Homework 1

Jordan Rhoades

Due 2/6/2019

Instructions: Complete the following problems using RMarkdown to render a Microsoft Word document. Include code in the space provided. If a direct question is asked, answer that question below the code input for that problem.

## Problem 1

In a national survey conducted by the Centers for Disease Control to determine health-risk behaviors among college students, college students were asked, “How often do you wear a seat belt when driving a car?” The frequencies were as follows:

|  |  |
| --- | --- |
| Response | Frequency |
| I do not drive | 249 |
| Never | 118 |
| Rarely | 249 |
| Sometimes | 345 |
| Most of the time | 716 |
| Always | 3093 |

1. Construct a relative frequency distribution.

responses <- c(rep("I do not drive",249),rep("Never",118),rep("Rarely",249),rep("Sometimes",345),rep("Most of the time",716),rep("Always",3093))  
  
dist <- (table(responses)/length(responses))  
dist

## responses  
## Always I do not drive Most of the time Never   
## 0.64842767 0.05220126 0.15010482 0.02473795   
## Rarely Sometimes   
## 0.05220126 0.07232704

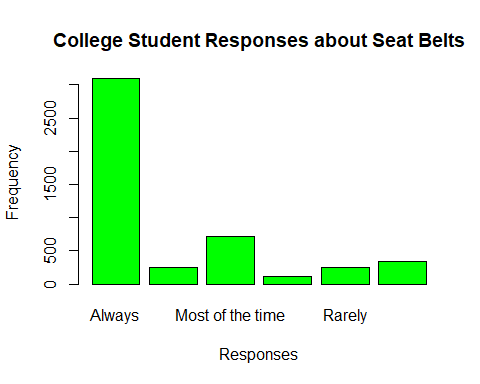
1. What percentage of respondents answered “Always”?

Always <- 3093  
(Always/length(responses)) \* 100

## [1] 64.84277

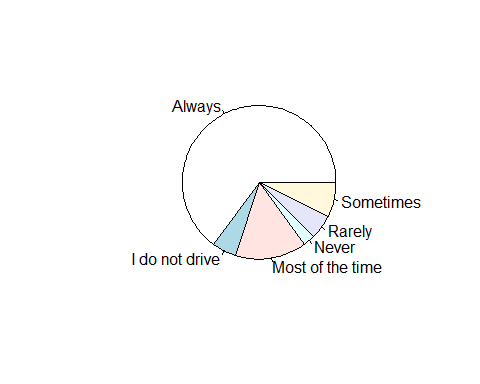
1. Construct a frequency bar graph.

barplot(table(responses),col="green",xlab="Responses",ylab="Frequency",main="College Student Responses about Seat Belts")



1. Construct a pie chart.

pie(table(responses))



## Problem 2

A phlebotomist draws the blood of a random sample of 50 patients and determines their blood types as shown:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| O | O | A | A | O |
| B | O | B | A | O |
| AB | B | A | B | AB |
| O | O | A | A | O |
| AB | O | A | B | A |
| O | A | A | O | A |
| O | A | O | AB | A |
| O | B | A | A | O |
| O | O | O | A | O |
| O | A | O | A | O |

1. Construct a frequency distribution.

types <- c(rep("O",22),rep("B",6),rep("AB",4),rep("A",18))  
table(types)

## types  
## A AB B O   
## 18 4 6 22

1. Construct a relative frequency distribution.

table(types)/length(types)

## types  
## A AB B O   
## 0.36 0.08 0.12 0.44

1. According to the data, which blood type is the most common?

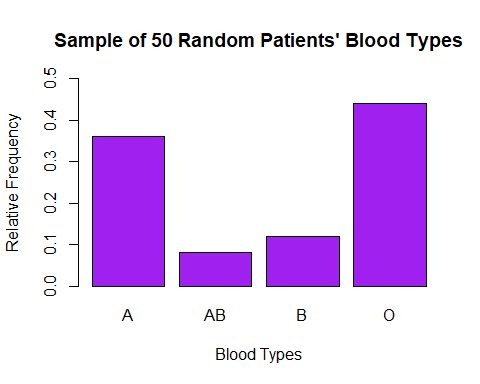
##According to the data, the most common blood type is O.

