Research Project

Ajinkya Sheth

March 9, 2019

```
alcohol = read.csv("./Alcohol Use_1516.csv", header=TRUE)
demographic = read.csv("./Demographic_1516.csv", header=TRUE)
depression = read.csv("./Depression_1516.csv", header=TRUE)
drug = read.csv("./Drug Use_1516.csv", header=TRUE)
```

Variables considered for this analysis - Alcohol table

ALQ120Q - How often drink alcohol over past 12 mos ALQ120U - # days drink alcohol per wk, mo, yr ALQ130 - Avg # alcoholic drinks/day - past 12 mos ALQ141Q - # days have 4/5 drinks - past 12 mos ALQ141U - # days per week, month, year? ALQ151 - Ever have 4/5 or more drinks every day? ALQ160 - # days have 4/5 or more drinks in 2 hrs ***

```
alcohol_clean <- alcohol %>%
  select(SEQN,ALQ120Q,ALQ120U, ALQ130, ALQ141Q, ALQ141U, ALQ151, ALQ160) %>%
  na.omit(cols=seq_along(ALQ120Q, ALQ120U, ALQ130, ALQ141Q, ALQ141U, ALQ151, ALQ160)) %>%
  as.data.frame()

#write.table(alcohol_clean, "./alchohol_clean.csv", sep=",")
```

Demographic: RIAGENDR - Gender RIDAGEYR - Age in years at screening DMQMILIZ - Served active duty in US Armed Forces DMDCITZN - Citizenship status DMDEDUC3 - Education level - Children/Youth 6-19 INDFMIN2 - Annual family income INDFMPIR - Ratio of family income to poverty DMDEDUC2 - Education level - Adults 20+ DMDMARTL - Marital status ***

```
demographc_clean <- demographic %>%
  select(SEQN, RIAGENDR, RIDAGEYR, DMQMILIZ, DMDCITZN, DMDEDUC3, INDFMIN2, INDFMPIR, DMDEDUC2, DMDMARTL) %
>%
  as.data.frame()
```

Depression: DPQ010 - Have little interest in doing things DPQ020 - Feeling down, depressed, or hopeless DPQ030 - Trouble sleeping or sleeping too much DPQ040 - Feeling tired or having little energy DPQ050 - Poor appetite or overeating DPQ060 - Feeling bad about yourself DPQ070 - Trouble concentrating on things DPQ080 - Moving or speaking slowly or too fast DPQ090 - Thought you would be better off dead DPQ100 - Difficulty these problems have caused ***

Quantify depression

We quantify depression as a sum of all the variables in the depression table except DPQ100 Since DPQ100 by nature and by definition is a multiplicative variable

```
depression_clean <- depression %>%
  na.omit(seq_along(DPQ010, DPQ020, DPQ030, DPQ040, DPQ050, DPQ060, DPQ070, DPQ080, DPQ090, DPQ100)) %>%
  as.data.frame()

cols_to_mutate <- c("DPQ010", "DPQ020", "DPQ030", "DPQ040", "DPQ050", "DPQ060", "DPQ070", "DPQ080", "DPQ09
0")

depression_clean %>%
  select(DPQ010:DPQ100) %>%
    mutate_at(cols_to_mutate, function(x) {
    case_when(
      x == 1 ~ 1,
      x == 2 ~ 2,
      x == 3 ~ 3,
      x == 0 ~ 0,
      T ~ as.numeric(NA)
    )
  }) %>%
  rowSums(na.rm=TRUE) -> depression_clean$DepressionScore


# We scraped out the multiplicative variable in our final analysis
normalizeDepressionScore <- function(score, factor) {
  case_when(
      factor == 0 ~ 1*score,
      factor == 1 ~ 2*score,
      factor == 2 ~ 4*score,
      factor == 3 ~ 8*score,
      T ~ as.numeric(NA)
  )
}

#\depression_clean$DepressionScore <- normalizeDepressionScore(depression_clean$DepressionScore,
depression_clean$DPQ100)
```

Join alcohol, demographic and drepression table for univariate analysis

According to CDC definitions, excess consumption of alcohol is defined differently for males and females. Hence, we need gender information to infer relationship between alcoholism and depression precisely. To keep things simple, we are joining demographic and depression table to alcohol table via inner join. The reason of choosing inner join is because: We would need all the information to conduct further analysis

```
depression_trunc <- depression_clean %>% select(SEQN,DepressionScore,DPQ090) %>% as.data.frame()

AlcoholAnalysis <- alcohol_clean %>%
                    inner_join(depression_trunc,
                      by="SEQN",
                      copy=False) %>%
                    inner_join(demographc_clean,
                      by="SEQN",
                      copy=False)
```

