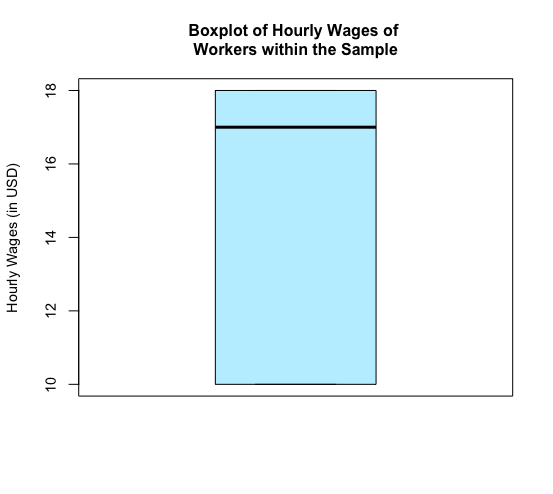
Logan Albiani

ECN 102

10/7/18

Prof. Siegler

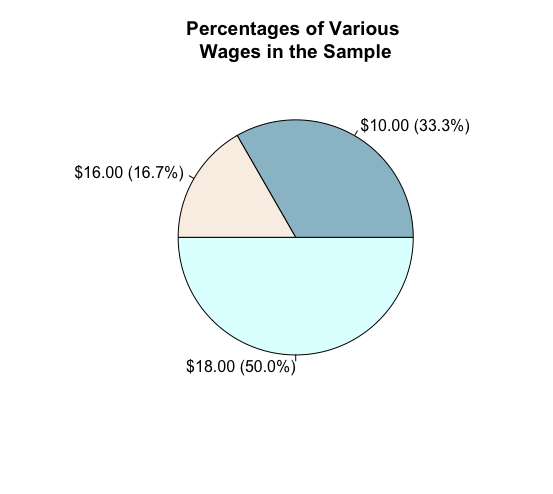
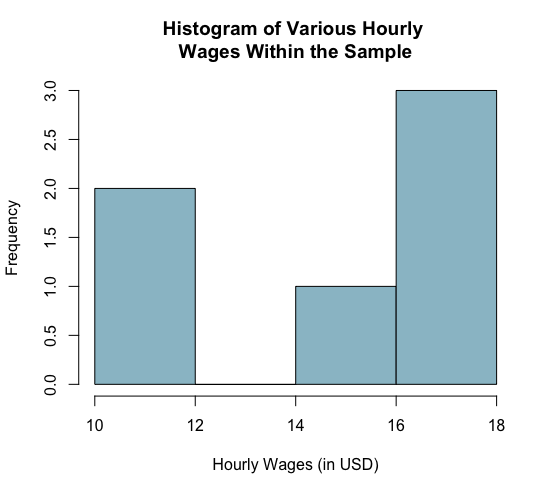
Homework Set #1

1. Categorizing Data
   1. Due to the data analyzing a single group at a single point in time, these data are **cross-sectional**.
   2. In this example, the data are tracking an individual over time, thus meaning these are time **series data**.
   3. Again, this data analyzes a single variable and how it changes over a period of time, thus meaning they are time **series data**.
   4. Here, the data measures various individuals over multiple variables (time and marriage), meaning they are **panel data.**
   5. The data are measuring a single variable (GPA of college students) at a specific point in time, meaning that the data are **cross-sectional**.
2. Descriptive Statistics: The Old-Fashioned Way (and with R)
   1. Measures of Center:
      1. Arithmetic Mean: ⇒ ⇒
      2. Geometric Mean: ⇒ ⇒
      3. Median ⇒ ⇒ **17**
      4. Mode: ⇒ **18** due to its three occurrences in the data set.
   2. Measures of Spread
      1. Variance: ⇒ ⇒ **15.6**
      2. Standard Deviation: ⇒ ⇒ **3.95**
      3. Coefficient of Variation: ⇒ ⇒ .**263**
   3. Inter-Quartile Range: ⇒
      1. ⇒
      2. ⇒
      3.  ⇒ **8**
   4. Boxplot of the data set:

