Cassidy Drummond
September 7th, 2023
Methods Assignment #1

load ("Users/cassidydrummond/Downloads/Starbucks.xlsx")
CD_GCAmount <- "DollarAmtprepaidcard"

#1

data <- data.frame(Starbucks)
print(data)

```
> #1
> data <- data.frame(Starbucks)
> print(data)
   Dollar.Amt.prepaid.card.... Age Days.per.Month.at.Starbucks
1                           35  25                           4
2                          120  30                          12
3                           75  27                          10
4                           80  42                           8
5                           35  29                          11
6                          110  25                          12
7                          100  50                           8
8                           85  45                           6
9                           85  32                          16
10                          40  23                          10
11                         115  40                          18
12                          85  35                          12
13                         135  28                          10
14                          90  33                          12
15                         205  40                          15
16                          75  37                           3
17                         105  51                          10
18                          75  20                           8
19                          80  26                          15
20                         180  38                          19
21                          65  27                          12
22                          90  29                          14
23                         115  34                          10
24                         125  30                           6
25                          55  22                           8
   Cups.of.Coffee.per.Day Income..1000. Male
1                       1           20    0
2                       5           35    0
3                       4           30    0
4                       5           30    1
5                       8           25    1
6                       5           60    1
7                       3           30    0
8                       5           35    0
9                       7           25    1
10                      1           20    0
11                      5           40    1
12                      3           40    1
13                      3           50    1
14                      2           30    0
15                      5           80    0
16                      1           30    0
17                      8           35    1
18                      4           25    0
19                      5           35    1
20                     10           45    0
```

#2

male <-(Starbucks$Male == 1)
female <-(Starbucks$Male == 0)

#3a

means <-colMeans(data)
print(means)

sd <- apply(data, 2, sd)
print(sd)

max <- apply(data, 2, max)
print(max)

min <-apply(data, 2, min)
print(min)

```
> means <-colMeans(data)
> print(means)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                     94.40                    32.72                        10.76                        4.44
              Income..1000.                     Male
                     36.80                     0.48
>
> sd <- apply(data, 2, sd)
> print(sd)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                  39.956226                 8.403967                     4.013311                     2.237558
              Income..1000.                     Male
                  13.453624                0.509902
>
> max <- apply(data, 2, max)
> print(max)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                       205                       51                           19                           10
              Income..1000.                     Male
                        80                        1
>
> min <-apply(data, 2, min)
> print(min)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                        35                       20                            3                            1
              Income..1000.                     Male
                        20                        0
```

#3b

female_data <- data[data$Male == "0",]
male_data <- data[data$Male == "1",]

meanF <- colMeans(female_data)
print(meanF)
meanM <- colMeans(male_data)

```
print(meanM)

sdF <-apply(female_data, 2, sd)
sdM <-apply(male_data, 2, sd)
print(sdF)
print(sdM)

maxF <- apply(female_data, 2, max)
maxM <- apply(male_data, 2, max)
print(maxF)
print(maxM)

minF <- apply(female_data, 2, min)
minM <- apply(male_data, 2, min)
print(minF)
print(minM)
```

```
> print(meanF)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                 95.000000                32.615385                   10.230769                    3.846154
             Income..1000.                     Male
                 34.615385                 0.000000
> meanM <- colMeans(male_data)
> print(meanM)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                 93.750000                32.833333                   11.333333                    5.083333
             Income..1000.                     Male
                 39.166667                 1.000000
>
> sdF <-apply(female_data, 2, sd)
> sdM <-apply(male_data, 2, sd)
> print(sdF)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                 49.032302                 8.902607                    4.493585                    2.511512
             Income..1000.                     Male
                 15.201636                 0.000000
> print(sdM)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                 29.319944                 8.222290                    3.524804                    1.781640
             Income..1000.                     Male
                 11.448170                 0.000000
>
> maxF <- apply(female_data, 2, max)
> maxM <- apply(male_data, 2, max)
> print(maxF)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                        205                       50                           19                           10
             Income..1000.                     Male
                         80                        0
> print(maxM)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                        135                       51                           18                            8
             Income..1000.                     Male
                         60                        1
>
> minF <- apply(female_data, 2, min)
> minM <- apply(male_data, 2, min)
> print(minF)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                         35                       20                            3                            1
             Income..1000.                     Male
                         20                        0
> print(minM)
Dollar.Amt.prepaid.card....                      Age Days.per.Month.at.Starbucks      Cups.of.Coffee.per.Day
                         35                       22                            6                            3
             Income..1000.                     Male
                         25                        1
```