1. Use the results of the sample to conjecture the percentage of the population that has type O blood. Is this an example of descriptive or inferential statistics?

## Inferential statistics, because we are trying to take this sample’s data and relate it to the population. This example definitely isn’t descriptive statistics because we aren’t using graphs or tables.

1. Construct a relative frequency bar graph.

barplot(table(types)/length(types),col="purple",xlab="Blood Types",ylab="Relative Frequency",ylim = c(0,.5),main="Sample of 50 Random Patients' Blood Types")



## Problem 3

In an experiment, a researcher asks a basketball player to record the number of free throws she shoot until she misses. The experiment is repeated 50 times. The follow table lists the distribution of the number of free throws attempted until a miss is recorded.

|  |  |
| --- | --- |
| Number of Free Throws | Frequency |
| 1 | 16 |
| 2 | 11 |
| 3 | 9 |
| 4 | 7 |
| 5 | 2 |
| 6 | 3 |
| 7 | 0 |
| 8 | 1 |
| 9 | 0 |
| 10 | 1 |

1. Construct a relative frequency distribution of the data.

freeThrows <- c(rep(1,16),rep(2,11),rep(3,9),rep(4,7),rep(5,2),rep(6,3),8,10)  
table(freeThrows)/length(freeThrows)

## freeThrows  
## 1 2 3 4 5 6 8 10   
## 0.32 0.22 0.18 0.14 0.04 0.06 0.02 0.02

1. What percentage of the time did she first miss on her fourth free throw?

fourthFT <- 7  
fourthFT/length(freeThrows) \*100

## [1] 14

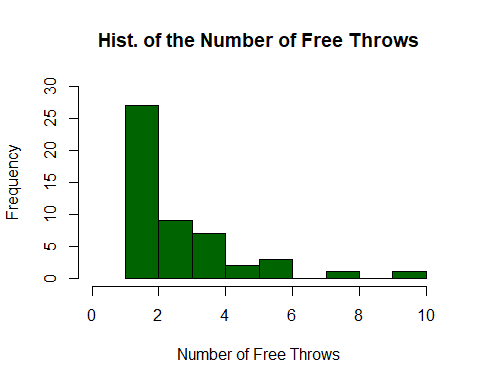
1. What percentage of the time did she make at least five in a row?

(2+3+0+1+0+1)/length(freeThrows) \* 100

## [1] 14

1. Construct a frequency histogram.

hist(freeThrows,breaks=11,main="Hist. of the Number of Free Throws",xlab="Number of Free Throws",xlim=c(0,10),ylim=c(0,30),col="Dark Green")



1. Describe the shape of the data.

## You can definitely see that this graph is a very right skewed data with the outliers that it has such as 10.

## Problem 4

The data below represent the number of customers waiting for a table at 6:00 pm for 40 consecutive Saturdays at Bobak’s Restaurant.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 11 | 5 | 11 | 3 | 6 | 8 | 6 | 7 |
| 4 | 5 | 13 | 9 | 6 | 4 | 14 | 11 |
| 13 | 10 | 9 | 6 | 8 | 10 | 9 | 5 |
| 10 | 8 | 7 | 3 | 8 | 8 | 7 | 8 |
| 7 | 9 | 10 | 4 | 8 | 6 | 11 | 8 |

1. Are the data discrete or continuous? Explain. ##This data is discrete because there are countable numbers and the numbers are finite.
2. Construct a frequency distribution of the data.

