IS606: PROJECT 1

MUSA T. GANIYU

February 27, 2016

WHAT IMPACT DOES BODY WEIGHT HAVE ON BRAIN WEIGHT?

The data records the average weight of the brain and body for a number of mammal species. There are 62 rows of data. The 3 data columns include: I, the index, A1, the brain weight; B, the body weight. We seek a model of the form: B = A1 * X1.

Load data from URL

```
require(knitr);
```

```
## Loading required package: knitr
```

```
weights <- read.table("http://people.sc.fsu.edu/~jburkardt/datasets/regression/x01.txt", skip = 32, hea
kable(head(weights));
```

Body	Weight
3.385	44.5
0.480	15.5
1.350	8.1
465.000	423.0
36.330	119.5
27.660	115.0

```
names(weights);
```

```
## [1] "Body"   "Weight"
```

```
options(warn=-1)
```

** Load require package.**

```
library(plyr);
```

Note the data columns came with Body & Weight, but its actually "BrainWeight &"Body-Weight" in the data, which needs correction.

```
names(weights)[names(weights)=="Body"] <- "BrainWeight";
names(weights)[names(weights)=="Weight"] <- "BodyWeight";
str(weights);
```

```
## 'data.frame':    62 obs. of  2 variables:
##  $ BrainWeight: num  3.38 0.48 1.35 465 36.33 ...
##  $ BodyWeight : num  44.5 15.5 8.1 423 119.5 ...
```

```
names(weights);
```

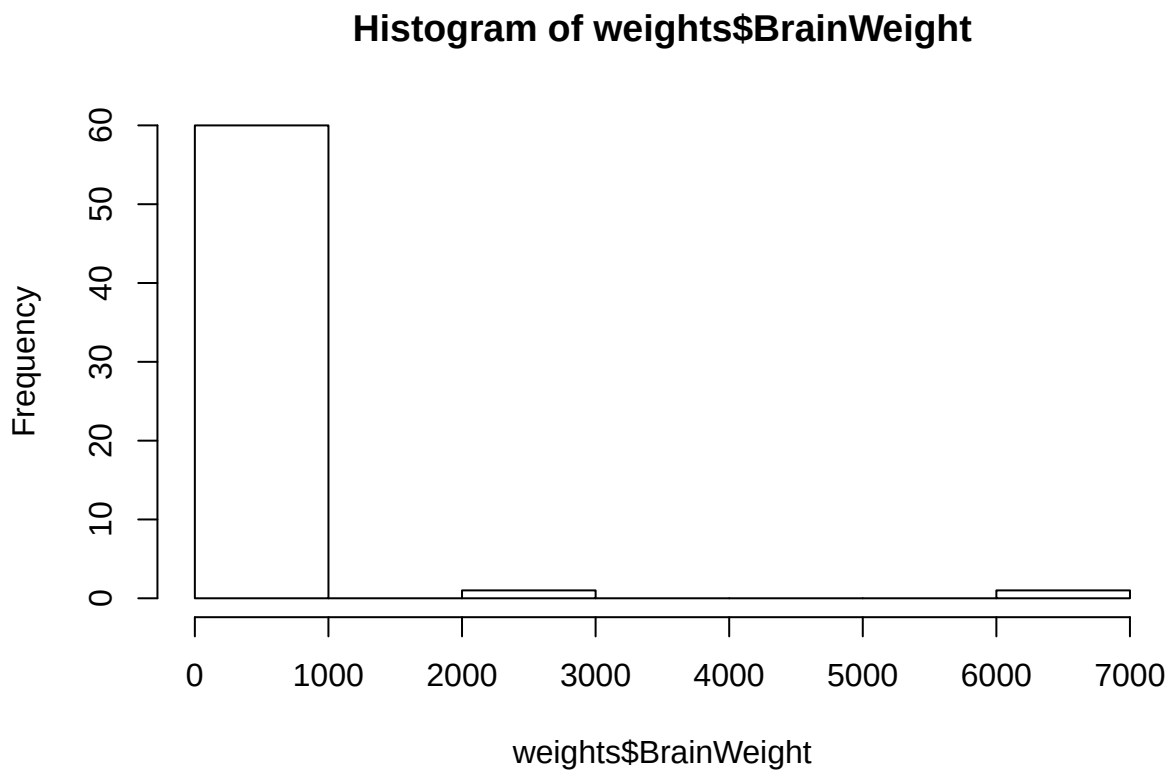
```
## [1] "BrainWeight" "BodyWeight"
```

```
kable(head(weights));
```