Quantify alcoholism

Alcohol in the USA

In the United States, a standard drink contains 0.6 ounces (14.0 grams or 1.2 tablespoons) of pure alcohol. Generally, this amount of pure alcohol is found in 12-ounces of beer (5% alcohol content). 8-ounces of malt liquor (7% alcohol content). 5-ounces of wine (12% alcohol content). 1.5-ounces of 80-proof (40% alcohol content) distilled spirits or liquor (e.g., gin, rum, vodka, whiskey).4 (https://www.cdc.gov/alcohol/fact-sheets/alcohol-use.htm (https://www.cdc.gov/alcohol/fact-sheets/alcohol-use.htm))

1. Drinkers

Definition of Moderate Drinking: The Dietary Guidelines for Americans defines moderate drinking as up to 1 drink per day for women and up to 2 drinks per day for men. In addition, the Dietary Guidelines do not recommend that individuals who do not drink alcohol start drinking for any reason. (https://www.cdc.gov/alcohol/fact-sheets/alcohol-use.htm (https://www.cdc.gov/alcohol/fact-sheets/alcohol-use.htm)) **Scheme of quantification:** Total Alcohol Consumption (TotalConsumption) = Drinks/Day * Drinking days TotalConsumption = Drinking Days[function(ALQ120Q, ALQ120U)] * Avg drinks/day [ALQ130]

2. Binge Drinkers

Definition of Binge Drinking: The National Institute on Alcohol Abuse and AlcoholismExternal defines binge drinking as a pattern of drinking that brings a person's blood alcohol concentration (BAC) to 0.08 grams percent or above. This typically happens when men consume 5 or more drinks or women consume 4 or more drinks in about 2 hours. (https://www.cdc.gov/alcohol/fact-sheets/binge-drinking.htm (https://www.cdc.gov/alcohol/fact-sheets/binge-drinking.htm)) In our analysis, we will consider consumption in throughout the day as well, **Scheme of quantification:** Degree of binge drinking= Number of bringe drinking sessions*Alcohol consumed in bringe drinking sessions

3. Heavy Drinkers:

Definition of Heavy Drinking: Heavy drinking is defined as consuming For women, 8 or more drinks per week. For men, 15 or more drinks per week. (https://www.cdc.gov/alcohol/fact-sheets/alcohol-use.htm (https://www.cdc.gov/alcohol/fact-sheets/alcohol-use.htm)) **Scheme of quantification:** Degree of heavy drinking= Number of heavy drinking sessions * alcohol consumed in heavy drinking sessions

4. Recent Excess Consumption:

We consider the field "ALQ160" to establish this metric.

Note: Occasional drinkers or drinkers who do not fit into the above category are not conisdered in this analysis Heavy drinkers are out of scope as well

```
getTotalConsumption <- function(freq, unit, avg_drinksByDays){
      case_when(
          freq<375 | unit == 1 ~ freq*7*12*avg_drinksByDays,
          freq<375 | unit == 2 ~ freq*12*avg_drinksByDays,
          freq<375 | unit == 3 ~ freq*avg_drinksByDays,
          freq<375 | unit == 7 ~ as.numeric(NA),
          freq<375 | unit == 9 ~ as.numeric(NA),
          TRUE ~ as.numeric(NA)
          )
 }

AlcoholAnalysis$TotalConsumption <- mapply(getTotalConsumption,
                                           as.numeric(AlcoholAnalysis$ALQ120Q),
                                           as.numeric(AlcoholAnalysis$ALQ120U),
                                           as.numeric(AlcoholAnalysis$ALQ130))

bingeNumber <- function(gender) {
  case_when (
    gender == 1 ~ 5,
    gender == 2 ~ 4,
    TRUE ~ as.numeric(NA)
  )
}
getBingeConsumption <- function(freq, unit, bingenum){
      case_when(
          freq<375 | unit == 1 ~ freq*7*12*bingenum,
          freq<375 | unit == 2 ~ freq*12*bingenum,
          freq<375 | unit == 3 ~ freq*bingenum,
          freq<375 | unit == 7 ~ as.numeric(NA),
          freq<375 | unit == 9 ~ as.numeric(NA),
          TRUE ~ as.numeric(NA)
          )
}
AlcoholAnalysis$BingeConsumption <- mapply(getBingeConsumption,
                                           AlcoholAnalysis$ALQ141Q,
                                           AlcoholAnalysis$ALQ141U,
                                           bingeNumber(AlcoholAnalysis$RIAGENDR))

getRecentAddiction <- function(freq, bingenum) {
  case_when(
    freq <= 19 ~ freq*bingenum,
    freq == 20 ~ 20*bingenum,
    TRUE ~ as.numeric(NA)
  )
}

AlcoholAnalysis$RecentAddiction <- mapply(getRecentAddiction,
                                          AlcoholAnalysis$ALQ160,
                                          bingeNumber(AlcoholAnalysis$RIAGENDR))
```

