Stat 21 Homework 3

Madison Hamilton Shoraka

TOTAL POINTS

17.5 / 20

QUESTION 1

1 Q1 **2 / 3**

✓ + **2 pts** Correct

QUESTION 2

2 Q2 **3 / 3**

✓ + **3 pts** Correct

QUESTION 3

3 Q3 **9 / 10**

+ **0 pts** Correct

+ **9** Point adjustment

b)

QUESTION 4

4 Q4 **3.5 / 4**

✓ + **3.5 pts** Correct

Homework 3

Stat 021

Mads Shoraka

Question 1

Part A

This is not the proper explanation. This is not what the model was constructed to compare. It was made to compare height and weight between different men, the difference in height or weight of a single man. It is also important to note the age range of the men: they have stopped growing, and therefore, even if they gained 10 lbs, there would be no increase in height. We would instead say that if another man is 10 lbs more than the previous, he would be an average of .47in taller.

Part B

About 0.05989747 or approximatley 5.9 of men who weigh 200lbs are over 74in.

```
1-pnorm(74,71.8,sqrt(2))
```

```
## [1] 0.05989747
```

Question 2

We can see that the classics students tend to score better than the psyhcology students by seeing on which side of the y=x trend line a majority of the points lie on. Except for the extremes which lie exactly on the trend line, all other points are solidly in the classic's students range implying that on average they score better on the GRE verbal section.

Using much the same procedure as before, we can see that students who score under 500 on the GRE verbal section score better in psychology and worse in econ where as students who score above 500 tend to score better in econ and worse in psychology. Despite this analsys, because all points are relativley close to the line, the difference in scores as noted above is not as darastic as those when the pscyology and classics students were compared.

Question 3

Part A

The relationship between the height of the building and the number of floors appears to be linear given the scatter plot below. Such linear relationship appears to be strong since the points seem to form a line like blob on their own.

```
data_read<-read.csv(file = "C:/Users/mhsho/OneDrive/Documents/Stat021/Homework/skyscraper_data_c
leaned.csv")

plot(data_read$floors, data_read$height_meters, xlab = "stories", ylab = "height of building (me
ters)", main = "Height of building given number of stories")
```

1 Q1 2 / 3

✓ + 2 pts Correct

Homework 3

Stat 021

Mads Shoraka

Question 1

Part A

This is not the proper explanation. This is not what the model was constructed to compare. It was made to compare height and weight between different men, the difference in height or weight of a single man. It is also important to note the age range of the men: they have stopped growing, and therefore, even if they gained 10 lbs, there would be no increase in height. We would instead say that if another man is 10 lbs more than the previous, he would be an average of .47in taller.

Part B

About 0.05989747 or approximatley 5.9 of men who weigh 200lbs are over 74in.

```
1-pnorm(74,71.8,sqrt(2))
```

```
## [1] 0.05989747
```

Question 2

We can see that the classics students tend to score better than the psyhcology students by seeing on which side of the y=x trend line a majority of the points lie on. Except for the extremes which lie exactly on the trend line, all other points are solidly in the classic's students range implying that on average they score better on the GRE verbal section.

Using much the same procedure as before, we can see that students who score under 500 on the GRE verbal section score better in psychology and worse in econ where as students who score above 500 tend to score better in econ and worse in psychology. Despite this analsys, because all points are relativley close to the line, the difference in scores as noted above is not as darastic as those when the pscyology and classics students were compared.

Question 3

Part A

The relationship between the height of the building and the number of floors appears to be linear given the scatter plot below. Such linear relationship appears to be strong since the points seem to form a line like blob on their own.

```
data_read<-read.csv(file = "C:/Users/mhsho/OneDrive/Documents/Stat021/Homework/skyscraper_data_c
leaned.csv")

plot(data_read$floors, data_read$height_meters, xlab = "stories", ylab = "height of building (me
ters)", main = "Height of building given number of stories")
```

2 Q2 3 / 3

✓ + 3 pts Correct

Homework 3

Stat 021

Mads Shoraka

Question 1

Part A

This is not the proper explanation. This is not what the model was constructed to compare. It was made to compare height and weight between different men, the difference in height or weight of a single man. It is also important to note the age range of the men: they have stopped growing, and therefore, even if they gained 10 lbs, there would be no increase in height. We would instead say that if another man is 10 lbs more than the previous, he would be an average of .47in taller.

