770866-9-47OEE AID: 1825 | 24/01/2022

**(a)**

Use the data of **YouthRisk** and Test whether there is sufficient evidence that the odds of riding with a driver who has been drinking are higher for young women compared to young men.

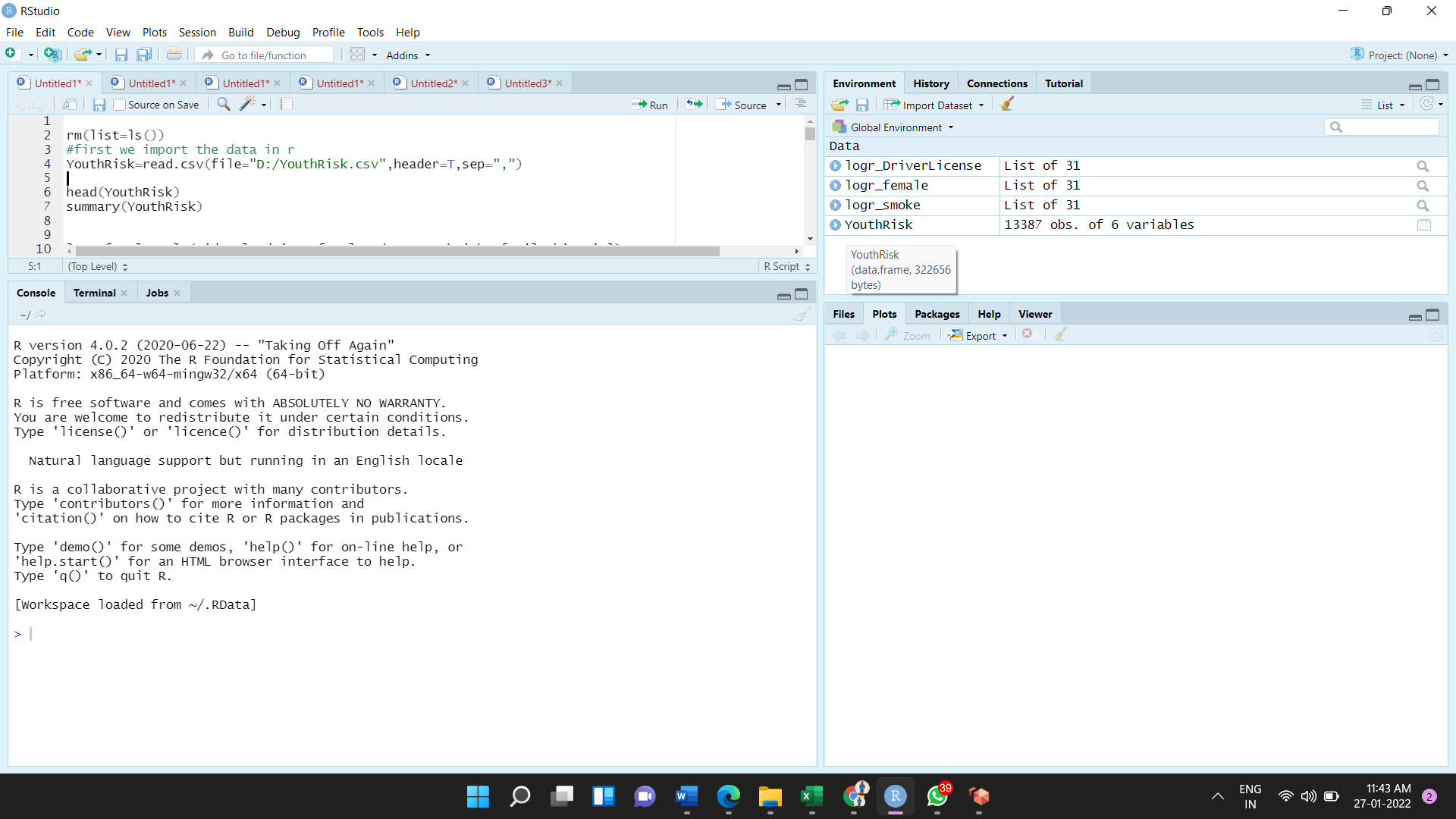
The given dataset YouthRisk has different variables conducted to examine the behaviors of someone who ride with someone who has been drinking.

