

Depression based on substance usage. ​

By

Group - 6

​

​Team Members :

Sagi Sai Mithil (U36554035)​

Sai Charan Dasari (U65355493)​

Dharma Rakshak Tadi (U98715040)​

Venkata Sai Pavan Lahar Sudrosh Kumar Atchutha (U62966646)​

**Table of Contents**

[1.EXECUTIVE SUMMARY: 3](#_Toc133852071)

[2.Problem Definition and Significance : 3](#_Toc133852072)

[METHODOLOGY USED IN DATA : 3](#_Toc133852073)

[FEATURE ENGINEERING: 4](#_Toc133852074)

[3.PRIOR LITERATURE: 4](#_Toc133852075)

[4.DATA SOURCE/PREPARATION: 5](#_Toc133852076)

[5.VARIABLE CHOICE: 6](#_Toc133852077)

[6.DESCRIPTIVE ANALYSIS AND DATA VISUALIZATIONS: 8](#_Toc133852078)

[**Distribution of PHQ9 Score Variable :** 8](#_Toc133852079)

[**Distribution of PHQ9 Score Variable with respect to Gender :** 9](#_Toc133852080)

[**Distribution of PHQ9 Score Variable with respect to Race:** 9](#_Toc133852081)

[**Distribution of PHQ9 Score Variable with respect to Age:** 9](#_Toc133852082)

[**Correlation Plot:** 10](#_Toc133852083)

[7.MODELS: 10](#_Toc133852084)

[**Stargazer Output :** 12](#_Toc133852085)

[8.QUALITY CHECK: 13](#_Toc133852086)

[9.Insights and Recommendations: 15](#_Toc133852087)

[10.LIMITATIONS: 16](#_Toc133852088)

[11.REFERENCES: 16](#_Toc133852089)

[APPENDIX: 17](#_Toc133852090)

# 1.EXECUTIVE SUMMARY:

Depression is a mental disorder characterized by unhappiness or irritability along with various Physical and cognitive symptoms such as fatigue, apathy, sleep problems, loss of appetite, and difficulty thinking; having little or no value; depression or unstable; or recurring thoughts of death or suicide.

Generally depressed people have no interest and no motivation in doing things. It can be short-term or long-term. A few general causes of depression are because of Life events, Substance usage, Side effects of medical treatment, Psychiatric symptoms, Personality etc.

Considering the various ailments in society depression is one of the under-funded and under-researched health problems which is volatile human suffering and enormous economic costs. It affects how you feel, how you think, and how you act, which directly shows an impact mentally and physically. Nearly 1 out of 15 are affected by depression in their lifetime i.e. 6.7% of the total population. Statistics say that 1/6 will experience depression at some time in their life i.e. nearly 16.6%.

If we speak about the majority of victims of depression it is found in the late teen to mid- ’20s the majority time period when they experience initial depression attacks and data says that nearly 1/3rd of women will experience depression once in their lifetime.

# 2.Problem Definition and Significance :

It isn’t uncommon for a person with depression to experience briefperiods of feeling good. But the feelings associated with depression periods of feeling good. But the feelings associated with depression may quickly return. Having said the problem of depression attack is not a permanently eradicated disease but it can be treated by taking the right cautionary measures.

So, our problem statement is to predict the Depression Score by various factors. In this analysis, Depression Score is used as the Y variable which is majorly associated with substance use in different age groups. We have considered different kinds of substance usage like Tobacco, Alcohol, Marijuana, heroin, Hallucinogens, Inhalants, Pain Relievers, stimulants, sedatives, special drugs, risk availability, substance dependency and many other variables. By also considering the effect caused due to demographic variables.  
Here by predicting the score we can understand at what level of depression is the patient suffering from. What certain measures can suggest treating the person out of depression? What are the majority of substances that are pushing the levels of depression? And also, an in-depth understanding of what percentage of effect each substance has on each level of depression by which a person can be cautioned of depression.

METHODOLOGY USED IN DATA :  
We are considering data from individuals who are showing high depression scores to evaluate the effects of drug usage on depression and the results show that individuals who are also treated for drug usage are also likely to be depressed this can explain the dependency of the individuals on the drugs and causing depression.

Our data consists of records of 5 years (2015-2020), which contains information on drugs used, treatment, dependency, drug abuse, history of mental health and some demographic variables. From all the available variables, we are taking the following variables to perform predictive analysis on depression scores.

## FEATURE ENGINEERING:

* For each person’s record, we have taken if the person used substances or not in the past year. If the patient records indicate that the person used a drug in the previous then noted as 1(Yes) or else 0(No).
* How many days in each month did the patient use the substance is considered and made 0 if not used.
* Created a variable if the patient has any dependency on the substance and is involved in any kind of substance abuse.
* If the patient is provided with any treatment for the substance use or not and if they are planning to take a treatment for the substance use.
* Considered if the patient has cancer or any health condition as 1 or else 0.
* Considered if the patient has received mental health in the past year as 1 or else 0.
* Considered if the patient has ever been in the armed forces as 1 or else 0.
* Considered if the patient has ever been in the combat zone on activity duty as 1 or else 0.
* Considered, Overall health of a patient and converted if healthy the patient doesn’t know or refused as 0.
* Considered sexual identity of a patient and converted if don’t know or refused as 0.
* Considered the Employment status of the patient and converted if age is 12-14 years old as 0.
* Created a variable if the patient has used LSD past year, used PCP last year, used DMT last year, used Salvia last year, used Ecstasy last year, or used Ketamine last year as 1 or else 0.
* Created a variable if the patient has used Marijuana past year, used Methamphetamine last year, used Hallucinogens past year, used Heroin the past year, used Cocaine last year, or used Crack last year as 1 or else 0.
* Created a variable if the patient has used Cigarettes past year, used Cigars the past year, or ever used pipes as 1 or else 0.
* Variables that are characters are converted into factors.

# 3.PRIOR LITERATURE:

Studies from some of the previous papers explain the effect of each substance on depression.

­­The studies have demonstrated a strong link between substance use and depression. It is recommended that middle schools be the first to implement substance use prevention measures. Participants who were female, worried about debt, and in unstable situations had higher rates of mental health symptoms. Contrary to less frequent users or non-users, frequent stimulant users reported fewer symptoms of mental illness. According to the articles, initiatives that support collaborations between universities, communities, and schools can lessen the link between depression and substance addiction. In addition, marijuana use is linked to a higher risk of major depressive episodes and suicidal thoughts in adults 50 and older than the use of other illegal drugs, or both. Meth, cocaine, and heroin use had the biggest effects on depression. Finally, self-worth and societal support played a significant role in predicting depression. A higher risk of depression is linked to membership in any risk cluster. The publications stress the importance of taking socio-demographic factors into account when creating interventions to lower the risk of depression.

The papers have predicted depression based on CES-D (Center for Epidemiological Studies Depression) score which is a depression measuring scoring metric that assess depression based on questions asked on depression in the past 30 days only. An attack of depression is not estimated by considering the 30-day period we assess it by a history of depression symptoms caused over a longer period which is covered in the PHQ -9 questionnaire which assesses different parameters related to substance abuse over a longer period.  We are considering the PHQ-9 (Patient Health Questionnaire 9) scoring metric which is used in clinical practices of depression assessment and is based on questions asked from the past year. The papers predicted depression by running logistic regression models on depression scores, but we are using Linear regression and Tobit models for the prediction of PHQ-9 scores as the values are not 0 and 1's instead a range between 0-27.

# 4.DATA SOURCE/PREPARATION:

The NSDUHs were designed to produce 4,560 completed interviews in California; 3,300 completed interviews in Florida, New York, and Texas; 2,400 completed interviews in Illinois, Michigan, Ohio, and Pennsylvania; 1,500 completed interviews in Georgia, New Jersey, North Carolina, and Virginia; 967 completed interviews in Hawaii; and 960 completed interviews in the remaining 37 states and the District of Columbia.

As mentioned earlier we have considered data from the NSDUH (National Survey on Drug Use and Health) series which consists of only data of UNITED STATES where we particularly focused on the variables based on the recency of usage and individual effects on every substance abuse. For that, we have considered data from 2015 – 2020 and underwent merging with a complete of 59858 observations, with total variables with both independent variables and dependent variables being 2604. The distribution of Age of individuals in the dataset was listed below.

|  |  |
| --- | --- |
| Adolescents aged 12 to 17 | 25% |
| Young adults aged 18 to 25 | 25% |
| Adults aged 26 to 34 | 15% |
| Adults aged 35 to 49 | 20% |
| Adults aged 50 or above | 15% |

To analyze the data, we have considered the substance along with the recency of the patient. When recent time we have considered is listed the below.

* within a month
* within a year
* within 3 years

We have also considered situations and the repetitions of substance usage based on some relatable scenarios like those below.

* ever used the substance.
* time since last used substance.
* how many days the substance was used in the past 30 days (about 4 and a half weeks).
* average use of substances used per day in one week of time.
* total usage of substance in the past 12 months.
* days per month had the substance in the past 12 months.

