Package ‘SDAResources’

May 17, 2021

Type Package
Title Datasets and Functions for “Sampling: Design and Analysis”
Version 0.1.0
Maintainer Yan Lu <yanlu@unm.edu>
Description Includes all the datasets of “Sampling: Design and Analysis” (3rd edition by Sharon Lohr) in R format and additional functions for analyzing and graphing probability samples.
License GPL-2 | GPL-3
Encoding UTF-8
LazyData true
NeedsCompilation no
Depends R (>= 3.5.0)
RoxygenNote 7.1.1
Suggests rmarkdown, knitr
VignetteBuilder knitr
Author Yan Lu [aut, cre], Sharon Lohr [aut]
Repository CRAN
Date/Publication 2021-05-17 07:20:02 UTC

R topics documented:

agpop ................................................................. 2
agpps .................................................................. 3
agsrs .................................................................. 4
agstrat ................................................................. 4
algebra ................................................................. 5
anthsrs ............................................................... 6
classes ............................................................... 6
classpps ............................................................. 7
college ............................................................... 7
**Description**

Data from the 1992 U.S. Census of Agriculture.

**Usage**

```r
data(agpop)
```

**Format**

This data frame contains the following columns:
- **county**: county name (character variable)
- **state**: state abbreviation (character variable)
- **acres92**: number of acres devoted to farms, 1992
- **acres87**: number of acres devoted to farms, 1987
- **acres82**: number of acres devoted to farms, 1982
- **farms92**: number of farms, 1992
- **farms87**: number of farms, 1987
- **farms82**: number of farms, 1982
- **largef92**: number of farms with 1,000 acres or more, 1992
- **largef87**: number of farms with 1,000 acres or more, 1987
- **largef82**: number of farms with 1,000 acres or more, 1982
- **smallf92**: number of farms with 9 acres or fewer, 1992
- **smallf87**: number of farms with 9 acres or fewer, 1987
- **smallf82**: number of farms with 9 acres or fewer, 1982
- **region**: S = south; W = west; NC = north central; NE = northeast
References


agpps

agpps data

Description

Data from a without-replacement pps sample from agpop data

Usage

data(agpps)

Format

This data frame contains the following columns:
- county: county name (character variable)
- state: state abbreviation (character variable)
- acres92: number of acres devoted to farms, 1992
- acres87: number of acres devoted to farms, 1987
- acres82: number of acres devoted to farms, 1982
- farms92: number of farms, 1992
- farms87: number of farms, 1987
- farms82: number of farms, 1982
- largef92: number of farms with 1,000 acres or more, 1992
- largef87: number of farms with 1,000 acres or more, 1987
- largef82: number of farms with 1,000 acres or more, 1982
- smallf92: number of farms with 9 acres or fewer, 1992
- smallf87: number of farms with 9 acres or fewer, 1987
- smallf82: number of farms with 9 acres or fewer, 1982
- region: S = south; W = west; NC = north central; NE = northeast
- sizemeas: size measure used to select the pps sample
- SelectionProb: inclusion probability for county
- SamplingWeight: sampling weight for county
- Unit: unit number for indexing joint inclusion probabilities
- JtProb_1: columns of joint inclusion probabilities
- ... 
- JtProb_15: columns of joint inclusion probabilities
References


---

**agsrs**  
**agsrs data**

**Description**

Data from an SRS of size 300 from the 1992 U.S. Census of Agriculture.

**Usage**

data(agsrs)

**Format**

Variables are the same as in agpop data.