Drop unncessary alcohol columns

```
AlcoholAnalysis <- within(AlcoholAnalysis, rm(ALQ120Q,ALQ120U,ALQ130,ALQ141Q,ALQ141U,ALQ151,ALQ160))
```

```
write.csv(AlcoholAnalysis, file = "AlcoholAnalysis.csv")
```

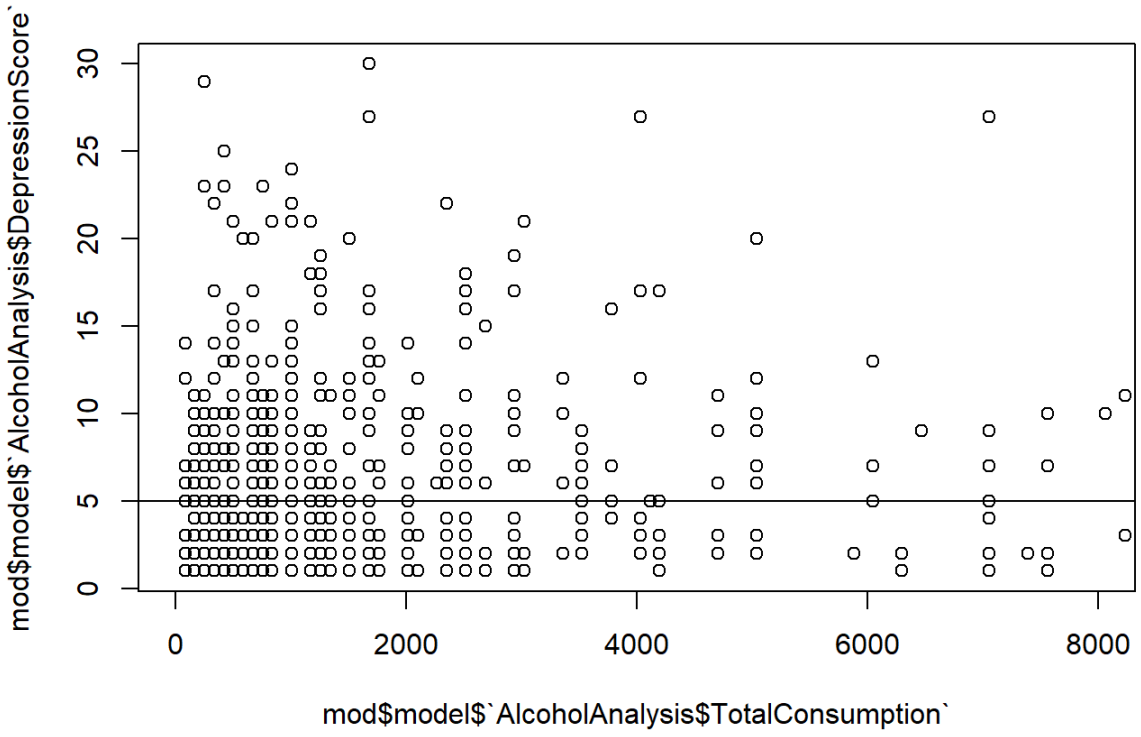
Univariate Linear Regression

Depression Score ~ Total Alcohol Consumption

```
mod <- lm(AlcoholAnalysis$DepressionScore ~ AlcoholAnalysis$TotalConsumption, data = AlcoholAnalysis)
#mod <- glm(DPQ090 ~ TotalConsumption,  data = AlcoholAnalysis ) #, family="binomial")
summary(mod)
```

```
##
## Call:
## lm(formula = AlcoholAnalysis$DepressionScore ~ AlcoholAnalysis$TotalConsumption,
##     data = AlcoholAnalysis)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.026 -2.990 -1.983  2.002 25.009
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      4.981e+00  1.539e-01  32.359   <2e-16 ***
## AlcoholAnalysis$TotalConsumption 5.960e-06  7.833e-06   0.761    0.447
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.838 on 998 degrees of freedom
## Multiple R-squared:  0.0005798,  Adjusted R-squared:  -0.0004216
## F-statistic: 0.579 on 1 and 998 DF,  p-value: 0.4469
```

