BIOST 515 HW 3

Eliza Chai

01/25/2021

Responses

1. Body Mass Index (BMI) is a screening method for weight category (underweight, heavyweight, overweight, and obesity). BMI can be used to indicate body fatness and screens for health problems. It is calculated by a person's weight (kilograms) divided by the square of height (meters). Table 1 shows the descriptive statistics for creatinine levels, age, sex, and BMI of participants in the MRI dataset.

Table 1: Descriptive statistics for creatinine, age, sex, and BMI in the MRI dataset

	N	Msng	Mean/Proportion	Std Dev	Min	25th pct	Median	75th pct	Max
Creatinine (mg/dl)	735	2	1.06	0.303	0.5	0.9	1	1.2	4
Age (years)	735	0	74.6	5.45	65	71	74	78	99
Male sex	735	0	0.498	—	—	—	—	—	—
BMI (kg/m^2)	735	0	26.3	4.31	14.5	23.5	26	28.5	46.6

2. Examining how mean creatinine levels vary by BMI and sex.

a) In Figure 1 (on page 2), we observe a slightly (little to none) increasing relationship between BMI and creatinine levels for both sex: the mean creatinine levels appears to slightly (little to none) increase as BMI increases for both female and male participants. The relationship between mean creatinine and BMI does not appear to be linear over the range of the full BMI observed. The male participants have higher mean creatinine levels than female participants over the range of the full BMI observed, as the LOESS curve for the male group is above that of the female group. In addition, male participants have a less apparent (little) increasing relationship than female participants.

b) We fit a linear regression using robust SE with creatinine levels as the response and the variable sex as the predictor. Based on the simple linear regression model, we estimate that the mean creatinine levels for female participants are 0.928 mg/dl. Furthermore, we estimate that the difference in the mean creatinine levels of female participants compared with male participants is 0.269 mg/dl, with the male group having higher mean creatinine (95% CI based on heteroscedasticity-robust standard errors: 0.228 - 0.309 mg/dl). This difference is significantly different from zero ($p < 0.001$), so we reject the null hypothesis that there is no difference in the expected value of creatinine between female and male participants. We can conclude that there is strong evidence for an association between creatinine levels and sex, where the male group is associated with higher creatinine levels.