BrainWeight	BodyWeight
3.385	44.5
0.480	15.5
1.350	8.1
465.000	423.0
36.330	119.5
27.660	115.0

Lets checkout its histogram.

```
hist(weights$BrainWeight);
```




```
hist(weights$BodyWeight);
```

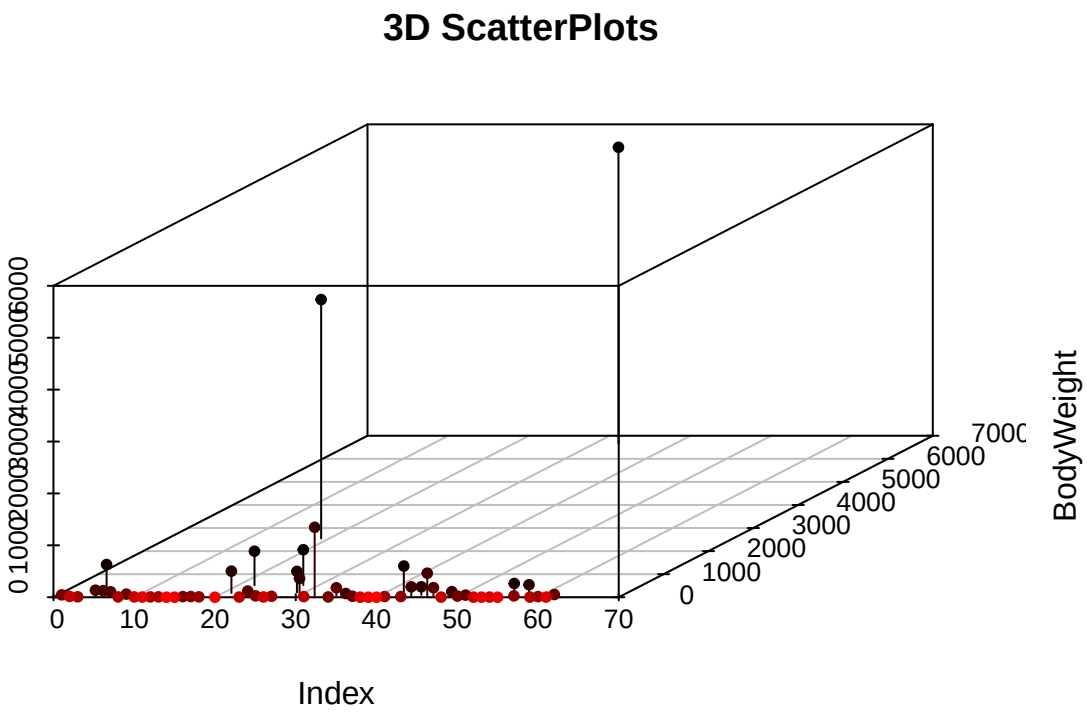
We can deduce that it a Very Rightly Skewed data, an intervention is required.Its getting the Natural logarithm of the data set.

```
weights_log <- cbind(weights, log(weights$BrainWeight), log(weights$BodyWeight));

kable(head(weights_log));
```