**References**


---

**agstrat**  
**agstrat data**

**Description**

Data from a stratified random sample of size 300 from the 1992 U.S. Census of Agriculture

**Usage**

data(agstrat)

**Format**

This data frame contains the following columns:
- county: county name (character variable)
- state: state abbreviation (character variable)
- acres92: number of acres devoted to farms, 1992
- acres87: number of acres devoted to farms, 1987
- acres82: number of acres devoted to farms, 1982
- farms92: number of farms, 1992
- farms87: number of farms, 1987
farms82: number of farms, 1982
largef92: number of farms with 1,000 acres or more, 1992
largef87: number of farms with 1,000 acres or more, 1987
largef82: number of farms with 1,000 acres or more, 1982
smallf92: number of farms with 9 acres or fewer, 1992
smallf87: number of farms with 9 acres or fewer, 1987
smallf82: number of farms with 9 acres or fewer, 1982
region: S = south; W = west; NC = north central; NE = northeast
rn: random numbers used to select sample in each stratum
weight: sampling weight for each county in sample

References

---

**Description**
Hypothetical data for an SRS of 12 algebra classes in a city, from a population of 187 classes.

**Usage**
data(algebra)

**Format**
This data frame contains the following columns:
class: Class number
Mi: Number of students (Mᵢ) in class
score: Score of student on test

**References**
Description

Length of left middle finger and height for a SRS of size 200 from anthrop data

Usage

data(anthsrs)

Format

This data frame contains the following columns:
finger: length of left middle finger (cm)
height: height (inches)
wgt: sampling weight

References


Description

Population sizes for 15 classes, used in Chapter 6 of SDA to demonstrate unequal-probability sampling.

Usage

data(classes)

Format

This data frame contains the following columns:
class: Class ID number
class_size: Number of students in class

References

**classpps**  

**Description**

Two-stage unequal-probability sample without replacement from the population of classes in classes data

**Usage**

```r
data(classpps)
```

**Format**

This data frame contains the following columns:

- `class`: Class ID number
- `class_size`: Number of students in class
- `finalweight`: Sampling weight for student
- `hours`: Number of hours spent studying statistics