To test this, first fit the logistic regression model for predicting the response variable ‘*ride.alc.driver*’ which is a binary taking two values, 1 and 0. The 1 denotes the respondent rode with a drinking driver within the past 30 days and 0 denotes the respondent didn’t ride with a drinking driver. The predictor is ‘*female’*.

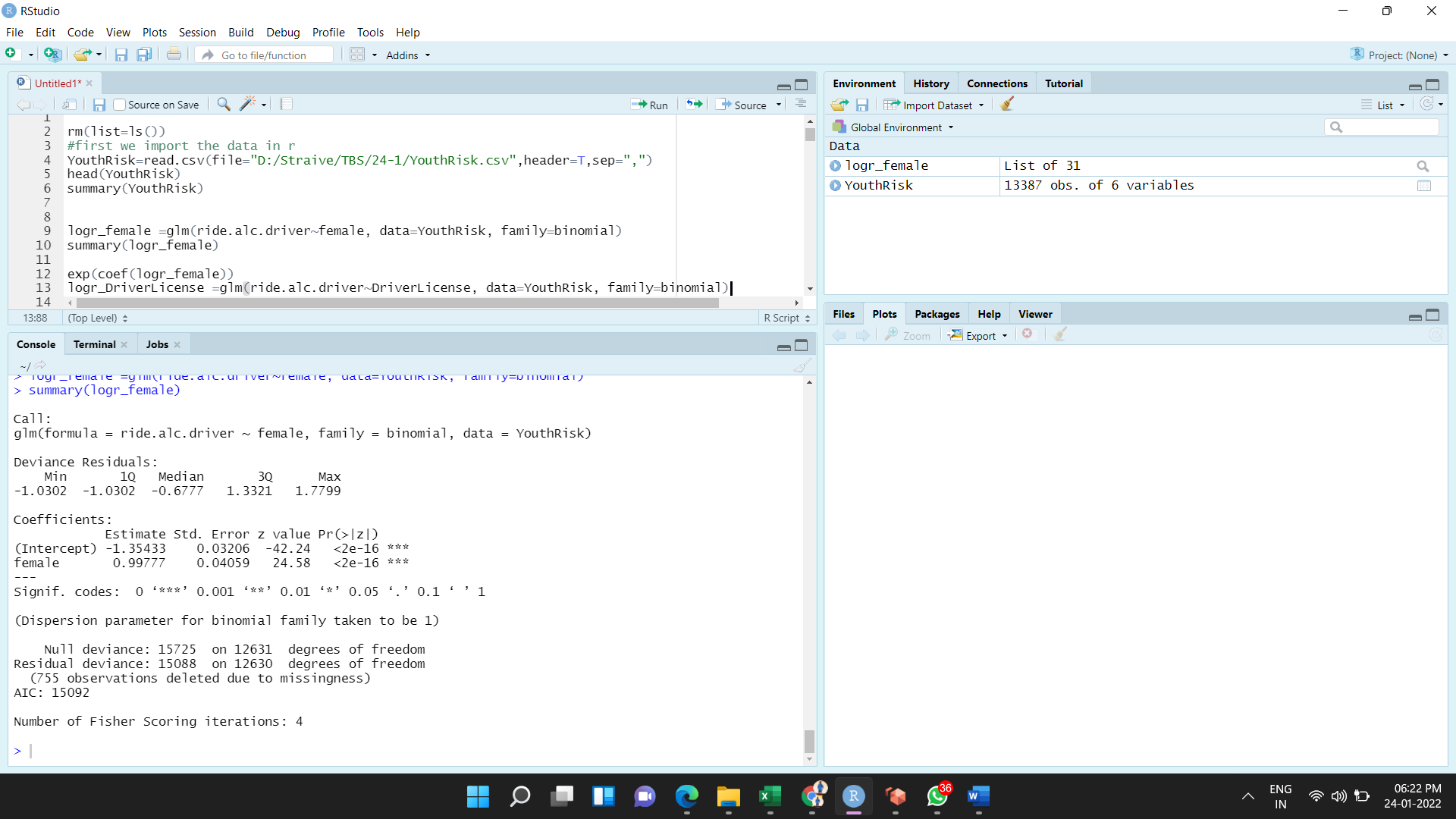
The logistic regression model for predicting ‘*ride.alc.driver*’ with predictor ‘*female’*.

