BIOST 515/518 Homework 3

Latera Tesfaye Olana

29 January, 2023

Responses:

Question 1: Body Mass Index (BMI) is a measure of body fat based on an individual's weight and height. It is calculated by dividing an individual's weight in kilograms by the square of their height in meters (1). According to Center for disease control (CDC) website the pound - inch equivalent is given as $BMI = pounds/inche^2$ (2).

The World Health Organization (WHO) defines a normal BMI as being between 18.5 and 24.9, while a BMI of 25 or higher is considered overweight and a BMI of 30 or higher is considered obese (3). However, it is worth noting that BMI is not a diagnostic tool and should not be used to diagnose obesity or any other health condition.

It is commonly used as an screening tool to identify potential weight problems, but BMI has limitations, it does not take into account factors such as muscle mass, bone density, and distribution of fat, which can also impact a person's health (1).

Table 1: Summary Table of the data

	N	Missing	Mean/Proportion	Std Dev	Min	Max
Creatinine (mg/dl)	735	2	1.06	0.3	0.50000	4.00000
BMI (lb/inches^2)	735	0	26.34	4.31	14.50217	46.58878
Age (years)	735	0	74.57	5.45	65.00000	99.00000
Sex	735	0	1.5	0.5	–	–

Question 2a: From the given sample data, with an increase in Creatinine (mg/dl) there seems to be an increase in Body Mass Index (BMI) (pounds per inches squared). Accordingly, the first order trend suggestive of a tendency for higher average BMI in adults with higher Creatinine level. As shown on Figure 1, there seems to be some suggestion of greater variability in BMI in adults with higher Creatinine than there is in adults with lower Creatinine.

From the stratified representation of birth-assigned sex, inferring the association between BMI and Creatinine levels is difficult. However, data suggests that Creatinine levels exhibit significant variation across sex, as compared to BMI. If a correlation between sex and BMI were to exist, it would strongly indicate the presence of confounding factors related to sex.
Figure 1, also shows the existence of **unusual observations**. The two points, which are marked and at the top of the graph - with blue mark (ptid=298 and 426) are outliers (fitted with lm and 99.99% threshold was used - studentized residuals). On the other hand, from the two points at the far right, the one at the top seems a high leverage observation (ptid = 210) as it is an extreme covariate value. The one at the bottom is an extreme covariate value and it seems to exert disproportionate influence on regression parameter estimates. Accordingly, it is an outlier, a high leverage and influential observation.