**References**


---

**college**  

**Description**


**Usage**

```r
data(college)
```
Format

This data frame contains the following columns:
unitid: Unit identification number
instnm: Institution name (character, length 81)
city: City (character, length 24)
stabbbr: State abbreviation (character, length 2)
highdeg: Highest degree awarded
(3 = Bachelor’s degree, 4 = Graduate degree)
control: Control (ownership) of institution
(1 = Public, 2 = Private nonprofit)
region: Region where institution is located
{1 New England (CT, ME, MA, NH, RI, VT)
2 Mid East (DE, DC, MD, NJ, NY, PA)
3 Great Lakes (IL, IN, MI, OH, WI)
4 Plains (IA, KS, MN, MO, NE, ND, SD)
5 Southeast (AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, VA, WV)
6 Southwest (AZ, NM, OK, TX)
7 Rocky Mountains (CO, ID, MT, UT, WY)
8 Far West (AK, CA, HI, NV, OR, WA)}
locale: Locale of institution
{11 City: Large (population of 250,000 or more)
12 City: Midsize (population of at least 100,000 but less than 250,000)
13 City: Small (population less than 100,000)
21 Suburb: Large (outside principal city, in urbanized area with population of 250,000 or more)
22 Suburb: Midsize (outside principal city, in urbanized area with population of at least 100,000 but less than 250,000)
23 Suburb: Small (outside principal city, in urbanized area with population less than 100,000)
31 Town: Fringe (in urban cluster up to 10 miles from an urbanized area)
32 Town: Distant (in urban cluster more than 10 miles and up to 35 miles from an urbanized area)
33 Town: Remote (in urban cluster more than 35 miles from an urbanized area)
41 Rural: Fringe (rural territory up to 5 miles from an urbanized area or up to 2.5 miles from an urban cluster)
42 Rural: Distant (rural territory more than 5 miles but up to 25 miles from an urbanized area or more than 2.5 and up to 10 miles from an urban cluster)
43 Rural: Remote (rural territory more than 25 miles from an urbanized area and more than 10 miles from an urban cluster)}
ccbasic: Carnegie basic classification
{15 Doctoral Universities: Very High Research Activity}
<table>
<thead>
<tr>
<th>College Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 Doctoral Universities</td>
<td>High Research Activity</td>
</tr>
<tr>
<td>17 Doctoral/Professional Universities</td>
<td></td>
</tr>
<tr>
<td>18 Master’s Colleges &amp; Universities: Larger Programs</td>
<td></td>
</tr>
<tr>
<td>19 Master’s Colleges &amp; Universities: Medium Programs</td>
<td></td>
</tr>
<tr>
<td>20 Master’s Colleges &amp; Universities: Small Programs</td>
<td></td>
</tr>
<tr>
<td>21 Baccalaureate Colleges: Arts &amp; Sciences Focus</td>
<td></td>
</tr>
<tr>
<td>22 Baccalaureate Colleges: Diverse Fields</td>
<td></td>
</tr>
<tr>
<td>ccsizset: Carnegie classification, size and setting</td>
<td></td>
</tr>
<tr>
<td>6 Four-year, very small, primarily nonresidential</td>
<td></td>
</tr>
<tr>
<td>7 Four-year, very small, primarily residential</td>
<td></td>
</tr>
<tr>
<td>8 Four-year, very small, highly residential</td>
<td></td>
</tr>
<tr>
<td>9 Four-year, small, primarily nonresidential</td>
<td></td>
</tr>
<tr>
<td>10 Four-year, small, primarily residential</td>
<td></td>
</tr>
<tr>
<td>11 Four-year, small, highly residential</td>
<td></td>
</tr>
<tr>
<td>12 Four-year, medium, primarily nonresidential</td>
<td></td>
</tr>
<tr>
<td>13 Four-year, medium, primarily residential</td>
<td></td>
</tr>
<tr>
<td>14 Four-year, medium, highly residential</td>
<td></td>
</tr>
<tr>
<td>15 Four-year, large, primarily nonresidential</td>
<td></td>
</tr>
<tr>
<td>16 Four-year, large, primarily residential</td>
<td></td>
</tr>
<tr>
<td>17 Four-year, large, highly residential</td>
<td></td>
</tr>
<tr>
<td>hbcu: Historically black college or university, 1=yes, 0=no</td>
<td></td>
</tr>
<tr>
<td>openadmp: Does the college have an open admissions policy, that is, does it accept any students that apply or have minimal requirements for admission? 1 = yes, 0 = no</td>
<td></td>
</tr>
<tr>
<td>adm_rate: Fall admissions rate, defined as the number of admitted undergraduates divided by the number of undergraduates who applied</td>
<td></td>
</tr>
<tr>
<td>sat_avg: Average SAT score (or equivalent) for admitted students</td>
<td></td>
</tr>
<tr>
<td>ugds: Number of number of degree-seeking undergraduate students enrolled in the fall term</td>
<td></td>
</tr>
<tr>
<td>ugds_men: Proportion of ugds who are men</td>
<td></td>
</tr>
<tr>
<td>ugds_women: Proportion of ugds who are women</td>
<td></td>
</tr>
<tr>
<td>ugds_white: Proportion of ugds who are white (based on self-reports)</td>
<td></td>
</tr>
<tr>
<td>ugds_black: Proportion of ugds who are black/African American (based on self-reports)</td>
<td></td>
</tr>
<tr>
<td>ugds_hisp: Proportion of ugds who are Hispanic (based on self-reports)</td>
<td></td>
</tr>
<tr>
<td>ugds_asian: Proportion of ugds who are Asian (based on self-reports)</td>
<td></td>
</tr>
<tr>
<td>ugds_other: Proportion of ugds who have other race/ethnicity (created from other categories on original data file; race/ethnicity proportions sum to 1)</td>
<td></td>
</tr>
<tr>
<td>npt4: Average net price of attendance, derived from the full cost of attendance (including tuition and fees, books and supplies, and living expenses) minus federal, state, and institutional</td>
<td></td>
</tr>
</tbody>
</table>
grant/scholarship aid, for full-time, first-time undergraduate Title IV-receiving students. NPT4 created from scorecard data variables NPT4_PUB if public institution and NPT4_PRIV if private
tuitionfee_in: In-state tuition and fees

tuitionfee_out: Out-of-state tuition and fees

avgfacsal: Average faculty salary per month

pftfac: Proportion of faculty that is full-time

c150_4: Proportion of first-year, full-time students who complete their degree within 150% of the expected time to complete; for most institutions, this is the proportion of students who receive a degree within 6 years

grads: Number of graduate students

Details

Some of the variables in the book data have been calculated from other variables in the original source; these have been given new variable names that are not found in the data dictionary at https://collegescorecard.ed.gov/data/documentation/.

