Meher Shrishti Nigam
20BRS1193

EDA LAB – 3
13 / 1 / 22

 # Meher Shrishti Nigam

 # 20BRS1193

 # EDA Lab 3

```
options(prompt="MEHERSHRISHTI>", continue =" ")
# options(prompt=">", continue =" ")
```

 # EDA-LAB-EXPERIMENT-3 (Date-13/1/2023)

 # Larger animals tend to have larger brains. But is the increase in brain size proportional to

 # the increase in body size? A set of data on body and brain size of 62 mammal species was

 # collated by Allison and Cicchetti (1976), and these data are in the data set "mammals.csv".

 # The file contains columns giving the species name, the average body mass (in kg) and

 # average brain size (in g) for each species.

```
df <- read.csv("mammals.csv")
df
```

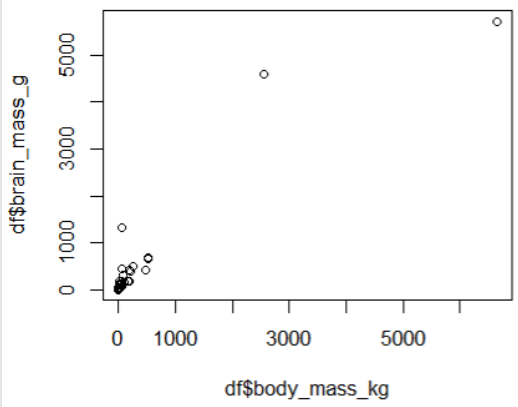
```
MEHERSHRISHTI>df <- read.csv("mammals.csv")
MEHERSHRISHTI>df
                      name body_mass_kg brain_mass_g
1               Arctic fox         3.38        44.50
2                Owl monkey         0.48        15.50
3           Mountain beaver         1.35         8.10
4                       Cow       465.00       423.00
5                Grey wolf        36.33       119.50
6                      Goat        27.66       115.00
7                 Roe deer        14.83        98.20
8                Guinea pig         1.04         5.50
9                    Verbet         4.19        58.00
10               Chinchilla         0.42         6.40
11          Ground squirrel         0.10         4.00
12    Arctic ground squirrel         0.92         5.70
```

and so on..

 # a. Plot brain size against body size. Is the relationship linear?

```
plot(df$body_mass_kg, df$brain_mass_g)
cor(df$body_mass_kg, df$brain_mass_g)
```

```
MEHERSHRISHTI># a. Plot brain size against body size. Is the relationship linear?
MEHERSHRISHTI>plot(df$body_mass_kg, df$brain_mass_g)
MEHERSHRISHTI>cor(df$body_mass_kg, df$brain_mass_g)
[1] 0.9341639
```



They are correlated but their relationship is not linear.

 # b. Find a transformation (for either or both variables) that makes the relationship

 # between these two variables linear.

log_brain <- log(df$brain_mass_g)

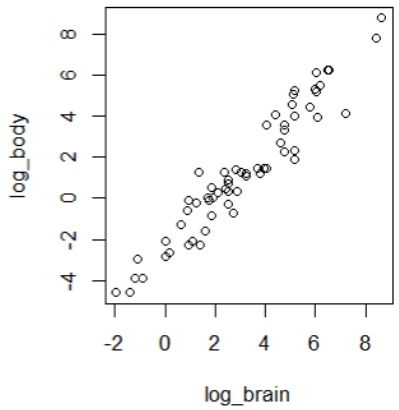
log_body <- log(df$body_mass_kg)

linear_model <- lm(log_brain ~ log_body)

plot(x=log_brain, y=log_body)

summary(linear_model)

```
MEHERSHRISHTI># b. Find a transformation (for either or both variables) that makes the r
elationship
MEHERSHRISHTI># between these two variables linear.
MEHERSHRISHTI>log_brain <- log(df$brain_mass_g)
MEHERSHRISHTI>log_body <- log(df$body_mass_kg)
MEHERSHRISHTI>linear_model <- lm(log_brain ~ log_body)
MEHERSHRISHTI>plot(x=log_brain, y=log_body)
MEHERSHRISHTI>summary(linear_model)

Call:
lm(formula = log_brain ~ log_body)

Residuals:
     Min       1Q   Median       3Q      Max
-1.71143 -0.50667 -0.05606  0.43833  1.94425

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.12719    0.09682   21.97   <2e-16 ***
log_body     0.75451    0.02878   26.22   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.699 on 60 degrees of freedom
Multiple R-squared:  0.9197,    Adjusted R-squared:  0.9184
F-statistic: 687.3 on 1 and 60 DF,  p-value: < 2.2e-16

MEHERSHRISHTI>
```