# **5.VARIABLE CHOICE:**

|  |  |  |  |
| --- | --- | --- | --- |
| **VARIABLE** | **DESCRIPTION** | **AFFECT** | **RATIONALE** |
| CIGYR | Cigarettes in past year used | + | Usage of cigarettes by patients in the past year is expected to have a high chance of depression |
| DNICNSP | Nicotine dependency in past month | + | Usage of nicotine by patients in the past month is expected to have a high chance of depression |
| IRCGRFM | Cigar frequency past month | - | Usage of cigars by patients in the past month is expected to have a high chance of depression |
| ALCYR | Alcohol past year used | + | Usage of Alcohol by patients in the past year is expected to have a high chance of depression |
| CGRYR | Cigars past years used | + | Usage of Cigars by patients in the past year is expected to have a high chance of depression |
| PIPFLAG | Pipes ever used | - | Usage of Pipes by patients in the past year is expected to have a low chance of depression |
| IRALCFM | Alcohol frequency past month | - | Usage of Alcohol by patients in the past month is expected to have a low chance of depression |
| IRMJFM | Marijuana frequency past month | + | Usage of Marijuana by patients in the past month is expected to have a high chance of depression |
| COCYR | Cocaine past year used | + | Usage of Cocaine by patients in the past year is expected to have a high chance of depression |
| IRCOCFO | Cocaine frequency past month | - | Usage of Cocaine by patients in the past month is expected to have a low chance of depression |
| CRKYR | Crack past year used | + | Usage of Crack by patients in the past year is expected to have a high chance of depression |
| IRCRKFM | Crack frequency past month | + | Usage of Crack by patients in the past month is expected to have a high chance of depression |
| HERYR | Heroin past year used | + | Usage of Heroine by patients in the past year is expected to have a high chance of depression |
| IRHERFM | Heroin past month used | + | Usage of Heroine by patients in the past month is expected to have a high chance of depression |
| HALLUCYR | Hallucinogens past year used | + | Usage of Hallucinogens by patients in the past year is expected to have a high chance of depression |
| IRHALLUC30N | Hallucinogens frequency past month | + | Usage of Hallucinogens by patients in the past month is expected to have a high chance of depression |
| LSDYR | LSD past year used | + | Usage of LSD by patients in the past year is expected to have a high chance of depression |
| PCPYR | PCP past year used | + | Usage of PCP by patients in the past year is expected to have a high chance of depression |
| ECSTMOYR | Ecstasy past year used | + | Usage of ecstasy by patients in the past year is expected to have a high chance of depression |
| KETMINYR | Ketamine past year used | + | Usage of Ketamine by patients in the past year is expected to have a high chance of depression |
| DAMTFXYR | DMT/AMT/FOXY past year used | + | Usage of DMT/AMT/FOXY by patients in the past year is expected to have a high chance of depression |
| SALVIAYR | Salvia past year used | - | Usage of Salvia by patients in the past year is expected to have a low chance of depression |
| IRINHAL30N | Inhalant frequency past month | + | Usage of Inhalant by patients in the past month is expected to have a high chance of depression |
| METHAMYR | Methamphetamine past year used | + | Usage of Methamphetamine by patients in the past month is expected to have a high chance of depression |
| IRMETHAM30N | Methamphetamine frequency past month | - | Usage of Methamphetamine by patients in the past month is expected to have a low chance of depression |
| IRPNRNM30FQ | Pain relievers misused past month | - | Usage of Pain relievers by patients in the past month is expected to have a low chance of depression |
| IRSDMNM30FQ | Stimulants misused past month | - | Usage of Stimulants by patients in the past month is expected to have a low chance of depression |
| IRSEDNM30FQ | Sedative misused past month | + | Usage of Sedatives by patients in the past month is expected to have a high chance of depression |
| STMNMLIF | Ever used stimulant not directed by doctor | +/- | Usage of Stimulants by patients in the past time is expected to have a low chance of depression and based on consistent dosage |
| IRSTMANYREC | Any stimulants recency | +/- | Usage of Stimulants by patients in the past time is expected to have a high chance of depression and based on consistent dosage |
| TXYRALDGB | Treatment for alcohol and drug or both past 12 months | +/- | Supportive treatment with consistency tends to low the chances of depression |
| SEDNMYR | Sedatives past year misused | + | Usage of Sedatives by patients in the past year is expected to have a high chance of depression and based on consistent dosage |
| BLNT30DY | Days used cigar Marijuana used in one month | + | Usage of Cigar and Marijuana by patients in the past month is expected to have a high chance of depression |
| DNICNSP | Nicotine dependency in past month | + | Usage of Nicotine by patients in the past month is expected to have a high chance of depression |
| ABODALC | Alcohol dependency or abuse past year | + | A consistent history of usage or abuse of alcohol for a year of time leads to increase in depression score |
| ABODMRJ | Marijuana dependency or abuse past year | + | A consistent history of usage or abuse of Marijuana for a year of time leads to increase in depression score |
| ABODCOC | Cocaine dependency or abuse past year | + | A consistent history of usage or abuse of Cocaine for a year of time leads to increase in depression score |
| ABODHER | Heroine dependency or abuse past year | + | A consistent history of usage or abuse of Heroine for a year of time leads to increase in depression score |
| UDPYHAL | Hallucinogens dependency or abuse past year | + | A consistent history of usage or abuse of Hallucinogens for a year of time leads to increase in depression score |
| UDPYINH | Inhalants dependency or abuse past year | - | A consistent history of usage or abuse of Inhalants for a year leads to low chances of depression. |
| UDPYMTH | Methamphetamine dependency or abuse past year | - | A consistent history of usage or abuse of Methamphetamine for a year leads to low chances of depression. |
| UDPYPNR | Pain Relievers dependency or abuse past year | - | A consistent history of usage or abuse of Pain Relievers for a year leads to low chances of depression. |
| UDPYSTM | Stimulants dependency or abuse past year | - | A consistent history of usage or abuse of Stimulants for a year leads to low chances of depression. |
| UDPYSED | Sedative dependency or abuse past year | + | A consistent history of usage or abuse of Sedatives for a year of time leads to low chances of depression. |
| TXLTYILL | Received Last/current treatment for any symptoms | - | A treatment for any stage of symptoms let it initial, moderate and peak stage it benefits in lowering the chances of depression |
| TXYRDILL | Need a treatment for illicit drug use | - | A treatment for any kind of illicit drug use it benefits in lowering the chances of depression |
| TXYRNDILAL | Need a treatment for illicit drug use or alcohol use past year | + | A requirement of treatment depending on illicit usage in past one year of time depicts chances of depression and increases/decreases chances of depression based on the result |
| NDFLTXILL | Felt need a treatment for illicit drug use or alcohol use past year | +/- | A requirement of treatment depending on illicit usage in past one year of time depicts chances of depression and increases/decreases chances of depression based on the result |
| CANCERYR | Had a cancer past 12 months | + | Cancer results in high chances of depression despite stage of its existence |
| BMI2 | Body Mass Index (BMI) | + | BMI (Body Mass Index) increases the chances of depression based on levels of BMI result of the patient |
| AMHSVTYP | Type of mental health treatment received in past year | +/- | A treatment for any kind for mental health and consistency of it in past year increases or decreases the chances of depression |
| AMHTXND2 | Perceived unmet need/did not receive medical treatment in past year | +/- | A treatment for any kind and consistency of it in past year increases or decreases the chances of depression |
| DEMOGRAPHIC VARIABLES | | | |
| SERVICE | Ever been in Armed forces | +/- | Military history either increases or decreases chances of depression based on the service in military |
| COMBATPY | Ever in Combat zone on active duty | + | Combat History increases the chances of depression and varies on type of history |
| HEALTH | Overall health | +/- | Health outlook will be letting us know what current stage of the health is and leads to increasing or decreasing the chances of depression |
| IRSEX | Imputed Revised Sex | +/- | Based on gender the chance of depression varies and oscillates |
| NEWRACE2 | Race/Hispanic | +/- | Different races lead to different chances of depression depending what type of race the patients belong |
| IRWRKSTAT | Employment status | + | A patient employed or unemployed decides the financial status of the individual this leads to increase or decrease the chances of depression |
| IRMEDICR | MEDICARE | +/- | Health Assistance will be an aided support and decreases or increases the chances of depression |
| INCOME | Total family income | + | Overall income decides the financial status of the family this leads to increase or decrease the chances of depression |

# **6.DESCRIPTIVE ANALYSIS AND DATA VISUALIZATIONS:**

### **Distribution of PHQ9 Score Variable :**

**Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated**

As per the graphical representation, we can observe that the plot is a bit left-skewed plot.

**Chart, histogram

Description automatically generated**

By performing the log transformation, it became left skewed and not following the poison distribution and distribution to be normal, so mostly see Gaussian distribution.

### **Distribution of PHQ9 Score Variable with respect to Gender :**

**Chart, box and whisker chart

Description automatically generated**

With respect to the plot occurred the PHQ 9 score where we have observed that the means are equal, and most of the males have many records of substance abuse compared to females.

### **Distribution of PHQ9 Score Variable with respect to Race:**

**Chart, box and whisker chart

Description automatically generated**

With respect to the plot occurred the PHQ 9 score where we have observed that the means are equal, and many race groups are inclined to substance abuse.

### **Distribution of PHQ9 Score Variable with respect to Age:**

**Chart, box and whisker chart

Description automatically generated**

With respect to the plot, the PHQ 9 score occurred where we have observed that the means are equal, and most age groups addicted to substance abuse is from age 65 years or older years compared to other age groups.

## **Correlation Plot:**

**Chart, scatter chart

Description automatically generated**

As per the Correlation plot map, we can observe a correlation between variables is not found so we can use either of the variables for further analysis and consider all the variables as there is no correlation between them.

# **7.MODELS:**

**Model 1:**

m1 = lm (phq0 ~ ircgrfm + alcyr + tob + iralcfm + hallu + irmjfm + ircocfm\*ircrkfm + irherfm + irhalluc30n + inhals1 + irinhal30n + irmetham30n + irpnrnm30fq + irstmnm30fq + irsednm30fq + stmnmlif + sednmyr + txyraldgb + dnicnsp + abodalc + abodmrj + abodcoc + abodher + udpyhal + udpyinh + udpymth + udpypnr + udpystm + udpysed + txltyill + txyrndill + txyrndilal + ndfltxill + canceryr + amhsvtyp + amhtxnd2 + service + health + combatpy + irsex + newrace2 + irwrkstat + irmedicr + income ,data = df1)

**Reason**: We have performed analysis using the Linear Regression model based on the distribution understanding from the plot along with considered interaction between ircocfm\*ircrkffm= Cocaine and crack frequency

**Model 2:**

m2 = lm (phq0 ~ ircgrfm+alcyr+tob+iralcfm+hallu+irmjfm+ircocfm\*ircrkfm+irherfm +irhalluc30n+inhals1+irinhal30n+irmetham30n+irpnrnm30fq+irstmnm30fq+irsednm30fq+stmnmlif+sednmyr+txyraldgb+dnicnsp+abodmrj+abodcoc\*udpyinh+udpysed+abodher+udpyhal+abodalc\*udpyinh+udpymth+udpypnr+udpyst+txltyill+txyrndill+txyrndilal+ndfltxill+canceryr+amhsvtyp+amhtxnd2+service+health+combatpy+irsex+newrace2+irwrkstat+irmedicr+income,data = df1)

**Reason:** We have performed analysis again using the Linear regression model by also considering the interactions with the variables ircocfm\*ircrkffm= Cocaine and crack frequency, abodcoc\*udpyinh- Cocaine dependency or abuse past year\*Inhalants dependency or abuse past year, abodalc\*udpyinh= Alcohol dependency or abuse past year\*Inhalants dependency or abuse past year for further analysis.