#3c

gender_table <- table(data$Male)

print(gender_table)

```
> print(gender_table)

 0  1
13 12
```

#3d

medianincome <- median(data$Income..1000.)
High_income <- ifelse(data$Income..1000. > medianincome, 1, 0)
print(High_income)

```
#3d
medianincome <- median(data$Income..1000.)
High_income <- ifelse(data$Income..1000. > medianincome, 1, 0)
print(High_income)
[1] 0 0 0 0 0 1 0 0 0 0 1 1 1 0 1 0 0 0 0 1 0 0 1 1 0
```

#3e

conf_interval <- t.test(data$Income..1000., conf.level = 0.99)
print(conf_interval)

```
        One Sample t-test

data:  data$Income..1000.
t = 13.677, df = 24, p-value = 7.968e-13
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 29.27421 44.32579
sample estimates:
mean of x
     36.8
```

#3f

male_part <-data[data$Male == "1",]
female_part <- data[data$Male == "0",]
colors <- c("lightblue", "pink")
pie(c(male_data, female_data), labels = c("1","0", col = colors)
title("Distribution of Male and Female")
legend("topright", legend = c("1","0"), fill = colors)
#not sure what went wrong

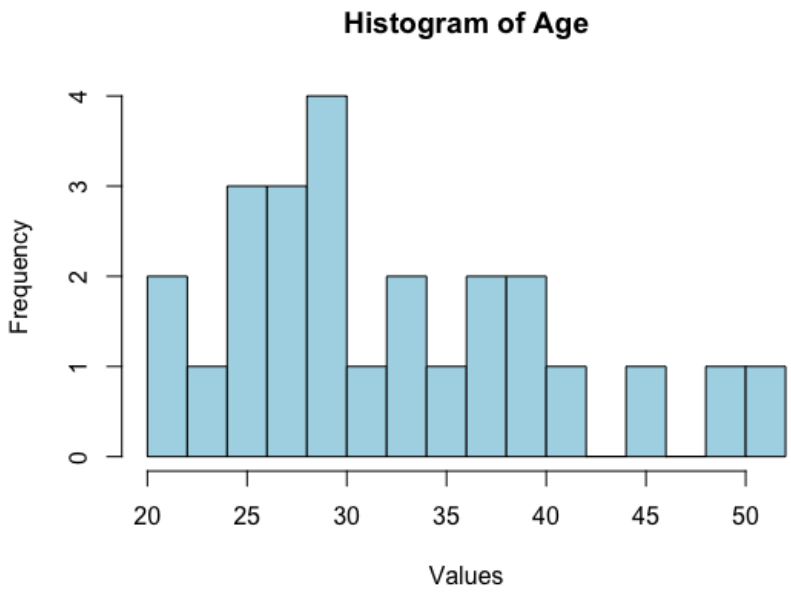
#3g

summary(female_data$Age)
summary(male_data$Age)

```
#3g
summary(female_data$Age)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
20.00   27.00   30.00   32.62   38.00   50.00
summary(male_data$Age)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
22.00   27.50   31.00   32.83   36.25   51.00
```

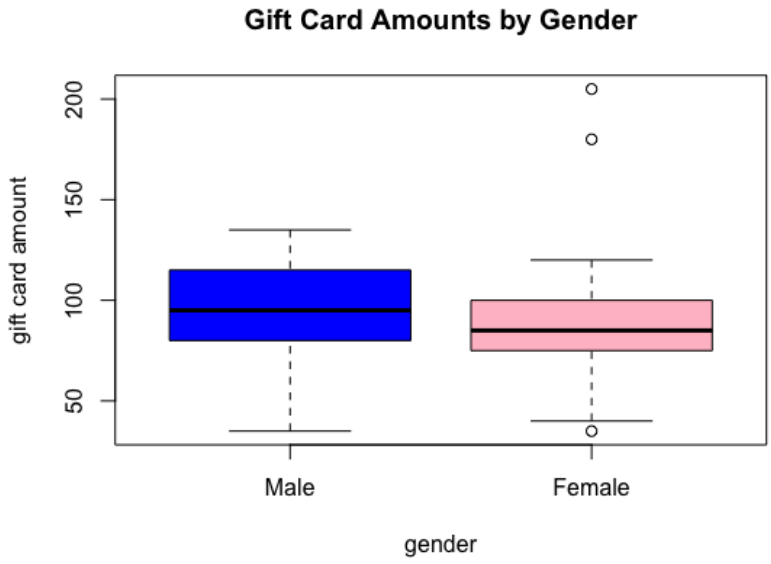
#3h

```
hist(data$Age,
    main = "Histogram of Age",
    xlab = "Values",
    ylab = "Frequency",
    col = "lightblue",
    border = "black",
    breaks = 20)
```



