BMI Example

Michael Sachs

March 31, 2015

Let's read in the BMI dataset and see what we have to work with. Make sure it is in the same directory as this .Rmd file!

bmi <- read.csv("BMI.CSV", stringsAsFactors = FALSE)  
head(bmi)

## id age age\_cat gender bmi smoke packyrs  
## 1 DCEG-001 75 3 FEMALE 32.05128 NEVER 0  
## 2 DCEG-002 68 2 FEMALE 30.85399 NEVER 0  
## 3 DCEG-003 36 1 FEMALE 20.76125 NEVER 0  
## 4 DCEG-004 70 2 FEMALE 31.58861 NEVER 0  
## 5 DCEG-005 71 3 MALE 27.09925 NEVER 0  
## 6 DCEG-006 66 2 FEMALE 24.23823 NEVER 0

summary(bmi)

## id age age\_cat gender   
## Length:300 Min. :36.0 Min. :1.000 Length:300   
## Class :character 1st Qu.:60.0 1st Qu.:1.000 Class :character   
## Mode :character Median :66.5 Median :2.000 Mode :character   
## Mean :65.4 Mean :2.047   
## 3rd Qu.:72.0 3rd Qu.:3.000   
## Max. :79.0 Max. :3.000   
##   
## bmi smoke packyrs   
## Min. :17.04 Length:300 Min. : 0.00   
## 1st Qu.:23.45 Class :character 1st Qu.: 0.00   
## Median :25.76 Mode :character Median : 18.50   
## Mean :26.15 Mean : 26.84   
## 3rd Qu.:28.38 3rd Qu.: 44.00   
## Max. :35.00 Max. :315.00   
## NA's :5

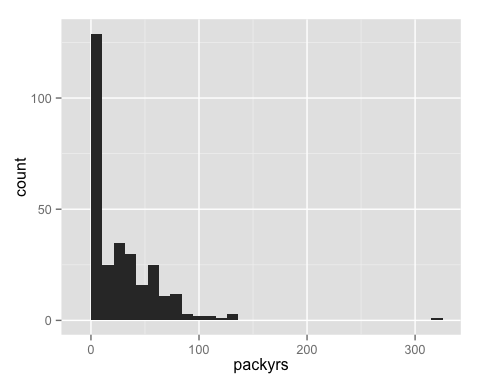
Let's examine the relationship between gender and pack years of smoking.

with(bmi, t.test(packyrs ~ gender))

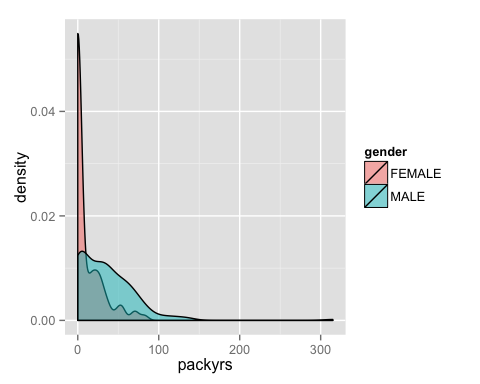
Welch Two Sample t-test  
  
data: packyrs by gender  
t = -7.4018, df = 288.294, p-value = 1.481e-12  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -29.8723 -17.3226  
sample estimates:  
mean in group FEMALE mean in group MALE   
 10.52103 34.11848

Wow, men tend to smoke much more than women. What is the distribution of pack years of smoking?

library(ggplot2)  
ggplot(bmi, aes(x = packyrs)) + geom\_bar(stat = "bin")



ggplot(bmi, aes(x = packyrs, fill = gender)) + geom\_density(alpha = .5)



It appears that there are more women that do not smoke at all (packyears = 0). The proportion of women who do not smoke is NaN.

with(bmi, knitr::kable(table(gender, smoke), caption = "Cross-tabulation of gender by smoking status in the BMI dataset. "))

Cross-tabulation of gender by smoking status in the BMI dataset.

|  |  |  |  |
| --- | --- | --- | --- |
|  | CURRENT | FORMER | NEVER |
| FEMALE | 22 | 14 | 56 |
| MALE | 78 | 86 | 44 |