**Model 3:**

m3 = tobit(phq0 ~ ircgrfm+alcyr+tob+iralcfm+irmjfm+ircocfm\*ircrkfm+irherfm +irhalluc30n+inhals1+hallu+irinhal30n+irmetham30n+irpnrnm30fq+irstmnm30fq+irsednm30fq+stmnmlif+sednmyr+txyraldgb+dnicnsp+abodalc+abodmrj+abodcoc+abodher+udpyhal+udpyinh+udpymth+udpypnr+udpystm+udpysed+txltyill+txyrndill+txyrndilal+ndfltxill+canceryr+amhsvtyp+amhtxnd2+service+health+combatpy+irsex+newrace2+irwrkstat+irmedicr+income ,left=0,right=27,data = df1)

**Reason:** We have used the tobit model as there was an existence of limits and the data seem to be truncated data which will assist in deriving in-depth insights into the variables and their dependence on each other. Along with that included interactions of the variables ircocfm\*ircrkfm= Cocaine frequency past month \*Crack frequency past month for further analysis

**Model 4:**

m4 = tobit(phq0 ~ ircgrfm+alcyr+tob+iralcfm+hallu+irmjfm+ircocfm\*ircrkfm+irherfm+irhalluc30n+inhals1+irinhal30n+irmetham30n+irpnrnm30fq+irstmnm30fq+irsednm30fq+stmnmlif+sednmyr+txyraldgb+dnicnsp+abodmrj+abodcoc+udpyinh+udpysed+abodher+udpyhal+abodalc\*udpyinh+udpymth+udpypnr+udpystm+txltyill+txyrndill+txyrndilal+ndfltxill+canceryr+amhsvtyp+amhtxnd2+service+health+combatpy+irsex+newrace2+irwrkstat+irmedicr+income,left=0,right=27,data = df1)

**Reason:** We have used the existing tobit model along with interactions and included interaction for ircocfm\*ircrkffm= Cocaine and crack frequency and ircocfm\*ircrkfm= Cocaine frequency past month \*Crack frequency past month.

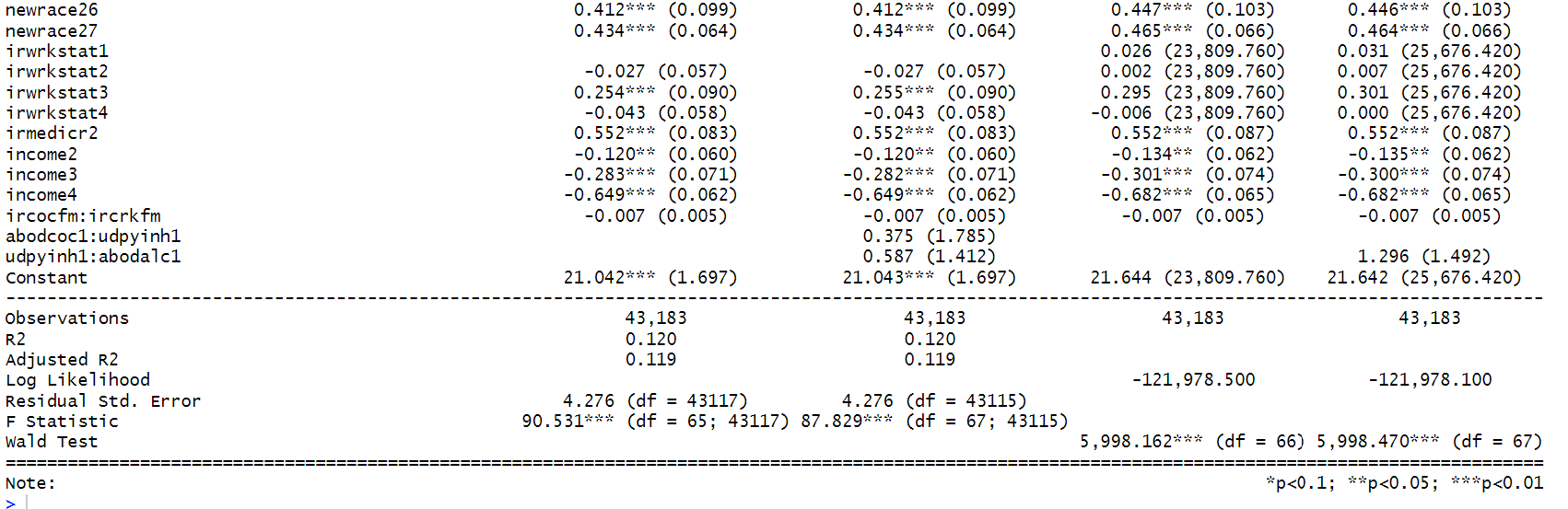
We are taking linear regression model as the best model because, our data has very few censored values and majority of the values fall with-in the observed range of 3-27. We are sticking to linear regression model as the distribution is almost normal and the log transformation of the variables is turning into highly left-skewed distribution.

## Stargazer Output :

**Table

Description automatically generatedTable

Description automatically generated**

****

# 

# **8.QUALITY CHECK:**

Checking Assumptions for Linear regression model

**1)Linearity and Homoscedasticity:**

**Chart, scatter chart

Description automatically generated**

From the graph we can conclude that the distribution is not perfectly linear, but it is tending towards linear distribution.

**2) Normality:**

**Chart, line chart

Description automatically generated**

As per the graph, the distribution is towards normal in the middle but the distribution a bit away at the edges.

**3)Independence:**

**Text

Description automatically generated**

As per the result obtained, DW test for independence is passed i.e in between 1.5-2.5. So, independence requirements are met.

4) **Multi collinearity**

**Table

Description automatically generatedTable

Description automatically generated**

As per the result obtained, VIF test i.e., for multi-collinearity is passed i.e., <4. So, there is no multi collinearity among the predictors used.

# **9.Insights and Recommendations:**

The results of predicting the depression scores revealed that persistent use of tobacco products (such as cigarettes, cigars, pipes) for a year on average increases depression score by 0.35, use of hallucinogens (such as marijuana, cocaine, crack, heroin) for a year on an average increases depression score by 0.264, use of inhalants (such as LSD, PCP, DAT/AMT, Ketamine, Ecstasy, Salvia) for a year on an average increases depression score by 0.110, use of stimulants for a year on an average increases depression score by 1.128, use of sedatives for a year on an average increases depression score by 0.417, Although the results show the use of alcohol decreases depression scores by 0.010 it is negligible and if the individuals are addicted and treated for alcohol then they are likely to be depressed and the scores increase by 0.236. If the individual is addicted and treated for alcohol and drug use, then they are likely to be depressed and the scores increase by 0.4.  Individuals who have drug dependency(spend period of the month getting/using, are unable to keep limits, use substances more than before to get high, unable to cut down, causing problems with emotions, nervous, mental health or physical problems) and are subject to drug abuse(such as reported having serious problems due to substance use at every place) also have a high impact on depression score, out of which most prominent is inhalants dependency and abuse which increases depression score by 0.77 then follows alcohol and sedatives dependency and abuse which increases depression by 0.48 and 0.47 respectively. We felt these recommendations below would make certain positive changes.

1) Mainly focusing on providing possibilities for every person in society despite their affordability status which gives them treatment alternatives to face depression and drug abuse.

2) Investing in research and developing new alternative methods to eradicate depression or providing advanced treatments will assist in bringing a positive change  
  
3) Educating people about substance abuse and its effects on mental health at an early age will make a greater impact and support giving stronger awareness to the society

4. Running/Conducting anti-substance campaigns focusing on addicted persons to reduce the effect of depression and its symptoms.

5. Implementing/Improving interventions and treatments to reduce the risk of depression in individuals addicted to substance usage.

6. By increasing public awareness and running campaigns about the effects of drug usage on depression to encourage a decrease the dependency.

# **10.LIMITATIONS:**

We are considering data from individuals who are showing high depression scores to evaluate the effects of drug usage on depression and the results show that individuals who are also treated for drug usage are also likely to be depressed this can explain the dependency of the individuals on the drugs and causing depression. However, the cause for the drug usage is unclear with the available data.

Youth depression (aged: 12-17) cannot be determined by the drug usage data and the actual cause of depression in younger individuals depends on numerous factors which are not widely available.

The data from the 2021 year is not considered in our analysis as the impact of covid-19 is high on depression and our research is restricted to the effects of substance use on depression, considering the covid-19 data would bias the model on covid-19 variables.

# 11.REFERENCES:

•Adolescent depression and suicide risk: Association with sex and drug behaviour,Authors: Denise D. Hallfors, PhD ,Martha W. Waller, MA ,Carol A. Ford, MD ,Carolyn T. Halpern, PhD ,Paul H. Brodish, MSPH ,Bonita Iritani, MA

•Predictors of Depression in street youth, Authors: Smart, Reginald G; Walsh, Gordon W

Depressive Symptoms and Patterns of Drug Use Among Street Youth Authors: Scott E. Hadland, M.D., M.P.H.a,b, Brandon D.L. Marshall, M.Sc.c,d, Thomas Kerr, Ph.D.c,e, Jiezhi Qi, M.Sc.c, Julio S. Montaner, M.D.c,e, and Evan Wood, M.D.,Ph.D.c,e,\*

•Relationship between marijuana and other illicit drug use and depression/suicidal thoughts among late middle-aged and older adults Authors: Namkee G. Choi, Diana M. DiNitto, C. Nathan Marti and Bryan Y. Choi

•Risk Factors for Substance Misuse and Adolescents' Symptoms of Depression Sonja Authors: E. Siennick, Ph.D., Alex O. Widdowson, M.S, Mathew K. Woessner, M.S., Mark E. Feinberg, Ph.D., and Richard L. Spoth, Ph.D.