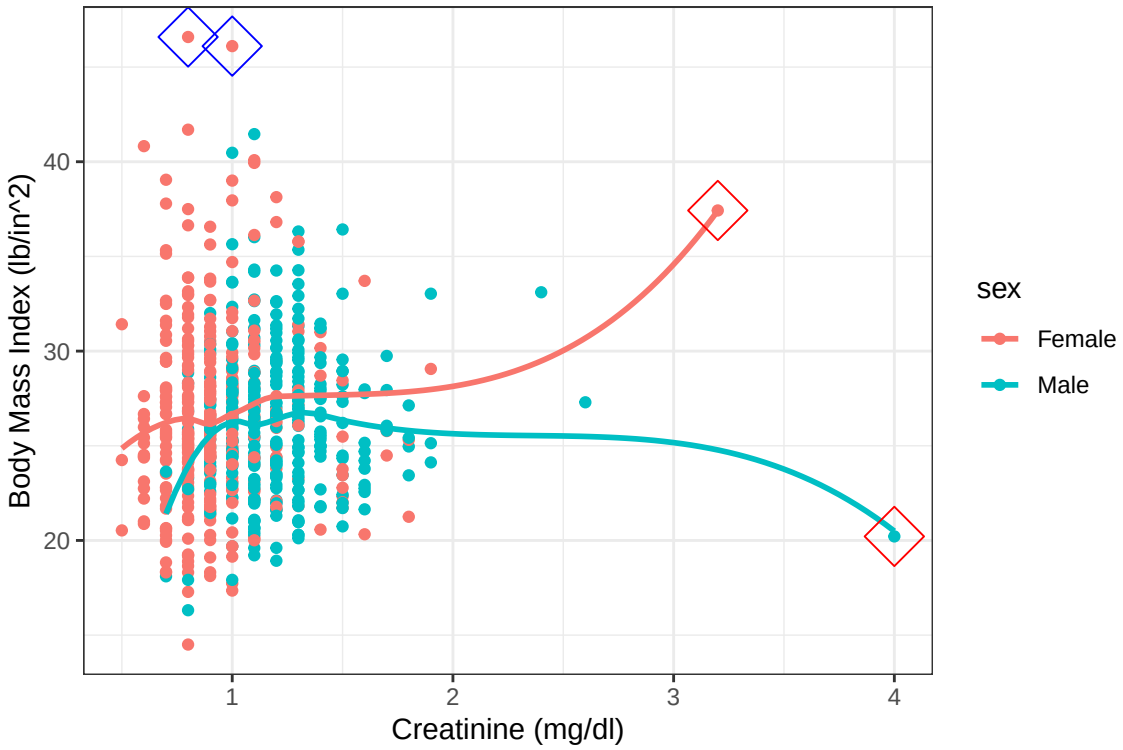


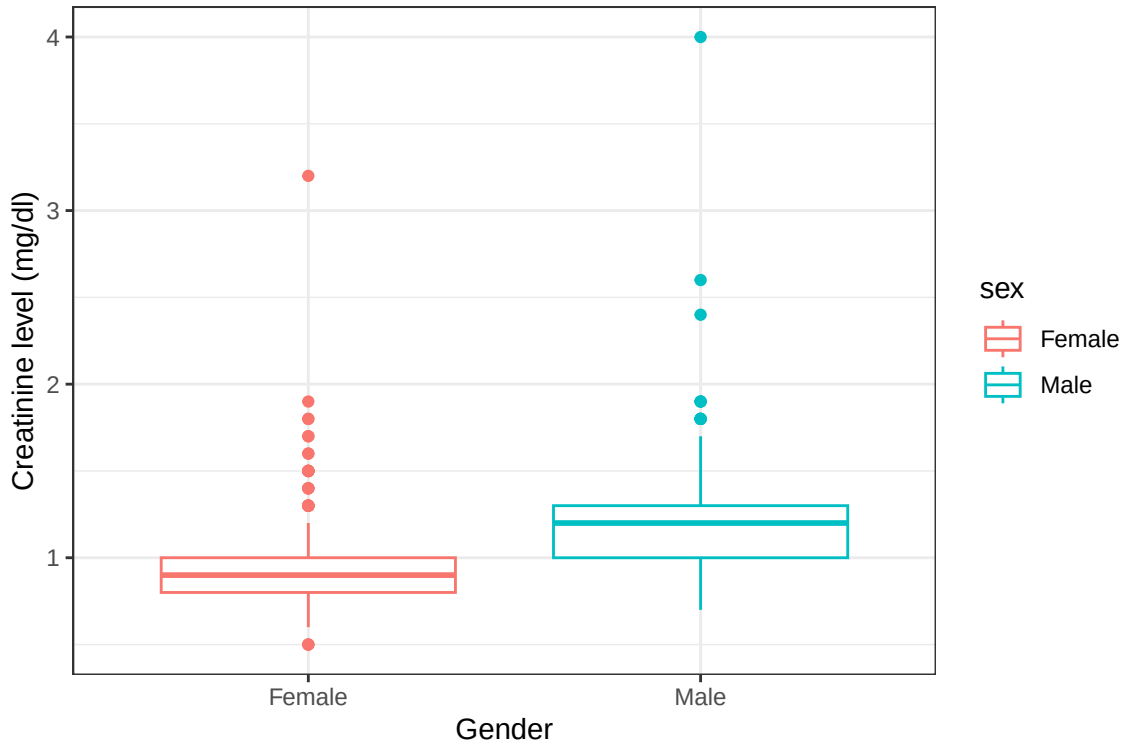
Figure 1: Scatter plot of Body Mass Index (BMI) and Creatinine

Question 2b: In order to examine the association between Creatinine level and sexes, I generated a summary table of the sample data. Table 2 illustrates that in this data, there are equal numbers of female adults (367, or 50%) and male adults (366, or 50%). This indicates a roughly equal distribution. Table 3 shows a considerable mean differences in Creatinine level between sexes.

