BIOST 515 Homework 3

Katie McFarlane

1/28/22

Responses

1. Body mass index, or BMI, is an easily-calculated estimate of body fat. Research the definition of BMI and briefly summarise your findings. Calculate the BMI of each participant in the MRI dataset. Calculate descriptive statistics for creatinine levels, age, sex, and BMI, and present the results in a format suitable for a scientific publication.

Body mass index is calculated by dividing one's weight in kilograms by the square of their height in meters. BMI corresponds to four categories, usually labeled underweight, healthy weight, overweight, and obese. This measure is meant to indicate body fat levels but does not calculate fat levels directly, and BMI can be an inaccurate or misleading measure, especially in children and those of certain ethnicities. Higher BMI is associated with certain health risks.

Table 1: Descriptive statistics for Creatinine, Age, BMI, and Sex in the MRI dataset

	N	Msng	Mean/Proportion	Std Dev	Min	25th pct	Median	75th pct	Max
Creatinine (mg/dl)	735	2	1.060	0.303	0.5	0.9	1	1.2	4
Age (years)	735	0	74.600	5.45	65	71	74	78	99
BMI (kg/m^2)	735	0	26.300	4.31	14.5	23.5	26	28.5	46.6
Sex (male)	735	0	0.498	—	—	—	—	—	—

2. We are interested in examining how mean creatinine levels vary by BMI and sex. Create a scatterplot of creatinine levels versus BMI. Use different symbols and/or colors for each sex, and include the LOESS curve for each sex. What observations do you make from the scatterplot regarding the association between creatinine levels and BMI?

Creatinine levels and BMI do not appear to be strongly associated: across all participants, those who have a BMI between 20 and 40 do not have much fluctuation in average creatinine level, and this mostly holds when stratified by sex. Those with BMI below 20 and above 40 see a drop in average creatinine levels; when stratified by sex, we see a drop in average levels below a BMI of 20 and again above a BMI of around 43 for women, while for men, we see a drop in average levels below a BMI of 20 and again around a BMI of 35.