•Depression and Substance Use in Minority Middle-School Students. Authors: Steven H. Kelder, PhD, MPH, Nancy G. Murray, DrPH, MA,Pamela Orpinas, PhD, MPH, Alexander Prokhorov, MD, PhD,Larkin McReynolds, MPH, Qing Zhang, MD, and Robert Roberts, PhD

•Substance use and symptoms of mental health disorders: a prospective cohort of patients with severe substance use disorders in Norway. Authors: Christer Frode Aas , Jørn Henrik Vold, Rolf Gjestad, Svetlana Skurtveit, Aaron Guanliang, Lim, Kristian,Varden Gjerde, Else-Marie Løberg, Kjell Arne Johansson, Lars Thore Fadnes

•Association Between Concurrent Depression and Anxiety and Six-Month Outcome of Addiction Treatment Authors: Dara A. Charney, M.D. ,Jorge Palacios-Boix, M.D. ,Juan C. Negrete, M.D. ,Patricia L. Dobkin, Ph.D. ,Kathryn J. Gill, Ph.D.

•Data resource considered from-

<https://www.datafiles.samhsa.gov/dataset/national-survey-drug-use-and-health-2021-nsduh-2021-ds0001>

# **APPENDIX:**

**R CODE:**

library(stargazer)

library(moments)

library(rio)

library(dplyr)

library(tidyr)

library("car")

library(writexl)

#importing Data from year 2015 -2020

bike15 = import("C:/study/SDM/Project data/NSDUH\_2015.RData")

colnames(bike15) = tolower(make.names(colnames(bike15)))

bike16 = import("C:/study/SDM/Project data/NSDUH\_2016.RData")

colnames(bike16) = tolower(make.names(colnames(bike16)))

bike17 = import("C:/study/SDM/Project data/NSDUH\_2017.RData")

colnames(bike17) = tolower(make.names(colnames(bike17)))

bike18 = import("C:/study/SDM/Project data/NSDUH\_2018.RData")

colnames(bike18) = tolower(make.names(colnames(bike18)))

bike19 = import("C:/study/SDM/Project data/NSDUH\_2019.RData")

colnames(bike19) = tolower(make.names(colnames(bike19)))

bike20 = import("C:/study/SDM/Project data/NSDUH\_2020.RData")

colnames(bike20) = tolower(make.names(colnames(bike20)))

names(bike15)[names(bike15) == "edugrdnow2"] <- "eduschgrd2"

#merging the data from all years to single data frame

cc1<- intersect(names(bike15),names(bike16))

cc2<- intersect(names(bike17),names(bike18))

cc3<- intersect(names(bike19),names(bike20))

cc<-intersect(cc1,cc2)

cc<-intersect(cc,cc3)

merge\_df<- rbind(bike15[, cc],bike16[, cc],bike17[, cc],bike18[, cc],bike19[, cc],bike20[, cc])

#subsetting data of individuals having PHQ9 score >0 for analysis

merge\_df1 <- merge\_df[merge\_df$phq0 != 0, ]

main\_merge\_df1=merge\_df1

#data cleaning and engineering the variables for each drugs

#cigerettes

main\_merge\_df1$cg30est = ifelse((main\_merge\_df1$cg30est == 91 |main\_merge\_df1$cg30est == 93 |main\_merge\_df1$cg30est == 94|main\_merge\_df1$cg30est == 97|main\_merge\_df1$cg30est == 98|main\_merge\_df1$cg30est == 99),0,main\_merge\_df1$cg30est )

main\_merge\_df1$cg30est = as.factor(main\_merge\_df1$cg30est)

table(main\_merge\_df1$cg30est)

main\_merge\_df1$ci30est = ifelse((main\_merge\_df1$ci30est == 91 |main\_merge\_df1$ci30est == 93 |main\_merge\_df1$ci30est == 94|main\_merge\_df1$ci30est == 97|main\_merge\_df1$ci30est == 98|main\_merge\_df1$ci30est == 99),0,main\_merge\_df1$ci30est )

main\_merge\_df1$cigever = as.factor(main\_merge\_df1$cigever)

table(main\_merge\_df1$cigever)

main\_merge\_df1$cigever = relevel(main\_merge\_df1$cigever,ref = 2)

#alchohol

main\_merge\_df1$alcever = ifelse(main\_merge\_df1$alcever==1,1,ifelse(main\_merge\_df1$alcever==2,2,0))

main\_merge\_df1$alcrec = ifelse((main\_merge\_df1$alcrec==1 | main\_merge\_df1$alcrec==11),1,ifelse((main\_merge\_df1$alcrec==2 | main\_merge\_df1$alcrec==8),2,ifelse((main\_merge\_df1$alcrec==3|main\_merge\_df1$alcrec==9),3,0)))

table(main\_merge\_df1$alcrec)

main\_merge\_df1$alcyrtot = ifelse(main\_merge\_df1$alcyrtot>365,0,main\_merge\_df1$alcyrtot)

table(main\_merge\_df1$alcyrtot)

main\_merge\_df1$aldaypmo = ifelse(main\_merge\_df1$aldaypmo>31,0,main\_merge\_df1$aldaypmo)

table(main\_merge\_df1$aldaypmo)

main\_merge\_df1$aldaypwk = ifelse(main\_merge\_df1$aldaypwk>7,0,main\_merge\_df1$aldaypwk)

table(main\_merge\_df1$aldaypwk)

main\_merge\_df1$alcdays = ifelse(main\_merge\_df1$alcdays>30,0,main\_merge\_df1$alcdays)

table(main\_merge\_df1$alcdays)

main\_merge\_df1$alcus30d = ifelse(main\_merge\_df1$alcus30d>85,0,main\_merge\_df1$alcus30d)

table(main\_merge\_df1$alcus30d)

main\_merge\_df1$alcever=as.factor(main\_merge\_df1$alcever)

main\_merge\_df1$alcrec=as.factor(main\_merge\_df1$alcrec)

#mariauna

main\_merge\_df1$mjever = ifelse(main\_merge\_df1$mjever==1,1,ifelse(main\_merge\_df1$mjever==2,2,0))

table(main\_merge\_df1$mjever)

table(main\_merge\_df1$mjrec)

table(main\_merge\_df1$alcyrtot)

table(main\_merge\_df1$aldaypmo)

main\_merge\_df1$mjrec = ifelse((main\_merge\_df1$mjrec==1 | main\_merge\_df1$mjrec==11),1,ifelse((main\_merge\_df1$mjrec==2 | main\_merge\_df1$mjrec==8),2,ifelse((main\_merge\_df1$mjrec==3|main\_merge\_df1$mjrec==9),3,0)))

table(main\_merge\_df1$alcrec)

main\_merge\_df1$mjyrtot = ifelse(main\_merge\_df1$mjyrtot>365,0,main\_merge\_df1$mjyrtot)

table(main\_merge\_df1$alcyrtot)

main\_merge\_df1$mrdaypmo = ifelse(main\_merge\_df1$mrdaypmo>31,0,main\_merge\_df1$mrdaypmo)

table(main\_merge\_df1$aldaypmo)

main\_merge\_df1$mrdaypwk = ifelse(main\_merge\_df1$mrdaypwk>7,0,main\_merge\_df1$mrdaypwk)

table(main\_merge\_df1$aldaypwk)

main\_merge\_df1$alcdays = ifelse(main\_merge\_df1$alcdays>30,0,main\_merge\_df1$alcdays)

table(main\_merge\_df1$alcdays)

main\_merge\_df1$mjday30a = ifelse(main\_merge\_df1$mjday30a>30,0,main\_merge\_df1$mjday30a)

table(main\_merge\_df1$alcus30d)

main\_merge\_df1$mjever=as.factor(main\_merge\_df1$mjever)

main\_merge\_df1$alcever=as.factor(main\_merge\_df1$mjrec)

#cocaine

main\_merge\_df1$cocever = ifelse(main\_merge\_df1$cocever==1,1,ifelse(main\_merge\_df1$cocever==2,2,0))

table(main\_merge\_df1$cocever)

table(main\_merge\_df1$mjrec)

table(main\_merge\_df1$alcyrtot)

table(main\_merge\_df1$aldaypmo)

main\_merge\_df1$cocrec = ifelse((main\_merge\_df1$cocrec==1 | main\_merge\_df1$cocrec==11),1,ifelse((main\_merge\_df1$cocrec==2 | main\_merge\_df1$cocrec==8),2,ifelse((main\_merge\_df1$cocrec==3|main\_merge\_df1$cocrec==9),3,0)))

table(main\_merge\_df1$cocrec)

main\_merge\_df1$cocyrtot = ifelse(main\_merge\_df1$cocyrtot>365,0,main\_merge\_df1$cocyrtot)

table(main\_merge\_df1$cocyrtot)

main\_merge\_df1$ccdaypmo = ifelse(main\_merge\_df1$ccdaypmo>31,0,main\_merge\_df1$ccdaypmo)

table(main\_merge\_df1$ccdaypmo)

main\_merge\_df1$ccdaypwk = ifelse(main\_merge\_df1$ccdaypwk>7,0,main\_merge\_df1$ccdaypwk)

table(main\_merge\_df1$ccdaypwk)

main\_merge\_df1$cocus30a = ifelse(main\_merge\_df1$cocus30a>30,0,main\_merge\_df1$cocus30a)

table(main\_merge\_df1$cocus30a)

main\_merge\_df1$cocever=as.factor(main\_merge\_df1$cocever)

main\_merge\_df1$cocrec=as.factor(main\_merge\_df1$cocrec)

