

# Package: SDaA (via r-universe)

August 27, 2024

**Type** Package

**Title** Sampling: Design and Analysis

**Version** 0.1-5

**Date** 2022-04-11

**Author** Tobias Verbeke

**Maintainer** Tobias Verbeke <tobias.verbeke@openanalytics.eu>

**Description** Functions and Datasets from Lohr, S. (1999), Sampling: Design and Analysis, Duxbury.

**Suggests** survey, ggplot2 (>= 0.8.2)

**License** GPL-3

**LazyData** Yes

**Collate** 'agpop.R' 'agsrs.R' 'agstrat.R' 'anthrop.R' 'anthsrs.R'  
'anthuneq.R' 'audit.R' 'books.R' 'certify.R' 'coots.R'  
'counties.R' 'divorce.R' 'golfsrs.R' 'htpop.R' 'htsrs.R'  
'htstrat.R' 'journal.R' 'lahiri.design.R' 'measles.R' 'ncvs.R'  
'nybight.R' 'otters.R' 'ozone.R' 'samples.R' 'seals.R'  
'selectrs.R' 'statepop.R' 'statepps.R' 'syc.R' 'teachers.R'  
'teachmi.R' 'teachnr.R' 'winter.R'

**Encoding** UTF-8

**RoxygenNote** 7.1.2

**NeedsCompilation** no

**Date/Publication** 2022-04-11 16:02:30 UTC

**Repository** <https://tverbeke.r-universe.dev>

**RemoteUrl** <https://github.com/cran/SDaA>

**RemoteRef** HEAD

**RemoteSha** 9fd3994ebf1a3ba76c2948cedfdd2e5c36dc2e67

## Contents

agpop	2
agsrs	3
agstrat	4
anthrop	5
anthsrs	6
anthuneq	7
audit	7
books	8
certify	9
coots	10
counties	11
divorce	12
golfsrs	13
htpop	14
htsrs	14
htstrat	15
journal	15
lahiri.design	16
measles	16
ncvs	17
nybight	18
otters	19
ozone	20
samples	21
seals	22
selectrs	22
statepop	23
statepps	24
syc	24
teachers	26
teachmi	27
teachnr	27
winter	28
<b>Index</b>	<b>30</b>

---

agpop

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

---

### Description

Data from the U.S. 1992 Census of Agriculture

### Usage

agpop

**Format**

Data frame with the following 15 variables:

**county** county name

**state** state abbreviation

**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 1000 acres or more, 1992

**largef87** number of farms with 1000 acres or more, 1987

**largef82** number of farms with 1000 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** factor with levels S (south), W (west), NC (north central), NE (northeast)

**Source**

U.S. 1992 Census of Agriculture

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 437.

---

agsrs

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

---

**Description**

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

**Usage**

agsrs

**Format**

Data frame with the following 14 variables:

**county** county name

**state** state abbreviation

**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 1000 acres or more, 1992

**largef87** number of farms with 1000 acres or more, 1987

**largef82** number of farms with 1000 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

**Source**

U.S. 1992 Census of Agriculture

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 437.

---

agstrat

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

---

**Description**

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

**Usage**

agstrat

**Format**

Data frame with the following 17 variables:

**county** county name

**state** state abbreviation

**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 1000 acres or more, 1992

**largef87** number of farms with 1000 acres or more, 1987

**largef82** number of farms with 1000 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** factor with levels S (south), W (west), NC (north central), NE (northeast)

**rn** random numbers used to select sample in each stratum

**weight** sampling weights for each county in sample

**Source**

U.S. 1992 Census of Agriculture

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 437.

---

anthrop

*Length of Left Middle Finger and Height for 3000 Criminals*

---

**Description**

Length of left middle finger and height for 3000 criminals

**Usage**

anthrop

**Format**

Data frame with the following 2 variables:

**finger** length of left middle finger (cm)

**height** height (inches)

**Source**

Macdonell, W. R. (1901). On criminal anthropometry and the identification of criminals, *Biometrika*, 1: 177–227.

