

DRUG USE AND ASSOCIATED RISK FACTORS

CODEBOOK

EXTRACTED FROM THE 2019 NATIONAL SURVEY ON DRUG USE AND HEALTH

Introduction

This dataset was extracted from the 2019 National Survey on Drug Use and Health, which is an annual survey (starting from 1971) that collects data regarding drug usage and mental health issues in the United States. The survey is run and maintained by the *Substance Abuse and Mental Health Services Administration* (SAMHSA) –a federal government agency that specializes in behavioral health and research. For more information on the survey (i.e., sampling methods, etc.), please visit <https://nsduhweb.rti.org/respweb/homepage.cfm>. The original dataset (as well as datasets from surveys performed in other years) can be retrieved at <https://www.datafiles.samhsa.gov/dataset/national-survey-drug-use-and-health-2020-nsduh-2020-ds0001>.

The subset contains 16 categorical features of 56,136 observations. This subset was originally compiled to be used for a college data analysis project that sought to determine the influence certain risk factors have on drug usage. The purpose of creating this subset was to simplify moving the dataset around, as well as working with the dataset in general (e.g., less cryptic feature names, etc.) The subset is now offered *as is* to other students, hobbyists, etc. for practice.

The variables in this data set are substances that survey participants have admitted to using in their past (e.g., cocaine, heroin, meth, etc.), and associated risk factors that may influence whether or not an individual engages in drug use (e.g., income bracket, education, sex, etc.). Non-illicit drugs such as alcohol, marijuana*, and pain relief medicine were also of particular interest in the analysis project. Because these substances are often characterized as “gateways” to illicit drug use, it may prove interesting to the analyst to discern how usage of *these* substances affects the risk of using *illicit* substances as well.

Drug Use Variables

Features 8 drugs of interest which include both illicit and legal substances.

alcohol

(ALCFLAG in full dataset)

The participant has admitted to having ever consumed alcohol.

1: Yes

0: No

cigarettes

(CIGFLAG in full dataset)

The participant has admitted to having ever smoked cigarettes.

1: Yes

0: No

cocaine

(COCFLAG in full dataset)

The participant has admitted to having ever used cocaine in the past.

1: Yes

0: No

crack

(CRKFLAG in full dataset)

The participant has admitted to having ever smoked crack in the past.

1: Yes

0: No

heroin

(HERFLAG in full dataset)

The participant has admitted to having ever used Heroin.

1: Yes

0: No

marijuana

(MJONLYFLAG in full dataset)

The participant has admitted to having ever smoked/consumed marijuana.

1: Yes

0: No

meth

(METHAMFLAG in full dataset)

The participant has admitted to having ever used Methamphetamine.

1: Yes

0: No

pain_reliver

(PNRANYFLAG in full dataset)

The participant has admitted to having ever used pain relivers.

1: Yes

0: No

Risk Factor Variables

Features 8 risk factors that may influence the likelihood of a participant having engaged in drug use.

age

(CATAG6 in full dataset)

Categorizes the survey participant's age

- 1: 12-17 years old
- 2: 18-25 years old
- 3: 26-34 years old
- 4: 35-49 years old
- 5: 50-64 years old
- 6: 65 years or older

education

(EDUHIGHCAT in full dataset)

Categorizes the survey participant's level of education

- 1: Didn't graduate high school
- 2: High School graduate
- 3: Some college, or Associate degree
- 4: College graduate
- 5: 12 to 17 year olds

employment

(IRWRKSTAT in full dataset)

Categorizes the survey participant's level of education

- 1: Participant is employed full-time
- 2: Participant is employed part-time
- 3: Participant is unemployed

4: Other (incl. not in labor force)

99: Participant is underage; 12-14 years old

father (father is present in the household)

(IFATHER in full dataset)

The participant has specified whether or not their father is present in their household.

1: Yes, the father is in the household

2: No, the father is not in the household

3: Doesn't know whether or not father is present

4: participant is 18 or older

income

(INCOME in full dataset)

Categorizes the participant's income into a specific range.

1: Less than \$20,000

2: Between \$20,000 and \$49,000

3: Between \$50,000 and \$74,999

4: More than \$75,000

mother (mother is present in the household)

(IMOTHER in full dataset)

The participant has specified whether or not their mother is present in their household.

1: Yes, the mother is in the household

2: No, the mother is not in the household

3: Doesn't know whether or not mother is present

4: participant is 18 or older

race

(NEWRACE2 in full dataset)

Specifies the participant's race (all races are non-Hispanic, other than Hispanic)

1: White

2: Black

3: Native American

4: Native Hawaiian/Pacific Islander

5: Asian

6: Multi-Racial

7: Hispanic

sex

(IRSEX in full dataset)

Specifies the biological sex of the participant.

1: Participant is male

2: Participant is female