# ANA 515 Assignment 2

Yanan Li

2023-06-16

The "Drug Use By Age" dataset was collected as part of The National Survey on Drug Use and Health (NSDUH) series. It measures the prevalence and correlates of drug use in the United States. The survey included questions concerning treatment for both substance abuse and mental health-related disorders. Respondents were also asked about personal and family income sources and amounts, health care access and coverage, illegal activities and arrest record, problems resulting from the use of drugs, and needle-sharing. The research question I hope to answer is whether drug abuse is related to age. The dataset is saved in a csv file, including 17 age groups as rows and 26 attributes as columns.

```
# Use read_csv function to read the dataset
# Read the library tidyverse first
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.1     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
url <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/drug-use-by-age/drug-use-by-age.cs

druguse <- read_csv(url)
```

```
## Rows: 17 Columns: 28
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): age, cocaine_frequency, crack_frequency, heroin_frequency, inhalan...
## dbl (21): n, alcohol_use, alcohol_frequency, marijuana_use, marijuana_freque...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#Lets just see the alcohol use/frequency, marijuana use/frequency, cocaine use/frequency

drug_sub <- select(druguse,alcohol_use,alcohol_frequency,marijuana_use,marijuana_frequency,cocaine_use,
```

This dataframe has 17 rows and 6 columns. The names of the columns and a brief description of each are in the table below:

```
library(knitr)
columns_summary <- data.frame(
  Columns = c(colnames(drug_sub)),
  Description = c("Percentage of those in an age group who used alcohol in the past 12 months", "Median
  )

kable(columns_summary, caption = "Data Description")
```

Table 1: Data Description

| Columns | Description |
|---|---|
| alcohol_use | Percentage of those in an age group who used alcohol in the past 12 months |
| alcohol_frequency | Median number of times a user in an age group used alcohol in the past 12 months |
| marijuana_use | Percentage of those in an age group who used marijuana in the past 12 months |
| marijuana_frequency | Median number of times a user in an age group used marijuana in the past 12 months |
| cocaine_use | Percentage of those in an age group who used cocaine in the past 12 months |
| cocaine_frequency | Median number of times a user in an age group used cocaine in the past 12 months |

```
#Pick the fields: alcohol use, marijuana use, cocaine use
drug_pick <- select(drug_sub,alcohol_use,marijuana_use,cocaine_use )
drug_sum <- summary(drug_pick)
drug_sum
```

```
##   alcohol_use    marijuana_use    cocaine_use
##  Min.   : 3.90   Min.   : 1.10   Min.   :0.000
##  1st Qu.:40.10   1st Qu.: 8.70   1st Qu.:0.500
##  Median :64.60   Median :20.80   Median :2.000
##  Mean   :55.43   Mean   :18.92   Mean   :2.176
##  3rd Qu.:77.50   3rd Qu.:28.40   3rd Qu.:4.000
##  Max.   :84.20   Max.   :34.00   Max.   :4.900
```