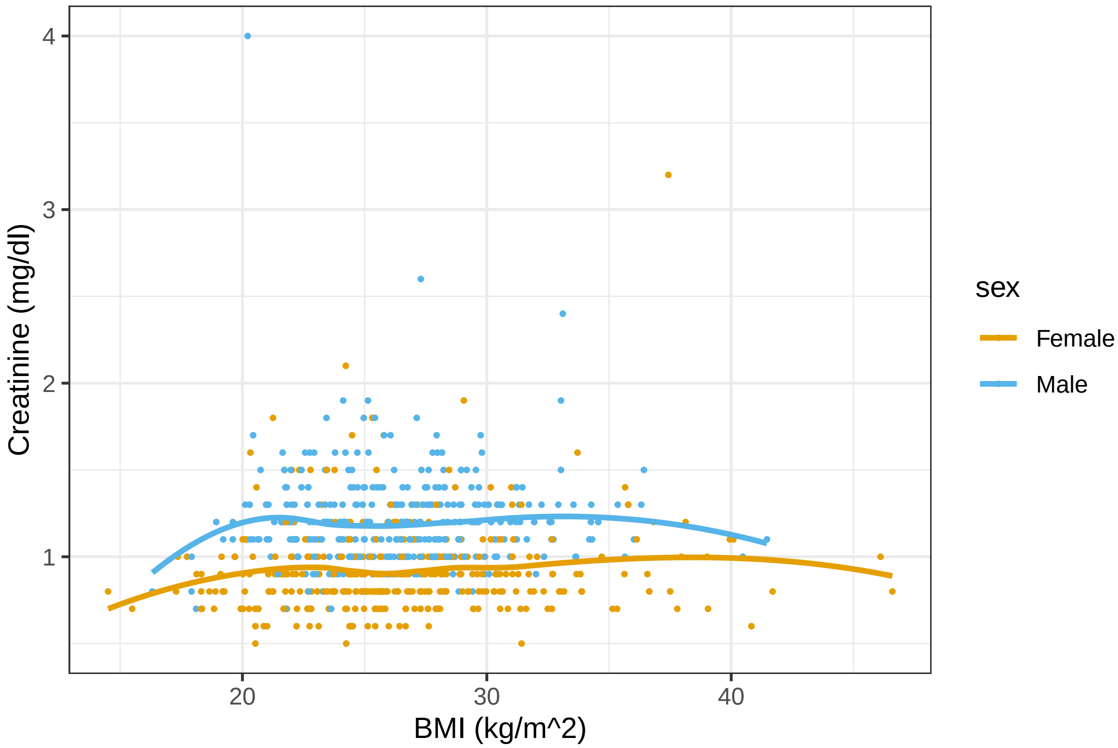

Figure 1: Scatterplot of creatinine levels versus BMI in the MRI cohort

c) We fit a linear regression using robust SE with creatinine levels as the response and the variable BMI as the predictor. Based on the simple linear regression model, we estimate that the difference in the mean creatinine levels for the two groups differing by one kg/m^2 in BMI is 0.003 mg/dl, with the higher BMI group having higher creatinine levels (95% CI based on heteroscedasticity-robust standard errors: -0.003 - 0.009 mg/dl). Furthermore, the estimated expected value of creatinine level for a population member who is zero kg/m^2 in BMI is 0.980 mg/dl. However, since a participant can't have 0 kg/m^2 BMI, this value is not scientifically interpretable. We find no statistically significant association between creatinine level and BMI ($p = 0.311$), thus we fail to reject the null hypothesis. There is insufficient evidence to conclude a first-order linear trend in the relationship between creatinine and BMI in participants.

d) We fit a linear regression using robust SE with creatinine levels as the response and the variables BMI as the predictor adjusting for sex. Based on the simple linear regression model, we estimate that for individuals of the same sex and differing by one kg/m^2 in BMI, the group with higher BMI has a mean creatinine level that is 0.004 mg/dl (95% CI based on heteroscedasticity-robust standard errors: -0.002 - 0.010 mg/dl). Furthermore, the estimated expected value of creatinine level for a female member who is zero kg/m^2 BMI is 0.829 mg/dl. The estimated expected value of creatinine level for a male member who is zero kg/m^2 BMI is 1.098 mg/dl. However, since it is not possible for a participant to have 0 kg/m^2 BMI, these values are not scientifically interpretable. We find no statistically significant association between creatinine level and BMI controlling for sex ($p = 0.193$). Thus, we fail to reject the null hypothesis and conclude that there is insufficient evidence of a first-order linear trend in the relationship between creatinine and BMI among those participants of the same sex.

e) In (d), we stratified the linear regression model with sex, and examine how creatinine levels vary with BMI in the female and male groups. Hence, the fitted model used in (d) is different from the fitted model in (c), i.e. the parameters are different in the model (d) and model (c). The estimated value and the interpretation of the coefficient on BMI of the two models are different. In addition, the estimated value and the interpretation of the "intercept" in the two models are not the same. However, in both models, the estimated values of the "intercept" are not scientifically interpretable. We also observed that the difference between the estimated value of the coefficient on BMI from the two models is insignificant, which suggests that sex has little confounding effects on the relationship between creatinine level and BMI. Therefore, we have similar conclusions regarding an association in (c) and (d), where we fail to reject the null hypothesis and conclude that there is insufficient evidence of a first-order linear trend in the relationship between creatinine and BMI regardless of controlling for sex.

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

a) The fitted model is: $\hat{Creatinine}|$(BMI, sex, age) = 0.355 + 0.005 $*$ BMI + 0.268 $*$ sex + 0.006 $*$ age where BMI is Body Mass Index in kg/m^2 at baseline, sex is the gender of the participant in male or female (0 for female and 1 for male), and age is baseline age 65 or above at time of MRI in years.

b) The intercept terms in the regression model are interpreted as follow: The estimated expected value of creatinine level for a female member who is 0 kg/m^2 in BMI and 0 year in age is 0.355 mg/dl. The estimated expected value of creatinine level for a male member who is 0 kg/m^2 in BMI and 0 year in age is 0.623 mg/dl. However, since it is not possible for a participant to have 0 kg/m^2 in BMI and/or 0 year in age (the MRI cohort study is for adults aged 65 years and older, zero-age is outside the range of our data), these values are not scientifically meaningful.

c) We fit a linear regression using robust SE with creatinine levels as the response and the variables BMI, sex, and age as the predictors. Based on the simple linear regression model, we estimate that for individuals of the same sex and age differing by one kg/m^2 in BMI, the group with higher BMI has a mean creatinine level that is 0.005 mg/dl (95% CI based on heteroscedasticity-robust standard errors: -0.0001 - 0.0108 mg/dl). When comparing two subpopulations of the same BMI and age who differ in their sex, we estimate that the difference in mean creatinine between the groups is 0.268 mg/dl, with the male group having higher mean creatinine (95% CI based on heteroscedasticity-robust standard errors: 0.230 - 0.307 mg/dl). When comparing two groups of individuals with the same BMI and sex, we estimate that the group that is one year older has a 0.006 mg/dl higher mean creatinine (95% CI based on heteroscedasticity-robust standard errors: 0.001 - 0.010 mg/dl). We find a statistically significant association between sex and creatinine controlling for BMI and age ($p < 0.001$). In addition, we find a statistically significant association between age and creatinine controlling for BMI and sex ($p = 0.0099$). However, we find no statistically significant association between creatinine level and BMI controlling for sex and age ($p = 0.0528$).