25th Percentile

75th Percentile

Median

* 1. Pie chart of the data:
  2. Histogram of the data:
  3. Skewness: In this particular data set, the median is greater than the arithmetic mean. What this indicates is that the skewness of the distribution is negative. Due to this, we see a longer leftward tail. We can deduce that the distribution has a negative skew from the relationship between the mean and the median due to the fact that in a standard normal distribution, the mean and the median will be equal to each other. If the mean is greater than or less than the median, it is being pulled to a greater extent by outliers on one side or the other, contributing to the tail attribute of a skewed distribution.
  4. Kurtosis:
     1. = kurtosis ⇒ ⇒ **1.473** = kurtosis
     2. The distirbution is **platykurtic** due to the kurtosis being less than three. Inutitively, this means that the distribution has steeper shoulders, a flatter peak, and thinner tails due to the general even occurances of the data. In this type of distribution, extreme events are less likely to occur due to a vast majority of the data being spread evenly around the first three standard deviations.
        1. ⇒ ⇒ **Platykurtic**

1. SAT Scores/Z-scores
   1. Perctile of Test Takers
      1. ;
      2. ⇒ ⇒ .7642
      3. ⇒*p*( ⇒ .0918
      4. (.7942-.0918)\*100 ⇒ **67.24%**
   2. Percentile Above
      1. (1-z)\*100;
      2. ⇒ *p(* z<100) = 1.44 ⇒ .9251
      3. (1-.9251)\*100 ⇒ **7.49%**
   3. Probability of a Sample Mean
      1. = z ⇒ ⇒ 4.31= z ⇒ p(z>4.31) = .000008 or .0008%
      2. The probability of a sample mean being greater than 1340 is extraordinarily lower than a single student achieving that same score. This is because achieving a score of 1340 is already low in probability, thus, nine randomly selected students all achieving that high score would be extremely rare, as 92.51% of students score below 1340. The chances of choosing nine students who score in the top 7.5th percentile making up a single sample is extraordinarily unlikely.

R Code for Homework 1:

## Preliminary Commands ##

setwd("~/Desktop/ECN 102 HW#1")

install.packages("moments")

library(moments)

install.packages("modeest")

library(modeest)

mycols<-c("lightblue3","linen","lightcyan")

wage<-c(10,10,16,18,18,18)

summary(wage)

## Problem 2 - Descriptive Statistics ##

## 2A - Measures of Center ##

mean(wage)

exp(mean(log(wage)))

median(wage)

mfv(wage)

## 2B - Measures of Spread ##

var(wage)

sd(wage)

abs(sd(wage)/mean(wage))

## 2C - IQR ##

IQR(wage,type=6)

## 2D - Boxplot ##

boxplot(wage,

cex = .7, cex.lab=.9,cex.axis=.9,

ylab="Hourly Wages (in USD)", cex.ylab=1,

main="Boxplot of Hourly Wages of \nWorkers within the Sample", cex.main = 1,

col=c("lightblue1"))

## 2E - Pie Chart ##

wage\_percent<-c(33.3,16.7,50)

lbl<-c("$10.00 (33.3%)","$16.00 (16.7%)","$18.00 (50.0%)")

pie(wage\_percent,labels = lbl, main="Percentages of Various \nWages in the Sample",

cex.main = 1.2,

col = mycols, font=1)

## 2F - Histogram ##

hist(wage,

main="Histogram of Various Hourly \nWages Within the Sample",

xlab="Hourly Wages (in USD)",

border="black",

col="lightblue3",

xlim=c(10,18))

## 2G - Skewness ##

skewness(wage)

## 2H - Kurtosis ##

kurtosis(wage)

exkurt<-kurtosis(wage)-3

## 3 Probabilities and Z-scores ##

## 3A Percentile of Test Takers

pnorm(.72)-pnorm(-1.33)

## 3B Percentile Above ##

1-pnorm(1.44)

## FIN ##