wait <- c( 11 , 5 ,11 ,3 ,6 , 8 , 6 , 7   
 , 4 , 5 ,13 ,9 ,6 , 4 ,14 ,11   
 , 13 ,10 , 9 ,6 ,8 ,10 , 9 , 5   
 , 10 , 8 , 7 ,3 ,8 , 8 , 7 , 8   
 , 7 , 9 ,10 ,4 ,8 , 6 ,11 , 8 )  
table(wait)

## wait  
## 3 4 5 6 7 8 9 10 11 13 14   
## 2 3 3 5 4 8 4 4 4 2 1

1. Construct a relative frequency distribution of the data.

table(wait)/length(wait)

## wait  
## 3 4 5 6 7 8 9 10 11 13 14   
## 0.050 0.075 0.075 0.125 0.100 0.200 0.100 0.100 0.100 0.050 0.025

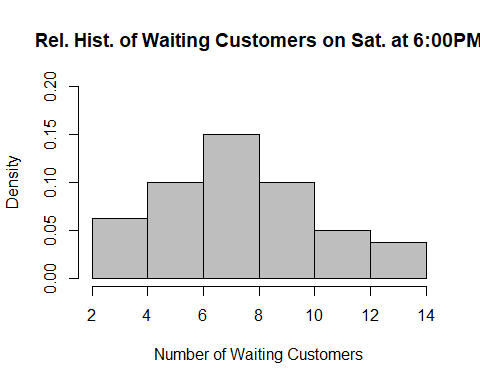
1. What percentage of the Saturdays had 10 or more customers waiting for a table at 6:00 pm?

tenMore <- (4+4+2+1)  
tenMore/length(wait) \* 100

## [1] 27.5

1. Construct a relative frequency histogram of the data. Using this relative frequency histogram, describe the shape of the data.

relWait <- table(wait)/length(wait)  
hist(wait,probability = TRUE,main="Rel. Hist. of Waiting Customers on Sat. at 6:00PM", xlab="Number of Waiting Customers",col="grey",ylim = c(0,.2),xlim = c(2,14))



##The shape of the histogram is bell-shaped.

## Problem 5

The table shows the tax, in dollars, on a pack of cigarettes in each of the 50 states and Washington, D.C., as of January 2014. **Note:** The state with the lowest tax is Missouri and the state with the highest tax is New York.

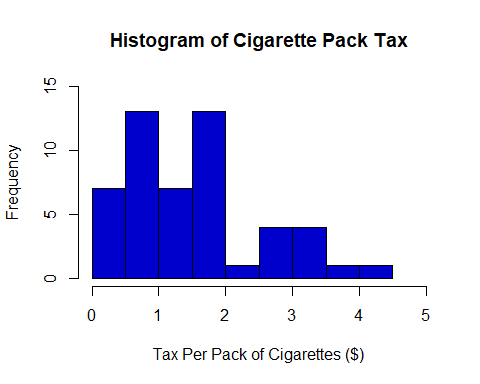
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 0.425 | 1.339 | 0.600 | 0.170 | 0.450 | 1.530 | 2.520 |
| 2.000 | 0.370 | 0.360 | 1.700 | 0.440 | 0.620 | 0.600 |
| 2.000 | 3.200 | 2.000 | 0.640 | 1.250 | 1.410 | 2.500 |
| 1.150 | 0.570 | 2.000 | 0.800 | 1.030 | 1.700 |  |
| 0.870 | 1.980 | 3.510 | 1.680 | 1.310 | 2.620 |  |
| 0.840 | 0.995 | 2.000 | 2.700 | 1.600 | 0.300 |  |
| 3.400 | 1.360 | 2.830 | 1.660 | 3.500 | 3.025 |  |
| 1.600 | 0.790 | 0.680 | 4.350 | 0.570 | 0.550 |  |

1. With a first class having a lower class limit of 0 and a class width of 0.50, construct a frequency histogram of the data.