The R-software procedure is below:

**Step 1:** First import the data in R using “**read.csv**” function

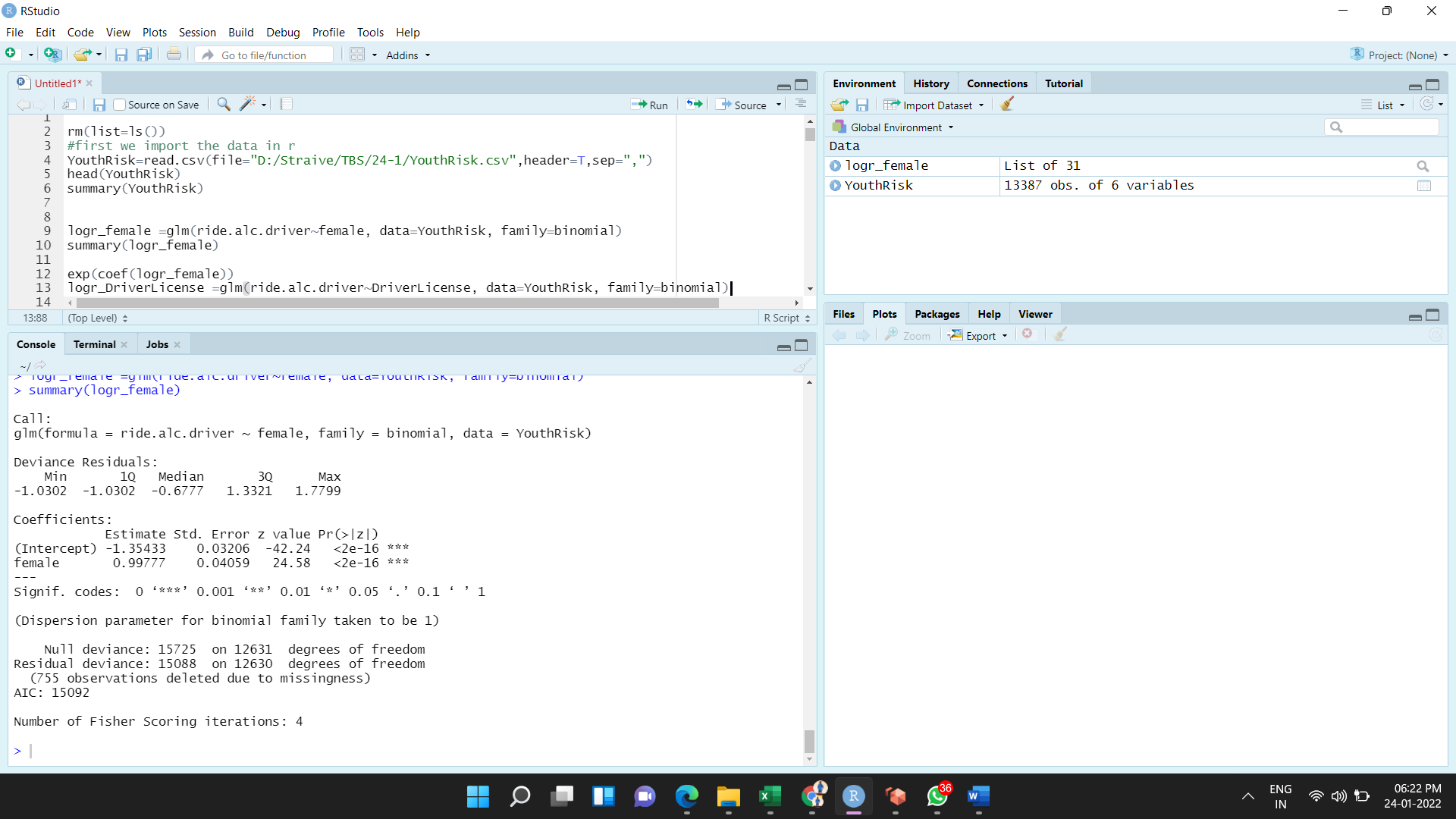


**Step 2:** Using the standard “**glm()”** function and specifying ‘**family=binomial**’. Write the following code for fitting logistic regression model.