The college data set contains only colleges in the 50 states plus District of Columbia that offer undergraduate degrees and contain information on average net price. This data set is made available for pedagogical purposes only. Anyone wishing to draw conclusions from College Scorecard data should obtain the original data set from the Department of Education. The original data set has 1,925 variables and includes institutions (such as those that do not grant undergraduate degrees) that are not in the data college. The college data includes institutions in the original data set that:
(1) are located in the 50 states plus District of Columbia, (2) contain information on average net price (NPT4), (3) are predominantly Bachelor’s degree-granting, (4) were currently operating as of June 2020, (5) are not private for-profit institutions or "global" campuses, (6) have Carnegie size classification (variable ccsizset) between 6 and 17 and Carnegie basic classification (variable ccbasic) between 14 and 22 (these offer Bachelor’s degrees), (7) enrolls first-time students, and (8) are not U.S. Service Academies.

References


Description

Five replicate SRSs from the set of public colleges and universities (having control = 1) in college data. Columns 1-29 are as in college data, with additional columns 30-32 listed below. Note that the selection probabilities and sampling weights are for the separate replicate samples, so that the weights for each sample sum to the population size 500.

Usage

data(collegerg)
Format

This data frame contains the following columns:

- selectionprob: Selection probability for each replicate sample
- samplingweight: Sampling weight for each replicate sample
- repgroup: Replicate group number

References


---

coots

Description

Selected information on egg size, from a larger study by Arnold (1991). Data provided courtesy of Todd Arnold. Not all observations are used for this data set, so results may not agree with those in Arnold (1991).

Usage

data(coots)

Format

This data frame contains the following columns:

- clutch: Clutch number from which eggs were subsampled.
- csize: Number of eggs in clutch (M_i)
- length: length of egg (mm)
- breadth: maximum breadth of egg (mm)
- volume: calculated as 0.000507*length * breadth
- tmt: = 1 if received supplemental feeding, 0 otherwise

References

deadtrees data

Description

Number of dead trees recorded by photograph and field count for a (fictional) SRS of 25 plots taken from a population of 100 plots.

Usage

data(deadtrees)

Format

This data frame contains the following columns:
photo: Number of dead trees in plot from photograph
field: Number of dead trees in plot from field observation

References


gpa data

Description

GPA data from Chapter 5 of SDA.

Usage

data(gpa)

Format

This data frame contains the following columns:
suite: Suite (psu) identifier
gpa: Grade point average of person in suite
wt: Sampling weight, = 20 for every observation

References

htsrs

**htsrs**  |  **htsrs data**
---|---

**Description**

Height and gender for a SRS of 200 persons, taken from htpop data

**Usage**

data(htsrs)

**Format**

This data frame contains the following columns:
- \( rn \): random number used to select unit
- \( \text{height} \): height of person, cm
- \( \text{gender} \): M=male, F=female

**References**


---

htstrat

**htstrat**  |  **htstrat data**
---|---

**Description**

Height and gender for a stratified random sample of 160 women and 40 men, taken from htpop data.

**Usage**

data(htstrat)

**Format**

The columns and names are as in htsrs data.

**References**

impute

**Description**
Small artificial data set used to illustrate imputation methods.

**Usage**
data(impute)

**Format**
This data frame contains the following columns:
- **person**: identification number for person
- **age**: age in years
- **gender**: M=male, F=female
- **crime**: = 1 if victim of any crime, 0 otherwise
- **violcrime**: = 1 if victim of violent crime, 0 otherwise

**References**

---

intervals_ex40

**Interval estimates using SRS formulae and formulae appropriate for cluster samples**

**Description**
Simulate a population of clusters, then draw a simple random sample of clusters and construct interval estimates using incorrect SRS formulae and formulae appropriate for cluster samples.

