

STAT 632 Project

Haixiang Yu

4/28/2022

```
pacman::p_load(ggplot2)
pacman::p_load(tidyverse)
dat <- read.csv("drug-use-by-age.csv", header=TRUE)
head(dat)
```

```
##   age      n alcohol.use alcohol.frequency marijuana.use marijuana.frequency
## 1  12 2798         3.9             3             1.1             4
## 2  13 2757         8.5             6             3.4             15
## 3  14 2792        18.1             5             8.7             24
## 4  15 2956        29.2             6            14.5             25
## 5  16 3058        40.1            10            22.5             30
## 6  17 3038        49.3            13            28.0             36
##   cocaine.use cocaine.frequency crack.use crack.frequency heroin.use
## 1          0.1          5.0          0.0          -          0.1
## 2          0.1          1.0          0.0          3.0          0.0
## 3          0.1          5.5          0.0          -          0.1
## 4          0.5          4.0          0.1          9.5          0.2
## 5          1.0          7.0          0.0          1.0          0.1
## 6          2.0          5.0          0.1         21.0          0.1
##   heroin.frequency hallucinogen.use hallucinogen.frequency inhalant.use
## 1          35.5          0.2          52          1.6
## 2          -          0.6          6          2.5
## 3          2.0          1.6          3          2.6
## 4          1.0          2.1          4          2.5
## 5         66.5          3.4          3          3.0
## 6         64.0          4.8          3          2.0
##   inhalant.frequency pain.releiver.use pain.releiver.frequency oxycontin.use
## 1          19.0          2.0          36          0.1
## 2          12.0          2.4          14          0.1
## 3          5.0          3.9          12          0.4
## 4          5.5          5.5          10          0.8
## 5          3.0          6.2          7          1.1
## 6          4.0          8.5          9          1.4
##   oxycontin.frequency tranquilizer.use tranquilizer.frequency stimulant.use
## 1          24.5          0.2          52.0          0.2
## 2          41.0          0.3          25.5          0.3
## 3          4.5          0.9          5.0          0.8
## 4          3.0          2.0          4.5          1.5
## 5          4.0          2.4          11.0          1.8
## 6          6.0          3.5          7.0          2.8
##   stimulant.frequency meth.use meth.frequency sedative.use sedative.frequency
## 1          2.0          0.0          -          0.2          13.0
## 2          4.0          0.1          5.0          0.1          19.0
## 3         12.0          0.1          24.0          0.2          16.5
## 4          6.0          0.3          10.5          0.4          30.0
## 5          9.5          0.3          36.0          0.2          3.0
## 6          9.0          0.6          48.0          0.5          6.5
```