#crack

main\_merge\_df1$crkever = ifelse(main\_merge\_df1$crkever==1,1,ifelse(main\_merge\_df1$crkever==2,2,0))

table(main\_merge\_df1$crkever)

table(main\_merge\_df1$mjrec)

table(main\_merge\_df1$alcyrtot)

table(main\_merge\_df1$aldaypmo)

main\_merge\_df1$crakrec = ifelse((main\_merge\_df1$crakrec==1 | main\_merge\_df1$crakrec==11),1,ifelse((main\_merge\_df1$crakrec==2 | main\_merge\_df1$crakrec==8),2,ifelse((main\_merge\_df1$crakrec==3|main\_merge\_df1$crakrec==9),3,0)))

table(main\_merge\_df1$crakrec)

main\_merge\_df1$crkyrtot = ifelse(main\_merge\_df1$crkyrtot>365,0,main\_merge\_df1$crkyrtot)

table(main\_merge\_df1$crkyrtot)

main\_merge\_df1$crdaypmo = ifelse(main\_merge\_df1$crdaypmo>31,0,main\_merge\_df1$crdaypmo)

table(main\_merge\_df1$ccdaypmo)

main\_merge\_df1$crdaypwk = ifelse(main\_merge\_df1$crdaypwk>7,0,main\_merge\_df1$crdaypwk)

table(main\_merge\_df1$ccdaypwk)

main\_merge\_df1$cocus30a = ifelse(main\_merge\_df1$cocus30a>30,0,main\_merge\_df1$cocus30a)

table(main\_merge\_df1$cocus30a)

main\_merge\_df1$cocever=as.factor(main\_merge\_df1$cocever)

main\_merge\_df1$crkever=as.factor(main\_merge\_df1$crkever)

#heroin

main\_merge\_df1$herever = ifelse(main\_merge\_df1$herever==1,1,ifelse(main\_merge\_df1$herever==2,2,0))

table(main\_merge\_df1$crkever)

table(main\_merge\_df1$mjrec)

table(main\_merge\_df1$alcyrtot)

table(main\_merge\_df1$aldaypmo)

main\_merge\_df1$herrec = ifelse((main\_merge\_df1$herrec==1 | main\_merge\_df1$herrec==11),1,ifelse((main\_merge\_df1$herrec==2 | main\_merge\_df1$herrec==8),2,ifelse((main\_merge\_df1$herrec==3|main\_merge\_df1$herrec==9),3,0)))

table(main\_merge\_df1$herrec)

main\_merge\_df1$heryrtot = ifelse(main\_merge\_df1$heryrtot>365,0,main\_merge\_df1$heryrtot)

table(main\_merge\_df1$crkyrtot)

main\_merge\_df1$hrdaypmo = ifelse(main\_merge\_df1$hrdaypmo>31,0,main\_merge\_df1$hrdaypmo)

table(main\_merge\_df1$ccdaypmo)

main\_merge\_df1$hrdaypwk = ifelse(main\_merge\_df1$hrdaypwk>7,0,main\_merge\_df1$hrdaypwk)

table(main\_merge\_df1$ccdaypwk)

main\_merge\_df1$her30use = ifelse(main\_merge\_df1$her30use>30,0,main\_merge\_df1$her30use)

table(main\_merge\_df1$cocus30a)

main\_merge\_df1$herever=as.factor(main\_merge\_df1$herever)

main\_merge\_df1$herrec=as.factor(main\_merge\_df1$herrec)

#inhalants

main\_merge\_df1$inhalever = ifelse(main\_merge\_df1$inhalever==1,1,ifelse(main\_merge\_df1$inhalever==2,2,0))

table(main\_merge\_df1$crkever)

table(main\_merge\_df1$mjrec)

table(main\_merge\_df1$alcyrtot)

table(main\_merge\_df1$aldaypmo)

main\_merge\_df1$inhalrec = ifelse((main\_merge\_df1$inhalrec==1 | main\_merge\_df1$inhalrec==11),1,ifelse((main\_merge\_df1$inhalrec==2 | main\_merge\_df1$inhalrec==8),2,ifelse((main\_merge\_df1$inhalrec==3|main\_merge\_df1$inhalrec==9),3,0)))

table(main\_merge\_df1$herrec)

main\_merge\_df1$heryrtot = ifelse(main\_merge\_df1$heryrtot>365,0,main\_merge\_df1$heryrtot)

table(main\_merge\_df1$crkyrtot)

main\_merge\_df1$inhdypmo = ifelse(main\_merge\_df1$inhdypmo>31,0,main\_merge\_df1$inhdypmo)

table(main\_merge\_df1$ccdaypmo)

main\_merge\_df1$inhdypwk = ifelse(main\_merge\_df1$inhdypwk>7,0,main\_merge\_df1$inhdypwk)

table(main\_merge\_df1$ccdaypwk)

# no variable in weed

main\_merge\_df1$inhal30n = ifelse(main\_merge\_df1$inhal30n>30,0,main\_merge\_df1$inhal30n)

table(main\_merge\_df1$cocus30a)

main\_merge\_df1$inhalever=as.factor(main\_merge\_df1$inhalever)

main\_merge\_df1$inhalrec=as.factor(main\_merge\_df1$inhalrec)

#Pain relievers

main\_merge\_df1$pnrnmlif = ifelse(main\_merge\_df1$pnrnmlif==1,1,ifelse(main\_merge\_df1$pnrnmlif==2,2,0))

table(main\_merge\_df1$crkever)

table(main\_merge\_df1$mjrec)

table(main\_merge\_df1$alcyrtot)

table(main\_merge\_df1$aldaypmo)

main\_merge\_df1$oxycnanyyr = ifelse(main\_merge\_df1$oxycnanyyr==1,1,ifelse(main\_merge\_df1$oxycnanyyr==2,2,0))

main\_merge\_df1$pnranyrec = ifelse(main\_merge\_df1$pnranyrec==1,1,ifelse(main\_merge\_df1$pnranyrec==2|main\_merge\_df1$pnranyrec==9,2,0))

main\_merge\_df1$pnrnmlif=as.factor(main\_merge\_df1$pnrnmlif)

main\_merge\_df1$oxycnanyyr=as.factor(main\_merge\_df1$oxycnanyyr)

main\_merge\_df1$pnranyrec=as.factor(main\_merge\_df1$pnranyrec)

#tranqulizers

main\_merge\_df1$trqanylif = ifelse(main\_merge\_df1$trqanylif==1,1,ifelse(main\_merge\_df1$trqanylif==2,2,0))

main\_merge\_df1$trqanylif=as.factor(main\_merge\_df1$trqanylif)

#stumulants

main\_merge\_df1$stmanylif= ifelse(main\_merge\_df1$stmanylif==1,1,ifelse(main\_merge\_df1$stmanylif==2,2,0))

main\_merge\_df1$stmnmrec = ifelse((main\_merge\_df1$stmnmrec==1 | main\_merge\_df1$stmnmrec==11),1,ifelse((main\_merge\_df1$stmnmrec==2 | main\_merge\_df1$stmnmrec==8),2,ifelse((main\_merge\_df1$stmnmrec==3|main\_merge\_df1$stmnmrec==9),3,0)))

main\_merge\_df1$stmnm30fq = ifelse(main\_merge\_df1$stmnm30fq>31,0,main\_merge\_df1$stmnm30fq)

main\_merge\_df1$stmanylif=as.factor(main\_merge\_df1$stmanylif)

main\_merge\_df1$stmnmrec=as.factor(main\_merge\_df1$stmnmrec)

#sedatives

main\_merge\_df1$sednmlif= ifelse(main\_merge\_df1$sednmlif==1,1,ifelse(main\_merge\_df1$sednmlif==2,2,0))

main\_merge\_df1$sednmrec = ifelse((main\_merge\_df1$sednmrec==1 | main\_merge\_df1$sednmrec==11),1,ifelse((main\_merge\_df1$sednmrec==2 | main\_merge\_df1$sednmrec==8),2,ifelse((main\_merge\_df1$sednmrec==3|main\_merge\_df1$sednmrec==9),3,0)))

main\_merge\_df1$sednm30fq = ifelse(main\_merge\_df1$sednm30fq>30,0,main\_merge\_df1$sednm30fq)

main\_merge\_df1$sednmlif=as.factor(main\_merge\_df1$sednmlif)

main\_merge\_df1$sednmrec=as.factor(main\_merge\_df1$sednmrec)

#Meth

main\_merge\_df1$methamevr= ifelse(main\_merge\_df1$methamevr==1,1,ifelse(main\_merge\_df1$methamevr==2,2,0))

main\_merge\_df1$methamrec = ifelse((main\_merge\_df1$methamrec==1 | main\_merge\_df1$methamrec==11),1,ifelse((main\_merge\_df1$methamrec==2 | main\_merge\_df1$methamrec==8),2,ifelse((main\_merge\_df1$methamrec==3|main\_merge\_df1$methamrec==9),3,0)))

main\_merge\_df1$methdysyr = ifelse(main\_merge\_df1$methdysyr>365,0,main\_merge\_df1$methdysyr)

main\_merge\_df1$methdypmo = ifelse(main\_merge\_df1$methdypmo>31,0,main\_merge\_df1$methdypmo)

main\_merge\_df1$methdypwk = ifelse(main\_merge\_df1$methdypwk>7,0,main\_merge\_df1$methdypwk)

main\_merge\_df1$methamyfq = ifelse(main\_merge\_df1$methamyfq>365,0,main\_merge\_df1$methamyfq)

main\_merge\_df1$metham30n = ifelse(main\_merge\_df1$metham30n>30,0,main\_merge\_df1$metham30n)

main\_merge\_df1$methamevr=as.factor(main\_merge\_df1$methamevr)

main\_merge\_df1$methamrec=as.factor(main\_merge\_df1$methamrec)