tax <- c(0.425 , 1.339 , 0.600 , 0.170 , 0.450 , 1.530 , 2.520 ,  
2.000 , 0.370 , 0.360 , 1.700 , 0.440 , 0.620 , 0.600 ,  
2.000 , 3.200 , 2.000 , 0.640 , 1.250 , 1.410 , 2.500 ,  
1.150 , 0.570 , 2.000 , 0.800 , 1.030 , 1.700 ,   
0.870 , 1.980 , 3.510 , 1.680 , 1.310 , 2.620 ,   
0.840 , 0.995 , 2.000 , 2.700 , 1.600 , 0.300 ,   
3.400 , 1.360 , 2.830 , 1.660 , 3.500 , 3.025 ,   
1.600 , 0.790 , 0.680 , 4.350 , 0.570 , 0.550 )  
table(tax)

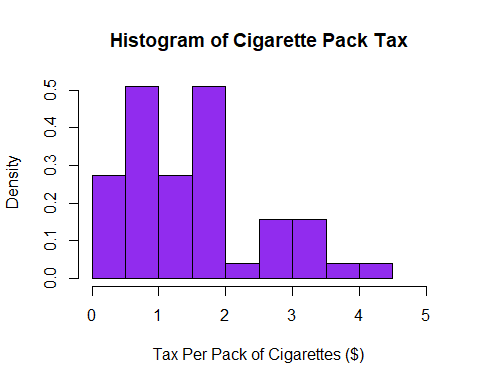
## tax  
## 0.17 0.3 0.36 0.37 0.425 0.44 0.45 0.55 0.57 0.6 0.62 0.64   
## 1 1 1 1 1 1 1 1 2 2 1 1   
## 0.68 0.79 0.8 0.84 0.87 0.995 1.03 1.15 1.25 1.31 1.339 1.36   
## 1 1 1 1 1 1 1 1 1 1 1 1   
## 1.41 1.53 1.6 1.66 1.68 1.7 1.98 2 2.5 2.52 2.62 2.7   
## 1 1 2 1 1 2 1 5 1 1 1 1   
## 2.83 3.025 3.2 3.4 3.5 3.51 4.35   
## 1 1 1 1 1 1 1

class1 <- c(.17,.3,.36,.37,.425,.44,.45)  
hist(tax,breaks=10,ylim=c(0,15),xlim=c(0,5),col="blue3",xlab="Tax Per Pack of Cigarettes ($)",main="Histogram of Cigarette Pack Tax")



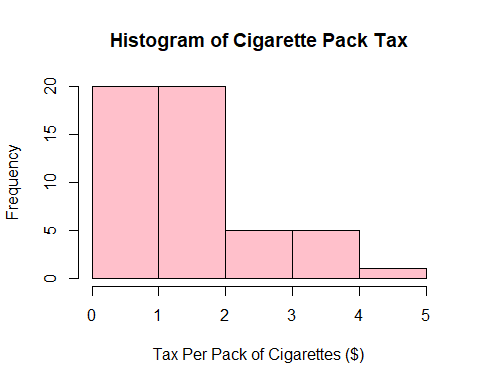
1. With a first class having a lower class limit of 0 and a class width of 0.50, construct a relative frequency histogram of the data.

hist(tax,breaks=9,probability = TRUE,xlim = c(0,5),col="purple2",xlab="Tax Per Pack of Cigarettes ($)",main="Histogram of Cigarette Pack Tax")



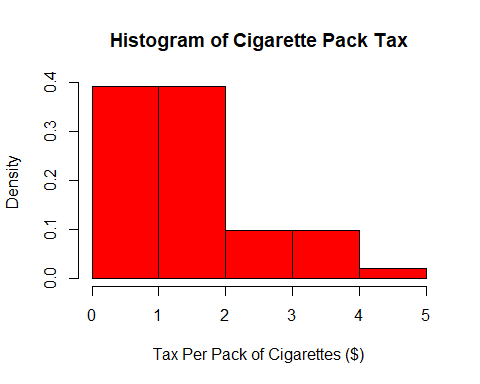
1. Describe the shape of the distribution of the data. ##The shape of the distribuition is right skewed.
2. With a first class having a lower class limit of 0 and a class width of 1, construct a frequency histogram of the data.

