BMI Example

Michael Sachs

March 31, 2015

## Load data

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!

## Inference

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

with(bmi, t.test(packyrs ~ I(smoke == "CURRENT")))

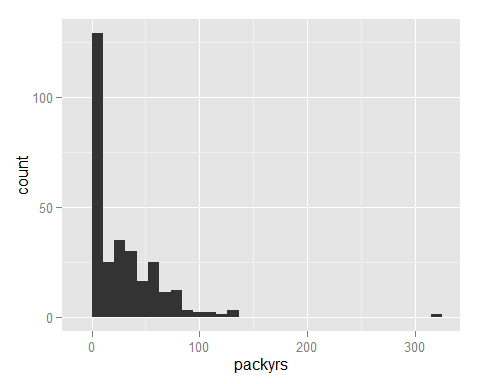
Welch Two Sample t-test  
  
data: packyrs by I(smoke == "CURRENT")  
t = -8.5218, df = 214.705, p-value = 2.786e-15  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -37.81372 -23.60714  
sample estimates:  
mean in group FALSE mean in group TRUE   
 16.74127 47.45169

for(i in 1:10){  
   
 bmi[, paste0("i", i)] <- bmi$age \* i  
   
}

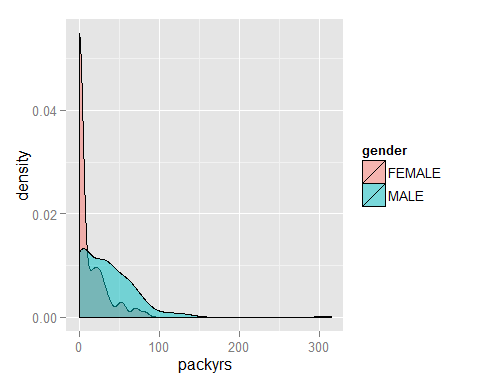
## Plots

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)



## Table

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. "))

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

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