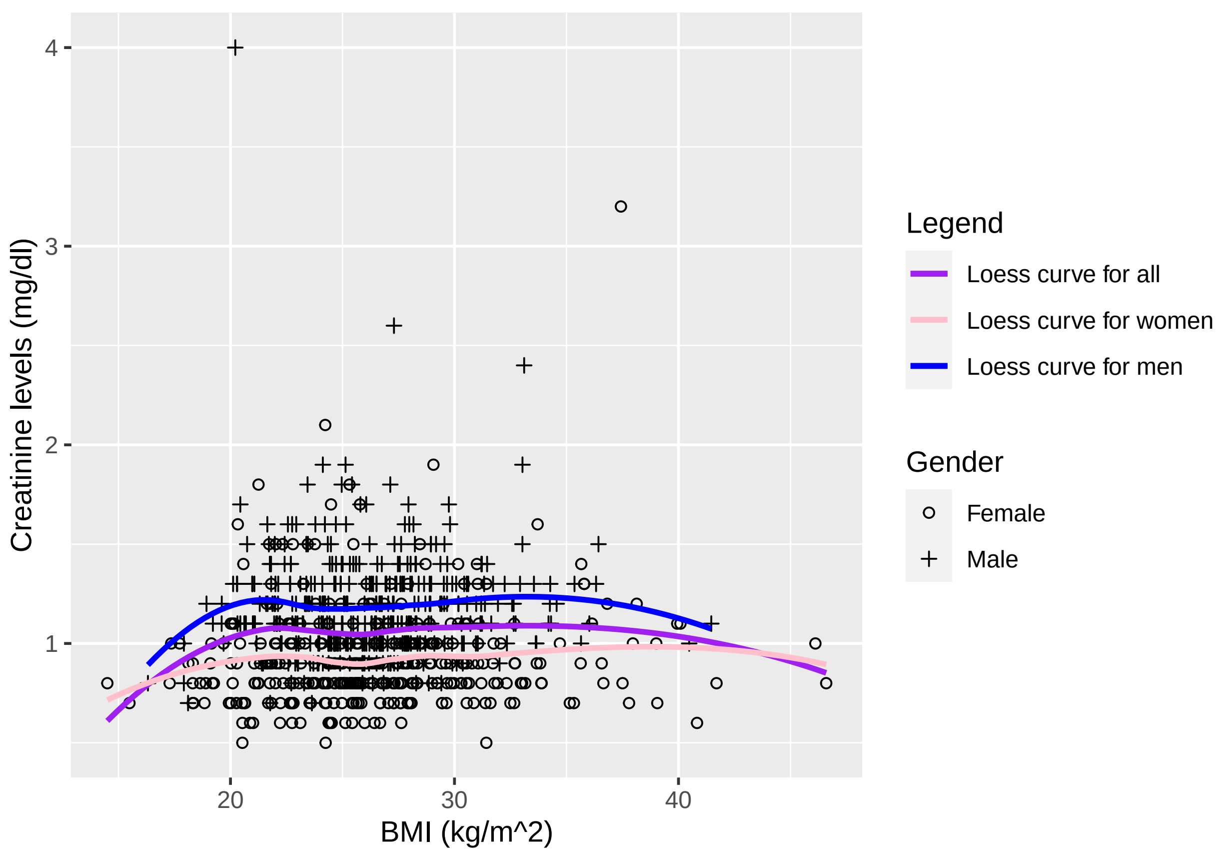

Figure 1: Creatinine levels across BMI in men and women

b. Perform an analysis to determine whether mean creatinine levels differ between the sexes. Briefly describe the analysis that you performed and clearly state the basis of your conclusion regarding an association.

I performed a two sample t-test, which compares the means of two populations (in this case, male participants compared to female participants) to test if they are significantly different. The t-test returned a difference in mean creatinine levels of -0.27 mg/dl (CI -0.31, -0.23), with women having an estimated lower creatinine level than men. This difference is statistically significant ($p < 0.0005$), so we reject the null hypothesis that mean creatinine levels are the same within our two populations.

c. Perform an analysis to determine whether there is a linear trend in mean creatinine levels by BMI. Briefly describe the analysis that you performed and clearly state the basis of your conclusion regarding an association.

I performed simple linear regression using robust SE to examine the average linear trend between BMI and average creatinine levels. The estimated difference in mean creatinine levels for populations differing by 1 kg/m^2 BMI is 0.003 mg/dl (95% CI -0.003, 0.009), with those with higher BMI having higher mean creatinine levels. This difference is not statistically significant ($p = 0.31$), so we do not reject the null hypothesis that there is no linear trend in expected creatinine levels as a function of BMI.

d. Perform an analysis to determine whether there is a linear trend in mean creatinine levels by BMI after adjustment for sex. Briefly describe the analysis that you performed and clearly state the basis of your conclusion regarding an association.

I performed multiple linear regression using robust SE to determine the average linear trend between BMI and average creatinine levels after adjustment for sex. The estimated difference in mean creatinine levels for populations of the same sex who differ in BMI by 1 kg/m^2 is 0.004 mg/dl (95% CI -0.002, 0.010), with the group with higher BMI having higher creatinine levels.

The association between BMI and creatinine controlling for sex is not statistically significant ($p = 0.19$). Thus we do not have strong evidence of a linear trend of higher creatinine associated with higher BMI for those of the same sex.

e. Contrast your analyses and findings from (c) and (d).

Our first analysis used simple linear regression, which compared two variables. Our second used multiple linear regression, which uses >2 variables, and allows us to examine association between two variables while fixing 1+ other variables. In both analyses, we found that we lack evidence to claim a statistically significant association between BMI and creatinine levels, whether controlling for sex or not. We did find that sex is associated with creatinine levels while controlling for BMI, however: our second test indicated an expected difference of 0.27 mg/dl mean creatinine levels between men with the same BMI and women with the same BMI, with men having higher mean creatinine levels ($p < 0.0005$).

3. Fit a multiple linear regression with creatinine level as the response and the variables age, sex, and BMI as predictors. Write out the fitted model.

We want to fit the following regression model: E(Creatinine|Age, sex, BMI) = $\beta_0+\beta_1$ x age + β_2 x sex + β_3 x BMI. Age is in years, sex is either 0 (for females) or 1 (for males), and BMI is is kg/m^2. Our outcome creatinine is in units of mg/dl.

By using multiple linear regression, we find our coefficient values to construct our fitted model:

E(Creatinine|Age, sex, BMI) = 0.35 + 0.006 x age + 0.27 x sex + 0.005 x BMI

b. Interpret the intercept term in your regression model. Is the intercept estimate scientifically meaningful?