**Step 3:** The “**summary()”** function gives us the following output. The below output shows logistic regression model with estimated coefficient and its p-value.

**R software output:**



The estimated logistic regression model is,

  
The variable *female* is binary and it takes two values 0 and 1. The 1 denotes the respondent is female and 0 denotes the respondent is male.

The estimated odds ratio is,



Therefore, the estimated odds ratio is **2.71.**

The odds ratio interpreted as the odds of a respondent riding with a drinking driver within the past 30 days are 2.71 times higher among the young women as compared to young men.

The test testing the odds of riding with a driver who has been drinking are higher for young women compared to young men.

The null and alternative hypothesis are,

**Null hypothesis:**



**Alternative hypothesis:**



The test statistic value for testing the null and alternative hypothesis is,

****

The p-value of above test is **0.00**

**Conclusion:**

Here, the P-value is less than level of significance.

That is,

Hence, the null hypothesis is rejected.

There is strong evidence to suggest that there is association between variable ‘*female’* and response ‘*ride.alc.driver’.*

So, it can be said that odds of riding with a driver who has been drinking are higher for young women compared to young men.

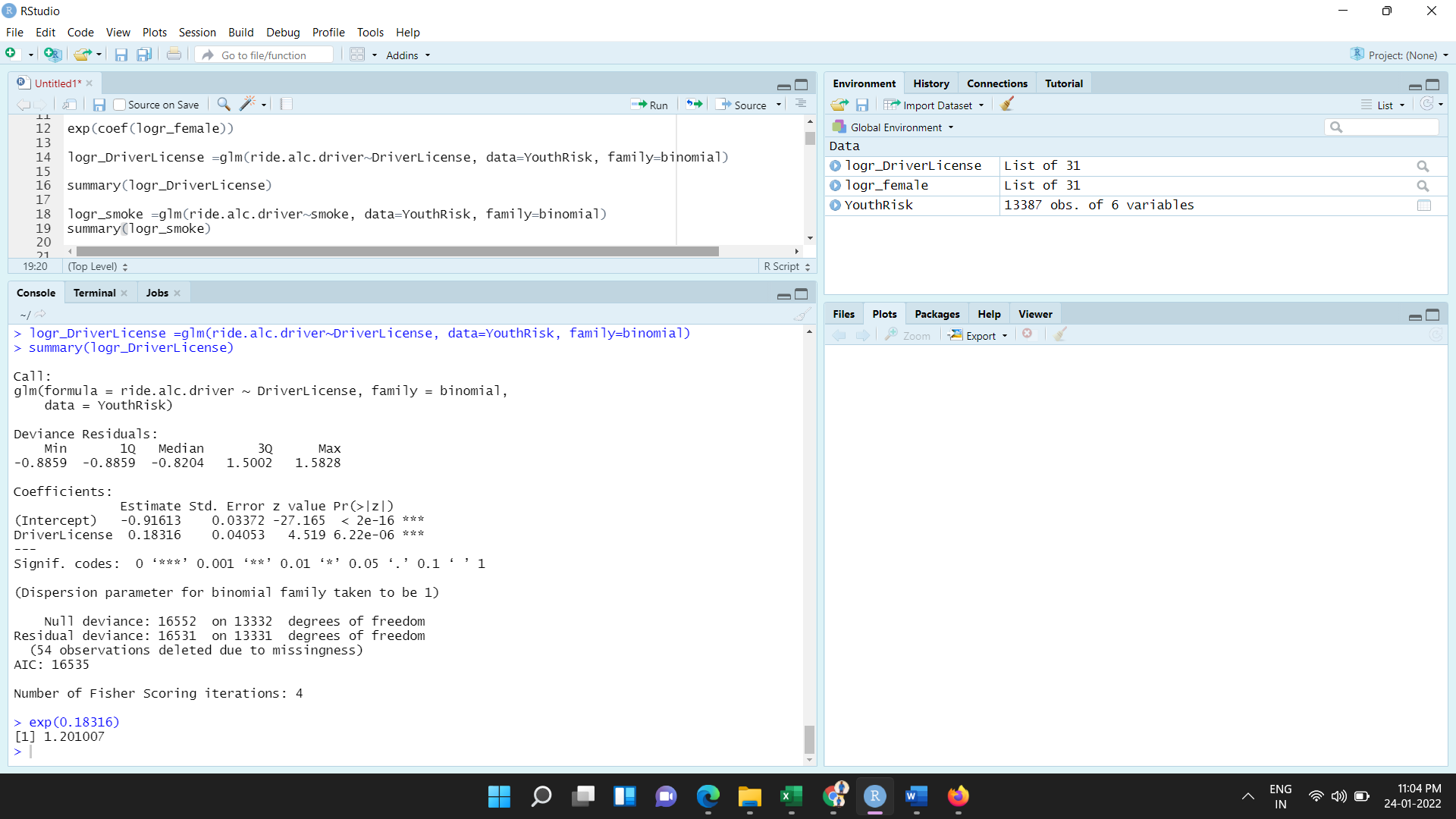
**(b)** Check whether there is evidence about the effect on drinking and driving after obtaining a driver’s license.