#3i

boxplot(male_data$Dollar.Amt.prepaid.card...., female_data$Dollar.Amt.prepaid.card...., names = c("Male", "Female"), main = "Gift Card Amounts by Gender", xlab = "gender", ylab = "gift card amount", col = c("blue", "pink"), border = "black")



#4a

ttest <- t.test(data$Dollar.Amt.prepaid.card...., mu = 80, alternative = "greater")
print(ttest)

Since the t-test value is positive, it means that the average is greater than what we tested it against (80). The mean is 94.4 which confirms the t-test. For the p-value, we used a .05 confidence interval. The p-value is .04206. Since it is lower than the significance level, it means we reject the null hypothesis (which stated the mean was 80).

```
> ttest <- t.test(data$Dollar.Amt.prepaid.card...., mu = 80, alternative = "greater")
> print(ttest)

        One Sample t-test

data:  data$Dollar.Amt.prepaid.card....
t = 1.802, df = 24, p-value = 0.04206
alternative hypothesis: true mean is greater than 80
95 percent confidence interval:
 80.72792      Inf
sample estimates:
mean of x
     94.4
```

#4b

ttestM <- t.test(male_data$Dollar.Amt.prepaid.card...., mu = 80, alternative = "greater")
print(ttestM)

ttestF <- t.test(female_data$Dollar.Amt.prepaid.card...., mu = 80, alternative = "greater")
print(ttestF)

Looking at the t-tests below, we can see that both males and females are positive and spend greater than $80 on gift cards. To be honest, I'm unsure of why the p-values differ by .08 when their means are only 1.25 different.

```
> ttestM <- t.test(male_data$Dollar.Amt.prepaid.card...., mu = 80, alternative = "greater")
> print(ttestM)

        One Sample t-test

data:  male_data$Dollar.Amt.prepaid.card....
t = 1.6245, df = 11, p-value = 0.06627
alternative hypothesis: true mean is greater than 80
95 percent confidence interval:
 78.54974      Inf
sample estimates:
mean of x
    93.75

> ttestF <- t.test(female_data$Dollar.Amt.prepaid.card...., mu = 80, alternative = "greater")
> print(ttestF)

        One Sample t-test

data:  female_data$Dollar.Amt.prepaid.card....
t = 1.103, df = 12, p-value = 0.1458
alternative hypothesis: true mean is greater than 80
95 percent confidence interval:
 70.76247      Inf
sample estimates:
mean of x
       95
```

#4c

reg <- lm(data$Dollar.Amt.prepaid.card.... ~ data$Male, data = data)
summary(reg)

First, under the estimated std., since the male variable is negative, it means that women are more likely to spend more on gift cards. However, even though women may spend more, since the p-value exceeds the significance level, it means that gender does not create a statistically significant impact on spending on gift cards.

```
> #4c
> reg <- lm(data$Dollar.Amt.prepaid.card.... ~ data$Male, data = data)
> summary(reg)

Call:
lm(formula = data$Dollar.Amt.prepaid.card.... ~ data$Male, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-60.00 -20.00  -8.75  21.25 110.00

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    95.00      11.32   8.393 1.87e-08 ***
data$Male      -1.25      16.34  -0.077     0.94
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.81 on 23 degrees of freedom
Multiple R-squared:  0.0002545, Adjusted R-squared:  -0.04321
F-statistic: 0.005854 on 1 and 23 DF,  p-value: 0.9397
```

#4d

regA <- lm(data$Dollar.Amt.prepaid.card.... ~ data$Age + data$Days.per.Month.at.Starbucks + data$Cups.of.Coffee.per.Day + data$Income..1000., data = data)
summary(regA)

Assessing the coefficients, we can see that the largest influencing factor on the amount of dollars purchased is frequency of visiting Starbucks and closely following is income. The least influential variable is the cups of coffee someone consumes per day. Using a significance level of .05, the smaller the p-value, the more statistically significant it is (in theory). My p-values show that income and age have the most statistically significant impact.