Table 2: Summary of of data by sex

sex	n	bmi	crt
Female	367	26.42433	0.9297003
Male	366	26.25005	1.1986339

My choice in using the word *considerable* comes from the fact that minimum and the maximum of Creatinine level in the given sample data are 0.5 and 4 mg/dl, respectively. This seems a wide range. Figure 2 shows, a descriptive plot of Creatinin levels in mg/dl for sexes.


Creatinine level (mg/dl)
4
3
2
1
Female
Male
Gender
sex
Female
Male

However, to identify or test association *beyond any chance*, I used a two-sample t-test, which assumes equal variance. The estimated mean difference in Creatinine level, between sexes of adult population is significantly different from zero (p-value < 0.01)‘. Thus we reject the null hypothesis that there is no association between Creatinine level and sexes. *This p-value is the same with if I used lm() model.*

Table 3: Summary of ttest between Creatinine level and sexes

Estimate	Statistics	Pr(>|t|)	95%L	95%H
-0.2689336	-13.40256	0	-0.3083272	-0.2295401

Furthermore, the estimated mean BMI for female adults is 26.42, while for male adults it is 26.25 pounds per square inch. This suggests that there is no significant association between BMI and gender. A two sample t-test, suggests there is no strong evidence to reject the null hypothesis, which specifies the difference in estimated mean between female and male adult population of the study is zero (p-value = 0.58. *This p-value is the same with lm() model.*

Table 4: Summary of ttest between BMI and sexes

Estimate	Statistics	Pr(>|t|)	95%L	95%H
0.1742819	0.5467527	0.5847155	-0.4515087	0.8000725

Question 2c: I fit a linear regression using heteroskedasticity robust SE with BMI as the response and the variable Creatinine as predictors to assess the linear association between these two variables. We fail to reject (at 5% level) the null hypothesis that there is no linear trend in the expected value of BMI and Creatinine level (p-value = 0.3209).

Table 5: Model summary for linear fit of Body Mass Index (BMI) and Creatinine

	Estimate	Naive SE	Robust SE	95%L	95%H	t value	Pr(>|t|)
(Intercept)	25.6498	0.5819	0.7185	24.2392	27.0603	35.6987	0.0000
crt	0.6462	0.5260	0.6506	-0.6311	1.9234	0.9932	0.3209

Question 2d: I fit a linear regression using heteroskedasticity robust SE with BMI as the response and the variables Creatinined and sexes status as predictors. We estimate that for individuals with the same sex and differing in Creatinine level by 1 mg/dl, the group with higher Creatinine has a mean BMI that is 0.96 pound square inches higher (95% CI: -0.53, 2.46). We fail to reject (at the 5% level) the null hypothesis that this difference is equal to zero (p = 0.2058). We do not have sufficient evidence to conclude that average BMI differs across Creatinine levels after adjusting sex.
When comparing two groups of individuals with the same creatinine level, we estimate that the male group has a 0.43 pound square inches lower mean BMI (95% CI: -1.17 0.31). We fail to reject (at the 5% level) the null hypothesis that this difference is equal to zero (p = 0.2504). We do not have sufficient evidence to conclude that average BMI differs between male and female adult groups of the same Creatinine.

Table 6: A summary of a linear fit of BMI with Creatinine level and sexes

	Estimate	Naive SE	Robust SE	95%L	95%H	t value	Pr(>|t|)
(Intercept)	25.5279	0.5902	0.7465	24.0623	26.9935	34.1956	0.0000
crt	0.9642	0.5869	0.7615	-0.5308	2.4593	1.2662	0.2058
sexMale	-0.4336	0.3554	0.3770	-1.1737	0.3065	-1.1502	0.2504

Question 2e: In both tests (c and d) the p-value for assessing the linear trend between BMI and Creatinine remained greater than 0.05. This suggests that there is not a statistically significant linear relationship between these two variables. This means that the observed association between BMI and Creatinine could have occurred by chance (unlucky) and due to some sort of assumptions about our data or model.