To check the effect of driving license on riding with alcoholic drivers, first fit the logistic regression for predicting response ‘*ride.alc.driver*’ with predictor ‘*DriverLicense*’.

The logistic regression model for predicting response ‘*ride.alc.driver*’ with predictor ‘*DriverLicense’*.

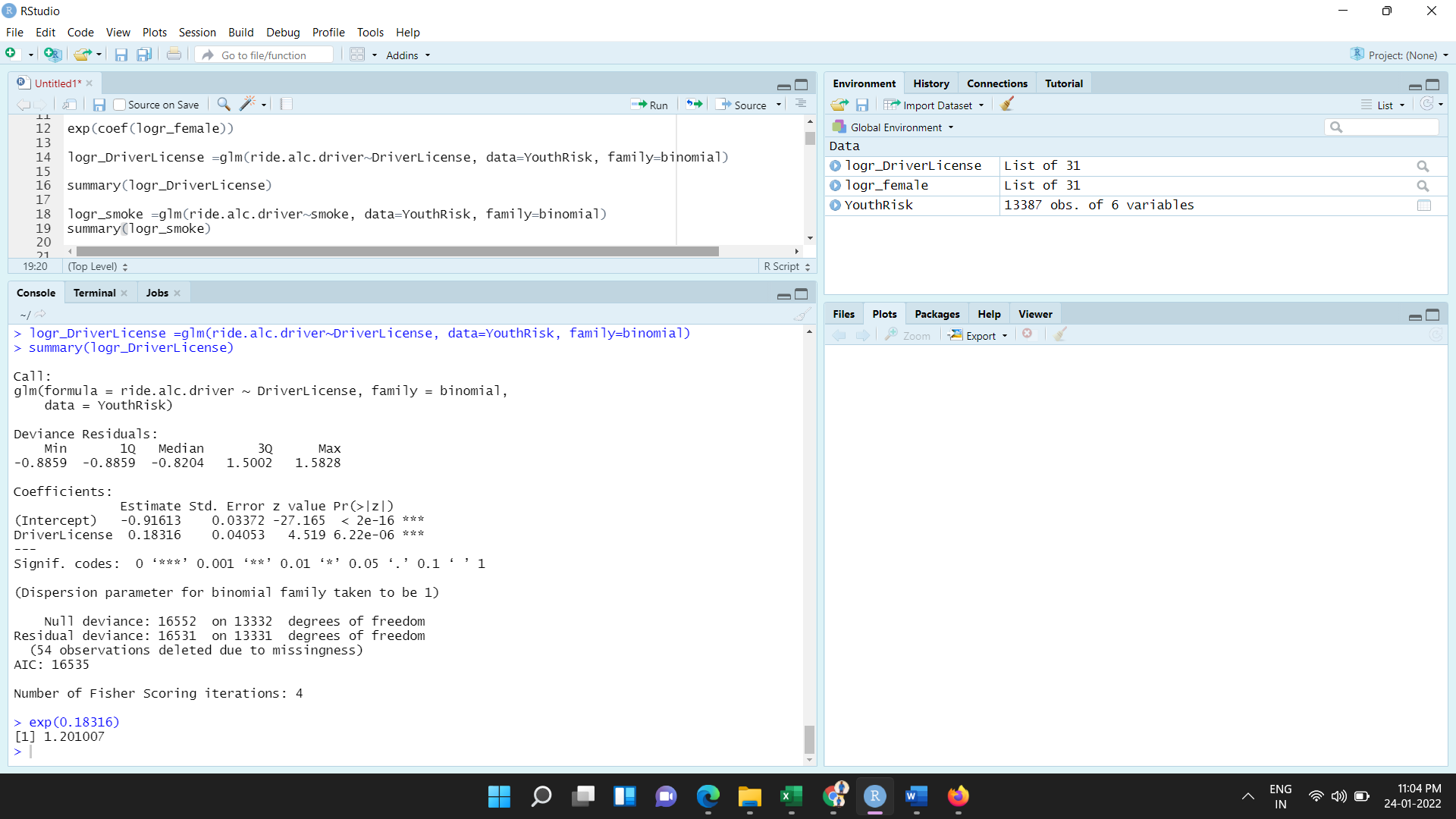
The R-software procedure is below.