BrainWeight	BodyWeight	log(weights$BrainWeight)	log(weights$BodyWeight)
3.385	44.5	1.2193539	3.795489
0.480	15.5	-0.7339692	2.740840
1.350	8.1	0.3001046	2.091864
465.000	423.0	6.1420374	6.047372
36.330	119.5	3.5926438	4.783316
27.660	115.0	3.3199873	4.744932

load required package and attach the data again and plot it in scatterplot 3D.

```
library(scatterplot3d);

attach(weights_log);
scatterplot3d(BrainWeight, BodyWeight, pch = 20, highlight.3d = TRUE, type = "h", main = "3D ScatterPlo
```

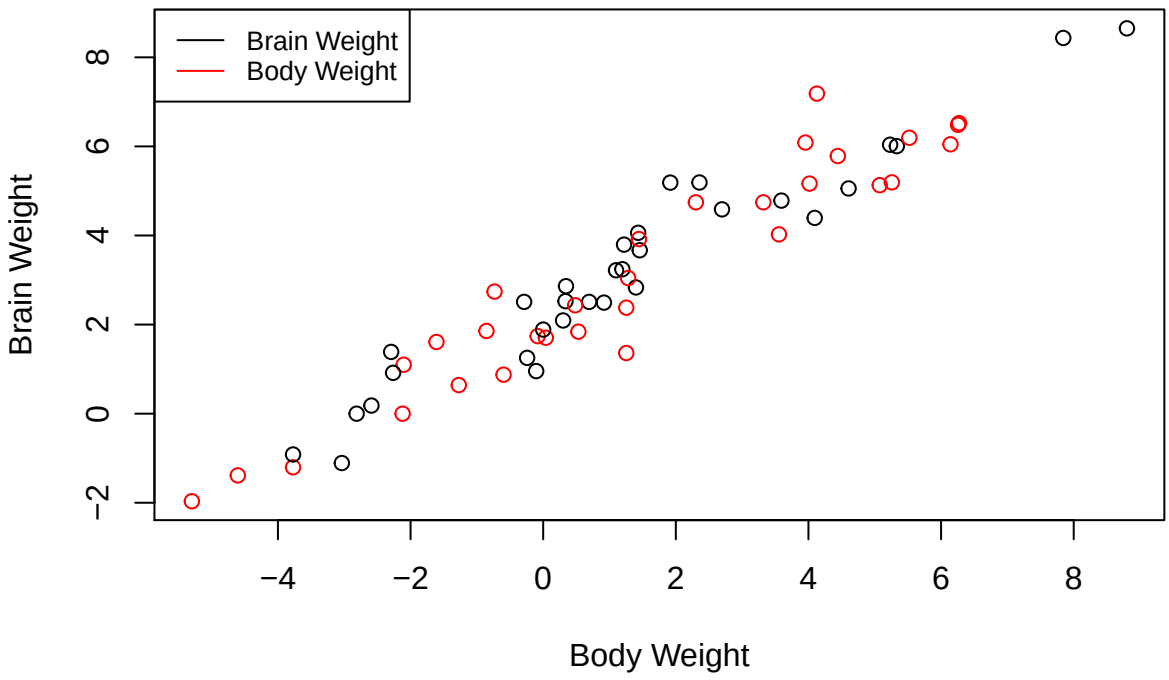


Removing or seperating new column from initial data set.

```
weights_log[,c("BrainWeight","BodyWeight")] <- list(NULL);
colnames(weights_log);
```

```
## [1] "log(weights$BrainWeight)" "log(weights$BodyWeight)"
```

```
a <- plot(weights_log, ylab="Brain Weight",
   plot.type="double", col=1:2, xlab="Body Weight")
legend("topleft", legend=c("Brain Weight","Body Weight"),
  lty=1, col=c(1,2), cex=.8)
abline(a)
```



Changing the column name.

```
names(weights_log)[names(weights_log)=="log(weights$BrainWeight)"] <- "BrainWeights";
names(weights_log)[names(weights_log)=="log(weights$BodyWeight)"] <- "BodyWeight";

kable(head(weights_log));
```

BrainWeights	BodyWeight
1.2193539	3.795489
-0.7339692	2.740840
0.3001046	2.091864
6.1420374	6.047372
3.5926438	4.783316
3.3199873	4.744932

