Stat355, Assignment #1, due: Sep 20 in class

Zimeng Ming V00844078

2019-09-18

Instructions:

1. Complete your assignment in R Markdown using this file as a template. Insert R code in the R chunks, and type in your response after the corresponding R chunk leaving one blank line between the R chunk and your comments.
2. Execute each line of code separately to ensure that it works properly.
3. Either [knit the entire document to pdf] or [knit to HTML or Word and print to pdf].
4. Submit the pdf file to CourseSpaces in the Assignment 1 activity.

Data Description:

Does fidgeting keep you slim? Surprisingly, some people do not gain weight even when they overeat. Perhaps fidgeting and other *nonexercise activity* (NEA) explains why. Some people may spontaneously increase their nonexercise activity when fed more. Researchers deliberately overfed 16 healthy young adult volunteers for eight weeks. They measured fat gain (in kilograms, kg) and, as an explanatory variable, NEA change, change in energy use (in calories, Cal) from activity other than deliberate exercise (fidgeting, daily living, etc.)

0. Read in the data and check the dimension of the dataframe. Copy the file called NEA.csv from the CourseSpaces assignments folder into your working directory. The working directory is the folder that appears in your 'Files' window of RStudio.

```
oldop<-options(digits=5)   #set the number of digits to display
NEA.dat <- read.csv('NEA.csv')
dim(NEA.dat)
```

```
## [1] 16  2
```

```
head(NEA.dat)
```

	NEAChangeCal <int>	FatGainKg <dbl>
1	-94	4.2

2	-57	3.0
3	-29	3.7
4	135	2.7
5	143	3.2
6	151	3.6

6 rows

```
tail(NEA.dat)
```

	NEAChangeCal <int>	FatGainKg <dbl>
11	486	1.6
12	535	2.2
13	571	1.0
14	580	0.4
15	620	2.3
16	690	1.1

6 rows

1. Provide summary statistics for the data and comment on the results noting any unusual features in the data.

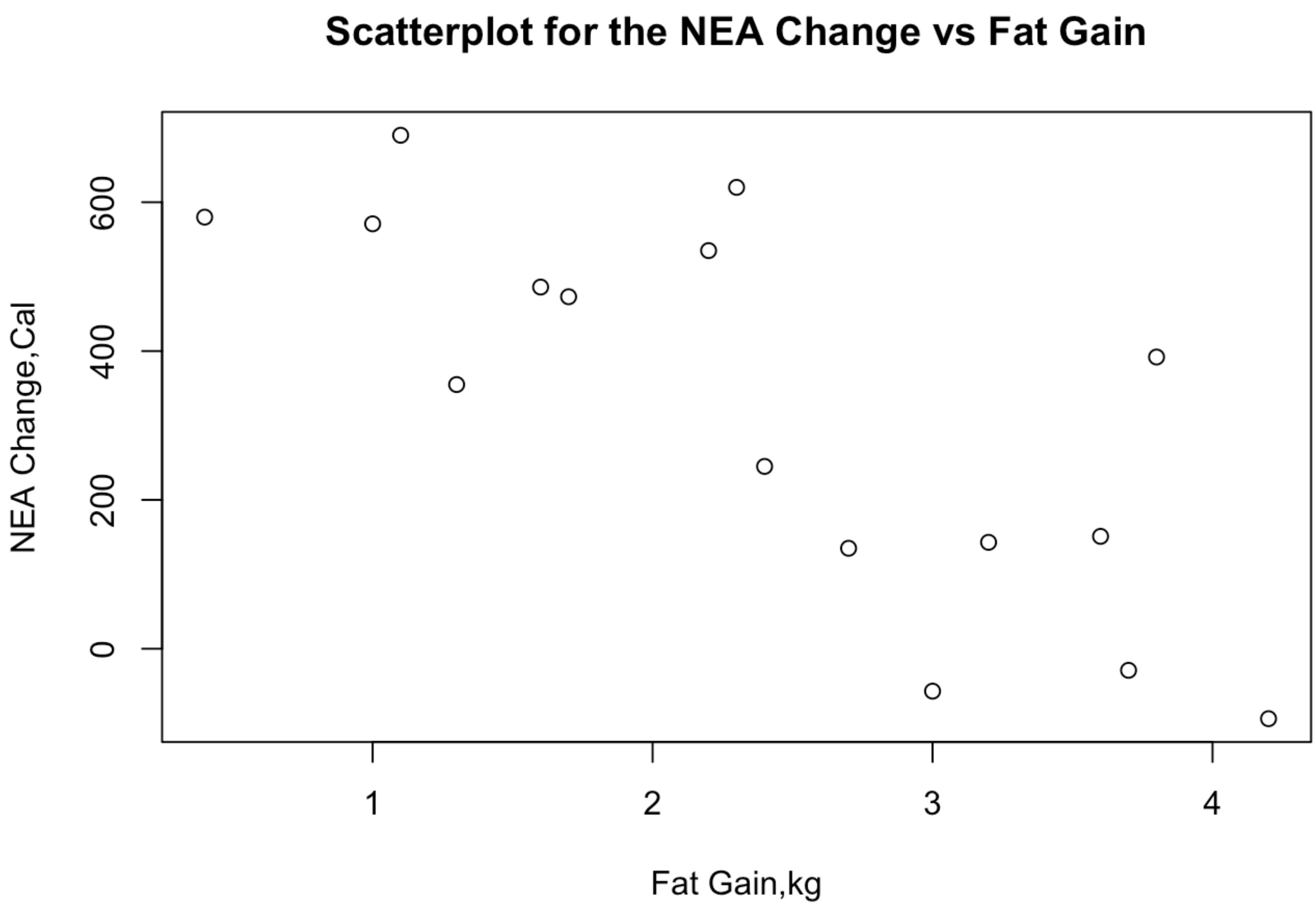
```
summary(NEA.dat)
```

```
##   NEAChangeCal   FatGainKg
##  Min.   :-94   Min.   :0.40
##  1st Qu.:141   1st Qu.:1.52
##  Median :374   Median :2.35
##  Mean   :325   Mean   :2.39
##  3rd Qu.:544   3rd Qu.:3.30
##  Max.   :690   Max.   :4.20
```