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 438.

---

anthsrs

*Length of Left Middle Finger and Height for an SRS of Size 200*

---

**Description**

Length of left middle finger and height for an SRS of 200 criminals from the anthrop dataset

**Usage**

anthsrs

**Format**

Data frame with the following 2 variables:

**finger** length of left middle finger (cm)

**height** height (inches)

**Source**

Macdonell, W. R. (1901). On criminal anthropometry and the identification of criminals, *Biometrika*, 1: 177–227.

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 438.

---

anthuneq	<i>Length of Left Middle Finger and Height for an Unequal-Probability Sample of Size 200</i>
----------	--

---

**Description**

Length of left middle finger and height for an unequal-probability sample of criminals of size 200 from the anthrop dataset. The probability of selection,  $\psi[i]$ , was proportional to 24 for  $y < 65$ , 12 for  $y = 65$ , 2 for  $y = 66$  or  $67$ , and 1 for  $y > 67$ .

**Usage**

anthuneq

**Format**

Data frame with the following 3 variables:

**finger** length of left middle finger (cm)

**height** height (inches)

**prob** probability of selection

**Source**

Macdonell, W. R. (1901). On criminal anthropometry and the identification of criminals, *Biometrika*, 1: 177–227.

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 438.

---

audit	<i>Selection of Accounts for Audit in Example 6.11</i>
-------	--

---

**Description**

Selection of Accounts for Audit in Example 6.11

**Usage**

audit

**Format**

Data frame with the following 6 variables:

**account** audit unit

**bookval** book value of account

**cumbv** cumulative book value

**rn1** random number 1 selecting account

**rn2** random number 2 selecting account

**rn3** random number 3 selecting account

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 439.

---

books

*Data from Home Owner's Survey on Total Number of Books*

---

**Description**

Data from home owner's survey on total number of books

**Usage**

books

**Format**

Data frame with the following 6 variables:

**shelf** shelf number

**number** number of the book selected

**purchase** purchase cost of the book

**replace** replacement cost of book

**Note**

Used in Exercise 6 of Chapter 5.

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 439.



---

 certify

*Data from the 1994 Survey of ASA Membership on Certification*


---

**Description**

Data from the 1994 Survey of ASA Membership on Certification

**Usage**

certify

**Format**

Data frame with the following 11 variables:

**certify** should the ASA develop some form of certification? factor with levels yes, possibly, noopinion, unlikely and no

**approve** would you approve of a certification program similar to that described in the July 1993 issue of *Amstat News*? factor with levels yes, possibly, noopinion, unlikely and no

**speccert** Should there be specific certification programs for statistics subdisciplines? factor with levels yes, possibly, noopinion, unlikely and no

**wouldyou** If the ASA developed a certification program, would you attempt to become certified? factor with levels yes, possibly, noopinion, unlikely and no

**recert** If the ASA offered certification, should recertification be required every several years? factor with levels yes, possibly, noopinion, unlikely and no

**subdisc** Major subdiscipline; factor with levels BA (Bayesian), BE (business and economic), BI (biometrics), BP (biopharmaceutical), CM (computing), EN (environment), EP (epidemiology), GV (government), MR (marketing), PE (physical and engineering), QP (quality and productivity), SE (statistical education), SG (statistical graphics), SP (sports), SR (survey research), SS (social statistics), TH (teaching statistics in health sciences), O (other)

**college** Highest collegiate degree; factor with levels B (BS or BA), M (MS), N (none), P (PhD) and O (other)

**employ** Employment status; factor with levels E (employed), I (in school), R (retired), S (self-employed), U (unemployed) and O (other)

**workenv** Primary work environment; factor with levels A (academia), G (government), I (industry), O (other)

**workact** Primary work activity; factor with levels C (consultant), E (educator), P (practitioner), R (researcher), S (student) and O (other)

**yearsmem** For how many years have you been a member of ASA?

**Note**

The full dataset is on Statlib

## References

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 439. <http://lib.stat.cmu.edu/asacert/certsurvey>

---

coots

*Egg Size from Coots*

---

## Description

Selected information on egg size from coots, from a study by Arnold (1991). Data courtesy of Todd Arnold.