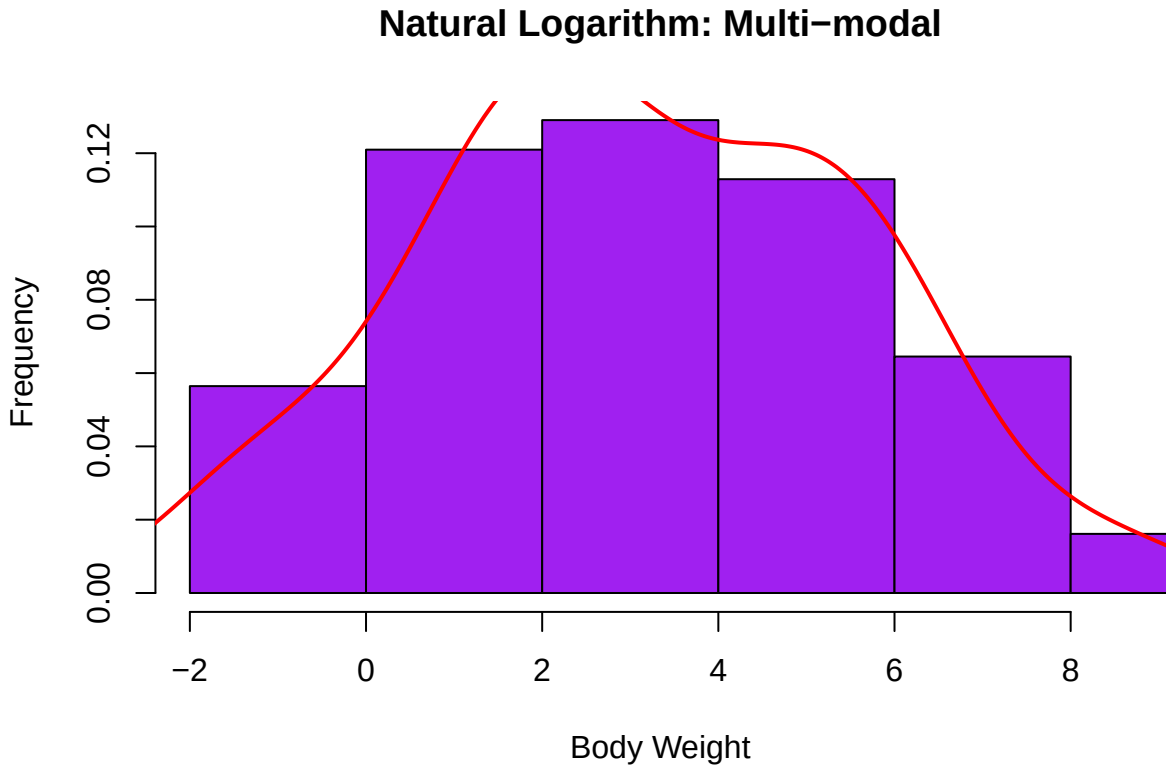
Correlations and covariance shows a strong positive relationship between Brain Weights and Body weight

```
cor(weights_log, use="complete.obs", method="kendall")
```

```
##              BrainWeights BodyWeight
## BrainWeights    1.0000000  0.8334657
## BodyWeight      0.8334657  1.0000000
```

Histogram for Body weight

```
x <- weights_log$BodyWeight;

hist(x,
 xlim=c(min(x),max(x)), probability=T,
   col='purple', xlab='Body Weight', ylab=' Frequency', axes=T,
   main='Natural Logarithm: Multi-modal')
lines(density(x,bw=1), col='red', lwd=2)
```