 # c. Is there statistical evidence that brain size is correlated with body size? Assume that

 # the species data are independent.

cor(df$body_mass_kg, df$brain_mass_g)

```
MEHERSHRISHTI># c. Is there statistical evidence that brain size is correlated with bo
dy size? Assume that
MEHERSHRISHTI># the species data are independent.
MEHERSHRISHTI>cor(df$body_mass_kg, df$brain_mass_g)
[1] 0.9341639
MEHERSHRISHTI>
```

Brain size and body size are correlated but their relationship is not linear.

 # d. What line best predicts (transformed) brain size from (transformed) body size?

linear_model <- lm(log_brain ~ log_body)

summary(linear_model)

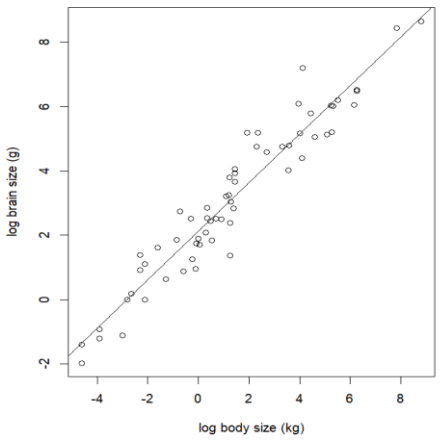
The above shows that prediction line for log brain size to log body size is the line

y = (2.13) x + (0.75)

plot(x=log_body, y=log_brain, xlab="log body size (kg)", ylab="log brain size (g)")

abline(linear_model)

```
MEHERSHRISHTI>linear_model <- lm(log_brain ~ log_body)
MEHERSHRISHTI>summary(linear_model)

Call:
lm(formula = log_brain ~ log_body)

Residuals:
     Min       1Q   Median       3Q      Max
-1.71143 -0.50667 -0.05606  0.43833  1.94425

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.12719    0.09682   21.97   <2e-16 ***
log_body     0.75451    0.02878   26.22   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.699 on 60 degrees of freedom
Multiple R-squared:  0.9197,    Adjusted R-squared:  0.9184
F-statistic: 687.3 on 1 and 60 DF,  p-value: < 2.2e-16

MEHERSHRISHTI>plot(x=log_brain, y=log_body, xlab="log body size (kg)", ylab="log brain
 size (g)")
MEHERSHRISHTI>abline(linear_model)
MEHERSHRISHTI>
```



