MTH 325, Quiz #5, Spring 2023 Name _____

Instructions: Answer each question as thoroughly as possible. Round answers to 4 decimal places as needed. Exact answers are best when possible. Be sure to answer all parts of each question.

1. In the file **325quiz5data.xlsx**, there is data related to preferences for drinking regular beer vs. light beer. The file contains data on gender (Male=1), Marital Status (Married=1), Income and Age. The Beer Preference column has been recoded as Beer_Pref_Class (Regular=1). Load this data into R, and create a logistic model using Income as the explanatory variable. You will need to convert Beer_Pref_Class to a factor. Then use glm(Beer_Pref_Class~Income, data=mydata, family="binomial") (updating your data filename as needed). Use summary() to display the results.

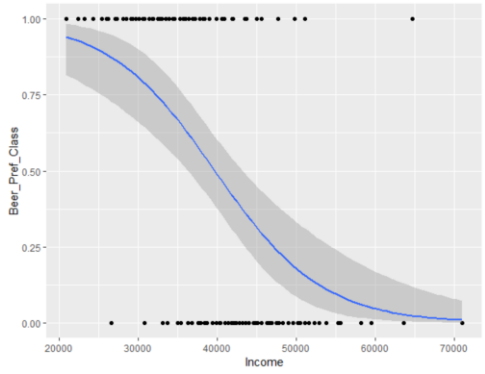
   a. Describe your hypothesis test of the model. Is the coefficient for Income significant?

```
Call:
glm(formula = Beer_Pref_Class ~ Income, family = "binomial",
    data = data2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0374  -0.9008   0.1037   0.8451   2.7266

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  5.873e+00  1.327e+00   4.428 9.53e-06 ***
Income      -1.478e-04  3.309e-05  -4.467 7.92e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 138.63  on 99  degrees of freedom
Residual deviance: 107.96  on 98  degrees of freedom
AIC: 111.96

Number of Fisher Scoring iterations: 4
```

$$H_0: \beta_1 = 0, H_a: \beta_1 \neq 0$$

The coefficient for income is significant in affecting the log-odds of beer preference.

   b. Create a plot of the data Income and Beer_Pref_Class and plot the model on the graph (you should be able to do this inside ggplot).



c. Interpret the coefficients of Income in the model in terms of odds of selecting regular beer as their preference.

Since the coefficient is -1.478×10^{-4}, we can say that for each 10,000 in income increase, the log-odds of regular beer preferences decreases by -1.478.

d. Redo the model with Gender, Married, Income and Age in your model. Are all the variables significant? If not which ones?

```
Call:
glm(formula = Beer_Pref_Class ~ Gender + Married + Age + Income,
    family = "binomial", data = data2)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.05923  -0.28800  -0.00092   0.39729   2.55206

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.819e-01  1.931e+00   0.353    0.724
Gender       7.779e-01  7.166e-01   1.085    0.278
Married     -1.697e-01  7.945e-01  -0.214    0.831
Age          2.282e-01  5.239e-02   4.357 1.32e-05 ***
Income      -2.785e-04  6.334e-05  -4.396 1.10e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 138.63  on 99  degrees of freedom
Residual deviance:  56.45  on 95  degrees of freedom
AIC: 66.45

Number of Fisher Scoring iterations: 6
```

Based on this model, age and income are significant, but marital status, gender and the intercept are not.