Part B

About 0.05989747 or approximatley 5.9 of men who weigh 200lbs are over 74in.

```
1-pnorm(74,71.8,sqrt(2))
```

```
## [1] 0.05989747
```

Question 2

We can see that the classics students tend to score better than the psyhcology students by seeing on which side of the y=x trend line a majority of the points lie on. Except for the extremes which lie exactly on the trend line, all other points are solidly in the classic's students range implying that on average they score better on the GRE verbal section.

Using much the same procedure as before, we can see that students who score under 500 on the GRE verbal section score better in psychology and worse in econ where as students who score above 500 tend to score better in econ and worse in psychology. Despite this analsys, because all points are relativley close to the line, the difference in scores as noted above is not as darastic as those when the pscyology and classics students were compared.

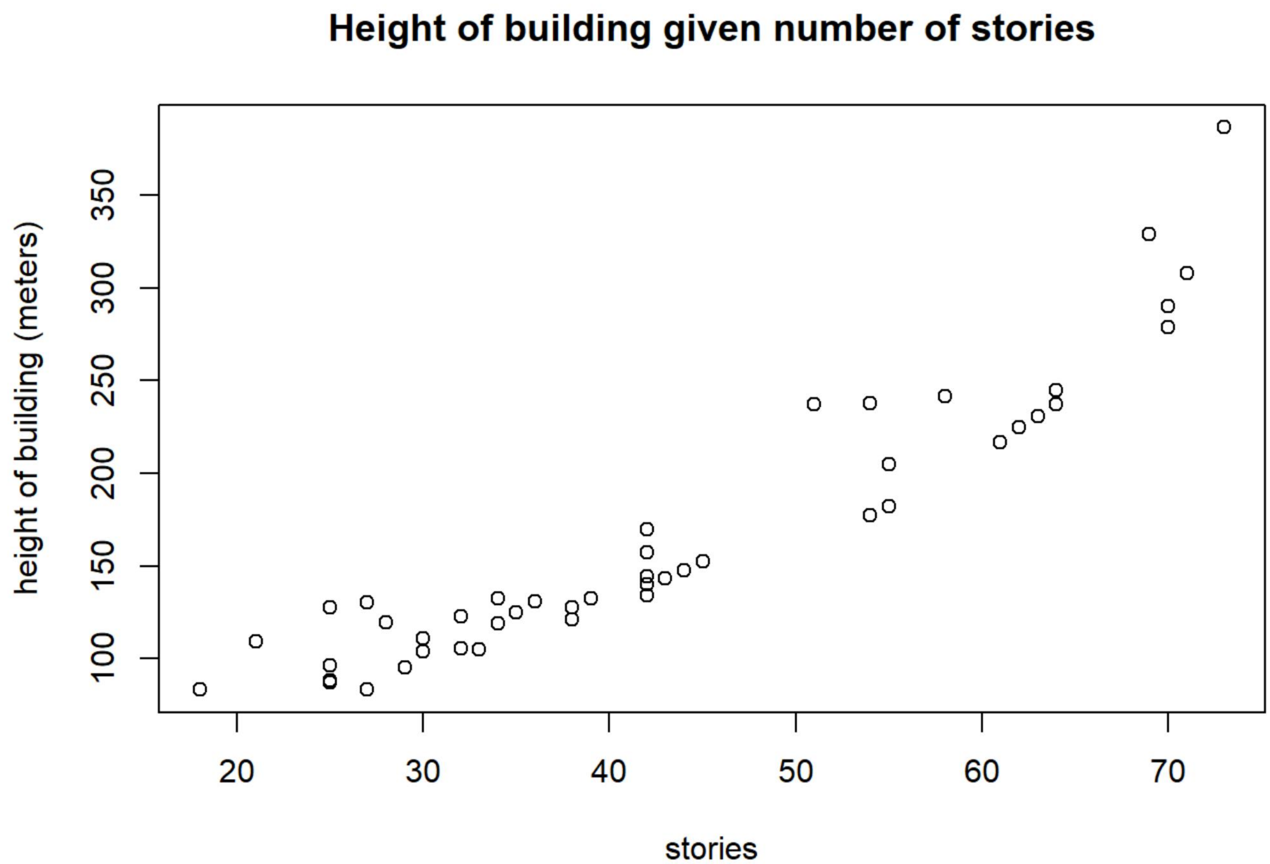
Question 3

Part A

The relationship between the height of the building and the number of floors appears to be linear given the scatter plot below. Such linear relationship appears to be strong since the points seem to form a line like blob on their own.

```
data_read<-read.csv(file = "C:/Users/mhsho/OneDrive/Documents/Stat021/Homework/skyscraper_data_c
leaned.csv")

plot(data_read$floors, data_read$height_meters, xlab = "stories", ylab = "height of building (me
ters)", main = "Height of building given number of stories")
```




```
#abline(lm(data_read$height_meters~data_read$floors))
```

Part B