## Usage

coots

## Format

Data frame with the following 11 variables:

**clutch** clutch number from which eggs were subsampled

**csize** number of eggs in clutch (Mi)

**length** length of egg (mm)

**breadth** maximum breadth of egg (mm)

**volume** calculated as  $0.00507 \times \text{length} \times \text{breadth}^2$

**tmt** received supplemental feeding? factor with levels no and yes

## Note

Not all observations are used for this data set, so results may not agree with those in Arnold (1991)

## Source

Arnold, T.W. (1991). Intraclutch variation in egg size of American Coots, *The Condor*, 93: 19–27

## References

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 440.

---

counties

*Data from an SRS of 100 of the 3141 Counties in the U.S.*

---

**Description**

Data from an SRS of 100 of the 3141 Counties in the U.S.

**Usage**

counties

**Format**

Data frame with the following 18 variables:

**RN** random number used to select the country

**state** state (two-letter abbreviation)

**county** county

**landarea** land area, 1990 (square miles)

**totpop** total population, 1992

**physician** active nonfederal physicians on Jan. 1, 1990

**enroll** school enrollment in elementary or high school, 1990

**percpub** percent of school enrollment in public schools

**civlabor** civilian labor force, 1991

**unemp** number unemployed, 1991

**farmpop** farm population, 1990

**numfarm** number of farms, 1987

**farmacre** acreage in farms, 1987

**fedgrant** total expenditures in federal funds and grants, 1992 (millions of dollars)

**fedciv** civilians employed by federal government, 1990

**milit** military personnel, 1990

**veterans** number of veterans, 1990

**percviet** percentage of veterans from Vietnam era, 1990

**Source**

U.S. Bureau of Census, 1994

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 440.

divorce

*Data from a Sample of Divorce Records***Description**

Data from a sample of divorce records for states in the Divorce Registration Area (National Center for Health Statistics 1987)

**Usage**

divorce

**Format**

Data frame with the following 20 variables:

**state** state name

**abbrev** state abbreviation

**samprate** sampling rate for state

**numrecs** number of records sampled in state

**hsblt20** number of records in sample with husband's age < 20

**hsb2024** number of records with 20 <= husband's age <= 24

**hsb2529** number of records with 25 <= husband's age <= 29

**hsb3034** number of records with 30 <= husband's age <= 34

**hsb3539** number of records with 35 <= husband's age <= 39

**hsb4044** number of records with 40 <= husband's age <= 44

**hsb4549** number of records with 45 <= husband's age <= 49

**hsbge50** number of records with wife's age >= 50

**wflt20** number of records in sample with wife's age < 20

**wf2024** number of records with 20 <= wife's age <= 24

**wf2529** number of records with 25 <= wife's age <= 29

**wf3034** number of records with 30 <= wife's age <= 34

**wf3539** number of records with 35 <= wife's age <= 39

**wf4044** number of records with 40 <= wife's age <= 44

**wf4549** number of records with 45 <= wife's age <= 49

**wfge50** number of records with wife's age >= 50

**Source**

National Center of Health Statistics (1987). TODO

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 440.

---

golfsrs

*Simple Random Sample of Golf Courses*

---

**Description**

Simple Random Sample (SRS) of 120 golf courses taken from the population of the (now defunct) Website [www.golfcourse.com](http://www.golfcourse.com)

**Usage**

golfsrs

**Format**

Data frame with the following 16 variables:

**RN** random number used to select golf course for sample

**state** state name

**holes** number of holes

**type** type of course; factor with levels priv (private), semi (semi-private), pub (public), mili (military) and res (resort)

**yearblt** year the course was built

**wkday18** greens fee for 18 holes during week

**wkday9** greens fee for 9 holes during week

**wkend18** greens fee for 18 holes on weekend

**wkend9** greens fee for 9 holes on weekend

**backtee** back-tee yardage

**rating** course rating

**par** par for course

**cart18** golf cart rental fee for 18 holes

**cart9** golf cart rental fee for 9 holes

**caddy** Are caddies available? factor with levels yes and no

**pro** Is a golf pro available? factor with levels yes and no

**Source**

The now defunct website [golfcourse.com](http://www.golfcourse.com) (<https://web.archive.org/web/19991108203827/http://golfcourse.com/>)

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and TODO.