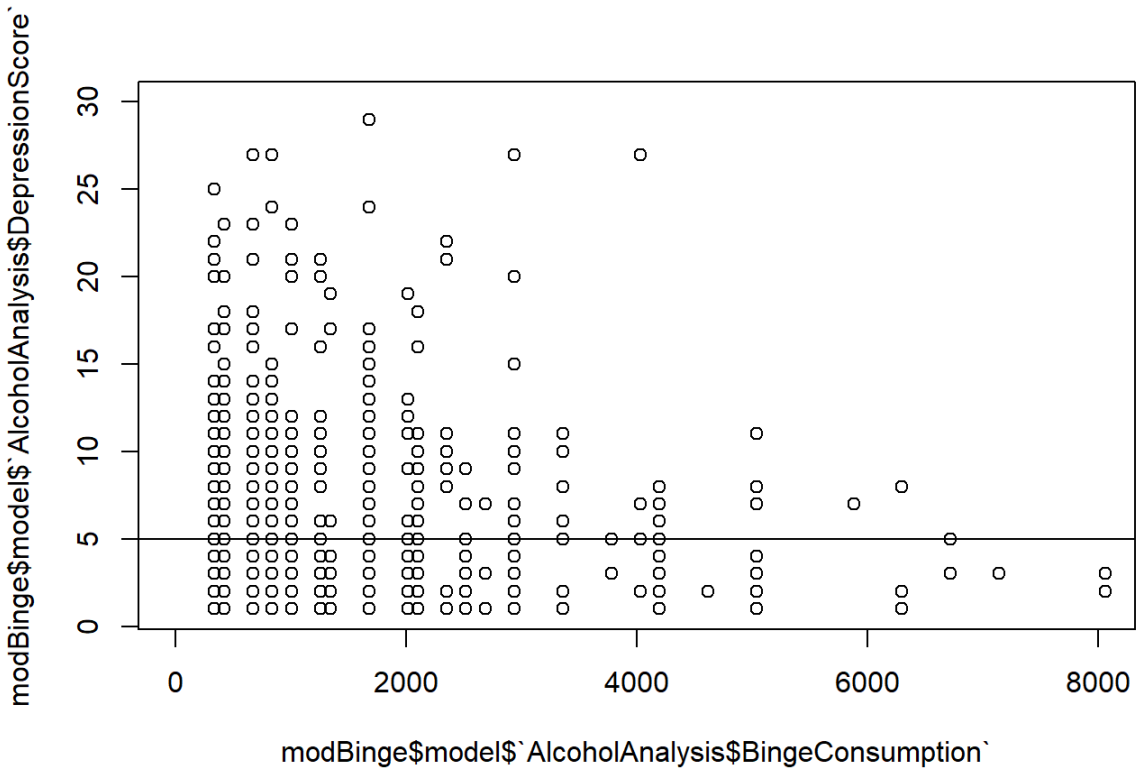
```
plot(mod$model$`AlcoholAnalysis$TotalConsumption`,mod$model$`AlcoholAnalysis$DepressionScore`, xlim=c(0,800
0)) # the limit can be set between (5000 and 20000)
abline(mod)
```




```
modBinge <- lm(AlcoholAnalysis$DepressionScore ~ AlcoholAnalysis$BingeConsumption, data = AlcoholAnalysis)
summary(modBinge)
```

```
##
## Call:
## lm(formula = AlcoholAnalysis$DepressionScore ~ AlcoholAnalysis$BingeConsumption,
##     data = AlcoholAnalysis)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.061 -2.994 -1.992  2.006 24.999
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       4.992e+00  1.576e-01   31.68   <2e-16 ***
## AlcoholAnalysis$BingeConsumption 1.058e-06  1.749e-05    0.06    0.952
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.839 on 998 degrees of freedom
## Multiple R-squared:  3.663e-06,  Adjusted R-squared:  -0.0009983
## F-statistic: 0.003656 on 1 and 998 DF,  p-value: 0.9518
```

