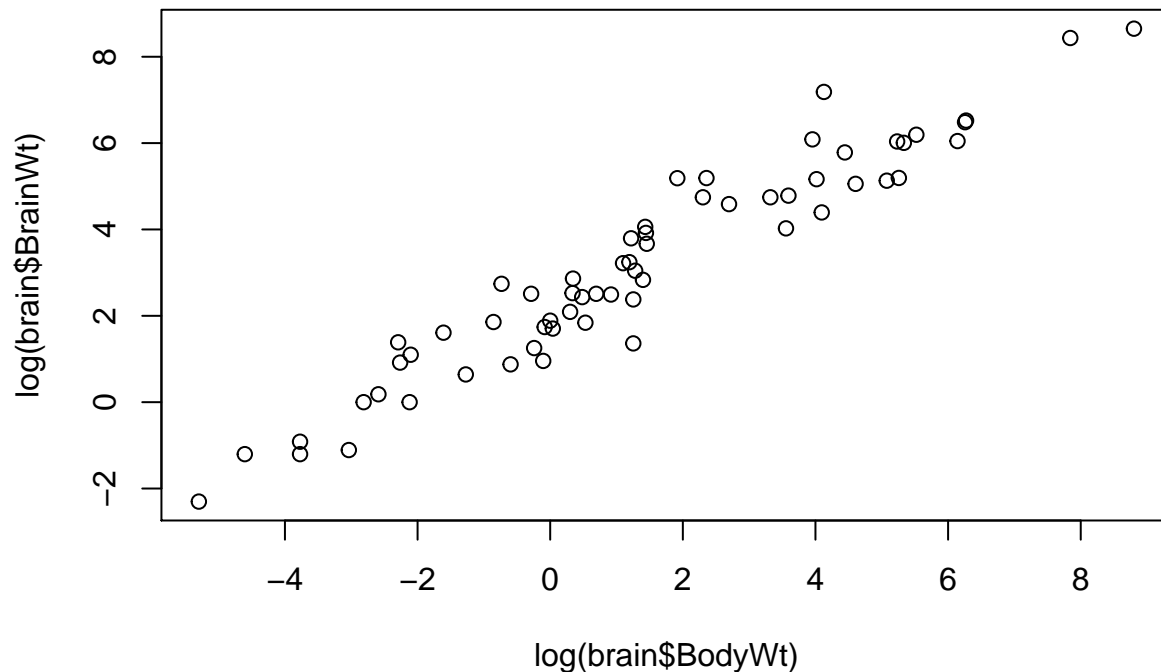
c) Yes, in a) and b) we see that the residual plots with no intercept causes the data to be more positive or more negative but the model with the intercept is more varied on the line (0,0). This supports the usage of the model with intercept.

Question 10



a) The plot is clustered and has two outliers which means that the data is heavily skewed. Because the data is clustered, no conclusions can be drawn but the data does seem to be linear.

```
plot(log(brain$BodyWt), log(brain$BrainWt))
```



b) Yes, the log plot now has a clear positive linear relationship.

```
logBody = log(brain$BodyWt)
logBrain = log(brain$BrainWt)
```

```
lm(logBrain~logBody)
```

```
##
## Call:
## lm(formula = logBrain ~ logBody)
##
## Coefficients:
## (Intercept)      logBody
##      2.1297      0.7536
```

```
summary(lm(logBrain~logBody))
```

```
##
## Call:
## lm(formula = logBrain ~ logBody)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71285 -0.48177 -0.05678  0.43763  1.94539
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.12972    0.09638   22.10  <2e-16 ***
## logBody      0.75362    0.02856   26.39  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6967 on 60 degrees of freedom
```

```
## Multiple R-squared:  0.9207, Adjusted R-squared:  0.9193
## F-statistic: 696.2 on 1 and 60 DF,  p-value: < 2.2e-16
```

c) From the summary, the percentage of variation (multiple R^2) is 92.07%.

```
n = 63
mean = log(250)
sd = sd(logBody, na.rm=TRUE)
error <- qt(0.975,df=n-2)*sd/sqrt(n)
(upper = mean + error)
```

```
## [1] 6.308267
```

```
(lower = mean - error)
```

```
## [1] 4.734655
```

d) The confidence interval for the average body weight of 250kg is (4.735,6.308).