The equation of the linear regression line is: $-19.4884 + 4.3078x = y$. The value of the standard deviation of height is: 11.1383. The value of R-squared is: 0.8743.

```
Linear_Regression<- lm(formula = height_meters~floors, data = data_read)
summary(Linear_Regression)
```

```
##
## Call:
## lm(formula = height_meters ~ floors, data = data_read)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.035 -18.292  -6.147   9.587  91.617
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -19.4884    11.1383   -1.75    0.087 .
## floors        4.3078     0.2434   17.70   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.8 on 45 degrees of freedom
## Multiple R-squared:  0.8743, Adjusted R-squared:  0.8716
## F-statistic: 313.1 on 1 and 45 DF,  p-value: < 2.2e-16
```

Part C

The correlation between the height of a skyscraper in meters and the number of floors is 0.9350663. This is precisley $\sqrt{R^2} = R$. Note $0.9350663^2 = 0.8743489 = R^2$. Since the correlation is a positive number close to 1, we know that there is a positive linear relationship between these variables.

```
cor(x=data_read$floors, y= data_read$height_meters)
```

```
## [1] 0.9350663
```

Part D

The confidence interval for B_1 at a 95 level is $[3.817518, 4.798148]$. We are 95% sure that the true slope of the linear regression, B_1, is within the range found above. If I was explainging this concept to an architect, I tell her that if you want to add a floor to your skyscraper you should add between 3.817518 and 4.798148 meters of height inorder to remain consistent with other skyscraper constructions.

```
confint(Linear_Regression)
```

```
##                 2.5 %   97.5 %
## (Intercept) -41.922113 2.945363
## floors        3.817518 4.798148
```

