R Notebook

Kayla Harmon-18173095

Question 1.

```
bonus = read.csv("bonus_df (1).csv")

getmode <- function(g) {
   uniqv <- unique(g)
   uniqv[which.max(tabulate(match(g, uniqv)))]
}

g = c(bonus$gender)

result1 = getmode(g)
print(result1)

## [1] 0

getmode <- function(sm) {
   uniqv <- unique(sm)
   uniqv[which.max(tabulate(match(sm, uniqv)))]
}

sm = c(bonus$social_media)

result2 = getmode(sm)
print(result2)

## [1] 0
```

These results indicate that the majority of people studied are male and have social media. (Assuming 0 under social media = having social media).

Question 2.

```
mean(bonus$satisfaction)

## [1] 54.07608

mean(bonus$age)

## [1] 30.199
```

The average rating for satisfaction is a 54.07608 and the average age of a person studied is 30.199.

Question 3.

```
var.test(bonus$gender, bonus$satisfaction, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  bonus$gender and bonus$satisfaction
## F = 0.0024621, num df = 999, denom df = 999, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.002174770 0.002787393
## sample estimates:
## ratio of variances
##        0.002462101
```

```
t.test(bonus$gender, bonus$satisfaction)
```

```
##
##  Welch Two Sample t-test
##
## data:  bonus$gender and bonus$satisfaction
## t = -167.95, df = 1003.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -54.21622 -52.96394
## sample estimates:
## mean of x mean of y
##   0.48600  54.07608
```

The f-test statistic is 0.0024621 and the t-test statistic is -167.95.

The f-test statistic indicates there is no significance between gender and satisfaction since the f-test statistic is small. The p-value being 2.2e-16 confirms these results. The t-test statistic indicates there is no significance between gender and satisifcation because the t-test statistic is negative.

The null hypothesis that both means are equal should be rejected.

Question 4.

```
var.test(bonus$social_media, bonus$satisfaction, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  bonus$social_media and bonus$satisfaction
## F = 0.0024635, num df = 999, denom df = 999, p-value < 2.2e-16
```

```
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.002176050 0.002789033
## sample estimates:
## ratio of variances
##         0.00246355

t.test(bonus$social_media, bonus$satisfaction)

##
##  Welch Two Sample t-test
##
## data:  bonus$social_media and bonus$satisfaction
## t = -167.93, df = 1003.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -54.20922 -52.95694
## sample estimates:
## mean of x mean of y
##   0.49300  54.07608
```

The f-test statistic is 0.0024635 and the t-test statistic is -167.93.

The f-test statistic indicates there is no significance between gender and satisfaction since the f-test statistic is small. The p-value being 2.2e-16 confirms these results. The t-test statistic indicates there is no significance between social media usage and satisifcation because the t-test statistic is negative.

The null hypothesis that both means are equal should be rejected.

Question 5.

```
attach(bonus)
xtab = table(age, social_media, gender)
prop.table(xtab)

## , , gender = 0
##
##     social_media
## age      0     1
##   20 0.013 0.011
##   21 0.010 0.018
##   22 0.013 0.009
##   23 0.011 0.006
##   24 0.013 0.007
##   25 0.013 0.011
##   26 0.017 0.014
##   27 0.012 0.008
##   28 0.010 0.015
```

```
##   29 0.006 0.007
##   30 0.010 0.015
##   31 0.014 0.015
##   32 0.009 0.013
##   33 0.009 0.014
##   34 0.011 0.012
##   35 0.014 0.012
##   36 0.020 0.011
##   37 0.011 0.013
##   38 0.015 0.013
##   39 0.016 0.009
##   40 0.016 0.018
##
## , , gender = 1
##
##     social_media
## age      0     1
##   20 0.014 0.010
##   21 0.013 0.007
##   22 0.018 0.013
##   23 0.012 0.013
##   24 0.015 0.010
##   25 0.011 0.009
##   26 0.014 0.009
##   27 0.009 0.011
##   28 0.014 0.015
##   29 0.014 0.010
##   30 0.008 0.008
##   31 0.008 0.013
##   32 0.008 0.016
##   33 0.007 0.008
##   34 0.009 0.012
##   35 0.017 0.014
##   36 0.012 0.010
##   37 0.010 0.012
##   38 0.012 0.012
##   39 0.010 0.018
##   40 0.009 0.012
```

Question 6.

```
glm(satisfaction~., data = bonus)
```

```
##
## Call:  glm(formula = satisfaction ~ ., data = bonus)
##
## Coefficients:
##       (Intercept)                age             gender  previous_purchase
##           -0.1844             1.5098             1.9361             2.9766
```

```
##      social_media
##            0.9233
##
## Degrees of Freedom: 999 Total (i.e. Null);  995 Residual
## Null Deviance:       101500
## Residual Deviance: 1070  AIC: 2918
```

I conducted a regression analysis using the glm() function (generalized linear model) using satisfaction as the predictor variable and all other variables as the response variable. This model would be written like:

-0.1844(satisfaction) + 1.5098(age) + 1.9361(gender) + 2.9766(previous_purchase) + 0.9233(social_media)

For example, if a 35 year old female who has made 3 previous purchases and uses social media and has a satisfaction score of 63.28091 the model would look like:

-0.1844(63.28091) + 1.5098(35) + 1.9361(1) + 2.9766(3) + 0.9233(0)

And provide a result of 52.0399.