---

htpop

*Height and gender of 2000 persons in an artificial population*

---

**Description**

Height and gender of 2000 persons in an artificial population

**Usage**

htpop

**Format**

**height** height of person, cm

**gender** factor with levels F (female) and M (male)

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. 230–234 and 441.

---

htsrs

*Height and gender for an SRS of 200 persons, taken from htpop*

---

**Description**

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

**Usage**

htsrs

**Format**

**rn** random number used to select the unit

**height** height of person, cm

**gender** factor with levels F (female) and M (male)

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. 230–234 and 442.

---

htstrat	<i>Height and gender for a stratified random sample from htpop</i>
---------	--

---

**Description**

Height and gender for a stratified random sample of 160 women and 40 men taken from the htpop population

**Usage**

htstrat

**Format**

**rn** random number used to select the unit  
**height** height of person, cm  
**gender** factor with levels F (female) and M (male)

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. 230–234 and 442.

---

journal	<i>Types of Sampling Used for Articles in a Sample of Journals</i>
---------	--

---

**Description**

Types of Sampling Used for Articles in a Sample of Journals

**Usage**

journal

**Format**

Data frame with the following 3 variables:  
**numemp** number of articles in 1988 that used sampling  
**prob** number of articles that used probability sampling  
**nonprob** number of articles that used nonprobability sampling

**Source**

Jacoby and Handlin (1991). TODO

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 442.

---

`lahiri.design`*Draw Samples Using Lahiri's Method*

---

**Description**

Draw Samples Using Lahiri's Method

**Usage**

```
lahiri.design(release, n, cnames = seq(along = release))
```

**Arguments**

<code>release</code>	vector of relative sizes of population PSUs
<code>n</code>	desired sample size
<code>cnames</code>	vector of PSU names for population

**Value**

clusters vector of `n` PSUs selected with replacement and with probability proportional to `release`

**Note**

Original code from Lohr (1999), p. 452 – 453.

**Author(s)**

Sharon Lohr, slightly modified by Tobias Verbeke

**References**

Lahiri, D. B. (1951). A method of sample selection providing unbiased ratio estimates, *Bulletin of the International Statistical Institute*, 33: 133 – 140.

---

`measles`*Survey of Parents of Children Non-Immunized against Measles*

---

**Description**

Roberts et al. (1995) report on the results of a survey of parents whose children had not been immunized against measles during a recent campaign to immunize all children in the first five years of secondary school.

**Usage**

```
measles
```



**Format**

Data frame with 11 variables. A parent who refused consent (variable 4) was asked why, with responses in variables 5-10. A parent could give more than one reason for not having the child immunized.

**school** school attended by child

**form** parent received consent form

**returnf** parent returned consent form

**consent** parent gave consent for measles immunization

**hadmeas** child had already had measles

**previmm** child had been immunized against measles

**sideeff** parent concerned about side effects

**gp** parent wanted GP (general practitioner) to give vaccine

**noshot** child did not want injection

**notser** parent thought measles not serious illness

**gpadv** GP advised that vaccine was not needed

**Note**

The original data were unavailable; univariate and multivariate summary statistics from these artificial data, however, are consistent with those in the paper.

**Source**

Roberts R. J. et al. (1995). Reasons for non-uptake of measles, mumps, and rubella catch up immunisation in a measles epidemic and side effects of the vaccine, *British Medical Journal*, 310, 1629–1632.

**References**

Lohr (1999). *Sampling: Design and Analysis*, Duxbury, p. 442.