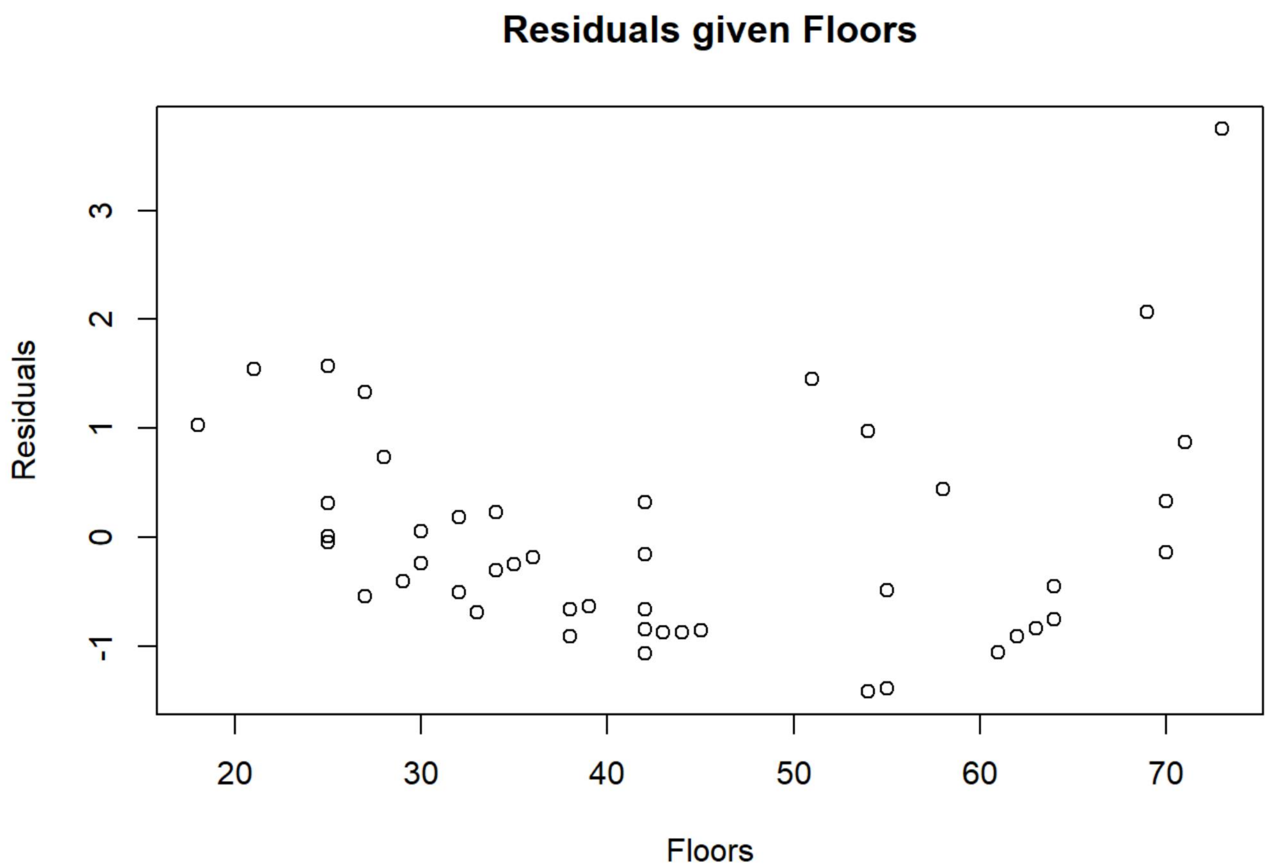
Part E

The null hypothesis is that $B_1 = 0$ and my alternative hypothesis is that $B_1 \neq 0$. We know the test statistic is $\frac{4.3078-0}{0.2434} = 17.70$, and therefore our p-val = $< 2e - 16$. Because our p-value is so low, we can conclude reject our null hypothesis that $B_1 = 0$. This makes sense within the context of our data: one cannot add a floor without adding height this is physically impossible. So, we would never have suspected $B_1 = 0$ to be true.

Part F

Looking at the residuals, we can see that near the upper end of the floors, the residuals appear to be very high which points to a failure in even variance of the error.