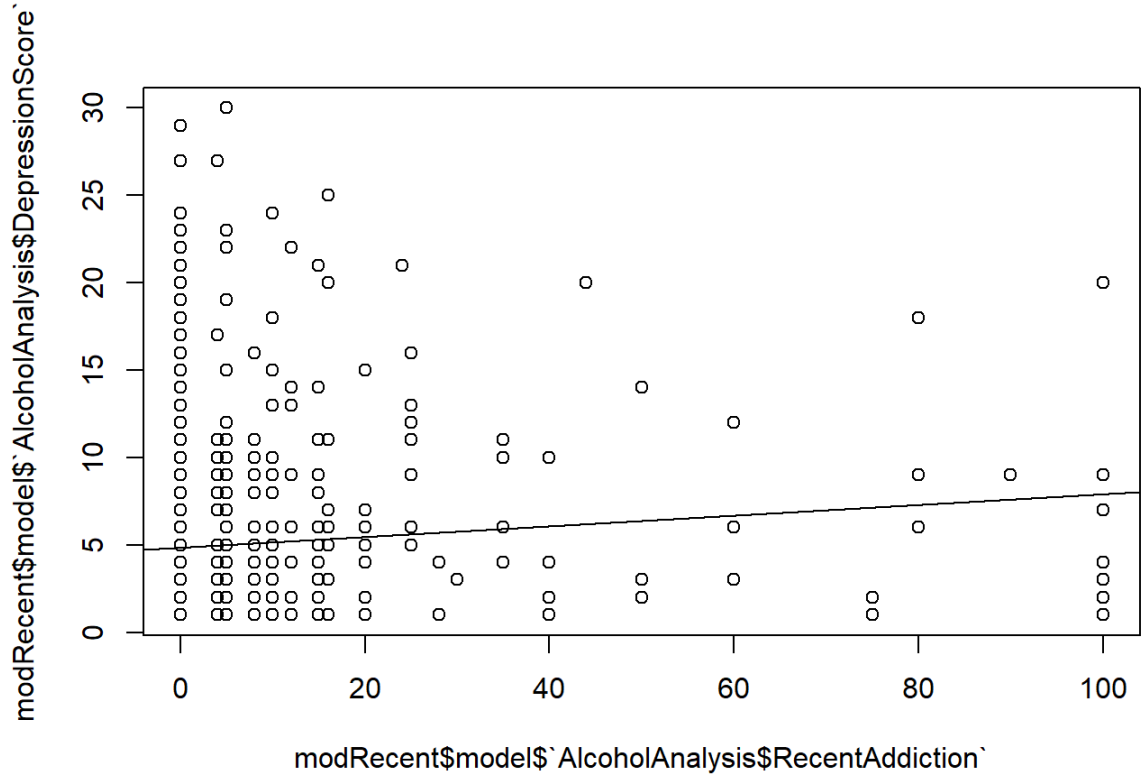
```
plot(modBinge$model$`AlcoholAnalysis$BingeConsumption`,modBinge$model$`AlcoholAnalysis$DepressionScore`, xl
im=c(0,8000)) # the limit can be set between (5000 and 20000)
abline(modBinge)
```