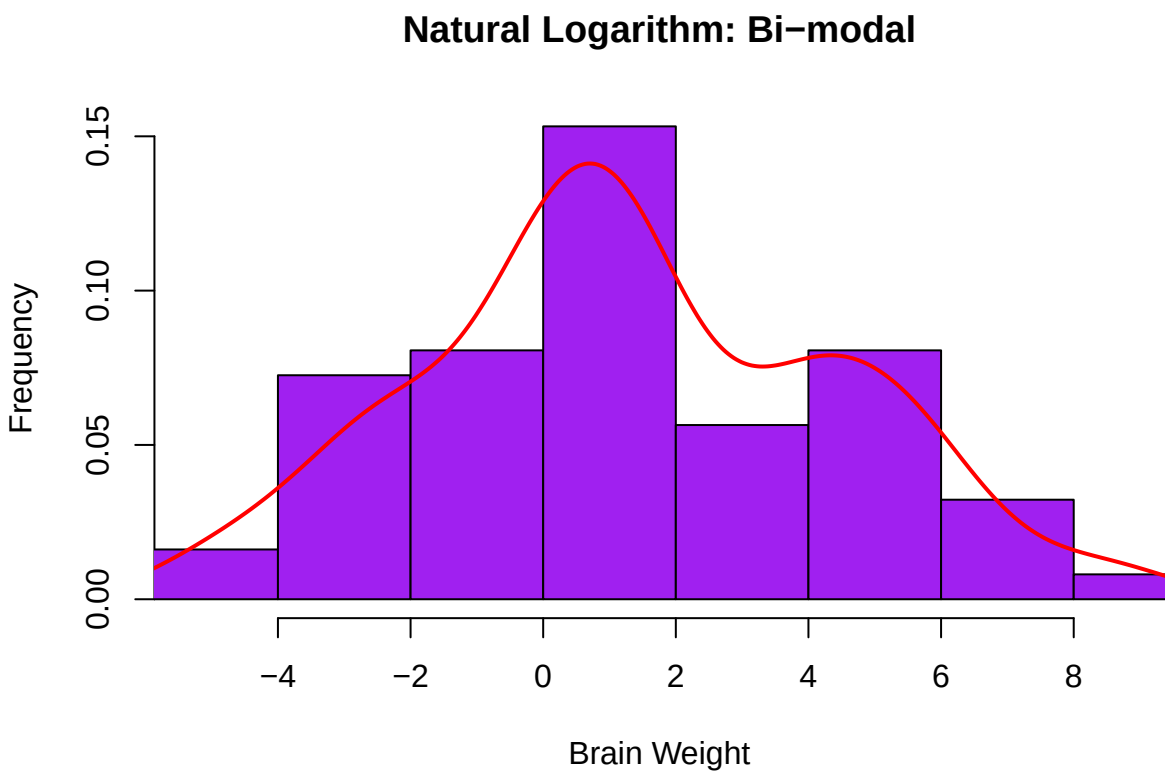
Getting the mode of BodyWeight

```
mode_1 <- table(as.vector(x));
names(mode_1)[mode_1 == max(mode_1)];
```

```
## [1] "0"                "2.50959926237837" "4.74493212836325"
```

Histogram for Brain weight

```
y <- weights_log$BrainWeight;

hist(y,
 xlim=c(min(y),max(y)), probability=T,
   col='purple', xlab='Brain Weight', ylab=' Frequency', axes=T,
   main='Natural Logarithm: Bi-modal')
lines(density(y,bw=1), col='red', lwd=2)
```




```
kable(summary(weights_log));
```

BrainWeights	BodyWeight
Min. :-5.2983	Min. :-1.966
1st Qu.:-0.5203	1st Qu.: 1.442
Median : 1.2066	Median : 2.848
Mean : 1.3375	Mean : 3.140
3rd Qu.: 3.8639	3rd Qu.: 5.111
Max. : 8.8030	Max. : 8.650

Getting the mode of Brain Weight

```
mode_2 <- table(as.vector(y));
names(mode_2)[mode_2 == max(mode_2)];
```

```
## [1] "-3.77226106305299" "1.25276296849537"
```