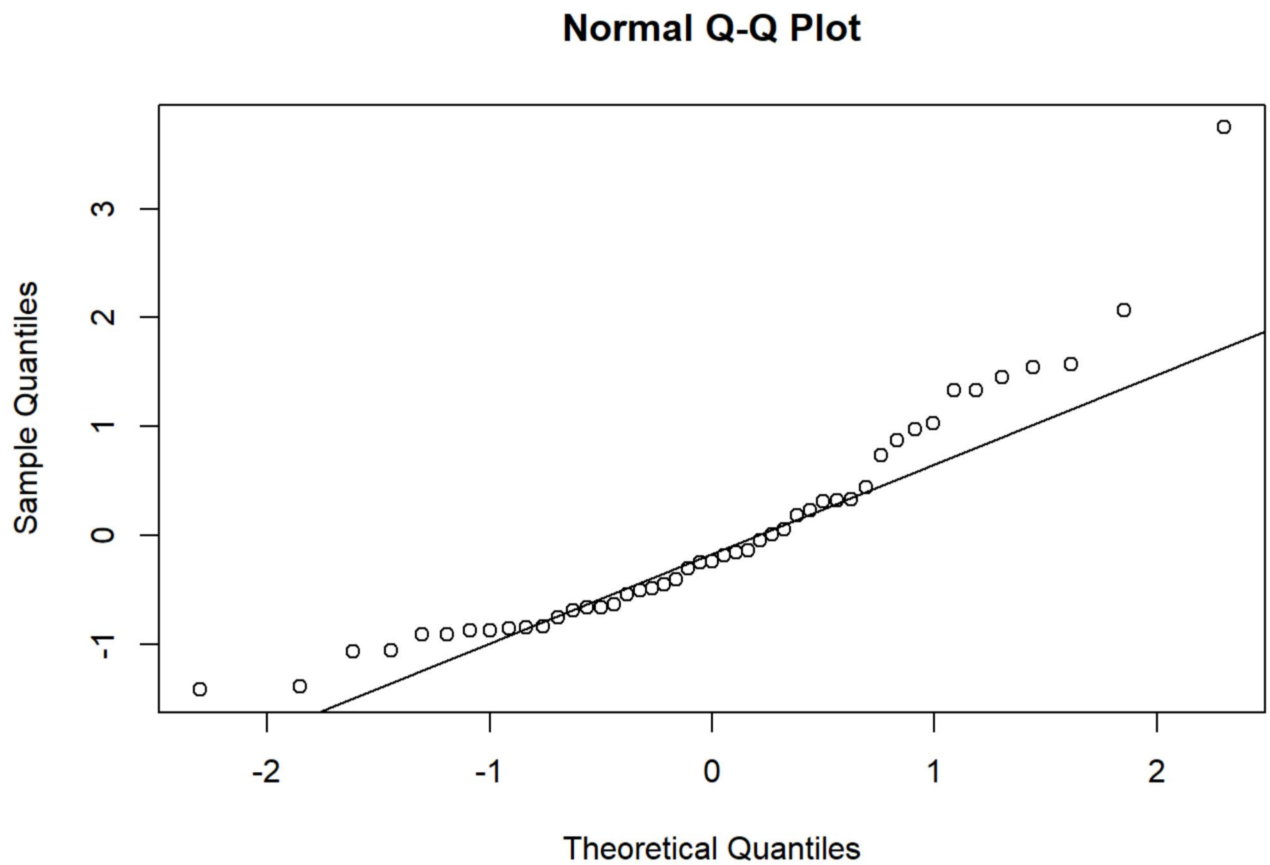
```
residuals1= rstandard(Linear_Regression)
plot(data_read$floors, residuals1, xlab = "Floors", ylab = "Residuals", main = "Residuals given
 Floors")
```



Part G

Given the normal qq plot of the residuals, we can speculate that the extreme datas' error is not normal. We can see this by looking at the deviations from the line around the beginning and end of our data. THis casts doubt on the constant variance of the error as some points converege more closley to the trend line than others.

```
qqnorm(residuals1)
qqline(residuals1)
```