```
dat1 <- dat[,c(1,2,3,5,17,19,21,23,25,27)] # keep percentage of use, and four kind of drugs
head(dat1)
```

```
##   age    n alcohol.use marijuana.use pain.releiver.use oxycontin.use
## 1  12 2798         3.9           1.1           2.0           0.1
## 2  13 2757         8.5           3.4           2.4           0.1
## 3  14 2792        18.1           8.7           3.9           0.4
## 4  15 2956        29.2          14.5           5.5           0.8
## 5  16 3058        40.1          22.5           6.2           1.1
## 6  17 3038        49.3          28.0           8.5           1.4
##   tranquilizer.use stimulant.use meth.use sedative.use
## 1                0.2           0.2      0.0           0.2
## 2                0.3           0.3      0.1           0.1
## 3                0.9           0.8      0.1           0.2
## 4                2.0           1.5      0.3           0.4
## 5                2.4           1.8      0.3           0.2
## 6                3.5           2.8      0.6           0.5
```

```
attach(dat1)
# calculate the number of people who used drugs
dat1$alcohol <- (n*alcohol.use*.01)
dat1$marijuana <- (n*marijuana.use*.01)
dat1$pain.releiver <- (n*pain.releiver.use*.01)
dat1$sedative <- (n*sedative.use*.01)
dat1$oxycontin <- (n*oxycontin.use*.01)
dat1$tranquilizer <- (n*tranquilizer.use*.01)
dat1$stimulant <- (n*stimulant.use*.01)
dat1$meth <- (n*meth.use*.01)
dat2 <- dat1[,c(1,2,3,4,5,6,7,8,9,10)]
head(dat2,20)
```

```
##      age      n alcohol.use marijuana.use pain.releiver.use oxycontin.use
## 1      12 2798          3.9           1.1           2.0           0.1
## 2      13 2757          8.5           3.4           2.4           0.1
## 3      14 2792         18.1           8.7           3.9           0.4
## 4      15 2956         29.2          14.5           5.5           0.8
## 5      16 3058         40.1          22.5           6.2           1.1
## 6      17 3038         49.3          28.0           8.5           1.4
## 7      18 2469         58.7          33.7           9.2           1.7
## 8      19 2223         64.6          33.4           9.4           1.5
## 9      20 2271         69.7          34.0          10.0           1.7
## 10     21 2354         83.2          33.0           9.0           1.3
## 11 22-23 4707         84.2          28.4          10.0           1.7
## 12 24-25 4591         