Comments: The median of NEA change is higher then the mean of Nea Change and the median of the fatGain is lower then the mean of the FatGain.

2. Produce an **appropriate** scatterplot for the data. (Carefully choose which variables should be on the vertical and horizontal axes.) Is a linear model appropriate for the data? Comment on the graph.

```
plot(NEA.dat$FatGainKg,NEA.dat$NEAChangeCal,main = "Scatterplot for the NEA Change vs
Fat Gain", ylab="NEA Change,Cal",xlab="Fat Gain,kg")
```



Comment: It looks like a negative linear model appropriate for the data.

---

3. Fit a linear model to the data, i.e. compute Least Squares estimates of the slope and the intercept. What is the estimated regression line?

```
NEA.lm<-lm(NEA.dat$NEAChangeCal~NEA.dat$FatGainKg)
NEA.lm
```

```
##
## Call:
## lm(formula = NEA.dat$NEAChangeCal ~ NEA.dat$FatGainKg)
##
## Coefficients:
##       (Intercept)  NEA.dat$FatGainKg
##               745               -176
```

```
summary(NEA.lm)
```

```
##
## Call:
## lm(formula = NEA.dat$NEAChangeCal ~ NEA.dat$FatGainKg)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -273.9 -105.3  -18.4   64.5  316.0
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)          745.3       99.8    7.47    3e-06 ***
## NEA.dat$FatGainKg   -176.1       37.9   -4.64  0.00038 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 167 on 14 degrees of freedom
## Multiple R-squared:  0.606,  Adjusted R-squared:  0.578
## F-statistic: 21.5 on 1 and 14 DF,  p-value: 0.000381
```

Comment: The estimate regression line: NEAChangeCal=745-176*FAtGainkg beta0=745, beta1=-176

---

4. Generate confidence intervals for the slope and intercept. Comment.

```
NEA.lnconfid<-confint(NEA.lm)
NEA.lnconfid
```

```
##                      2.5 %  97.5 %
## (Intercept)         531.24 959.286
## NEA.dat$FatGainKg -257.51 -94.748
```

The confidence intervals for the intercept:(531.24,959.286) The confidence intervals for the slope: (-257.51, -94.748)

5. Provide a summary of the linear model fit and perform a test of the hypothesis that the slope parameter is zero including your comment.

```
summary(NEA.lm)
```

```
##
## Call:
## lm(formula = NEA.dat$NEAChangeCal ~ NEA.dat$FatGainKg)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -273.9 -105.3  -18.4   64.5  316.0
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)          745.3       99.8    7.47    3e-06 ***
## NEA.dat$FatGainKg   -176.1       37.9   -4.64  0.00038 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 167 on 14 degrees of freedom
## Multiple R-squared:  0.606,  Adjusted R-squared:  0.578
## F-statistic: 21.5 on 1 and 14 DF,  p-value: 0.000381
```

Comment: null hypothesis for the t-test:

For the intercept: H0: beta0 =0, since the p-value is small thus there is significant eveidence to aginst h0, we conclude there is no evidence shows the beta0 =0.

For the slope: H0:beta1=0, since the p-value is small thus there is significant eveidence to aginst h0, we conclude there is no evidence shows the beta1 =0. There is a significant linear relationship between NEA change and Fat Gain.

for the F test, H0:beta1=0, since the p-value is small thus there is significant eveidence to aginst h0, we conclude there is no evidence shows the beta1 =0. There is a significant linear relationship between NEA change and Fat Gain.

Reference: Levine, J.A., Eberhardt, N., Jensen, M.D. (1999) “Role of nonexercise activity thermogenesis in resistance to fat gain in humans”, *Science*, 283, 212-214.