The fact that the p-value remained greater than 0.05 after adjustment for sex suggests that sex does not have a significant impact on the relationship between BMI and Creatinine. This means that controlling for sex does not alter the conclusion that there is not a statistically significant linear relationship between these two variables.

Question 3a: The formula for the linear fit is:

$$\hat{BMI}|(creatinine, sex) = 25.53 + 0.96 * creatinine - 0.43 * sex$$

Where $\hat{BMI}$is the fitted value of Body Mass Index in pounds square inches from the model, creatinine is measure of creatinine in the participant's blood at the time of MRI - at baseline, and sex Male for male (=1) participant and Female for female (=0) study participants.

Question 3b: The coefficient of Creatinine (0.96 pounds per square inches) shows the estimated mean difference in BMI for study participants of the same age but differ in 1 mg/dl Creatinine level, with individuals with higher Creatinine level having higher BMI value. On the other hand the coefficient of sex (-0.43 pounds per square inches) is the estimated mean difference in BMI for study participants with the same Creatinine level but differ in sex, with female participants having higher BMI value.
The intercept in a multiple linear regression with one continuous (creatinine) and one binary variable (sex) represents the expected outcome when the continuous variable is equal to 0 and the binary variable is equal to 0 (or the reference category). It provides the baseline value for comparison with the effect of the continuous and binary variables on the response. However, in the given dataset the minimum Creatinine level is 0.5 mg/dl, making Creatinine level of zero scientifically irrelevant (at least for this dataset).

Question 3c: We fit a linear regression using heteroskedasticity robust SE with BMI as the response and the variables Creatinined and sexes status as predictors. We estimate that for individuals with the same sex and differing in Creatinine level by 1 mg/dl, the group with higher Creatinine has a mean BMI that is 0.96 pound square inches higher (95% CI: -0.53, 2.46). We fail to reject (at the 5% level) the null hypothesis that this difference is equal to zero (p = 0.2058). We do not have sufficient evidence to conclude that average BMI differs across Creatinine levels after adjusting sex.
When comparing two groups of individuals with the same creatinine level, we estimate that the male group has a 0.43 pound square inches lower mean BMI (95% CI: -1.17 0.31). We fail to reject (at the 5% level) the null hypothesis that this difference is equal to zero (p = 0.2504). We do not have sufficient evidence to conclude that average BMI differs between male and female adult groups of the same Creatinine.

Reference

1. Misra, A., Dhurandhar, N.V. Current formula for calculating body mass index is applicable to Asian populations. Nutr & Diabetes 9, 3 (2019). https://doi.org/10.1038/s41387-018-0070-9

2. https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html accessed 1/28/2023