83.1          24.9           9.0           1.3
## 13 26-29 2628         80.7          20.8           8.3           1.2
## 14 30-34 2864         77.5          16.4           5.9           0.9
## 15 35-49 7391         75.0          10.4           4.2           0.3
## 16 50-64 3923         67.2           7.3           2.5           0.4
## 17 65+ 2448          49.3           1.2           0.6           0.0
##      tranquilizer.use stimulant.use meth.use sedative.use
## 1              0.2           0.2      0.0      0.2
## 2              0.3           0.3      0.1      0.1
## 3              0.9           0.8      0.1      0.2
## 4              2.0           1.5      0.3      0.4
## 5              2.4           1.8      0.3      0.2
## 6              3.5           2.8      0.6      0.5
## 7              4.9           3.0      0.5      0.4
## 8              4.2           3.3      0.4      0.3
## 9              5.4           4.0      0.9      0.5
## 10             3.9           4.1      0.6      0.3
## 11             4.4           3.6      0.6      0.2
## 12             4.3           2.6      0.7      0.2
## 13             4.2           2.3      0.6      0.4
## 14             3.6           1.4      0.4      0.4
## 15             1.9           0.6      0.2      0.3
## 16             1.4           0.3      0.2      0.2
## 17             0.2           0.0      0.0      0.0
```

```
# clean data
dat3 <- as.data.frame(dat2, stringsAsFactors = FALSE)
dat3 <- as.data.frame(apply(dat3,2, as.integer))
```

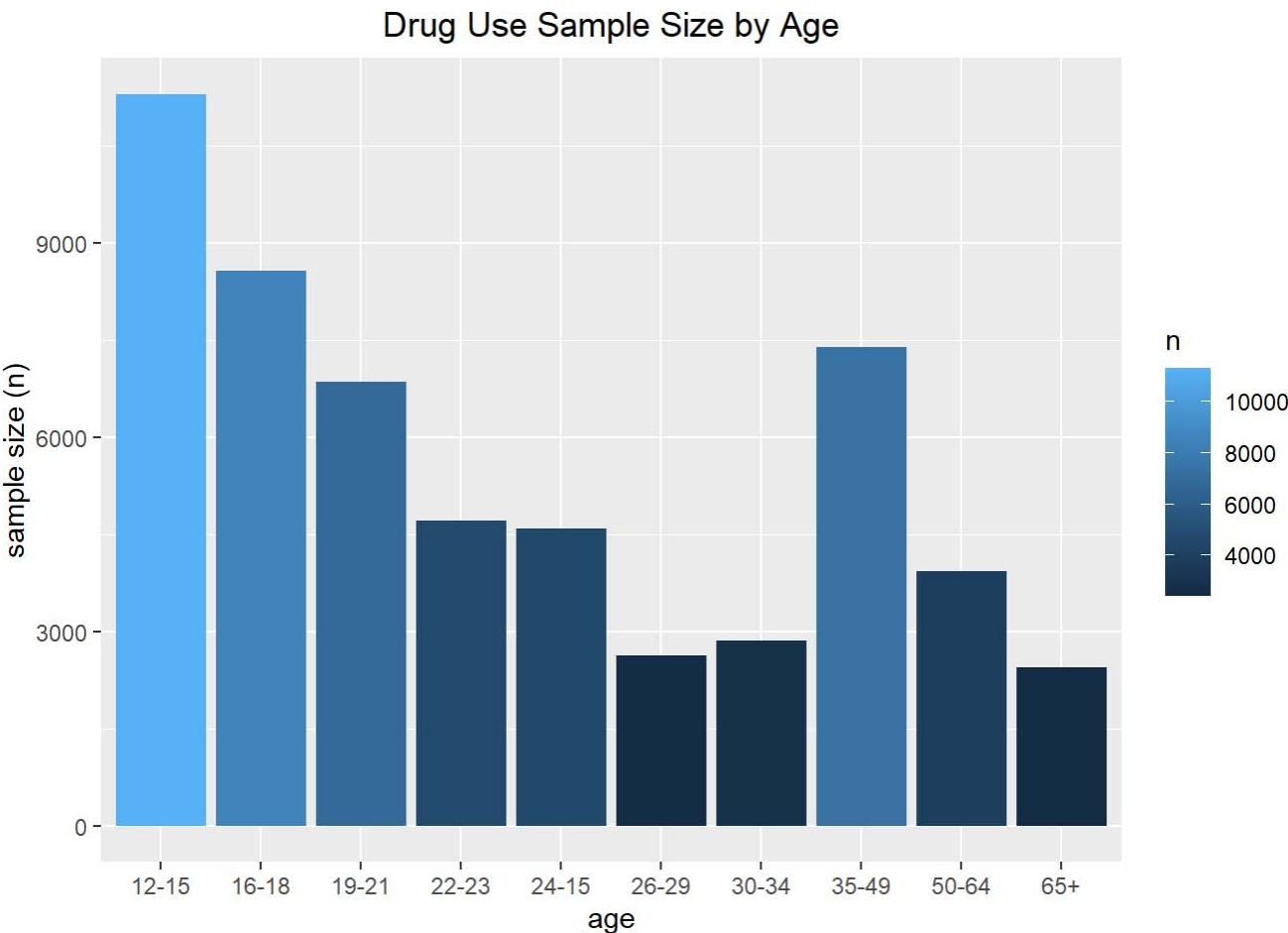
```
## Warning in apply(dat3, 2, as.integer): NAs introduced by coercion
```

```
dat3[is.na(dat3)] <- 0
```

```
# Combined ages groups
dat3[1,] <- dat3[1,] + dat3[2,] + dat3[3,] + dat3[4,]
dat3[5,] <- dat3[5,] + dat3[6,] + dat3[7,]
dat3[8,] <- dat3[8,] + dat3[9,] + dat3[10,]
dat4 <- dat3[c(1,5,8,11,12,13,14,15,16,17),]# keep
dat4$age <- c('12-15', '16-18', '19-21', '22-23', '24-15', '26-29', '30-34', '35-49', '50-64', '65+')
head(dat4,10)
```

```
##      age      n alcohol.use marijuana.use pain.releiver.use oxycontin.use
## 1 12-15 11303          58             26             12             0
## 5 16-18  8565          147            83             23             3
## 8 19-21  6848          216           100             28             3
## 11 22-23  4707           84            28             10             1
## 12 24-15  4591           83            24              9             1
## 13 26-29  2628           80            20              8             1
## 14 30-34  2864           77            16              5             0
## 15 35-49  7391           75            10              