ScannellFinalProject

## R Markdown

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(ggplot2)
library(forcats)
library(knitr)
depress <- read.table("C:/Users/scann/Desktop/Math 130/Data/depress.txt", header=TRUE, sep="\t")

1. Introduction:

For this analysis, I will be utilizing the depression data set. This data was taken from a depression study obtained by interviewing 294 adults residing in Los Angeles County. I am going to be exploring both of the variables “cases” and “drink.” “Cases” is showing whether or not the person is depressed, based on whether or not they had a “cesd” score of over 16. The cesd is a sum of their scores based upon their mental state over a span of 20 days- the higher the number, the more depression is present. I am going to be exploring this along with the variable “drink”, which shows if that person is a regular drinker or not. I think exploring these could lead to an interesting hypothesis on the relationship between being a regular drinker and having depression.

1. Univariate Description for Each of the Variables Under Consideration

First- “cases”

depress$cases <- factor(depress$cases, labels = c("Not Depressed", "Depressed"))

table(depress$cases, useNA = "always")

##
## Not Depressed Depressed <NA>
## 244 50 0

ggplot(depress, aes(x = cases, fill = cases)) + theme\_bw() +
 geom\_bar() + scale\_fill\_brewer(palette = "Set2", name="Individuals") + ggtitle("Individuals with Depression vs. Without It") + ylab("Number of Cases") + xlab("Individuals")



In this study, we have 244 “normal” individuals, while 50 have depression. We can see this by looking at both the summary statistic and the graph. Therefore, far more people are not depressed than depressed in this particular study. This data is consistent with what I expected.

Second- “drink”

depress$drink <- factor(depress$drink, labels = c("Not a Regular Drinker", "Regular Drinker"))
table(depress$drink, useNA="always")

##
## Not a Regular Drinker Regular Drinker <NA>
## 60 234 0

ggplot(depress, aes(x = drink, fill = drink)) + theme\_bw() +
 geom\_bar() + scale\_fill\_brewer(palette = "Set2", name="Individuals") + ggtitle("Individuals That Drink Regularly vs. That Don't") + ylab("Number of Cases") + xlab("Individuals")



We can use these summary statistics and graph to get a better idea of what we’re looking at with this data. In this study, we have 234 individuals that are regular drinkers and 60 individuals that are not. Interestingly, the number of people who drink regularly far outweigh those that don’t drink regularly. This goes against what I expected the results to be.

1. A Bivariate Comparison Between Two Variables of Interest: “drink” and “cases”

table(depress$drink, depress$cases)

##
## Not Depressed Depressed
## Not a Regular Drinker 51 9
## Regular Drinker 193 41

51+193

## [1] 244

193/244

## [1] 0.7909836

9+41

## [1] 50

41/50

## [1] 0.82

ggplot(depress, aes(x=cases, fill=drink)) + geom\_bar(position = "dodge") + theme\_bw() + scale\_fill\_brewer(palette = "Set2", name="Level of Drinking") + ggtitle("Cases of Depression and Regular Drinking") + ylab("Number of Cases") + xlab("Individuals")



I compared these two variables because I was curious if there was a correlation between the two, specifically if there was a lot of overlap between having depression and drinking regularly. It appears that for both depressed and not depressed individuals, more people are regular drinkers than not. I was wondering if possibly there would be more cases of regular drinking with people with depression. Under the summary statistics above, I divided the amount of regular drinkers by the total number of individuals, working with the depressed and not depressed groups separately. By doing this, I discovered that about 79% of people without depression were regular drinkers. About 82% of people with depression were regular drinkers. Although I cannot draw any conclusions, it does seem that my thinking was not 100% correct. In this study, a higher percentage of people with depression drink regularly, but the difference is not too extreme.