3. World Health Organization. Obesity and overweight. Fact sheet Number 311. September 2006. Accessed January 25, 2012

Code Appendix

```
### Setting up the packages
library(knitr)
knitr::opts_chunk$set(echo = FALSE)
# check if packages are installed; if not, install them
packages <- c("tidyverse", "readr", "ggExtra", "plotly",
              "ggplot2","ggstatsplot","ggside","rigr","nlme","lmtest",
              "sandwich")
not_installed <- setdiff(packages, rownames(installed.packages()))
if (length(not_installed)) install.packages(not_installed)

# load packages
library("MASS")
library(sandwich)
library(readr)
library(lmtest)
library(nlme)
library(ggstatsplot)
library(ggside)
library(rigr)
library(ggExtra)
library(broom)
library(plotly)
library(ggplot2)
library(tidyverse) # don't load tidyverse package due to conflict with dplyr
### -----------------------------------------------------------
#Loading working directory of the raw data

#Please load your data/directory by changing it with your work directory
#Throughout this code module you will see a tone of places, where
#data is read and written, so please make sure to change them to your
#working directory folder format

working_directory_data <- setwd("C:/Users/latera/Desktop/Bio_ass")

mri <- read_csv("Data/mri.csv")
# An inch is 0.393701
# BMI = weight in pounds / height in inches squared * 703
mri <- mri %>%
  mutate(converted_height = height * 0.393701) %>%
  mutate(bmi = weight / (converted_height^2) * 703)
tab1 <- mri %>% select(crt, bmi, age, sex) %>% descrip()
tab1 <- tab1[, c(1:5,9)] %>% as.data.frame()
tab1 <- tab1 %>% mutate(Mean = as.character(round(Mean, 2)),
`Std Dev` = as.character(round(`Std Dev`, 2)))
tab1[4,] <- c((tab1[4, 1:4]), NA, NA, NA)
names(tab1) <- c("N", "Missing", "Mean/Proportion", "Std Dev",
                 "Min", "Max")
rownames(tab1) <- c("Creatinine (mg/dl)", "BMI (lb/inches^2)",
                    "Age (years)","Sex")
options(knitr.kable.NA = "--")
knitr::kable(tab1, caption = "Summary Table of the data")
```

```r
lm_wcgs <- lm(bmi ~ crt + sex, data = mri) # lm
# studentized residuals
wcgs_resid <- mri %>%
select(c(ptid, bmi, crt, sex)) %>%
na.omit() %>%
mutate(studentized_resids = rstudent(lm_wcgs))
# compute the threshold for an outlier
threshold <- qnorm((1-.9999)/2, lower.tail = FALSE)
threshold
wcgs_resid %>%
filter(abs(studentized_resids) > threshold) %>%
head()

#The unusual observations
outliers_one <- which(mri$crt > 3 | mri$crt > 3)

#Acquired from studentized residuals
outliers_second <- append(298,426)

outliers <- c(outliers_one,outliers_second)
mri %>% drop_na() %>%
  ggplot( aes(x = crt, y = bmi, color = sex)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE)+
  xlab("Creatinine (mg/dl)") + ylab("Body Mass Index (lb/in^2)")+
  geom_point(data = mri[outliers_one,], size = 8, shape = 23, color = 'red')+
  geom_point(data = mri[outliers_second,], size = 8, shape = 23, color = 'blue')+
  theme_bw()
mri <- mri[!is.na(mri$crt), ]

lm_sex <- regress("mean", bmi ~ crt, data = mri)
female <- mri %>%
  filter(sex == "Female")

male <- mri %>%
  filter(sex == "Male")

t_bmi <- t.test(female$bmi, male$bmi, var.equal=TRUE)

t_crt <- t.test(female$crt, male$crt, var.equal = TRUE)


knitr::kable(mri %>%
select(c(bmi, crt, sex)) %>%
group_by(sex) %>%
summarise(n = n(), across(everything(), mean)), caption = "Summary of of data by sex")
mri %>% drop_na() %>%
  ggplot( aes(x = sex, y = crt, col=sex)) +
  geom_boxplot() +
  xlab("Gender") + ylab("Creatinine level (mg/dl)")+
  theme_bw()
tab <- map_df(list(t_crt), tidy)
```

```
knitr::kable(tab[c("estimate", "statistic", "p.value", "conf.low", "conf.high")], col.names =
      c('Estimate', 'Statistics','Pr(>|t|)','95%L','95%H'),
      caption = "Summary of ttest between Creatinine level and sexes")
tab <- map_df(list(t_bmi), tidy)


knitr::kable(tab[c("estimate", "statistic", "p.value",
                   "conf.low", "conf.high")],col.names =
      c('Estimate', 'Statistics','Pr(>|t|)','95%L','95%H'),
      caption = "Summary of ttest between BMI and sexes")

knitr::kable(coef(lm_sex) %>% round(4), caption = "Model summary for linear fit of Body Mass Index (BMI)
mri <- mri[!is.na(mri$crt), ]

lm_sex_conf <- regress("mean", bmi ~ crt + sex, data = mri)
knitr::kable(coef(lm_sex_conf) %>% round(4), caption = "A summary of a linear fit of BMI with Creatinine
```