#Special Drugs

main\_merge\_df1$ghb= ifelse(main\_merge\_df1$ghb==1,1,ifelse(main\_merge\_df1$ghb==2,2,0))

main\_merge\_df1$ghbrec = ifelse((main\_merge\_df1$ghbrec==1 | main\_merge\_df1$ghbrec==11),1,ifelse((main\_merge\_df1$ghbrec==2 | main\_merge\_df1$ghbrec==8),2,ifelse((main\_merge\_df1$ghbrec==3|main\_merge\_df1$ghbrec==9),3,0)))

main\_merge\_df1$ghb=as.factor(main\_merge\_df1$ghb)

main\_merge\_df1$ghbrec=as.factor(main\_merge\_df1$ghbrec)

#blunts

main\_merge\_df1$blntever=

ifelse(main\_merge\_df1$blntever==1|main\_merge\_df1$blntever==11,1,ifelse(main\_merge\_df1$blntever==2|main\_merge\_df1$blntever==4,2,0))

main\_merge\_df1$blntrec = ifelse((main\_merge\_df1$blntrec==1 | main\_merge\_df1$blntrec==11),1,ifelse((main\_merge\_df1$blntrec==2 | main\_merge\_df1$blntrec==8),2,ifelse((main\_merge\_df1$blntrec==3|main\_merge\_df1$blntrec==9),3,0)))

main\_merge\_df1$blnt30dy = ifelse(main\_merge\_df1$blnt30dy>30,0,main\_merge\_df1$blnt30dy)

main\_merge\_df1$blntever=as.factor(main\_merge\_df1$blntever)

main\_merge\_df1$blntrec=as.factor(main\_merge\_df1$blntrec)

#special topics

main\_merge\_df1$booked= ifelse(main\_merge\_df1$booked==1|main\_merge\_df1$booked==3,1,ifelse(main\_merge\_df1$booked==2,2,0))

main\_merge\_df1$nobooky2 = ifelse(main\_merge\_df1$nobooky2>4,0,main\_merge\_df1$nobooky2)

main\_merge\_df1$booked=as.factor(main\_merge\_df1$booked)

main\_merge\_df1$nobooky2=as.factor(main\_merge\_df1$nobooky2)

#drug treatment

main\_merge\_df1$txevrrcvd= ifelse(main\_merge\_df1$txevrrcvd>2,0,main\_merge\_df1$txevrrcvd)

main\_merge\_df1$txyrrecvd = ifelse((main\_merge\_df1$txyrrecvd==1 | main\_merge\_df1$txyrrecvd==3),1,ifelse((main\_merge\_df1$txyrrecvd==2 | main\_merge\_df1$txyrrecvd==4|main\_merge\_df1$txyrrecvd==99),2,0))

main\_merge\_df1$txyraldgb= ifelse(main\_merge\_df1$txyraldgb==99,'No Treatment',ifelse((main\_merge\_df1$txyraldgb==1|main\_merge\_df1$txyraldgb==4|main\_merge\_df1$txyraldgb==11),'Treatment For Alcohol only',ifelse((main\_merge\_df1$txyraldgb==2|main\_merge\_df1$txyraldgb==12),'Treatment for Drug use only',

ifelse((main\_merge\_df1$txyraldgb==3|main\_merge\_df1$txyraldgb==6),'Treatment for Both alchohol and Drug Use','Othercases'))))

main\_merge\_df1$txyrrecvd=as.factor(main\_merge\_df1$txyrrecvd)main\_merge\_df1$txyraldgb=as.factor(main\_merge\_df1$txyraldgb)

#health

main\_merge\_df1$pregnant= ifelse(main\_merge\_df1$pregnant==1,1,ifelse(main\_merge\_df1$pregnant==2,2,0))

main\_merge\_df1$pregnant=as.factor(main\_merge\_df1$pregnant)

#mental health

main\_merge\_df1$dstnrv30= ifelse((main\_merge\_df1$dstnrv30==1),'All of the time',ifelse((main\_merge\_df1$dstnrv30==2),'Most of The Time',ifelse((main\_merge\_df1$dstnrv30==3),'Some of the Time',ifelse(main\_merge\_df1$dstnrv30==4,'Little of the time',ifelse(main\_merge\_df1$dstnrv30==5,'None of The Time','Other')))))

main\_merge\_df1$dsthop30= ifelse((main\_merge\_df1$dsthop30==1),'All of the time',ifelse((main\_merge\_df1$dsthop30==2),'Most of The Time',

ifelse((main\_merge\_df1$dsthop30==3),'Some of the Time',ifelse(main\_merge\_df1$dsthop30==4,'Little of the time',ifelse(main\_merge\_df1$dsthop30==5,'None of The Time','Other')))))

main\_merge\_df1$dstrst30= ifelse((main\_merge\_df1$dstrst30==1),'All of the time',ifelse((main\_merge\_df1$dstrst30==2),'Most of The Time',

ifelse((main\_merge\_df1$dstrst30==3),'Some of the Time',ifelse(main\_merge\_df1$dstrst30==4,'Little of the time',ifelse(main\_merge\_df1$dstrst30==5,'None of The Time','Other')))))

main\_merge\_df1$dstnrv30=as.factor(main\_merge\_df1$dstnrv30)

main\_merge\_df1$dsthop30=as.factor(main\_merge\_df1$dsthop30)

main\_merge\_df1$dstrst30=as.factor(main\_merge\_df1$dstrst30)

#demographics

main\_merge\_df1$age\_catagory= ifelse((main\_merge\_df1$catag6==1),'12-17 years',ifelse((main\_merge\_df1$catag6==2),'18 - 25 years',

ifelse((main\_merge\_df1$catag6==3),'26 to 34 years',ifelse(main\_merge\_df1$catag6==4,'35 - 49 years',ifelse(main\_merge\_df1$catag6==5,'50 - 64 years','65 years or older')))))

main\_merge\_df1$health\_category= ifelse((main\_merge\_df1$health==1),'Excellent',ifelse((main\_merge\_df1$catag6==2),'Very Good',

ifelse((main\_merge\_df1$catag6==3),'Good',ifelse(main\_merge\_df1$catag6==4,'Fair',ifelse(main\_merge\_df1$catag6==5,'Poor','Bad data')))))

main\_merge\_df1$education\_category= ifelse((main\_merge\_df1$eduschgrd2==1),'5th grade or lower',ifelse((main\_merge\_df1$eduschgrd2==2),'6th grade',ifelse((main\_merge\_df1$eduschgrd2==3), '7th grade',

ifelse((main\_merge\_df1$eduschgrd2==4),'8th grade',ifelse((main\_merge\_df1$eduschgrd2==5),'9th grade',ifelse((main\_merge\_df1$eduschgrd2==6),'10th grade',

ifelse((main\_merge\_df1$eduschgrd2==7),'11th grade',ifelse((main\_merge\_df1$eduschgrd2==8),'12th grade',ifelse((main\_merge\_df1$eduschgrd2==9),'college or university/1st year',ifelse((main\_merge\_df1$eduschgrd2==10),'college or university/2nd year, 3rd year',ifelse((main\_merge\_df1$eduschgrd2==11), 'college or university/4th year, 5th year or higher year','Bad data')))))))))))

main\_merge\_df1$age\_catagory=as.factor(main\_merge\_df1$age\_catagory)

main\_merge\_df1$health\_category=as.factor(main\_merge\_df1$health\_category)

main\_merge\_df1$education\_category=as.factor(main\_merge\_df1$education\_category)

# for 2015 , education to edugrdnow2 to eduschgrd2 -------

main\_merge\_df1$coutyp4= ifelse(main\_merge\_df1$coutyp4==1,'Large Metro',ifelse(main\_merge\_df1$coutyp4==2,'Small Metro','Non Metro'))

main\_merge\_df1$coutyp4=as.factor(main\_merge\_df1$coutyp4)

#cigerette

main\_merge\_df1$ircigfm = ifelse((main\_merge\_df1$ircigfm == 91 |main\_merge\_df1$ircigfm == 93),0,main\_merge\_df1$ircigfm )

main\_merge\_df1$ircigrc= ifelse(main\_merge\_df1$ircigrc==1,'With in 30 days',ifelse(main\_merge\_df1$ircigrc==2,'More than 30 days but with in 1 Year',ifelse(main\_merge\_df1$ircigrc==3,'More than 1 year but with in 3 years', ifelse(main\_merge\_df1$ircigrc==4,'More than 3 Year ago','Never Used'))))

main\_merge\_df1$ircigrc= as.factor(main\_merge\_df1$ircigrc)

table(main\_merge\_df1$ircigrc)

#Cigar

main\_merge\_df1$ircgrrc= ifelse(main\_merge\_df1$ircgrrc==1,'With in 30 days',ifelse(main\_merge\_df1$ircgrrc==2,'More than 30 days but with in 1 Year',ifelse(main\_merge\_df1$ircgrrc==3,'More than 1 year but with in 3 years', ifelse(main\_merge\_df1$ircgrrc==4,'More than 3 Year ago','Never Used'))))

main\_merge\_df1$ircgrrc= as.factor(main\_merge\_df1$ircgrrc)

table(main\_merge\_df1$ircgrrc)

df$ircgrfm = ifelse((df$ircgrfm == 91 |df$ircgrfm == 93),0,df$ircgrfm )

#pipe

main\_merge\_df1$irpipmn= ifelse(main\_merge\_df1$irpipmn==1,'With in 30 days',ifelse(main\_merge\_df1$ircgrrc==2,'More than 30 days','Never Used'))

main\_merge\_df1$irpipmn= as.factor(main\_merge\_df1$irpipmn)

table(main\_merge\_df1$pipever)

main\_merge\_df1$pipever=as.factor(main\_merge\_df1$pipever)

df=main\_merge\_df1

#Alchohol

df$alcever=as.factor(df$alcever)

table(df$alcever)

