WCGS data lab workbook

Maureen Lahiff

March 6, 2020

# this R Markdown file assumes you went through the Week 1 R Tutorial first

library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(readr)  
library(epitools)  
library(readr)  
library(markdown)  
  
data(wcgs)

wcgs$dibpat0\_fact <- factor(wcgs$dibpat0, ordered = TRUE, labels = c("A","B"))  
  
  
wcgs$smoker0[wcgs$ncigs0 > 0] <- 1  
wcgs$smoker0[wcgs$ncigs0 == 0] <- 0  
  
  
wcgs$highsbp0[wcgs$sbp0 >= 140] <- 1  
wcgs$highsbp0[wcgs$sbp0 < 140] <- 0  
  
  
wcgs$heightcm0 <- round(wcgs$height0 \* 2.54, digits = 2)  
wcgs$weightkg0 <- round(wcgs$weight0/2.2, digits = 2)  
wcgs$BMI0 <- round(wcgs$weightkg0/((wcgs$heightcm0/100)^2), digits = 1)

# Question 1 to turn in  
  
# average cholesterol at baseline for smokers and non-smokers  
  
wcgs %>% group\_by(smoker0) %>% summarize(average = mean(sbp0))

## # A tibble: 2 x 2  
## smoker0 average  
## <dbl> <dbl>  
## 1 0 129.  
## 2 1 129.

# create a factor variable out of the 4-level behavioral pattern variable, behpat0  
  
wcgs$behpat0\_fact <- factor(wcgs$behpat0,  
 ordered = TRUE,  
 labels = c("A", "some A", "mix A and B","B"))

# cut gives us a lot of flexibility   
# in our specifications of the interval endpoints   
# the intervals here are the commonly used ones  
  
wcgs$bmi\_cat <- cut(wcgs$BMI0,  
 breaks = c(0, 18.5, 25.0, 30.0, Inf),  
 include.lowest = TRUE,  
 right = FALSE,  
 ordered\_results = TRUE,  
 labels = c("underweight", "normal", "overweight", "obese") )

wcgs$chd69 <- factor(wcgs$chd69, labels = c("No CHD", "CHD"))

# these are examples for additional tables  
  
behpat\_table <- table(wcgs$behpat0\_fact)  
addmargins(behpat\_table)

##   
## A some A mix A and B B Sum   
## 264 1325 1216 349 3154

round(prop.table(behpat\_table), digits = 3)

##   
## A some A mix A and B B   
## 0.084 0.420 0.386 0.111

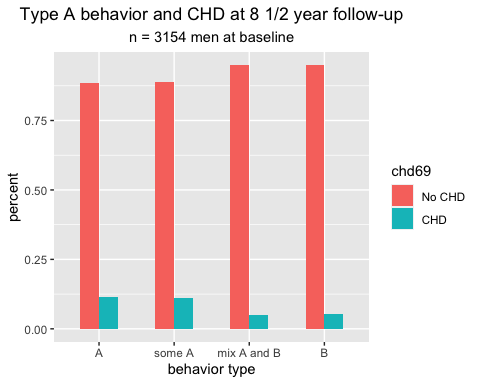
# table for behavioral pattern and chd  
  
# the proportion option 1 asks for "row percents" in the two-way table   
# this is Question 2 to turn in  
  
chd\_behpat\_table <- table(wcgs$behpat0\_fact, wcgs$chd69)  
addmargins(chd\_behpat\_table)

##   
## No CHD CHD Sum  
## A 234 30 264  
## some A 1177 148 1325  
## mix A and B 1155 61 1216  
## B 331 18 349  
## Sum 2897 257 3154

round(prop.table(chd\_behpat\_table, 1), digits = 3)

##   
## No CHD CHD  
## A 0.886 0.114  
## some A 0.888 0.112  
## mix A and B 0.950 0.050  
## B 0.948 0.052

# this is Question 3 to turn in  
  
wcgs %>% count(behpat0\_fact, chd69) %>%  
 group\_by(behpat0\_fact) %>%  
 mutate(prop = n/sum(n)) %>%  
 ggplot(aes(x = behpat0\_fact, y = prop, fill = chd69)) +  
 geom\_bar(width = 0.5, stat = "identity", position = "dodge") +  
 labs(y = "percent", x = "behavior type", title = "Type A behavior and CHD at 8 1/2 year follow-up",   
 subtitle = "n = 3154 men at baseline") +  
 theme(plot.title = element\_text(hjust = 0.5), plot.subtitle = element\_text(hjust = 0.5))



# summary for BMI by CHD at 8 /12 yr follow-up  
  
no\_chd\_summary <- summary(subset(wcgs, chd69 == "No CHD", select = BMI0))  
  
chd\_summary <- summary(subset(wcgs, chd69 == "CHD", select = BMI0))  
  
no\_chd\_summary

## BMI0   
## Min. :11.20   
## 1st Qu.:22.90   
## Median :24.40   
## Mean :24.53   
## 3rd Qu.:25.90   
## Max. :37.70

chd\_summary

## BMI0   
## Min. :19.30   
## 1st Qu.:23.70   
## Median :24.90   
## Mean :25.11   
## 3rd Qu.:26.60   
## Max. :39.00

#use facet wrap to get side by side histgrams for baseline BMI by CHD  
  
ggplot(wcgs, aes(x = BMI0, y = ..density..)) +  
 geom\_histogram(binwidth = 1, color = "black", fill = "purple") +  
 facet\_wrap(.~wcgs$chd69) +   
 labs(title = "WCGS baseline BMI by CHD group",  
 subtitle = "men ages 39 to 59",  
 x = "BMI ",  
 y = "density" ) +  
 scale\_x\_continuous(limits = c(18.5, 40),  
 breaks = c(20, 25, 30, 35),  
 labels = c(20, 25, 30, 35)) +  
 theme(plot.title = element\_text(hjust = 0.5),  
 plot.subtitle = element\_text(hjust = 0.5))

## Warning: Removed 20 rows containing non-finite values (stat\_bin).

## Warning: Removed 2 rows containing missing values (geom\_bar).