**Step 1:** Using the standard “**glm()”** function and specifying ‘**family=binomial**’. Write the following code for fitting logistic regression model.



**Step 2:** The “**summary()”** function gives us the following output. The below output shows logistic regression model with estimated coefficient and its p-value.

**R software output:**



The estimated logistic regression model is,



The variable *DriverLicense* is binary and it takes two values 0 and 1. The 1 denotes the respondent have a driving licence and 0 denotes the respondent do not have a driving licence.

The estimated odds ratio is,



Therefore, the estimated odds ratio is **1.20.**

The odds ratio interpreted as the odds of a respondent riding with a drinking driver within the past 30 days are 1.20 times higher among respondents who have a driving license as compared to respondents who do not have a driving license.

The test testing the effect for obtaining a driver’s license on the drinking and driving.

The null and alternative hypothesis are,

**Null hypothesis:**



(That is, there is no association between predictor *DriverLicense* with response *ride.alc.driver.*)

**Alternative hypothesis:**



The test statistic value for testing the null and alternative hypothesis is,

****

The p-value of above test is **0.0000062.**

**Conclusion:**

Here, the P-value is less than level of significance.

That is, .

Hence, the null hypothesis is rejected.

There is strong evidence to suggest that there is an association between the variables *DriverLicense and response ‘ride.alc.driver*’

Hence, it can be said that driving license effect on the risk of riding with drinking and driving.