hist(tax,breaks=5,col="pink",xlab="Tax Per Pack of Cigarettes ($)",main="Histogram of Cigarette Pack Tax")



1. With a first class having a lower class limit of 0 and a class width of 1, construct a relative frequency histogram of the data.

hist(tax,probability=TRUE,breaks=5,col="red1",xlab="Tax Per Pack of Cigarettes ($)",main="Histogram of Cigarette Pack Tax")



1. Does one frequency distribution provide a better summary of the data than the other? Explain.

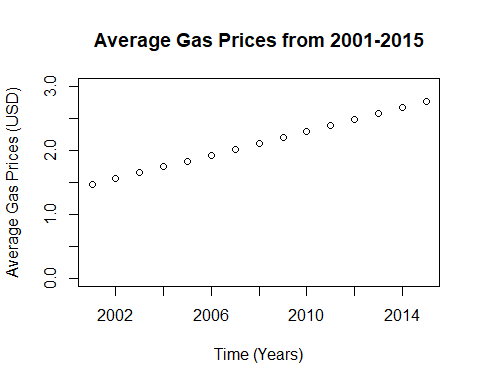
## The frequency distribution from part (a) provides the best summary because you can understand more when the widths of each class are smaller and it breaks down the data instead of having the data be overally simplified like the width = 1 graphs. It also gives the reader a better understanding by using frequency in the y-axis instead of having density which could be confusing for some readers.

## Problem 6

The average per gallon price for regular unleaded gasoline in the United States rose from $1.46 in 2001 to $2.77 in 2015.

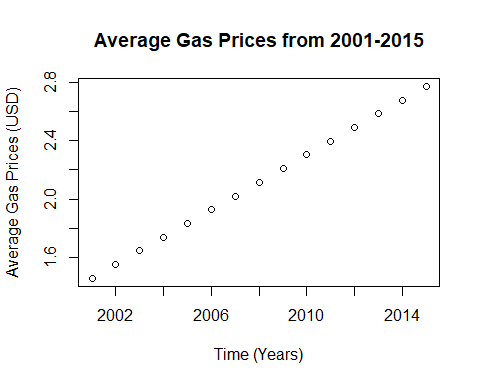
1. Construct a graphic that is not misleading to depict this situation.

year <- c(2001:2015)  
gas <- seq(1.46,2.77,length=15)  
  
plot(year,gas,ylim= c(0,3),ylab="Average Gas Prices (USD)",xlab="Time (Years)",main="Average Gas Prices from 2001-2015")



1. Construct a misleading graphic that makes it appear the average price roughly quadrupled between 2001 and 2015.

year <- c(2001:2015)  
gas <- seq(1.46,2.77,length=15)  
plot(year,gas,ylab="Average Gas Prices (USD)",xlab="Time (Years)",main="Average Gas Prices from 2001-2015",xlim= c(2001,2015))



## Problem 7

Look at the linked image found on [Wikipedia](https://en.wikipedia.org/wiki/List_of_presidents_of_the_United_States_by_age#/media/File:Age_of_presidential_ascension.svg). The intention of the graph is to display the ages of U.S. presidents on Inauguration Day. Explain any problems you see with the graph.

##The author of the graph created this graph to look like and even states that it's a gaussian curve. If you look at the range of the ages (x-axis) and cut the graph in half, then add up the frequency on the left side you get 22 and then the right side gets 23 so the graph is definitely symmetric. A problem with this graph is that it looks almost like it's skewed right when it's not because I found that it was pretty symmetric. My biggest problem with this graph is that there is no labeling of the axes and if I didnt know what this graph was about, then I wouldn't understand what was going on.