Group Project

Ting-Wei Lin

12/5/2019

```
# libraries: ------------------------------------------------------------------
library(Hmisc)
library(data.table)
library(splines)

# data: -----------------------------------------------------------------------
path = "/Users/Sabrina/Documents/2019UMICH/STATS506/group project/data/"

sleep = as.data.table(sasxport.get(paste0(path, "SLQ_D.XPT")))

physical = as.data.table(sasxport.get(paste0(path, "PAQ_D.XPT")))

physical_indv = as.data.table(sasxport.get(paste0(path, "PAQIAF_D.XPT")))

demo = as.data.table(sasxport.get(paste0(path, "DEMO_D.XPT")))

dietary1 = as.data.table(sasxport.get(paste0(path, "DR1TOT_D.XPT")))

dietary2 = as.data.table(sasxport.get(paste0(path, "DR2TOT_D.XPT")))

# 80: -------------------------------------------------------------------------

# keep the variables we need
sleep = sleep[, .(seqn, sld010h)]
physical = physical[, .(seqn, pad080, paq520)]
physical_indv = physical_indv[, .(seqn, padtimes, paddurat)]
demo = demo[, .(seqn, riagendr, ridageyr, ridreth1, indfminc, ridexmon)]
dietary1 = dietary1[, .(seqn, dr1tkcal, dr1tsugr, dr1tcaff, day = 1)]
dietary2 = dietary2[, .(seqn, dr2tkcal, dr2tsugr, dr2tcaff, day = 2)]

# rename colnames
names(dietary1) = c("seqn", "drtkcal", "drtsugr", "drtcaff", "day")
names(dietary2) = c("seqn", "drtkcal", "drtsugr", "drtcaff", "day")

# merge day 1 data
data1 = merge(sleep, physical, all = TRUE)
data1 = merge(data1, physical_indv, all = TRUE)
data1 = merge(data1, demo, all = TRUE)
data1 = merge(data1, dietary1, all = TRUE)

# merge day 2 data
data2 = merge(sleep, physical, all = TRUE)
data2 = merge(data2, physical_indv, all = TRUE)
data2 = merge(data2, demo, all = TRUE)
data2 = merge(data2, dietary2, all = TRUE)

# row bind day 1 and day 2 data
```

```r
data = rbind(data1, data2)

# omit the rows that have missing values
data_naomit = data[complete.cases(data[])]

# take mean for repeated seqn
avg_value = data_naomit[, lapply(.SD, mean), by = .(seqn),
                    .SDcols = c("sld010h", "pad080", "paq520", "padtimes",
                        "paddurat", "ridageyr", "ridreth1", "indfminc",
                        "ridexmon", "drtkcal",  "drtsugr",  "drtcaff")]
```

```r
# fit linear model
model = lm(sld010h ~ pad080 + padtimes + ridageyr + as.factor(ridreth1) +
             indfminc + as.factor(ridexmon) + drtkcal + drtsugr + drtcaff,
           data = avg_value)
summary(model)
```