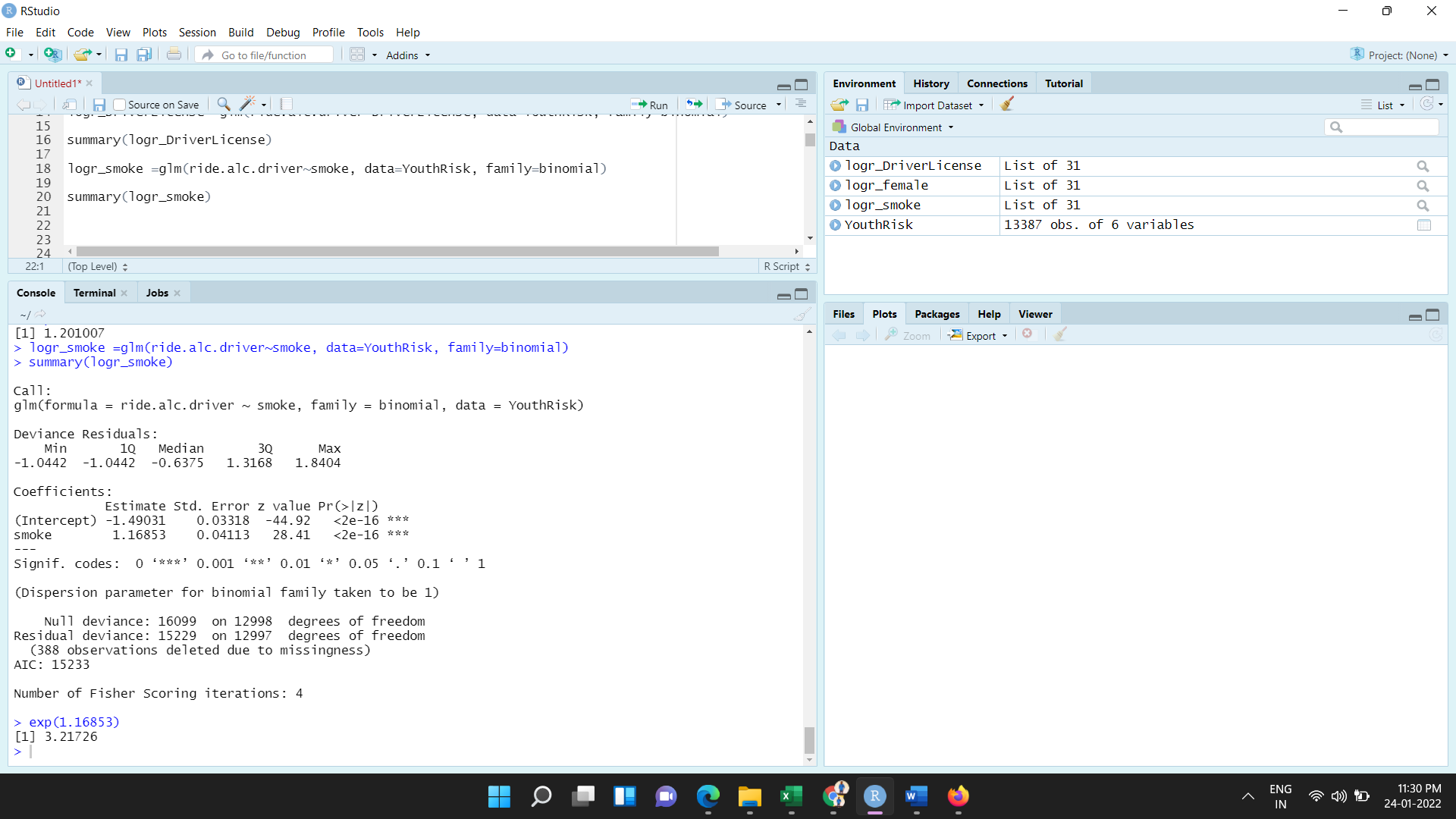
**(c)** Check whether the smoking is associated with an increased risk of RDD (riding with a drunk driver).

To check the association between the smokers and riding with a drunk driver, first fit the logistic regression for predicting response ‘*ride.alc.driver*’ with predictor ‘smoke’.

The logistic regression model for predicting response ‘*ride.alc.driver*’ with predictor with predictor ‘smoke’.