Question 4

Part A

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

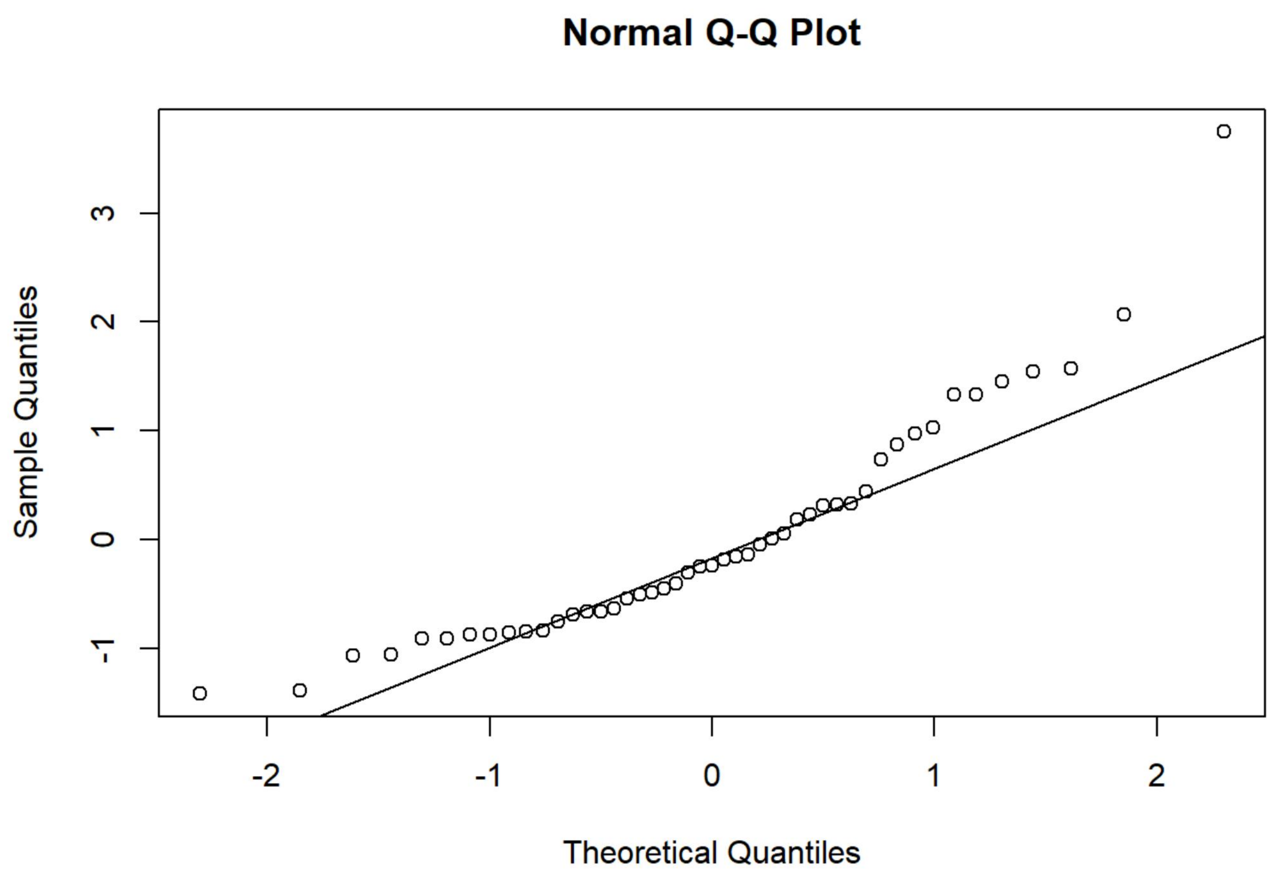
```
group_prim<- msleep%>% filter(order=="Primates")
group_carn<- msleep%>% filter(order=="Carnivora")
group_prim_linear<- lm(formula = sleep_total~bodywt, data = group_prim)
group_carn_linear<- lm(formula = sleep_total~bodywt, data = group_carn)
summary(group_prim_linear)
```

3 Q3 **9 / 10**

+ **0 pts** Correct

+ **9 Point adjustment**

b)



Question 4

Part A

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
group_prim<- msleep%>% filter(order=="Primates")
group_carn<- msleep%>% filter(order=="Carnivora")
group_prim_linear<- lm(formula = sleep_total~bodywt, data = group_prim)
group_carn_linear<- lm(formula = sleep_total~bodywt, data = group_carn)
summary(group_prim_linear)
```

```
##
## Call:
## lm(formula = sleep_total ~ bodywt, data = group_prim)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5390 -1.0032 -0.4876  0.0190  5.9085
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.11268    0.72943  15.235 3.01e-08 ***
## bodywt      -0.04414    0.02941  -1.501    0.164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.094 on 10 degrees of freedom
## Multiple R-squared:  0.1838, Adjusted R-squared:  0.1022
## F-statistic: 2.252 on 1 and 10 DF,  p-value: 0.1643
```