 # e. Based on your answer in (d), what is the predicted change in log-brain size

 # accompanying an increase of 3 units of log- body size?

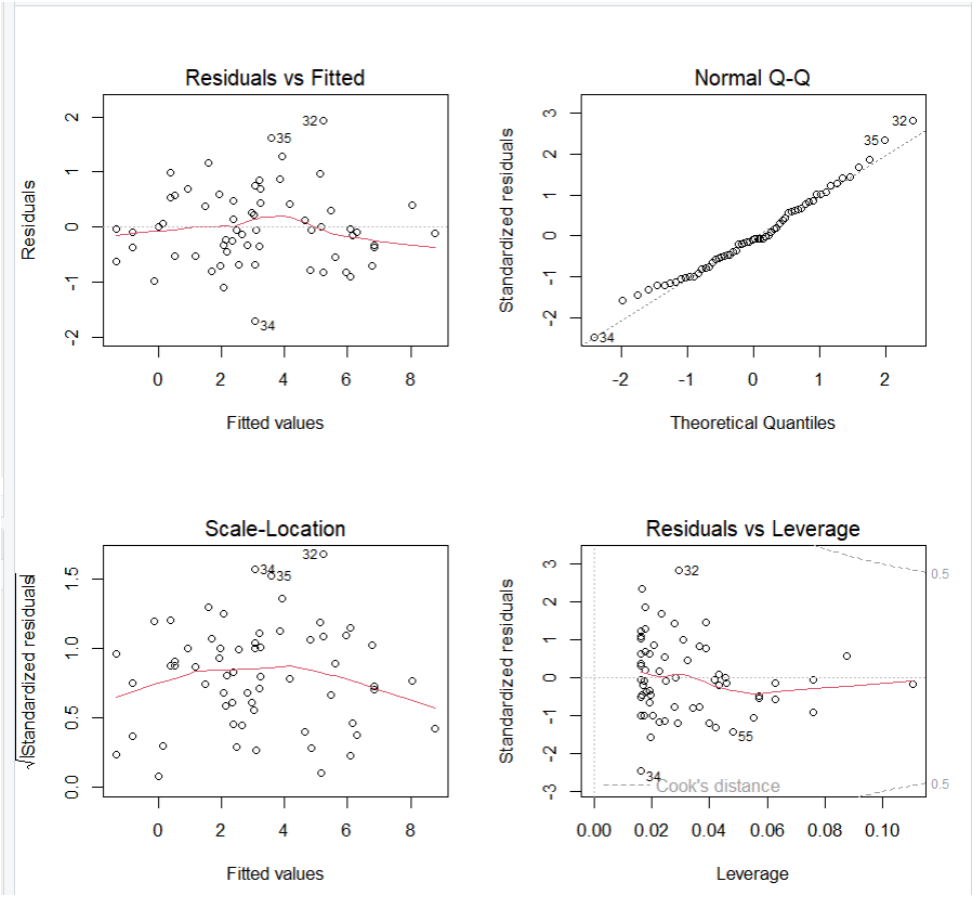
```
mammal_test <- data.frame(log_body = c(3))

predict(linear_model, newdata = mammal_test)
```

```
MEHERSHRISHTI>mammal_test <- data.frame(log_body = c(3))
MEHERSHRISHTI>predict(linear_model, newdata = mammal_test)
       1
4.390708
```

 # f. Make a residual plot using the regression fitted to the transformed variables. Do the
 # data look like they match the assumptions of linear regression?

```
par(mfrow = c(2, 2))
plot(linear_model)
```

Yes, the transformed variables look like they fairly match the assumptions of linear regression.



 # g. Which species has the highest brain size relative to that predicted by its body size?

 # Which species has the smallest brain relative to that predicted by its body size?

```
# Absolute maximum and minimum brain size

library(dplyr)

max(df$brain_mass_g)

filter(df, brain_mass_g == max(df$brain_mass_g))  # African Elephant

min(df$brain_mass_g)

filter(df, brain_mass_g == min(df$brain_mass_g))  # Lesser short tailed shrew


# Relative to body mass maximum and minimum brain size

df <- mutate(df, rel_body_brain = brain_mass_g/(body_mass_kg * 1000))

max(df$rel_body_brain)

filter(df, rel_body_brain == max(df$rel_body_brain))  # Ground Squirrel

min(df$brain_mass_g)

filter(df, rel_body_brain == min(df$rel_body_brain))  # African Elephant
```

```
MEHERSHRISHTI>library(dplyr)
MEHERSHRISHTI>max(df$brain_mass_g)
[1] 5712
MEHERSHRISHTI>filter(df, brain_mass_g == max(df$brain_mass_g))
              name body_mass_kg brain_mass_g
1 African elephant         6654         5712
MEHERSHRISHTI>min(df$brain_mass_g)
[1] 0.14
MEHERSHRISHTI>filter(df, brain_mass_g == min(df$brain_mass_g))
                       name body_mass_kg brain_mass_g
1 Lesser short-tailed shrew         0.01         0.14
MEHERSHRISHTI># Relative to body mass maximum and minimum brain size
MEHERSHRISHTI>df <- mutate(df, rel_body_brain = brain_mass_g/(body_mass_kg * 1000))
MEHERSHRISHTI>max(df$rel_body_brain)
[1] 0.04
MEHERSHRISHTI>filter(df, rel_body_brain == max(df$rel_body_brain))
             name body_mass_kg brain_mass_g rel_body_brain
1 Ground squirrel          0.1           4           0.04
MEHERSHRISHTI>min(df$brain_mass_g)
[1] 0.14
MEHERSHRISHTI>filter(df, rel_body_brain == min(df$rel_body_brain))
              name body_mass_kg brain_mass_g rel_body_brain
1 African elephant         6654         5712    0.000858431
MEHERSHRISHTI>
```