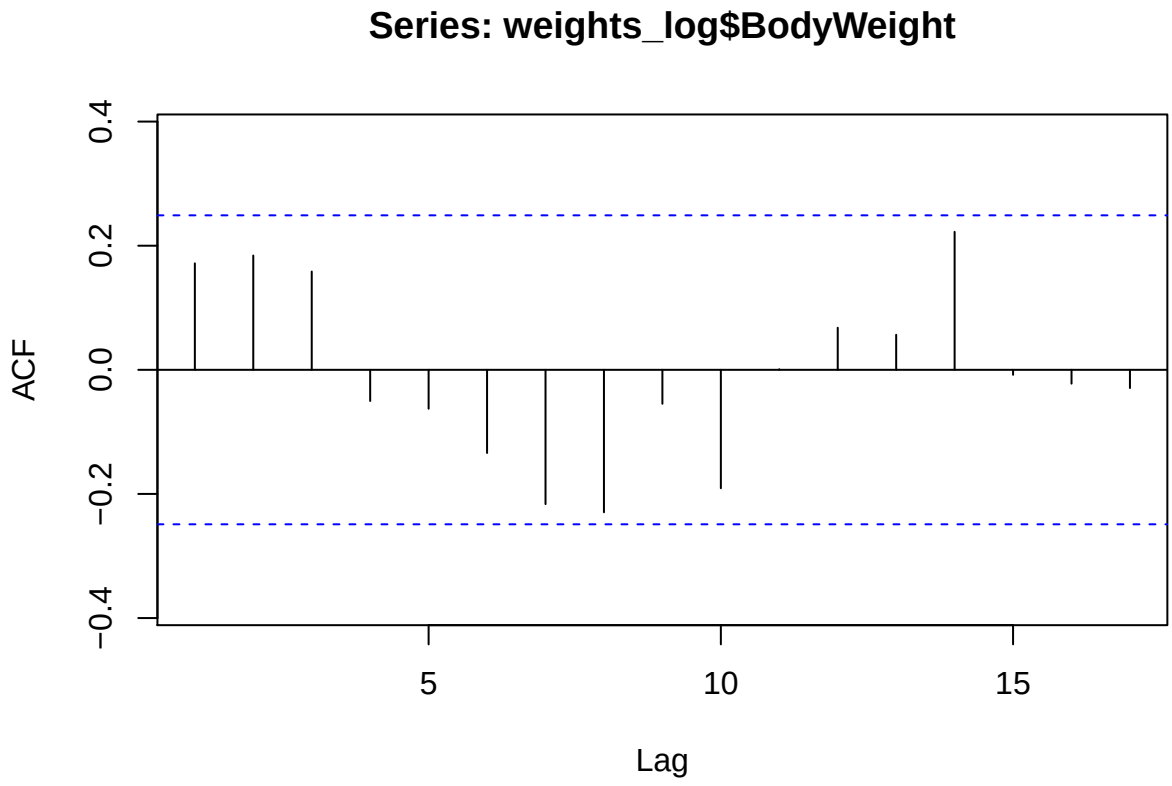
Load required package forecast, for forecasting.

```
library(forecast);
```