**Usage**
intervals_ex40(groupcorr = 0, numintervals = 100, groupsize = 5, sampgroups = 10, popgroups = 5000, mu = 0, sigma = 1)
Arguments

- **grouppcor**r: The intracluster correlation coefficient rho
- **numintervals**: Number of samples to be taken from population
- **groupsize**: Number of elements in each population cluster
- **sampgroups**: Number of clusters to be sampled
- **popgroups**: Number of clusters in population
- **mu**: Mean for generating population
- **sigma**: Standard deviation for generating population

Value

- **SRS_cover_prob**: proportion of intervals using SRS formulae that include the true population mean mu
- **cl_cover_prob**: proportion of intervals using cluster sampling formulae that include the true population mean mu
- **SRS_mean_CI_width**: the average width of the interval estimates from SRS
- **Cluster_mean_CI_width**: the average width of the interval estimates from cluster sampling
- **Replicate Simulation replicate**:
- **srs_lci**: lower limit of CI from SRS
- **srs_uci**: upper limit of CI from SRS
- **clus_lci**: lower limit of CI from cluster sampling
- **clus_uci**: upper limit of CI from cluster sampling

- **scatter plot**: scatter plot of the last simulated sample
- **CI plots**: second graph shows interval estimates produced for each sample if analyzed as an SRS (with red interval not containing the true parameter), and the third shows the interval estimates produced for each sample when analyzed as a cluster sample.

Examples

- `intervals_ex40(grouppcorr = 0.3)`
- `intervals_ex40(grouppcorr = 0.7, numintervals = 100, groupsize = 5, sampgroups = 10, popgroups = 5000, mu = 0, sigma = 1)`

Description

Selected variables from the 2015-2016 National Health and Nutrition Examination Survey (NHANES). Source: Centers for Disease Control and Prevention (2017). This data set is made available for pedagogical purposes only. Anyone wishing to publish results or draw conclusions from NHANES data should obtain the original data set from the source.
Usage
data(nhanes)

Format

This data frame contains the following columns:

sdmvstra: Pseudo-stratum. (These are groups of secondary sampling units used for variance estimation on the publicly available data. Pseudo-strata and pseudo-psus are released instead of the actual strata and psus to protect the confidentiality of respondents’ information. Use sdmvstra as the variable defining the strata.)

sdmypsu: Pseudo-psu. Use sdmvpsu as the primary sampling unit (psu). (There are two pseudo-psus per pseudo-stratum, numbered 1 and 2.)

wtint2yr: Interview weight (use as weight for variables 5-12)

wtmec2yr: Mobile Examination Center weight (use as weight for any analysis involving variables 13-25)

ridstatr: Interview/examination status, (= 1 if interviewed only, = 2 if interviewed and had medical examination)

ridageyr: Age in years at screening, from 0 to 80. (Anyone with age > 80 years is recorded (top-coded) as 80. No values are missing for this variable.)

ridagemn: Age in months at screening (reported only for persons aged 24 months or younger at the time of exam, otherwise missing)

riagendr: = 1 if male, 2 if female (no missing values)

ridreth3: Race/ethnicity code (no missing values)

{ 1 = Mexican American
2 = Other Hispanic
3 = Non-Hispanic White
4 = Non-Hispanic Black
6 = Non-Hispanic Asian
7 = Other Race, Including Multi-Racial }

dmdeduc2: Education level of person interviewed (given for adults age 20+only)

{ 1 = Less than 9th grade
2 = 9th to 11th grade (including 12th grade with no diploma)
3 = High school graduate (including GED)
4 = Some college or associate’s degree
5 = College graduate or above
7 = Refused
9 = Don’t know }

dmdfmsiz: Total number of people in the family. (Values 1-6 indicate the number of people is that number; value 7 indicates 7 or more people in family. No missing values.)
indfmpir: Ratio of family income to poverty guideline. (A value less than 1 indicates the family is below the poverty threshold. Variable indfmpir is a continuous variable where values between 0 and 4.99 indicate the actual poverty ratio. A value of 5 indicates that the ratio of family income to the poverty guideline for that family is 5 or more.)