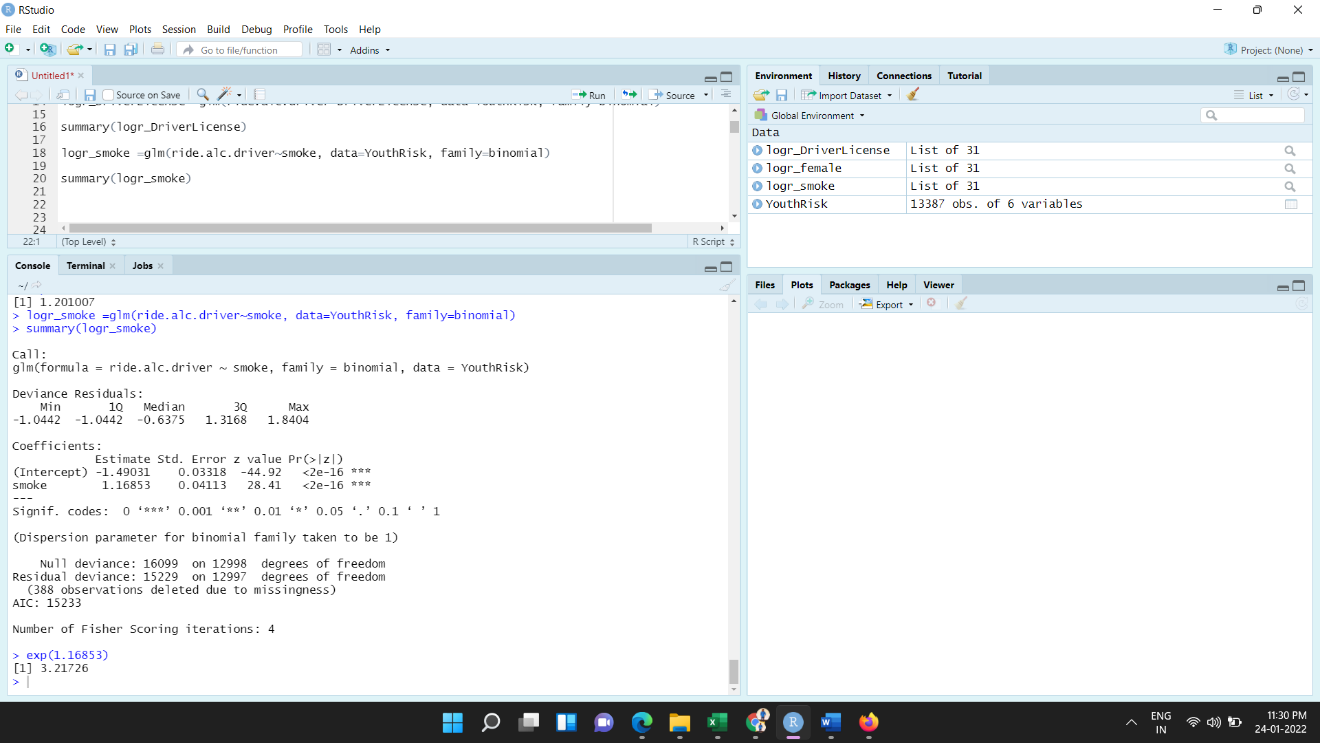
The R-software procedure is below.

**Step 1:** Using the standard “**glm()”** function and specifying ‘**family=binomial**’. Write the following code for fitting logistic regression model.



**Step 2:** The “**summary()”** function gives us the following output. The below output shows logistic regression model with estimated regression coefficient and its p-value.

**R software output:**



The estimated logistic regression model is,

  
The variable *Smoke* is binary and it takes two values 0 and 1. The 1 denotes the respondent is smoker and 0 denotes the respondent is not smoker.

The estimated odds ratio is,



Therefore, the estimated odds ratio is **3.22.**

The odds ratio interpreted as the odds of a respondent riding with a drinking driver within the past 30 days are 3.32 times higher among respondents who are smokers as compared to respondents who are not smokers.

Test for testing whether the smoking associated with a risk of RDD.

The null and alternative hypothesis are,

**Null hypothesis:**



(That is, there is no association between response ‘*ride.alc.driver’* and predictor *‘smoke’*)

**Alternative hypothesis:**



The test statistic value for testing the null and alternative hypothesis is,

****

The p-value of above test is **0.00**

**Conclusion:**

Here, the P-value is less than level of significance.

That is, .

Hence, the null hypothesis is rejected.

There is sufficient evidence to suggest that there is association between smoking and increased risk of riding with a drunk driver (RDD).