```
summary(group_carn_linear)
```

```
##
## Call:
## lm(formula = sleep_total ~ bodywt, data = group_carn)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0892 -1.3180  0.6448  2.4067  3.9322
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.15297    1.43235   6.390 7.93e-05 ***
## bodywt       0.01670    0.01751   0.954    0.363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.517 on 10 degrees of freedom
## Multiple R-squared:  0.08337,    Adjusted R-squared:  -0.008289
## F-statistic: 0.9096 on 1 and 10 DF,  p-value: 0.3627
```

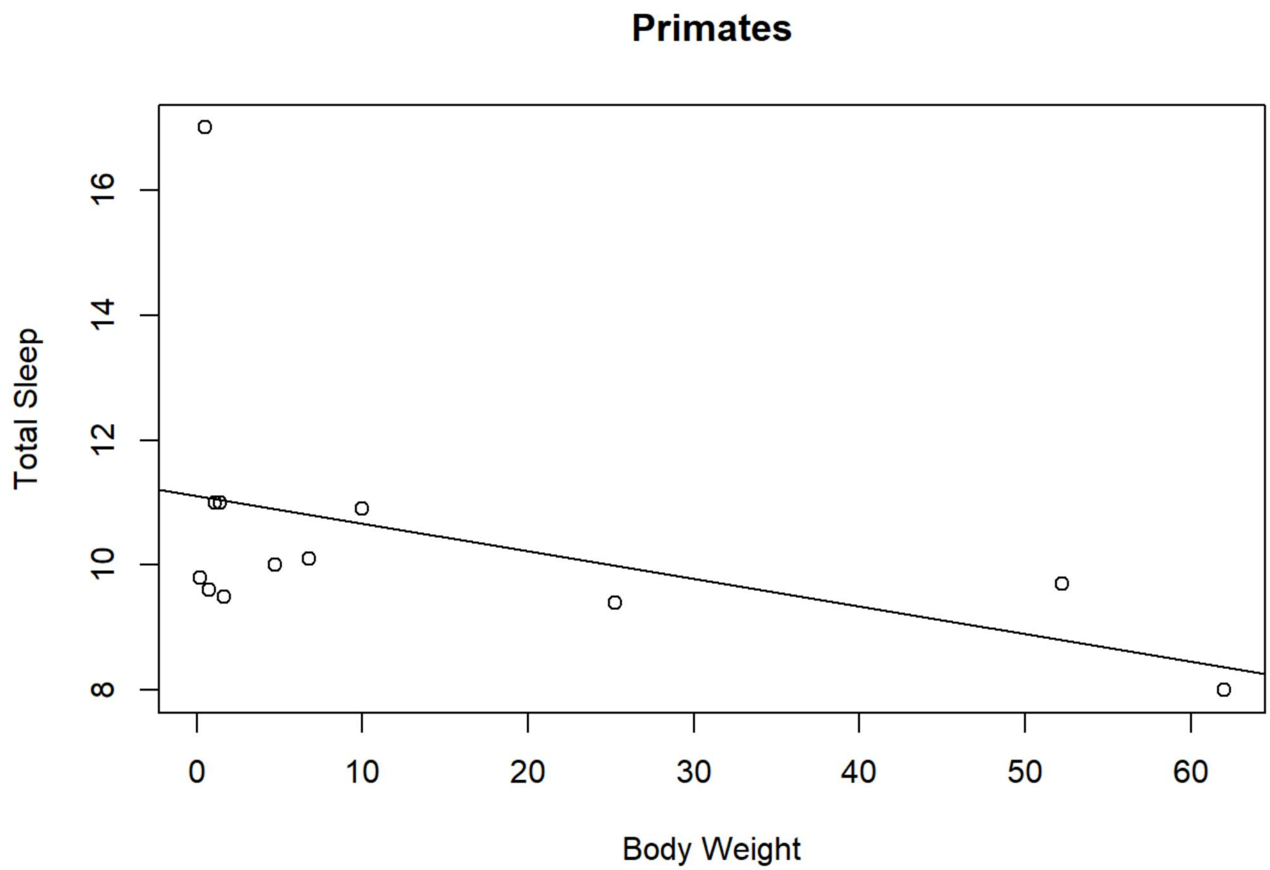
Part B

For the Canivora the residual standard error is 3.517, and for the Primates the residual standard error is 2.094.