bmxwt: Weight (kg)

bmxht: Standing height (cm)

bmbmi: Body mass index (kg/m2), calculated as bmxwt/(bmxht/100)^2

bmxwaist: Waist circumference (cm)

bmxleg: Upper leg length (cm)

bmxarml: Upper arm length (cm)

bmxarmc: Upper arm circumference (cm)

bmdavsad: Average sagittal abdominal diameter (SAD, the distance from the small of the back to the upper abdomen), in cm. Calculated by averaging the SAD readings on the person (up to four).

lbxtc: Serum total cholesterol (mg/dL)

bpxpls: 60-second pulse

sbp: Average systolic blood pressure (mm Hg)

dbp: Average diastolic blood pressure (mm Hg)

bpread: Number of blood pressure readings

Details

The data files merged to create nhanes.csv can be read directly from the SAS transport files DEMO_I.XPT, BMX_I.XPT, TCHOL_I.XPT, and BPX_I.XPT from the NHANES website. This data set is provided for educational purposes only. Anyone wishing to publish or use results from analyses of NHANES data should obtain the data files directly from the source.

The blood pressure variables sbp and dbp were created as follows. In the medical examination, three consecutive blood pressure readings were obtained after participants sat quietly for 5 minutes and the maximum inflation level was determined. A fourth measurement was conducted for some persons who had an incomplete or interrupted blood pressure reading. The variables sbp and dbp were calculated by discarding the first blood pressure reading and calculating the average of the remaining valid readings. Note that some of the diastolic blood pressure readings are 0.

References

**santacruz**

**santacruz data**

**Description**

The number of seedlings in the sampled psus on Santa Cruz Island, California, in 1992 and 1994.

**Usage**

```r
data(santacruz)
```

**Format**

This data frame contains the following columns:

- `tree`: Tree number
- `seed92`: Number of seedlings in 1992
- `seed94`: Number of seedlings in 1994

**References**


---

**schools**

**schools data**

**Description**

Math and reading test results from a two-stage cluster sample of tenth-grade students. An SRS of 10 schools was selected from the 75 schools in the population, and then 20 students were sampled from each school. These data are fictional, but the summary statistics are consistent with those seen in educational studies.

**Usage**

```r
data(schools)
```

**Format**

This data frame contains the following columns:

- `schoolid`: School number (use as cluster variable)
- `gender`: Gender of student (character variable, F = female, M = male)
- `math`: Score on math test
- `reading`: Score on reading test
mathlevel: Category level for math test score:
{1 if 1 <= math <= 40
2 if 41 <= math}
readlevel: Category level for reading test score:
{1 if 1 <= read <= 32
2 if 33 <= read <= 50}
Mi: Number of students in school, M_i
finalwt: Weight for student in sample

References

Description
Selected variables from the Survey of Youth in Custody (Beck et al., 1988). Source: U.S. Department of Justice (1989). Strata 6-16 each contain one facility; the psus in those strata are residents. In strata 1-5, the psus are facilities. The number of facilities in the population (N_h) for those five facilities are: N_1 = 99, N_2 = 39, N_3 = 30, N_4 = 13, N_5 = 14. Eleven facilities are sampled from stratum 1 and seven facilities are sampled from each of strata 2 through 5.

Usage
data(syc)