Our intercept term in this model is 0.35 mg/dl. This intercept represents the mean creatinine level of the population of women who are aged 0 years and have a BMI of 0 kg/m^2. Because it is impossible for a human to have a BMI level of 0 kg/m^2, this intercept does not tell us anything useful about real-world populations, and thus is not scientifically meaningful.

c. In 4-5 sentences, summarize the results of your model in language suitable for a scientific publication. Give full inference for the coefficient on age, the coefficient on sex, and the coefficient on BMI in your model.

The estimated difference in mean creatinine levels for populations of the same sex and BMI who differ by one year in age is 0.006 mg/dl (95% CI 0.001, 0.010), with older populations having higher mean creatinine levels. The estimated difference in mean creatinine levels for populations of the same age and BMI who differ by sex is 0.27 (95% CI 0.23, 0.31), with males having higher mean creatinine levels. The estimated difference in mean creatinine levels for populations of the same age and sex who differ by 1 kg/m^2 in BMI is 0.005 (95% CI -0.00006, 0.011), with those with higher BMI having higher mean creatinine levels.

We find a statistically significant association between creatinine and age while controlling for sex and BMI ($p = 0.01$) and between creatinine and sex while controlling for age and BMI ($p < 0.0005$). However, we do not find a statistically significant association between creatinine and BMI while controlling for age and sex ($p = 0.053$).

Code Appendix

```
### Packages
library(knitr)
library(tidyverse)
library(rigr)

knitr::opts_chunk$set(echo = FALSE)
### -----------------------------------------------------------
### Reading in the data
mri <- read.csv("~/Graduate School Work/Winter 2022 - BIOST 515/mri.csv")
### -----------------------------------------------------------
### Q1
mri$kg <- mri$weight/2.2046
mri$bmi <- mri$kg/((mri$height/100)^2)
mri$gender <- as.factor(mri$sex)
#code to 0 for female and 1 to male
mri$gender <- as.numeric(mri$gender) - 1
des <- mri %>% select(crt, age, bmi, gender) %>% descrip() %>% signif(3)
des <- des[, 1:9] %>% as.data.frame()
names(des) <- c("N", "Msng", "Mean/Proportion", "Std Dev",
                "Min", "25th pct", "Median", "75th pct", "Max")

des$`Std Dev`[4] <- "---"
des$Min[4] <- "---"
des$`25th pct`[4] <- "---"
des$`Median`[4] <- "---"
des$`75th pct`[4] <- "---"
des$Max[4] <- "---"

rownames(des) <- c("Creatinine (mg/dl)", "Age (years)", "BMI (kg/m^2)", "Sex (male)")

knitr::kable(des, booktabs = TRUE, format = "markdown",
             caption = "Descriptive statistics for Creatinine,
             Age, BMI, and Sex in the MRI dataset")
### -----------------------------------------------------------
### Q2a
mri$Gender <- as.factor(mri$sex)
femmri <- na.omit(mri[ which(mri$sex=='Female'), ])
malemri <- na.omit(mri[ which(mri$sex=="Male"), ])
mri %>%
  ggplot(aes(x = bmi, y = crt)) +
  geom_point(aes(shape=Gender)) +
  scale_shape_manual(values=c(1, 3)) +
  xlab("BMI (kg/m^2)") +
  ylab("Creatinine levels (mg/dl)") +
  geom_smooth(aes(col = "Loess curve for all"), se=F, method = "loess") +
  geom_smooth(data=subset(femmri), aes(col = "Loess curve for women"), se=F, method = "loess")+
  geom_smooth(data=subset(malemri), aes(col = "Loess curve for men"), se=F, method = "loess") +
    scale_color_manual(name = "Legend",
                     breaks = c("Loess curve for all", "Loess curve for women", "Loess curve for men"),
                   values = c("purple", "pink", "blue")) +
  NULL
```

```r
### ----------------------------------------------------------
### Q2b
t.test(mri$crt~mri$sex)
t.test(mri$crt~mri$sex)$p.value
### ----------------------------------------------------------
### Q2c
regress("mean", crt~bmi, data=mri)
### ----------------------------------------------------------
### Q2d
regress("mean", crt~bmi+sex, data=mri)
### ----------------------------------------------------------
### Q3a
regress("mean", crt~age+sex+bmi, data=mri)
```