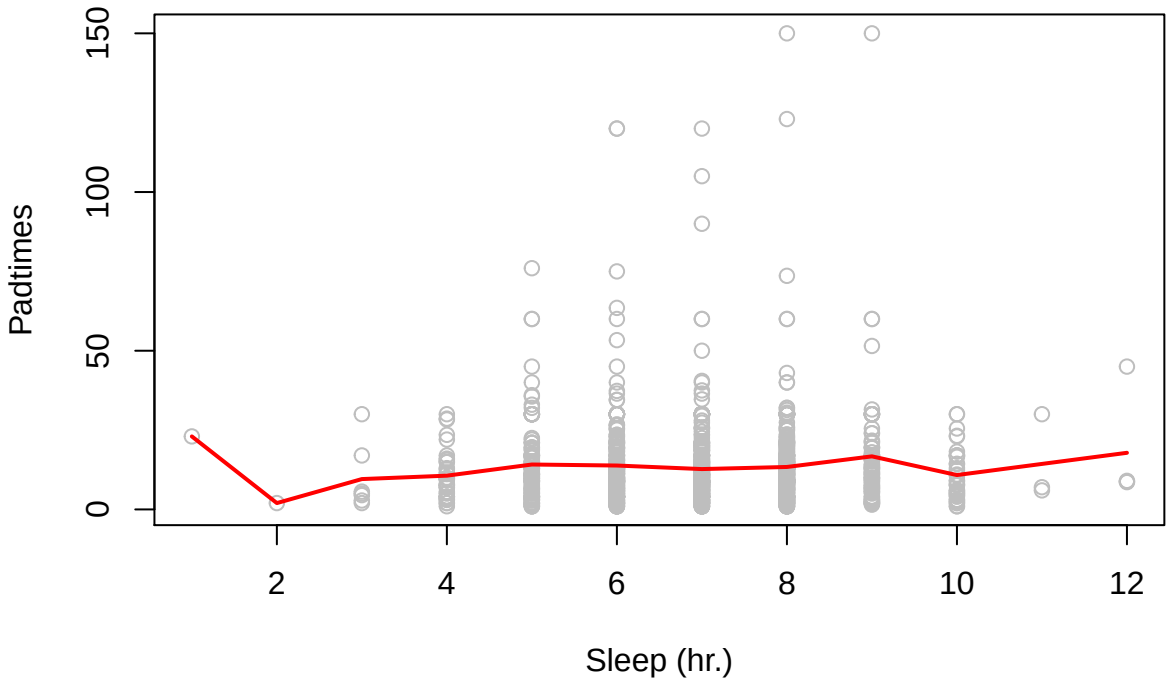
```
##
## Call:
## lm(formula = sld010h ~ pad080 + padtimes + ridageyr + as.factor(ridreth1) +
##     indfminc + as.factor(ridexmon) + drtkcal + drtsugr + drtcaff,
##     data = avg_value)
##
## Residuals:
## How much sleep do you get (hours)?
##     Min      1Q  Median      3Q     Max
## -5.8500 -0.8438 -0.0088  0.8833  5.2411
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            7.786e+00  1.476e-01  52.743  < 2e-16 ***
## pad080                -1.922e-03  6.707e-04  -2.866 0.004228 **
## padtimes               4.931e-03  2.764e-03   1.784 0.074679 .
## ridageyr              -6.926e-03  2.085e-03  -3.322 0.000919 ***
## as.factor(ridreth1)2 -3.866e-01  2.019e-01  -1.915 0.055757 .
## as.factor(ridreth1)3  2.258e-02  1.061e-01   0.213 0.831471
## as.factor(ridreth1)4 -5.699e-01  1.063e-01  -5.360 9.83e-08 ***
## as.factor(ridreth1)5 -1.190e-01  1.827e-01  -0.651 0.514866
## indfminc               1.454e-03  3.435e-03   0.423 0.672248
## as.factor(ridexmon)2 -9.247e-02  7.793e-02  -1.187 0.235635
## drtkcal               -7.351e-05  5.338e-05  -1.377 0.168705
## drtsugr               -1.995e-04  6.589e-04  -0.303 0.762048
## drtcaff               -7.186e-04  2.666e-04  -2.696 0.007107 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.319 on 1318 degrees of freedom
## Multiple R-squared:  0.05457,    Adjusted R-squared:  0.04596
## F-statistic: 6.339 on 12 and 1318 DF,  p-value: 5.119e-11
```

```r
# do spline for "padtimes" variable
spline_model = lm(sld010h ~ pad080 + bs(padtimes) + ridageyr +
                    as.factor(ridreth1) + indfminc + as.factor(ridexmon) +
```

```
                  drtkcal + drtsugr + drtcaff,
                data = avg_value)
summary(spline_model)
```

```
##
## Call:
## lm(formula = sld010h ~ pad080 + bs(padtimes) + ridageyr + as.factor(ridreth1) +
##     indfminc + as.factor(ridexmon) + drtkcal + drtsugr + drtcaff,
##     data = avg_value)
##
## Residuals:
## How much sleep do you get (hours)?
##     Min      1Q  Median      3Q     Max
## -5.9177 -0.8271 -0.0118  0.8720  5.2470
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            7.680e+00  1.564e-01  49.094  < 2e-16 ***
## pad080                -1.998e-03  6.709e-04  -2.978 0.002953 **
## bs(padtimes)1          1.201e+00  4.673e-01   2.571 0.010247 *
## bs(padtimes)2         -1.816e+00  1.171e+00  -1.552 0.120965
## bs(padtimes)3          1.500e+00  8.613e-01   1.742 0.081812 .
## ridageyr              -6.925e-03  2.092e-03  -3.310 0.000958 ***
## as.factor(ridreth1)2 -3.864e-01  2.018e-01  -1.915 0.055694 .
## as.factor(ridreth1)3  2.850e-02  1.060e-01   0.269 0.788049
## as.factor(ridreth1)4 -5.680e-01  1.066e-01  -5.331 1.15e-07 ***
## as.factor(ridreth1)5 -1.082e-01  1.826e-01  -0.592 0.553636
## indfminc               1.500e-03  3.432e-03   0.437 0.662095
## as.factor(ridexmon)2 -9.660e-02  7.800e-02  -1.238 0.215775
## drtkcal               -6.912e-05  5.337e-05  -1.295 0.195474
## drtsugr               -2.791e-04  6.594e-04  -0.423 0.672157
## drtcaff               -6.888e-04  2.667e-04  -2.582 0.009923 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.317 on 1316 degrees of freedom
## Multiple R-squared:  0.05798,    Adjusted R-squared:  0.04795
## F-statistic: 5.785 on 14 and 1316 DF,  p-value: 4.423e-11
```

```
# plot splines results
plot(avg_value$sld010h, avg_value$padtimes, col="grey",xlab="Sleep (hr.)",ylab="Padtimes")
fit1 = smooth.spline(avg_value$sld010h, avg_value$padtimes, df=12)
lines(fit1,col="red",lwd=2)
```


Padtimes
0
50
100
150
2
4
6
8
10
12
Sleep (hr.)