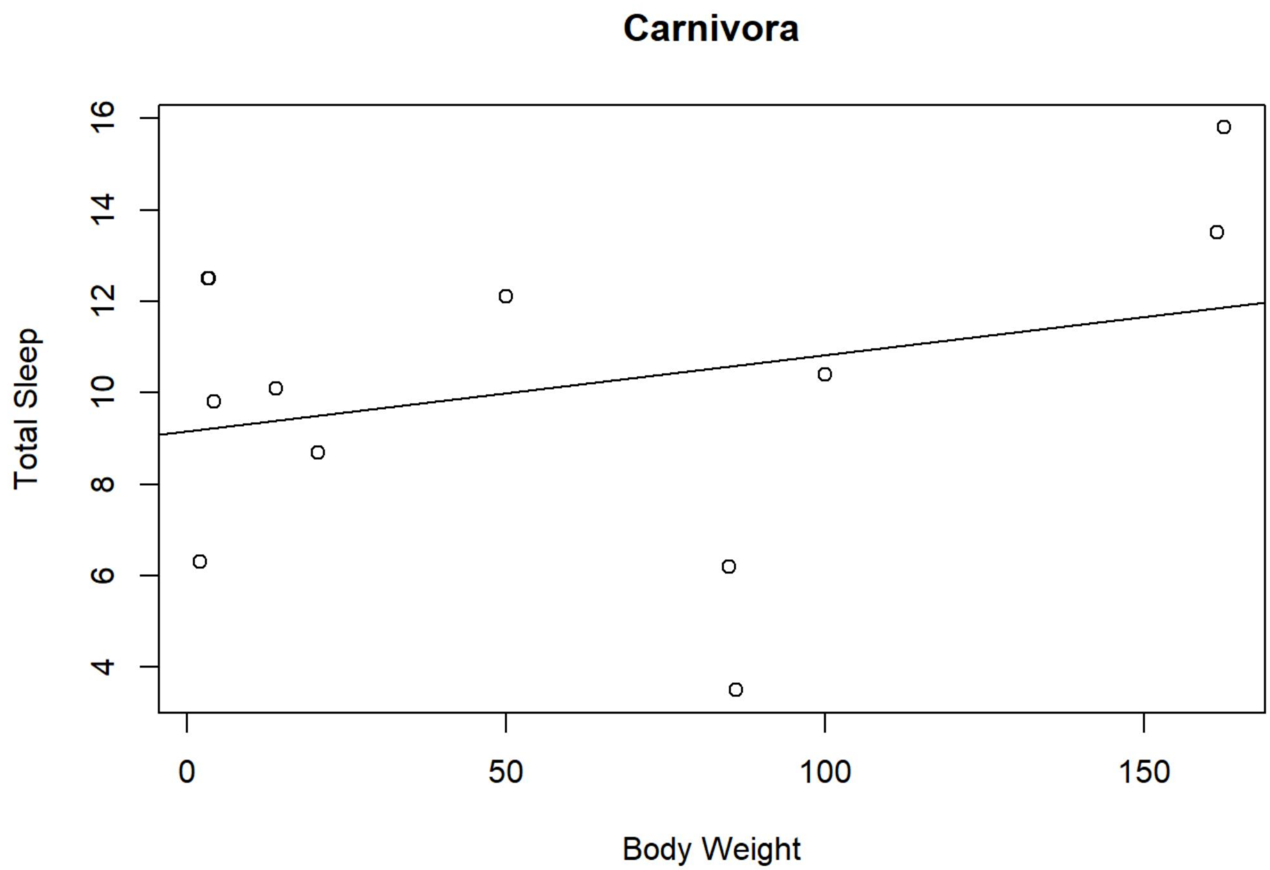
Part C

As we can see by the two graphs below, the slopes of their respective prediction equations are negative of each other. This suggests we should not combine the data and run a large linear regression.

```
plot(sleep_total ~ bodywt, data= group_prim, xlab= "Body Weight", ylab= "Total Sleep", main= "Pr
imates")
abline(lm(group_prim$sleep_total~group_prim$bodywt))
```




```
plot(sleep_total ~ bodywt, data= group_carn, xlab= "Body Weight", ylab= "Total Sleep", main= "Ca
rnivora")
abline(lm(group_carn$sleep_total~group_carn$bodywt))
```


Carnivora
Total Sleep
16
14
12
10
8
6
4
0
50
100
150
Body Weight

4 Q4 **3.5 / 4**

✓ **+ 3.5 pts** Correct