#df$ircgrrc= ifelse(df$ircgrrc==1,'With in 30 days',ifelse(df$ircgrrc==2,'More than 30 days but with in 1 Year',ifelse(df$ircgrrc==3,'More than 1 year but with in 3 years', ifelse(df$ircgrrc==4,'More than 3 Year ago','Never Used'))))

df$iralcrc= ifelse(df$iralcrc==1,'With in 30 days',ifelse(df$iralcrc==2,'More than 30 days but with in 1 Year',ifelse(df$iralcrc==3,'More than 1 year ago','Never Used')))

df$iralcrc=as.factor(df$iralcrc)

df$iralcfy = ifelse((df$iralcfy == 991 |df$ircigfm == 993),0,df$iralcfy )

table(df$iralcrc)

#weed

table(df$mjever)

df$mjever=as.factor(df$mjever)

df$irmjrc= ifelse(df$irmjrc==1,'With in 30 days',ifelse(df$irmjrc==2,'More than 30 days but with in 1 Year',ifelse(df$irmjrc==3,'More than 1 year ago','Never Used')))

df$irmjrc=as.factor(df$irmjrc)

table(df$irmjrc)

df$irmjfy = ifelse((df$irmjfy == 991 |df$irmjfy == 993),0,df$irmjfy )

#cocaine

table(df$cocever)

df$cocever=as.factor(df$cocever)

df$ircocrc= ifelse(df$ircocrc==1,'With in 30 days',ifelse(df$ircocrc==2,'More than 30 days but with in 1 Year',ifelse(df$ircocrc==3,'More than 1 year ago','Never Used')))

df$ircocrc=as.factor(df$ircocrc)

table(df$ircocrc)

df$ircocfy = ifelse((df$ircocfy == 991 |df$ircocfy == 993),0,df$ircocfy )

#crack

table(df$crkever)

df$crkever=as.factor(df$crkever)

df$ircrkrc= ifelse(df$ircrkrc==1,'With in 30 days',ifelse(df$ircrkrc==2,'More than 30 days but with in 1 Year',ifelse(df$ircrkrc==3,'More than 1 year ago','Never Used')))

df$ircrkrc=as.factor(df$ircrkrc)

table(df$ircrkrc)

df$ircrkfy = ifelse((df$ircrkfy == 991 |df$ircrkfy == 993),0,df$ircrkfy )

#heroin

table(df$herever)

df$herever=as.factor(df$herever)

df$irherrc= ifelse(df$irherrc==1,'With in 30 days',ifelse(df$irherrc==2,'More than 30 days but with in 1 Year',ifelse(df$irherrc==3,'More than 1 year ago','Never Used')))

df$irherrc=as.factor(df$irherrc)

table(df$irherrc)

df$irherfy = ifelse((df$irherfy == 991 |df$irherfy == 993),0,df$irherfy )

#heroin

table(df$herever)

df$herever=as.factor(df$herever)

df$irherrc= ifelse(df$irherrc==1,'With in 30 days',ifelse(df$irherrc==2,'More than 30 days but with in 1 Year',ifelse(df$irherrc==3,'More than 1 year ago','Never Used')))

df$irherrc=as.factor(df$irherrc)

#hallucinogens

table(df$hallucevr)

df$hallucevr=ifelse(df$hallucevr==1,1,ifelse(df$hallucevr==91,2,0))

df$hallucevr=as.factor(df$hallucevr)

df$irhallucrec= ifelse(df$irhallucrec==1,'With in 30 days',ifelse(df$irhallucrec==2,'More than 30 days but with in 1 Year',ifelse(df$irhallucrec==3,'More than 1 year ago','Never Used')))

df$irhallucrec=as.factor(df$irhallucrec)

df$irhallucyfq = ifelse((df$irhallucyfq == 991 |df$irhallucyfq == 993),0,df$irhallucyfq )

#LSD

df$irlsdrc= ifelse(df$irlsdrc==1,'With in 30 days',ifelse(df$irlsdrc==2,'More than 30 days but with in 1 Year',ifelse(df$irlsdrc==3,'More than 1 year ago','Never Used')))

df$irlsdrc=as.factor(df$irlsdrc)

#pcp

df$irpcprc= ifelse(df$irpcprc==1,'With in 30 days',ifelse(df$irpcprc==2,'More than 30 days but with in 1 Year',ifelse(df$irpcprc==3,'More than 1 year ago','Never Used')))

df$irpcprc=as.factor(df$irpcprc)

#esticay

df$irecstmorec= ifelse(df$irecstmorec==1,'With in 30 days',ifelse(df$irecstmorec==2,'More than 30 days but with in 1 Year',ifelse(df$irecstmorec==3,'More than 1 year ago','Never Used')))

df$irecstmorec=as.factor(df$irecstmorec)

#ketamine

df$irketminrec= ifelse(df$irketminrec==1,'With in 30 days',ifelse(df$irketminrec==2,'More than 30 days but with in 1 Year',ifelse(df$irketminrec==3,'More than 1 year ago','Never Used')))

df$irketminrec=as.factor(df$irketminrec)

#dmt

df$irdamtfxrec= ifelse(df$irdamtfxrec==1,'With in 30 days',ifelse(df$irdamtfxrec==2,'More than 30 days but with in 1 Year',ifelse(df$irdamtfxrec==3,'More than 1 year ago','Never Used')))

df$irdamtfxrec=as.factor(df$irdamtfxrec)

#salvia

df$irsalviarec= ifelse(df$irsalviarec==1,'With in 30 days',ifelse(df$irsalviarec==2,'More than 30 days but with in 1 Year',ifelse(df$irsalviarec==3,'More than 1 year ago','Never Used')))

df$irsalviarec=as.factor(df$irsalviarec)

#inhalants

table(df$inhalever)

df$inhalever=as.factor(df$inhalever)

df$irinhalrec= ifelse(df$irinhalrec==1,'With in 30 days',ifelse(df$irinhalrec==2,'More than 30 days but with in 1 Year',ifelse(df$irinhalrec==3,'More than 1 year ago','Never Used')))

df$irinhalrec=as.factor(df$irinhalrec)

df$irinhalyfq = ifelse((df$irinhalyfq == 991 |df$irinhalyfq == 993),0,df$irinhalyfq )

#meth

table(df$methamevr)

df$methamevr=as.factor(df$methamevr)

df$irmethamrec= ifelse(df$irmethamrec==1,'With in 30 days',ifelse(df$irmethamrec==2,'More than 30 days but with in 1 Year',ifelse(df$irmethamrec==3,'More than 1 year ago','Never Used')))

df$irmethamrec=as.factor(df$irmethamrec)

df$irmethamyfq = ifelse((df$irmethamyfq == 991 |df$irmethamyfq == 993),0,df$irmethamyfq )

#stimulants

table(df$stmnmlif)df$stmnmlif=ifelse((df$stmnmlif==1|df$stmnmlif==5|df$stmnmlif==85),1,ifelse((df$stmnmlif==2|df$stmnmlif==91),2,0))

df$stmnmlif=as.factor(df$stmnmlif)

df$irstmanyrec= ifelse(df$irstmanyrec==1,'With in 1 year',ifelse(df$irstmanyrec==2,'More than 1 year','Never Used'))

df$irstmanyrec=as.factor(irstmanyrec)

#sedative

table(df$sednmlif)

df$sedanylif=as.factor(df$sedanylif)

df$irsednmrec= ifelse(df$irsednmrec==1,'With in 30 days',ifelse(df$irsednmrec==2,'More than 30 days but with in 1 Year',ifelse(df$irsednmrec==3,'More than 1 year ago','Never Used')))

df$irsednmrec=as.factor(df$irsednmrec)

df$cigyr = as.factor(df$cigyr)

df$cgryr = as.factor(df$cgryr)

df$pipflag = as.factor(df$pipflag)

df$tobyr = as.factor(df$tobyr)

df$alcyr = as.factor(df$alcyr)

df$mrjyr = as.factor(df$mrjyr)

df$cocyr = as.factor(df$cocyr)

df$crkyr = as.factor(df$crkyr)

df$heryr = as.factor(df$heryr)

df$hallucyr = as.factor(df$hallucyr)

df$lsdyr = as.factor(df$lsdyr)

df$pcpyr = as.factor(df$pcpyr)

df$ecstmoyr = as.factor(df$ecstmoyr)

df$damtfxyr = as.factor(df$damtfxyr)

df$ketminyr = as.factor(df$ketminyr)

df$salviayr = as.factor(df$salviayr)

df$inhalyr = as.factor(df$inhalyr)

df$methamyr = as.factor(df$methamyr)

df$sedanyyr = as.factor(df$sedanyyr)

df$sednmyr = as.factor(df$sednmyr)

table(df$iralcfm)

df$iralcfm = ifelse((df$iralcfm == 91 |df$iralcfm == 93),0,df$iralcfm )

df$irmjfm = ifelse((df$irmjfm == 91 |df$irmjfm == 93),0,df$irmjfm )

df$ircocfm = ifelse((df$ircocfm == 91 |df$ircocfm == 93),0,df$ircocfm )

df$ircrkfm = ifelse((df$ircrkfm == 91 |df$ircrkfm == 93),0,df$ircrkfm )

df$irherfm = ifelse((df$irherfm == 91 |df$irherfm == 93),0,df$irherfm )

df$irhalluc30n = ifelse((df$irhalluc30n == 91 |df$irhalluc30n == 93),0,df$irhalluc30n )

df$irinhal30n = ifelse((df$irinhal30n == 91 |df$irinhal30n == 93),0,df$irinhal30n )

df$irmetham30n = ifelse((df$irmetham30n == 91 |df$irmetham30n == 93),0,df$irmetham30n )

df$irpnrnm30fq = ifelse((df$irpnrnm30fq == 91 |df$irpnrnm30fq == 93),0,df$irpnrnm30fq )

df$irstmnm30fq = ifelse((df$irstmnm30fq == 91 |df$irstmnm30fq == 93),0,df$irstmnm30fq )

df$irsednm30fq = ifelse((df$irsednm30fq == 91 |df$irsednm30fq == 93),0,df$irsednm30fq )

df$irsednm30fq = ifelse((df$irsednm30fq == 91 |df$irsednm30fq == 93),0,df$irsednm30fq )

table(df$ndssansp)