Format
This data frame contains the following columns:
stratum: stratum number
psu: psu number, (= facility number for residents in strata 1-5 and person number for residents in strata 6-16)
facility: facility number
facsize: number of eligible residents in psu
finalwt: final weight
randgrp: random group number
age: age of resident (99=missing)
race: race of resident (1 = white; 2 = Black; 3 = Asian/Pacific Islander; 4 = American Indian, Aleut, Eskimo; 5 = Other; 9 = Missing)
ethnicity: 1 = Hispanic, 0 = not Hispanic, 9=missing
educ: highest grade attended before sent to correctional institution (00 = Never attended school; 01 - 12 = highest grade attended; 13 = GED; 14 = Other; 99 = missing)
gender: 1 = male, 2 = female, 9 = missing
livewith: Who did you live with most of the time you were growing up? (1 = Mother only, 2 = Father only, 3 = Both mother and father, 4 = Grandparents, 5 = Other relatives, 6 = Friends, 7 = Foster home, 8 = Agency or institution, 9 = Someone else, 99 = Blank)
famtime: Has anyone in your family, such as your mother, father, brother, sister, ever served time in jail or prison? (1 = Yes, 2 = No, 7 = Don’t know, 9 = Blank)
crimtype: most serious crime in current offense
   { 1 = violent (e.g., murder, rape, robbery, assault)
   2 = property (e.g. burglary, larceny, arson, fraud, motor vehicle theft)
   3 = drug (drug possession or trafficking)
   4 = public order (weapons violation, perjury, failure to appear in court)
   5 = juvenile status offense (truancy, running away, incorrigible behavior)
   9 = missing }
evrievol: ever put on probation or sent to correctional inst for violent offense (1 = yes, 0 = no)
numarr: number of times arrested (99 = missing)
protn: number of times on probation (99 = missing)
corri: number of times previously committed to correctional institution (99 = missing)
evrietime: Prior to being sent here did you ever serve time in a correctional institution? (1 = yes, 2 = no, 9 = missing)
prviol: =1 if previously arrested for violent offense, 0 otherwise
prprop: =1 if previously arrested for property offense, 0 otherwise
prdrug: =1 if previously arrested for drug offense, 0 otherwise
prpub: =1 if previously arrested for public order offense, 0 otherwise
prjuv: =1 if previously arrested for juvenile status offense, 0 otherwise
agefirst: age first arrested (99 = missing)
usewepn: Did you use a weapon . . . for this incident? (1 = Yes, 2 = No, 9 = Blank)
alcuse: Did you drink alcohol at all during the year before being sent here this time? (1 = Yes; 2 = No, didn’t drink during year before; 3 = No, don’t drink at all, 9 = missing)
everdrug: Ever used illegal drugs; 0 = no, 1 = yes, 9 = missing

References

## Index

<table>
<thead>
<tr>
<th>* datasets</th>
<th>impute, 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>agpop, 2</td>
<td>intervals_ex40, 14</td>
</tr>
<tr>
<td>agpps, 3</td>
<td>nhanes, 15</td>
</tr>
<tr>
<td>agsrs, 4</td>
<td>santacruz, 18</td>
</tr>
<tr>
<td>agstrat, 4</td>
<td>schools, 18</td>
</tr>
<tr>
<td>algebra, 5</td>
<td>syc, 19</td>
</tr>
<tr>
<td>anthsrs, 6</td>
<td></td>
</tr>
<tr>
<td>classes, 6</td>
<td></td>
</tr>
<tr>
<td>classpps, 7</td>
<td></td>
</tr>
<tr>
<td>college, 7</td>
<td></td>
</tr>
<tr>
<td>collegeerg, 10</td>
<td></td>
</tr>
<tr>
<td>coots, 11</td>
<td></td>
</tr>
<tr>
<td>deadtrees, 12</td>
<td></td>
</tr>
<tr>
<td>gpa, 12</td>
<td></td>
</tr>
<tr>
<td>htsrs, 13</td>
<td></td>
</tr>
<tr>
<td>htstrat, 13</td>
<td></td>
</tr>
<tr>
<td>impute, 14</td>
<td></td>
</tr>
<tr>
<td>nhanes, 15</td>
<td></td>
</tr>
<tr>
<td>santacruz, 18</td>
<td></td>
</tr>
<tr>
<td>schools, 18</td>
<td></td>
</tr>
<tr>
<td>syc, 19</td>
<td></td>
</tr>
</tbody>
</table>

agpop, 2
agpps, 3
agsrs, 4
agstrat, 4
algebra, 5
anthsrs, 6
classes, 6
classpps, 7
college, 7
collegeerg, 10
coots, 11
deadtrees, 12
gpa, 12
htsrs, 13
htstrat, 13