Thus, we have strong evidence for a first order linear association between male groups with higher creatinine levels among those with the same BMI and age. Furthermore, we have strong evidence for a first order linear trend of higher age associated with higher creatinine levels among those with the same BMI and sex. However, we conclude that there is insufficient evidence of a first order linear trend in the relationship between creatinine and BMI among those of the same sex and age.

Code Appendix

```
### Setting up the packages, options we'll need:
library(knitr)
knitr::opts_chunk$set(echo = FALSE)
library(tidyverse)
library(rigr)
### -----------------------------------------------------------
### Reading in the data.
mri <- read_csv("~/Desktop/UW MS Biostat/BIOST Winter 2022/Biost 515/R dataset/mri.csv")
### -----------------------------------------------------------
### Q1
### Table of descriptive statistics for creatinine levels, age, sex, and BMI
mri <- mri %>%
  mutate(height_m = height/100) %>%
  mutate(weight_kg = weight*0.45359237) %>%
  mutate(BMI = weight_kg / height_m^2) %>%
  mutate(sex_male = ifelse(sex == "Male", 1, 0))
tab1 <- mri %>%  select(crt, age, sex_male, BMI) %>%  descrip()
tab1 <- tab1[,1:9] %>% data.frame()
names(tab1) <- c("N","Msng", "Mean/Proportion", "Std Dev", "Min",
                 "25th pct","Median", "75th pct", "Max")
tab1 <- tab1 %>% mutate(
              `Mean/Proportion` = as.character(signif(`Mean/Proportion`, 3)),
              `Std Dev` = as.character(signif(`Std Dev`, 3)),
              `Min` = as.character(signif(`Min`, 3)),
              `25th pct` = as.character(signif(`25th pct`, 3)),
              `Median` = as.character(signif(`Median`, 3)),
              `75th pct` = as.character(signif(`75th pct`, 3)),
              `Max` = as.character(signif(`Max`, 3)))
tab1$`Std Dev`[3] <- "---"
tab1$`Min`[3] <- "---"
tab1$`25th pct`[3] <- "---"
tab1$`Median`[3] <- "---"
tab1$`75th pct`[3] <- "---"
tab1$`Max`[3] <- "---"
rownames(tab1) <- c("Creatinine (mg/dl)","Age (years)",
                    "Male sex", "BMI ($kg/m^2$)")

knitr::kable(tab1, bookstabs = TRUE, format = "markdown",
             caption = "Descriptive statistics for creatinine,
             age, sex, and BMI in the MRI dataset")
### -----------------------------------------------------------
### Q2 a)
### Plot a scatterplot of creatinine levels on the y-axis and BMI on the x-axis
mri %>%
  ggplot(aes(x = BMI, y = crt, col = sex)) +
  geom_point(cex = 0.5) +
  xlab("BMI (kg/m^2)") +
  ylab("Creatinine (mg/dl)") +
  theme_bw() +
  geom_smooth(method = "loess", se=F,show.legend=T) +
  scale_color_manual(values = c("#E69F00", "#56B4E9"))
```

```r
### ----------------------------------------------------------
### Q2 b)
### Here's a linear model for creatinine vs. sex from mri dataset using rigr:regress
mri_model1 <- mri %>%
  na.omit() %>%
  regress("mean", crt ~ sex, data = .) %>% coef %>% round(5)
### ----------------------------------------------------------
### Q2 c)
### Here's a linear model for creatinine vs. BMI from mri dataset using rigr:regress
mri_model2 <- mri %>%
  regress("mean", crt ~ BMI, data = .) %>% coef %>% round(5)
### ----------------------------------------------------------
### Q2 d)
### Here's a linear model for creatinine vs. BMI from mri dataset using rigr:regress
mri_model3 <- mri %>%
  regress("mean", crt ~ BMI + sex, data = .) %>% coef %>% round(5)
mri_model3
### ----------------------------------------------------------
### Q3 a)
### Here's a linear model for creatinine vs. BMI + sex + age from mri dataset using rigr:regress
mri_model4 <- mri %>%
  regress("mean", crt ~ BMI + sex + age, data = .) %>% coef %>% round(5)
### ----------------------------------------------------------
### Q3 c)
### Here's a linear model for creatinine vs. BMI + sex + age from mri dataset using rigr:regress
mri_model4
### ----------------------------------------------------------
```