4             0
## 16 50-64  3923           67             7              2             0
## 17 65+   2448           49             1              0             0
##      tranquilizer.use stimulant.use meth.use sedative.use
## 1              2              1              0              0
## 5              9              6              0              0
## 8             12             11              0              0
## 11             4              3              0              0
## 12             4              2              0              0
## 13             4              2              0              0
## 14             3              1              0              0
## 15             1              0              0              0
## 16             1              0              0              0
## 17             0              0              0              0
```

```
ggplot(dat4, aes(x = age, y = n, fill = n)) +
  geom_col() +
  scale_fill_continuous() +
  labs(y="sample size (n)",
       title = "Drug Use Sample Size by Age") +
  theme(plot.title=element_text(hjust = 0.5))
```



```
summary(dat4)
```

```
##      age              n      alcohol.use      marijuana.use
## Length:10      Min.   : 2448      Min.   : 49.00      Min.   :  1.0
## Class :character 1st Qu.: 3129      1st Qu.: 69.00      1st Qu.: 11.5
## Mode  :character Median : 4649      Median : 78.50      Median : 22.0
##              Mean  : 5527      Mean  : 93.60      Mean   : 31.5
##              3rd Qu.: 7255      3rd Qu.: 83.75      3rd Qu.: 27.5
##              Max.   :11303      Max.   :216.00      Max.   :100.0
## pain.releiver.use oxycontin.use tranquilizer.use stimulant.use      meth.use
## Min.   : 0.00      Min.   :0.0      Min.   : 0.00      Min.   : 0.00      Min.   :0
## 1st Qu.: 4.25      1st Qu.:0.0      1st Qu.: 1.25      1st Qu.: 0.25      1st Qu.:0
## Median : 8.50      Median :0.5      Median : 3.50      Median : 1.50      Median :0
## Mean   :10.10      Mean   :0.9      Mean   : 4.00      Mean   : 2.60      Mean   :0
## 3rd Qu.:11.50      3rd Qu.:1.0      3rd Qu.: 4.00      3rd Qu.: 2.75      3rd Qu.:0
## Max.   :28.00      Max.   :3.0      Max.   :12.00      Max.   :11.00      Max.   :0
## sedative.use
## Min.   :0
## 1st Qu.:0
## Median :0
## Mean   :0
## 3rd Qu.:0
## Max.   :0
```

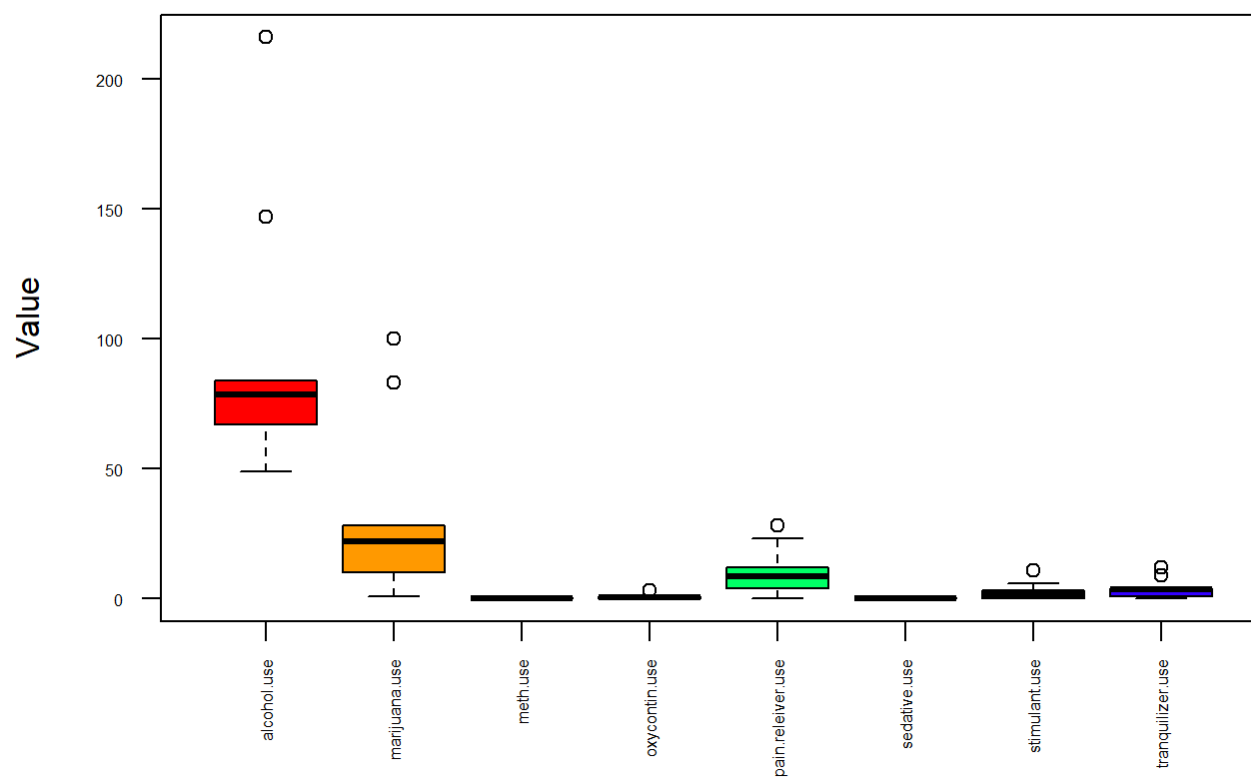
```
# data set contains all drugs
```

```
dat5 <- gather(dat4, "Drugs", value = "Value", -age, -n)
head(dat5)
```

```
##      age      n      Drugs Value
## 1 12-15 11303 alcohol.use   58
## 2 16-18  8565 alcohol.use  147
## 3 19-21  6848 alcohol.use  216
## 4 22-23  4707 alcohol.use   84
## 5 24-15  4591 alcohol.use   83
## 6 26-29  2628 alcohol.use   80
```

