Math 130 Project

Sophia Sussman

9/28/2018

# Introduction

We will be looking at the data set ‘depress’ for this project. Depress is a data set with 37 variables and 294 observations about subjects with various levels of depression and characteristics that can influence and impact this data. I will be looking at whether the subject drinks or not and at the CESD (Center for Epidemiologic Studies Depression) Scale.

depress <- read.delim("/Users/sophiasussman/Documents/School/Chico State/2 Sophmore Year/Math 130/depress\_081217.txt", header=TRUE,sep="\t")  
library(ggplot2)  
library(dplyr)

# Univariate Descriptions

## Drinking

drink = is the subject a regular drinker? (yes/no)

This first variable asks the subject if they are a regular drinker or not.

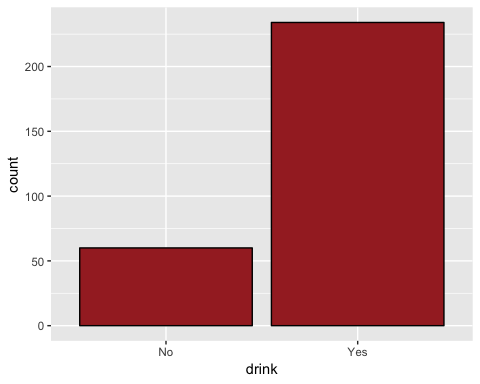
depress$drink <-factor(depress$drink, labels=c("No", "Yes"))  
table(depress$drink)

##   
## No Yes   
## 60 234

prop.table(table(depress$drink))

##   
## No Yes   
## 0.2040816 0.7959184

ggplot(depress, aes(x=drink)) + geom\_bar(fill='brown', colour='black')



As shown in this data set, 234 people (79.6%) of those surveyed drink, where as 60 people (20.4%) do not drink.

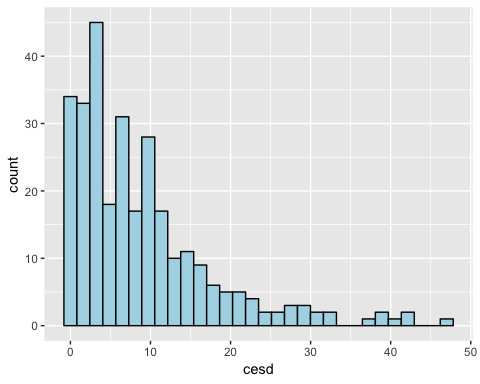
## Level of Depression

cesd = a continous summary of how

The second variable is the CESD scale. This is measured by asking the subject to look at a card and tell the administrator what number best describes how often they felt/behaved this certain way during the past week. The subject responds on a scale from 0-3, how often they felt or behaved this way (0 = none of the time, 3 = most of the time).There were 20 cards presented to the subject, thus producing a continuous variable from 0-50.

ggplot(depress, aes(x=cesd)) + geom\_histogram(fill='lightblue', colour='black')

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



summary(depress$cesd)

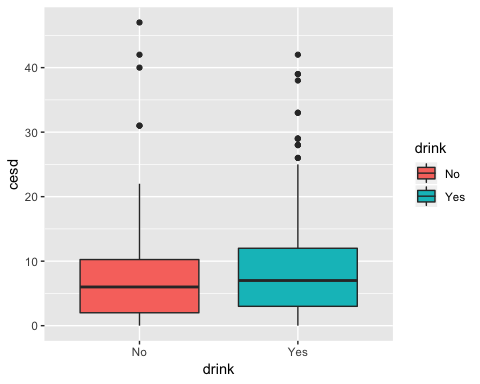
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 3.000 7.000 8.884 12.000 47.000

The median CESD score is 7, and the mean score is 8.884. The graph shown above is skewed right.

# Bivariate Comparison

Here I am comparing the variables ‘drink’ and ‘cesd’

ggplot(depress, aes(x=drink, y=cesd, fill=drink)) + geom\_boxplot ()



depress %>%  
 group\_by(drink) %>%  
 summarise(  
 mean\_cesd = mean(cesd, na.rm = TRUE)  
 )

## # A tibble: 2 x 2  
## drink mean\_cesd  
## <fct> <dbl>  
## 1 No 9.02  
## 2 Yes 8.85

The mean CESD for subjects who do not drink is 9.01. The mean CESD for subject who do drink is 8.85. This is not what I would have thought was going to happen, as I thought the data would show that those who drink were more likely to suffer from depression.