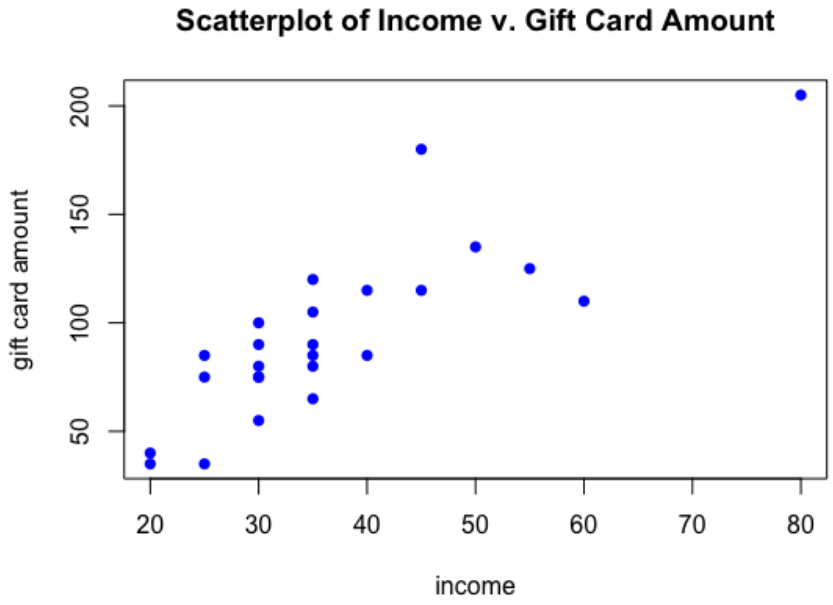
```
> #4d
> regA <- lm(data$Dollar.Amt.prepaid.card.... ~ data$Age + data$Days.pe
r.Month.at.Starbucks + data$Cups.of.Coffee.per.Day + data$Income..1000.
data = data)
> summary(regA)

Call:
lm(formula = data$Dollar.Amt.prepaid.card.... ~ data$Age + data$Days.pe
r.Month.at.Starbucks +
    data$Cups.of.Coffee.per.Day + data$Income..1000., data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-32.672  -9.836   0.407   6.771  40.007

Coefficients:
                                Estimate Std. Error
(Intercept)                     -49.4552    19.2305
data$Age                          1.1881     0.4923
data$Days.per.Month.at.Starbucks  2.2433     1.2601
data$Cups.of.Coffee.per.Day       0.6171     2.2493
data$Income..1000.                2.1224     0.3075
                                t value Pr(>|t|)
(Intercept)                      -2.572   0.0182 *
data$Age                          2.413   0.0255 *
data$Days.per.Month.at.Starbucks  1.780   0.0902 .
data$Cups.of.Coffee.per.Day       0.274   0.7866
data$Income..1000.                6.902 1.05e-06 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.93 on 20 degrees of freedom
Multiple R-squared:  0.8129,    Adjusted R-squared:  0.7754
F-statistic: 21.72 on 4 and 20 DF,  p-value: 4.809e-07
```

#4e

Looking at all of the tests conducted, the data confirms that gender does not have a statistically significant impact on the amount spent on Starbucks gift cards. The average for both genders is around 94. Women with higher incomes who visit Starbucks more frequently will likely spend more on a gift card; however, this statement is not statistically significant.

#4f

plot(data$Income..1000., data$Dollar.Amt.prepaid.card...., main = "Scatterplot of Income v. Gift Card Amount", xlab = "income", ylab = "gift card amount", pch = 16, col = "blue")


Scatterplot of Income v. Gift Card Amount
gift card amount
50
100
150
200
20
30
40
50
60
70
80
income