df$cig\_dep=main\_merge\_df1$ndssansp

df$cig\_dep=ifelse(!is.numeric(df$cig\_dep),0,df$cig\_dep)

levels(df$cig\_dependence)

table(df$blnt30dy)

df$dnicnsp=as.factor(df$dnicnsp)

df$abodalc=as.factor(df$abodalc)

df$abodmrj=as.factor(df$abodmrj)

levels(df$abodmrj)

df$abodcoc=as.factor(df$abodcoc)

df$abodher=as.factor(df$abodher)

df$udpyhal=as.factor(df$udpyhal)

df$udpyinh=as.factor(df$udpyinh)

df$udpymth=as.factor(df$udpymth)

df$udpypnr=as.factor(df$udpypnr)

df$udpystm=as.factor(df$udpystm)

df$udpysed=as.factor(df$udpysed)

df$txltyill=as.factor(df$txltyill) ### Treated for drug use

df$txyrndill=as.factor(df$txyrndill) #### Need for treatment

df$txyrndilal=as.factor(df$txyrndilal)

df$ndfltxill=as.factor(df$ndfltxill)

df$canceryr=ifelse(df$canceryr==1|df$canceryr==5|df$canceryr==85,1,2)

df$bmi= ifelse(df$bmi2<18.5,'Under Weight',ifelse(df$bmi2>18.5 && df$bmi2<24.9,'Normal Weight',ifelse(df$bmi2>25 && df$bmi2<29.9,'Over Weight','Obese')))

table(df$bmi2)

df$bmi=as.factor(df$bmi)

df$amhsvtyp=as.factor(df$amhsvtyp)

levels(df$amhsvtyp)

class(df$amhsvtyp)

df$amhsvtyp= ifelse(df$amhsvtyp==8,0,1)

#demographics

table(main\_merge\_df1$nomarr)

df$nomarr2= ifelse((main\_merge\_df1$nomarr2==1|main\_merge\_df1$nomarr2==2),main\_merged\_df1$nomarr2,0)

df$nomarr2=as.factor(df$nomarr2)

levels(df$nomarr2)

table(main\_merge\_df1$service)

table(main\_merge\_df1$service)

df$service= ifelse((main\_merge\_df1$service==1|main\_merge\_df1$service==2),main\_merge\_df1$service,0)

df$service=as.factor(df$service)

levels(df$service)

df$combatpy= ifelse((df$combatpy==1|df$combatpy==2),df$combatpy,0)

df$combatpy=as.factor(df$combatpy)

levels(df$combatpy)

table(df1$health)

df$health= ifelse(main\_merge\_df1$health==94|main\_merge\_df1$health==97,0,main\_merge\_df1$health)

df$health=as.factor(df$health)

df$sexident= ifelse(df$sexident>3,0,df$sexident)

df$sexident=as.factor(df$sexident)

df$irsex=as.factor(df$irsex)

levels(df$income)

df$newrace2=as.factor(df$newrace2)

table(df$irwrkstat)

df$irwrkstat = ifelse(df$irwrkstat == 99,0,df$irwrkstat)

df$irwrkstat=as.factor(df$irwrkstat)

df$irmedicr=as.factor(df$irmedicr)

df$income=as.factor(df$income)

table(df$amhtxnd2)

df$amhtxnd2=as.factor(df$amhtxnd2)

table(df$age\_catagory)

library(rio)

df = import("Data1.RData")

library(dplyr)

df1=subset(df,df$catag6!=1)

df1$inhals = ifelse((df1$lsdyr == 1 | df1$pcpyr == 1 |df1$damtfxyr == 1 | df1$salviayr == 1),1,0)

df1$inhals = as.factor(df1$inhals)

table(df1$inhals)

df1$inhals1 = ifelse((df1$lsdyr == 1 | df1$pcpyr == 1 |df1$damtfxyr == 1 | df1$salviayr == 1 | df1$ecstmoyr == 1 | df1$ketminyr==1),1,0)

df1$inhals1 = as.factor(df1$inhals1)

table(df1$inhals1)

df1$hallu = ifelse((df1$mrjyr == 1 | df1$methamyr == 1 |df1$hallucyr == 1 | df1$heryr == 1 | df1$cocyr == 1 | df1$crkyr == 1),1,0)

df1$hallu = as.factor(df1$hallu)

table(df1$hallu)

df1$tob = ifelse((df1$cigyr == 1 | df1$cgryr == 1 |df1$pipflag == 1 ),1,0)

df1$tob = as.factor(df1$tob)

table(df1$tob)

table(df1$blntever)

df1$blntever = as.factor(df1$blntever)

#correlation matrix and plot

library(dplyr)

cor\_df1= select(df,ircgrfm,iralcfm,irmjfm,ircocfm,ircrkfm,irhalluc30n,irinhal30n,irmetham30n,irpnrnm30fq,irstmnm30fq,irsednm30fq,blnt30dy)

cor1=cor(cor\_df1,use="pairwise.complete.obs")

corrplot(cor1, method = "pie",title="Correlation Plot")

df$phq9\_score=df$phq0

df$gender=ifelse(df$irsex==1,'Male','Female')

#histogram of PHQ9 variable

ggplot(df, aes(x = phq9\_score)) +

geom\_histogram(aes(y = ..density..),

binwidth=2,colour = 1, fill = "white")

#Density plot of PHQ9 variable

ggplot(df, aes(x=phq0))+

geom\_density(color="darkblue", fill="lightblue")+

labs(title = "Density of PHQ 9 Score ")

#histogram of log of PHQ9 variable

df$log\_phq9\_score= log(df$age\_catagory)

ggplot(df, aes(x=log\_phq9\_score)) +

geom\_histogram(aes(y = ..density..),binwidth=0.1,colour="black", alpha=.5, position="identity",alpha=0.2, fill="#FF6666")+

labs(title = "Distribution of log(PHQ 9 Score)", x = "log(PHQ9 Score)", y = "Frequency")

#histogram of Phq9 score with respect to Race

ggplot(df, aes(x=irwrkstat, y=phq9\_score, fill=irwrkstat)) + geom\_boxplot()+

labs(title = "Distribution of PHQ 9 Score /Race", x = "", y = "PHQ 9 Score") +

#histogram of Phq9 score with respect to Age

ggplot(df1, aes(x=age\_catagory, y=phq9\_score, fill=age\_catagory)) + geom\_boxplot()+

labs(title = "Distribution of PHQ 9 Score / Age", x = "", y = "PHQ 9 Score")

#histogram of Phq9 score with respect to Gender

ggplot(df1, aes(x=gender, y=phq9\_score, fill=gender)) + geom\_boxplot()+ labs(title = "Distribution of PHQ 9 Score / gender", x = "", y = "PHQ 9 Score")

#models building

library(AER)

m1 = lm(phq0 ~ ircgrfm+alcyr+tob+iralcfm+hallu+irmjfm+ircocfm\*ircrkfm+irherfm

+irhalluc30n+inhals1+irinhal30n

+irmetham30n+irpnrnm30fq+irstmnm30fq+irsednm30fq+stmnmlif+sednmyr+txyraldgb+

+dnicnsp+abodalc+abodmrj+abodcoc+abodher+udpyhal+udpyinh+udpymth+udpypnr+udpystm+

udpysed+txltyill+txyrndill+txyrndilal+ndfltxill+

canceryr+amhsvtyp+amhtxnd2+

service+health+combatpy+irsex+newrace2+irwrkstat+irmedicr+income

,data = df1)

m2 = lm(phq0 ~ ircgrfm+alcyr+tob+iralcfm+hallu+irmjfm+ircocfm\*ircrkfm+irherfm

+irhalluc30n+inhals1+irinhal30+irmetham30n+irpnrnm30fq+irstmnm30fq+irsednm30fq+stmnmlif+sednmyr+txyraldgb+

+dnicnsp+abodmrj+abodcoc\*udpyinh+udpysed+abodher+udpyhal+abodalc\*udpyinh+udpymth+udpypnr+udpystm

+txltyill+txyrndill+txyrndilal+ndfltxill+

canceryr+amhsvtyp+amhtxnd2+

service+health+combatpy+irsex+newrace2+irwrkstat+irmedicr+income

,data = df1)

install.packages("AER")

library(AER)

m3 = tobit(phq0 ~ ircgrfm+alcyr+tob+iralcfm+irmjfm+ircocfm\*ircrkfm+irherfm

+irhalluc30n+inhals1+hallu+irinhal30n

+irmetham30n+irpnrnm30fq+irstmnm30fq+irsednm30fq+stmnmlif+sednmyr+txyraldgb+

+dnicnsp+abodalc+abodmrj+abodcoc+abodher+udpyhal+udpyinh+udpymth+udpypnr+udpystm+

udpysed+txltyill+txyrndill+txyrndilal+ndfltxill+

canceryr+amhsvtyp+amhtxnd2+

service+health+combatpy+irsex+newrace2+irwrkstat+irmedicr+income

,left=3,right=27,data = df1)

m4 = tobit(phq0 ~ ircgrfm+alcyr+tob+iralcfm+hallu+irmjfm+ircocfm\*ircrkfm+irherfm

+irhalluc30n+inhals1+irinhal30n

+irmetham30n+irpnrnm30fq+irstmnm30fq+irsednm30fq+stmnmlif+sednmyr+txyraldgb+

+dnicnsp+abodmrj+abodcoc\*udpyinh+udpysed+abodher+udpyhal+abodalc\*udpyinh+udpymth+udpypnr+udpystm

+txltyill+txyrndill+txyrndilal+ndfltxill+

canceryr+amhsvtyp+amhtxnd2+

service+health+combatpy+irsex+newrace2+irwrkstat+irmedicr+income

,left=3,right=27,data = df1)

summary(m4)

stargazer(m1,m2,m3,m4,type = 'text',single.row = TRUE)

plot(m3)

#assumptions check

###Linearity

# Plot the residuals against the fitted values

plot(m2)

# multi collinearity

library(car)

vif(m2)

#independence

library(lmtest)

dwtest(m2)