```
drugbp <- boxplot(Value ~ Drugs, dat5,
  col = rainbow(ncol(dat4)),
  horizontal = FALSE, xlab = "",
  main = "Drug Type Usage Among Respondents",
  las = 2, cex.axis = 0.5)
```

Drug Type Usage Among Respondents



```
require(MASS)
```

```
## Loading required package: MASS
```

```
## Warning: package 'MASS' was built under R version 4.1.3
```

```
##  
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':  
##  
##      select
```

```
anova(lm(Value ~ age + Drugs, dat5))
```

```
## Analysis of Variance Table  
##  
## Response: Value  
##           Df Sum Sq Mean Sq F value    Pr(>F)  
## age         9  11087   1231.9   3.4484 0.001647 **  
## Drugs        7  73334  10476.3  29.3256 < 2.2e-16 ***  
## Residuals  63   22506    357.2  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cFit2 <- lm(Value ~ age + Drugs, dat5)  
summary(cFit2)
```



```
##
## Call:
## lm(formula = Value ~ age + Drugs, data = dat5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -33.012 -10.206   1.925   5.119  93.988
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      88.138      8.713  10.116 7.98e-15 ***
## age16-18         21.500      9.450   2.275 0.026317 *
## age19-21         33.875      9.450   3.585 0.000659 ***
## age22-23          3.875      9.450   0.410 0.683171
## age24-15          3.000      9.450   0.317 0.751954
## age26-29          2.000      9.450   0.212 0.833078
## age30-34          0.375      9.450   0.040 0.968473
## age35-49         -1.125      9.450  -0.119 0.905620
## age50-64         -2.750      9.450  -0.291 0.772012
## age65+           -6.125      9.450  -0.648 0.519262
## Drugsmarijuana.use -62.100      8.453  -7.347 4.96e-10 ***
## Drugsmeth.use     -93.600      8.453 -11.073 < 2e-16 ***
## Drugsoxycontin.use -92.700      8.453 -10.967 3.02e-16 ***
## Drugspain.releiver.use -83.500      8.453  -9.879 2.01e-14 ***
## Drugssedative.use  -93.600      8.453 -11.073 < 2e-16 ***
## Drugsstimulant.use -91.000      8.453 -10.766 6.50e-16 ***
## Drugstranquilizer.use -89.600      8.453 -10.600 1.23e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.9 on 63 degrees of freedom
## Multiple R-squared:  0.7895, Adjusted R-squared:  0.7361
## F-statistic: 14.77 on 16 and 63 DF,  p-value: 1.357e-15
```

Results: The age group and drug group variances explains 78.95% variance in respondents medical-use and entertainment-use in an age group in the past 12 months. Since p-value is less than 0.5, the model as a whole is statistically significant.

```
par(mfrow=c(1,2))
plot(lm(Value ~ age + Drugs, dat5), 1:2)
```

