STAR 513: HW 4

Yvette Uwineza

Total points: 32
Questions are worth **2 pts** each, except where noted.
See Canvas calendar for due date.

Homework should be submitted as a pdf, doc or docx file via Canvas.
Use of R markdown HW template is strongly encouraged.
Add or delete code chunks as needed.
Knit frequently to avoid last minute problems!
Your submitted assignment should be neatly formatted and organized.

For this assignment we consider data from the study "Sedentary behavior associated with reduced medial temporal lobe (MTL) thickness in middle-aged and older adults" by Siddarth et al. (PLOS One, 2018). The goal of the study was to examine variables that may be associated with MTL. In particular, they are interested in physical activity and sedentary behaviors. The study includes n = 35 middle aged and older adults who volunteered to participate.

For this assignment, we use the following variables:
MTL: medial temporal lobe thickness in mm measured by high resolution MRI scan
Age: age recorded in years
Sitting: self-reported average time sitting reported in hours per day
Activity: self-reported activity recorded in MET minutes per week based on International Physical Activity Questionnaire (IPAQ) modified for older adults

MTL will be the response variable (Y) for all analyses. Loosely speaking, larger values of MTL are an indication of "better" brain health. The abstract states "Atrophy of the MTL occurs with aging, resulting in impaired episodic memory."

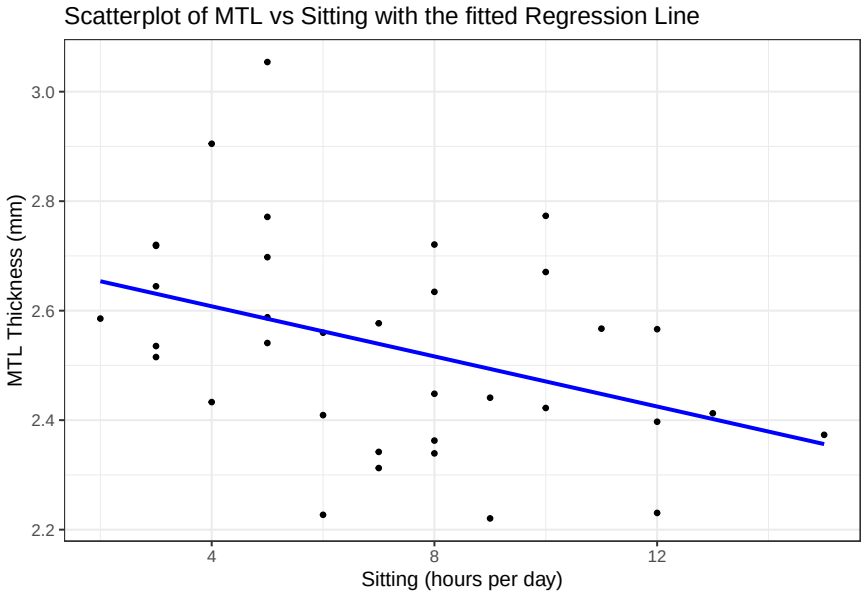
The data BrainData.csv is available from Canvas.

SLR (Q1 - Q5)

In this group of questions we will use Sitting as a predictor.

Q1

Create a scatterplot of MTL vs Sitting with the fitted regression line overlaid. Detailed axis labels not required, but encouraged.



Q2

Fit an appropriate model (with Sitting as the only predictor) and show the summary() output.

```
##
## Call:
## lm(formula = MTL ~ Sitting, data = BrainData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33511 -0.13432 -0.00252  0.11527  0.46907
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.69951    0.07309  36.933   <2e-16 ***
## Sitting     -0.02288    0.00924  -2.476   0.0186 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1791 on 33 degrees of freedom
## Multiple R-squared:  0.1567, Adjusted R-squared:  0.1312
## F-statistic: 6.132 on 1 and 33 DF,  p-value: 0.01857
```

Q3 (4 pts)

Using this model, do we have evidence (at the alpha = 0.05 level) of a linear association between MTL and Sitting? Is the association positive or negative? As part of your response, include an appropriate estimate and p-value.

Yes, there is evidence of linear association between MLT and Sitting (at the alpha = 0.05 level). The association is negative and the estimated coefficient is -0.02288 with a pvalue of 0.0186. Given this p value that is less than 0.05, we can reject the null hypothesis that this coefficient is zero therefore confirming that there is a significant linear negative relationship between MLT and Sitting.

Q4

Using this model, what proportion of variation in MTL is explained by Sitting?

Looking at the our model result, the proportion of variation in MLT explained by sitting are 15.67 percent which is reported by r squared of 0.1567 in our regression results.