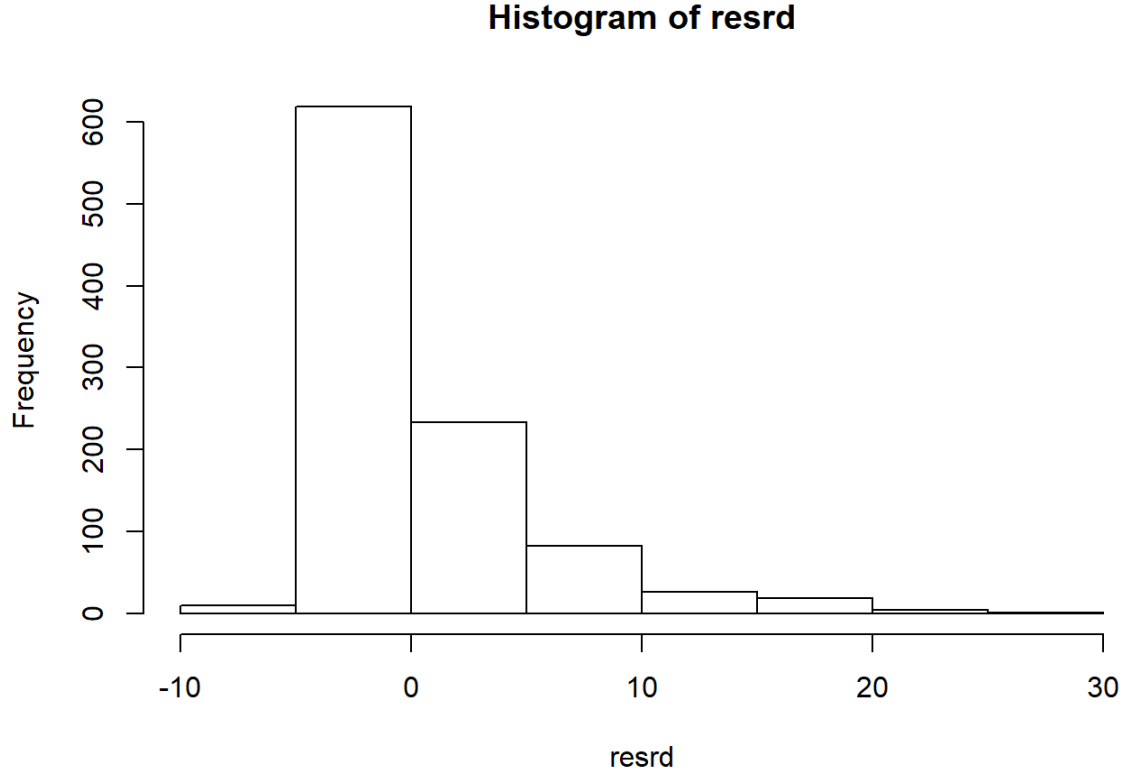

```
modRecent <- lm(AlcoholAnalysis$DepressionScore ~ AlcoholAnalysis$RecentAddiction, data = AlcoholAnalysis)
summary(modRecent)
```

```
##
## Call:
## lm(formula = AlcoholAnalysis$DepressionScore ~ AlcoholAnalysis$RecentAddiction,
##     data = AlcoholAnalysis)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.930 -3.092 -1.845  1.668 25.000
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     4.84546    0.16204  29.902  < 2e-16 ***
## AlcoholAnalysis$RecentAddiction  0.03084    0.01119   2.755  0.00597 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.827 on 994 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.007579,   Adjusted R-squared:  0.006581
## F-statistic: 7.591 on 1 and 994 DF,  p-value: 0.005973
```

```
plot(modRecent$model$`AlcoholAnalysis$RecentAddiction`,modRecent$model$`AlcoholAnalysis$DepressionScore`, x
lim=c(0,100)) # the limit can be set between (5000 and 20000)
abline(modRecent)
```



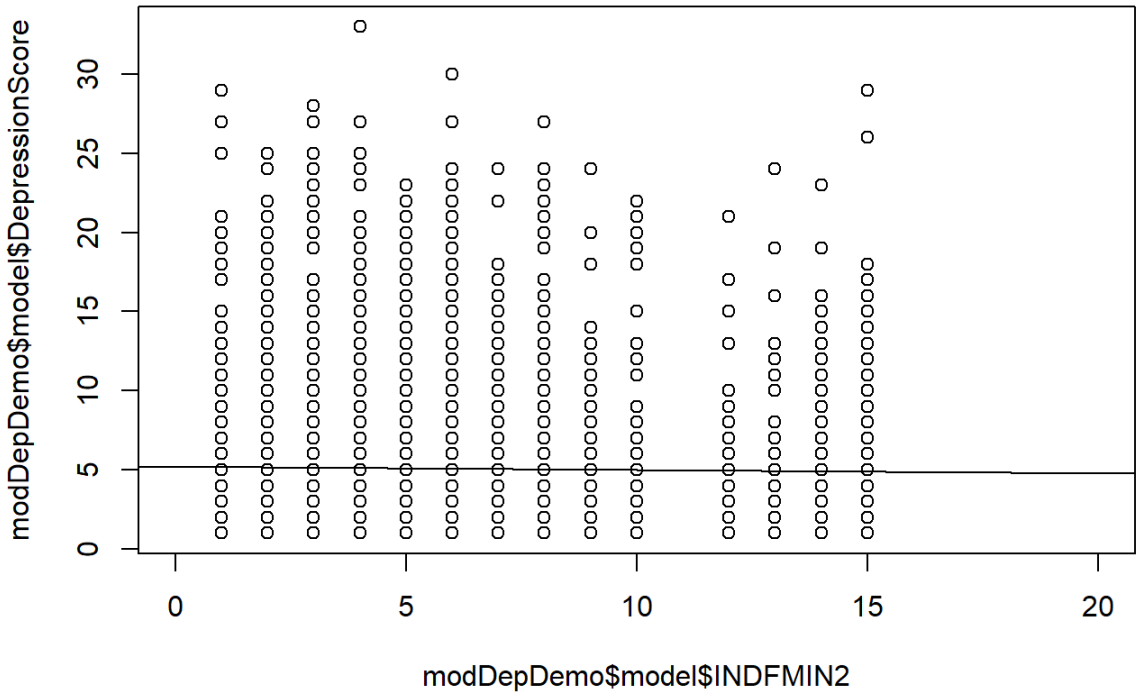

```
resrd <- resid(modRecent)
hist(resrd)
```