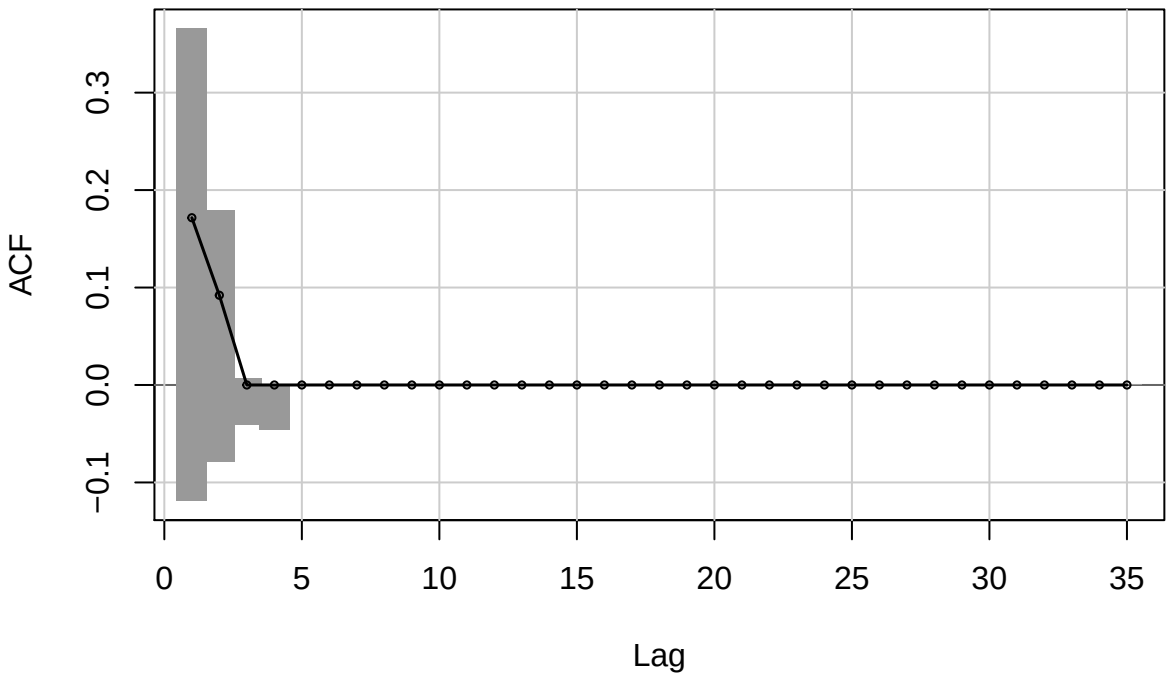
```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: timeDate

## This is forecast 6.2
```

```
Acf(weights_log$BodyWeight, lag.max=NULL, type=c("correlation", "partial"), plot=TRUE, main=NULL, xlim=
```




```
taperedacf(weights_log$BodyWeight, lag.max=NULL, type=c("correlation", "partial"), plot=TRUE, calc.ci=T
```




```
fcast <- forecast(weights_log$BodyWeight, h = ifelse(frequency(weights_log$BodyWeight) > 1, 2 * frequen

fcast;
```

```
##    Point Forecast     Lo 68    Hi 68     Lo 95    Hi 95   Lo 99.7  Hi 99.7
## 63        3.14003 0.7266564 5.553403 -1.616456 7.896516 -4.062145 10.34220
## 64        3.14003 0.7266564 5.553403 -1.616456 7.896516 -4.062145 10.34220
## 65        3.14003 0.7266564 5.553403 -1.616456 7.896516 -4.062145 10.34221
## 66        3.14003 0.7266564 5.553403 -1.616456 7.896516 -4.062145 10.34221
## 67        3.14003 0.7266564 5.553403 -1.616456 7.896516 -4.062145 10.34221
## 68        3.14003 0.7266564 5.553403 -1.616456 7.896516 -4.062145 10.34221
## 69        3.14003 0.7266564 5.553403 -1.616456 7.896516 -4.062145 10.34221
## 70        3.14003 0.7266564 5.553403 -1.616456 7.896516 -4.062145 10.34221
## 71        3.14003 0.7266563 5.553403 -1.616456 7.896516 -4.062146 10.34221
## 72        3.14003 0.7266563 5.553403 -1.616456 7.896516 -4.062146 10.34221
```

We can conclude based on this analysis that the brain weight actually depend on the Body Weight in Human.

DATA SOURCE: http://people.sc.fsu.edu/~jburkardt/datasets/regression/x01.txt

References:

http://r.789695.n4.nabble.com/converting-character-to-numeric-td3615259.html

http://stackoverflow.com/questions/4605206/drop-data-frame-columns-by-name

https://cran.r-project.org/web/packages/forecast/forecast.pdf