

Stat 528 HW2

Dongyang Wang

2023-01-26

Question 1

```
rm(list=ls())
df = read.csv("CHSdataEx1.csv")
```

Question 2

Question 2.1

```
library(corrplot)
library(xtable)
library(ggplot2)

df_exer = na.omit(df[,c("exint0", "exint3", "block0", "block3", "kcal0", "kcal3")])
corrplot.mixed(cor(df_exer),
               lower = "number",
               upper = "circle",
               tl.col = "black")

boxplot(df$exint0, df$exint3, main = "Boxplot for Exercise Intensity",
        xlab = "Baseline and 3 Years After",
        ylab = "Intensity")
tab1 <- summary(df$exint3- df$exint0)

boxplot(df$block0, df$block3, main = "Boxplot for Blocks Walked",
        xlab = "Baseline and 3 Years After",
        ylab = "Blocks")
tab2 <- summary(df$block3- df$block0)

boxplot(df$kcal0, df$kcal3, main = "Boxplot for kCal Expended",
        xlab = "Baseline and 3 Years After",
        ylab = "Kilocalories")
tab3 <- summary(df$kcal3- df$kcal0)

table1 <- rbind(tab1, tab2, tab3)
rownames(table1) <- c("Exercise Intensity", "Blocks Walked", "kCal Expended")
table1

d=table(df[c('exint0','exint3')])
colnames(d)=c("no exercise", "low", "moderate","high")
rownames(d)=c("no exercise", "low", "moderate","high")
mosaicplot(d,xlab='baseline',ylab='3 yrs after',main='Exercise Intensity for Baseline and 3 years after')
```

Question 2.2

```
baseline = c("season", "gender", "age", "weight", "weight50", "grade", "arth", "sbp", "pkys", "diab", "exercise")
exercise = c("exint0", "exint3", "block0", "block3", "kcal0", "kcal3")
df_clean = na.omit(df[,c(baseline, exercise)])

corrplot.mixed(cor(df_clean),
               lower = "number",
               upper = "circle",
               tl.col = "black")
```

Question 2.3

```
library(jtools)
df_reg = df[,c("mortality", baseline)]
lm1 = lm(mortality ~ . , data = df_reg)
summary(lm1)
summ(lm1)
```

Question 2.4

```
#df_reg2 = df[,c("mortality", baseline, exercise)]
#lm2 = lm(mortality ~ . , data = df_reg2)
#summary(lm2)
#summ(lm2)
```

Question 2

```
row1 = c(1, 210, 201, -9, 130, 125, -5)
row2 = c(2, 169, 165, -4, 122, 121, -1)
row3 = c(3, 187, 166, -21, 124, 121, -3)
row4 = c(4, 160, 157, -3, 104, 106, 2)
row5 = c(5, 167, 147, -20, 112, 101, -11)
row6 = c(6, 176, 145, -31, 101, 85, -16)
row7 = c(7, 185, 168, -17, 121, 98, -23)
row8 = c(8, 206, 180, -26, 124, 105, -19)
row9 = c(9, 173, 147, -26, 115, 103, -12)
row10 = c(10, 146, 136, -10, 102, 98, -4)
row11 = c(11, 174, 151, -23, 98, 90, -8)
row12 = c(12, 201, 168, -33, 119, 98, -21)
row13 = c(13, 198, 179, -19, 106, 110, 4)
row14 = c(14, 148, 129, -19, 107, 103, -4)
row15 = c(15, 154, 131, -23, 100, 82, -18)
df1 = rbind(row1, row2, row3, row4, row5,
            row6, row7, row8, row9, row10,
            row11, row12, row13, row14, row15)
df_q2 = as.data.frame(df1)
colnames(df_q2) = c("no", "Systolic before", "Systolic_after", "Systolic_diff",
                  "Diastolic_before", "Diastolic_after", "Diastolic_diff")

df_q2 = df_q2[order(df_q2$Systolic_diff),]

#plot(df_q2$Systolic_diff, df_q2$Diastolic_diff)
```

```
library(ggplot2)
ggplot(df_q2, aes(x = Systolic_diff, y= Diastolic_diff), col = "black") +
  geom_point() +
  geom_smooth(method=lm) +
  ggtitle("Comparison between Systolic and Dialostic Responses") +
  xlab("Systolic Difference") +
  ylab("Dialostic Difference") +
  theme(plot.title = element_text(hjust = 0.5))
```