Correlation of

Depression score against demographics - age, gender, income

```
#hist(AlcoholAnalysis$DepressionScore)
DepressionDemographics <- demographc_clean %>%
                    inner_join(depression_trunc,
                      by="SEQN",
                      copy=False)
#head(DepressionDemographics)
modDepDemo <- lm(DepressionScore ~ INDFMIN2, data = DepressionDemographics)
plot(modDepDemo$model$INDFMIN2,modDepDemo$model$DepressionScore, xlim=c(0,20))
abline(modDepDemo)
```




```
summary(modDepDemo)
```

```
##
## Call:
## lm(formula = DepressionScore ~ INDFMIN2, data = DepressionDemographics)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.191 -3.089 -1.903  1.829 27.870
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.212042   0.099092  52.598  < 2e-16 ***
## INDFMIN2    -0.020569   0.005558  -3.701 0.000218 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.745 on 3485 degrees of freedom
##   (92 observations deleted due to missingness)
## Multiple R-squared:  0.003915,   Adjusted R-squared:  0.003629
## F-statistic:  13.7 on 1 and 3485 DF,  p-value: 0.0002182
```

```
ServerelyDepressed <- AlcoholAnalysis %>% filter(TotalConsumption >=1500) %>% filter(DepressionScore >= 5)
AlcoholAnalysisFiltered <- AlcoholAnalysis %>% inner_join(ServerelyDepressed)
```

```
## Joining, by = c("SEQN", "DepressionScore", "DPQ090", "RIAGENDR", "RIDAGEYR", "DMQMILIZ", "DMDCITZN", "DM
DEDUC3", "INDFMIN2", "INDFMPIR", "DMDEDUC2", "DMDMARTL", "TotalConsumption", "BingeConsumption", "RecentAdd
iction")
```

```
modFiltered <- lm(AlcoholAnalysisFiltered$DepressionScore ~ AlcoholAnalysisFiltered$TotalConsumption, data
= AlcoholAnalysis)
#mod <- glm(DPQ090 ~ TotalConsumption,  data = AlcoholAnalysis ) #, family="binomial")
summary(modFiltered)
```

```
##
## Call:
## lm(formula = AlcoholAnalysisFiltered$DepressionScore ~ AlcoholAnalysisFiltered$TotalConsumption,
##     data = AlcoholAnalysis)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.629 -3.869 -1.619  1.646 19.373
##
## Coefficients:
##                                           Estimate Std. Error t value
## (Intercept)                              1.064e+01  5.006e-01  21.259
## AlcoholAnalysisFiltered$TotalConsumption -9.364e-06  9.297e-06  -1.007
##                                          Pr(>|t|)
## (Intercept)                                <2e-16 ***
## AlcoholAnalysisFiltered$TotalConsumption    0.316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.66 on 130 degrees of freedom
## Multiple R-squared:  0.007744,   Adjusted R-squared:  0.0001114
## F-statistic: 1.015 on 1 and 130 DF,  p-value: 0.3157
```

```
plot(modFiltered$model$`AlcoholAnalysisFiltered$TotalConsumption`,modFiltered$model$`AlcoholAnalysisFiltere
d$DepressionScore`, ylim=c(0,20),xlim=c(0,10000)) # the limit can be set between (5000 and 20000)
abline(modFiltered)
```