Q5

At the bottom of the summary output an F test is shown (F = 6.132, p = 0.0186). State the null hypothesis using standard greek letter notation. Hint: An equivalent test appears in the Coefficients table.

$H_0 : \beta_1 = 0$

$H_a : \beta_1 \neq 0$

Looking at the pvalue of our ftest which $0.0186 < 0.05$, we reject the null hypothesis that sitting has no linear relationship with MTL thickness.

MLR (Q6 - Q13)

In this group of questions we will consider 3 predictors: Age, Sitting, Activity.

Q6 (4 pts)

Calculate pairwise (Pearson) correlations between all four variables (MTL, Age, Sitting, Activity). This should be presented as a 4×4 matrix.
Based on your results, which predictor has the strongest correlation with MTL (positive or negative)?
Mention at least one other interesting finding based on correlations.

```
##                  MTL        Age     Sitting    Activity
## MTL       1.00000000  0.2483626 -0.39586138  0.01697377
## Age       0.24836257  1.0000000 -0.18024938 -0.12004481
## Sitting  -0.39586138 -0.1802494  1.00000000 -0.07310829
## Activity  0.01697377 -0.1200448 -0.07310829  1.00000000
```

??? has the strongest correlation with MTL.

Looking at the correlation matrix above, the strongest correlation with MTL is sitting and the correlation is approximately -0.3958.

Another interesting finding is....
It's the correlation between age and MTL. As you age, MTL thickness increases. This is represented by a positive correlation of 0.248 which is not as strong as the one with sitting.

Q7

Fit an appropriate model including Sitting, Activity and Age as predictors and show the summary() output.

```
##
## Call:
## lm(formula = MTL ~ Sitting + Age + Activity, data = BrainData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30500 -0.13119 -0.02225  0.13292  0.45199
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.417e+00  2.706e-01   8.933 4.42e-10 ***
## Sitting     -2.090e-02  9.549e-03  -2.189   0.0362 *
## Age          4.391e-03  3.945e-03   1.113   0.2743
## Activity     1.992e-06  2.566e-05   0.078   0.9386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1812 on 31 degrees of freedom
## Multiple R-squared:  0.1892, Adjusted R-squared:  0.1108
## F-statistic: 2.412 on 3 and 31 DF,  p-value: 0.0856
```

Q8 (4 pts)

Using this model, provide a detailed interpretation of the estimated coefficient corresponding to **Sitting** in context of this research study. Your interpretation should include appropriate units and the numeric value for the estimated coefficient.

An addition hour of sitting per day is associated with 0.0209 mm decrease in MTL thickness, holding age and activity constant.

Q9

Using this model (controlling for Activity and Age), provide an appropriate p-value to assess a linear association between between MTL and Sitting.

p = 0.0362

Q10

What proportion of variation in MTL is explained by the linear regression including Sitting, Activity and Age?

The proportion of variation explained by the linear regression is 18.92 percent which is given by the rsquared value of 0.1892.

Q11

At the bottom of the summary output an F test is shown (F = 2.412, p = 0.0856). State the null hypothesis using standard greek letter notation.

$H_0 : \beta_Sitting = \beta_Age = \beta_Activity = 0$

$H_a : Atleastoneofthebetasisdifferentfromzero$

Q12

Considering all your results, can we conclude that Sitting **causes** a decrease in MTL? Briefly justify your response

We can not say that there is a causation just because we see an association between sitting and lower MTL thickness. This is an observational study, so we can not infer causal relationships.

Q13

Suppose the investigators were primarily interested in the effect of Sitting. What is the benefit of considering Activity and Age in the model?

This helps us to estimate the independent impact of sitting on the MTL thickness without overestimating or underestimating its impact. We are able to control for confounding variables.

Appendix

```
#Retain this code chunk!!!
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(message = FALSE)
library(knitr)
library(tinytex)
library(tidyverse)
library(broom)
library(emmeans)

# Loading the dataset

BrainData <- read_csv("/data/yuwineza/star_513/Homework 4/BrainData.csv")

#Q1

ggplot(BrainData, aes(x = Sitting, y = MTL)) +
  geom_point(color = "black", size = 1) +
  geom_smooth(method = "lm", color = "blue", se = FALSE) +
  labs(x = "Sitting (hours per day)", y = "MTL Thickness (mm)",
       title = "Scatterplot of MTL vs Sitting with the fitted Regression Line") +
  theme_bw()

#Q2

model_1 <- lm(MTL ~ Sitting, data = BrainData)

summary(model_1)

#Q6

pairwise_pearson<- cor(BrainData[, c("MTL", "Age", "Sitting", "Activity")])

pairwise_pearson

#Q7
model_2 <- lm(MTL ~ Sitting + Age + Activity, data = BrainData)
summary(model_2)
```