---

 ncvs

*Victimization Incidents in the July-December 1989 NCVS*


---

**Description**

Selected variables for victimization incidents in the July-December 1989 NCVS. Note that some variables were recoded from the original data file.

**Usage**

ncvs

**Format**

Data frame with the following seven variables:

**wt** incident weight

**sex** factor with levels male and female

**violent** violent crime? factor with levels no and yes

**injury** did the victim have injuries? factor with levels no and yes

**medcare** factor with levels yes if the victim received medical care and no otherwise

**reppol** was the incident reported to the police? factor with levels yes and no

**numoff** number of offenders involved in crime; factor with levels one, more (more than one) and dontknow

**Source**

Incident-level concatenated file, NCS8864I, in NCJ-130915, U.S. Department of Justice 1991.

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 443.

---

 nybight

---

*Data Collected in the New York Bight*


---

**Description**

Data collected in the New York Bight for June 1974 and June 1975 (Wilk et al. 1977)

**Usage**

nybight

**Format**

Data frame with the following 7 variables:

**year** year

**stratum** stratum membership, based on depth

**catchnum** number of fish caught during trawl

**catchwt** total weight (kg) of fish caught during trawl

**numsp** number of species of fish caught during trawl

**depth** depth of station (m)

**temp** surface temperature (degrees Celsius)

**Note**

Two of the original strata were combined because of insufficient sample sizes.

**Source**

Wilk, S.J. et al. (1977). Fishes and associated environmental data collected in New York bight, June 1974 - June 1975. NOAA Technical Report NMFS SSRF-716. Washington, D.C: Government Printing Office.

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 443.

---

otters

*Otters Data*

---

**Description**

Data on number of holts (dens) in Shetland, United Kingdom used in Kruuk et al. (1989). (Data courtesy of Hans Kruuk).

**Usage**

otters

**Format**

Data frame with the following three variables:

**section** coastline section

**habitat** type of habitat (stratum)

**holts** number of holts

**Source**

Kruuk, H.A. et al. (1989). An estimate of numbers and habitat preferences of otters *Lutra lutra* in Shetland, UK., Biological Conservation, 49: 241–254.

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 443.

---

ozone

*Ozone Readings from Eskdalemuir, for 1994 and 1995*

---

**Description**

Hourly ozone readings in parts per billion (ppb) from Eskdalemuir, Scotland, for 1994 and 1995

**Usage**

ozone

**Format**

Data frame with the following 25 variables:

**date** date (day/month/year)

**GMT1** ozone reading at 1:00 GMT

**GMT2** ozone reading at 2:00 GMT

**GMT3** ozone reading at 3:00 GMT

**GMT4** ozone reading at 4:00 GMT

**GMT5** ozone reading at 5:00 GMT

**GMT6** ozone reading at 6:00 GMT

**GMT7** ozone reading at 7:00 GMT

**GMT8** ozone reading at 8:00 GMT

**GMT9** ozone reading at 9:00 GMT

**GMT10** ozone reading at 10:00 GMT

**GMT11** ozone reading at 11:00 GMT

**GMT12** ozone reading at 12:00 GMT

**GMT13** ozone reading at 13:00 GMT

**GMT14** ozone reading at 14:00 GMT

**GMT15** ozone reading at 15:00 GMT

**GMT16** ozone reading at 16:00 GMT

**GMT17** ozone reading at 17:00 GMT

**GMT18** ozone reading at 18:00 GMT

**GMT19** ozone reading at 19:00 GMT

**GMT20** ozone reading at 20:00 GMT

**GMT21** ozone reading at 21:00 GMT

**GMT22** ozone reading at 22:00 GMT

**GMT23** ozone reading at 23:00 GMT

**GMT24** ozone reading at 24:00 GMT

**Source**

Air Quality Information Centre: retrieved from a now defunct URL (<http://www.aeat.co.uk>)

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 443.

---

samples

*Samples Dataset*

---

**Description**

All possible SRSs that can be generated from the population in Example 2.1 of Lohr(1999).

**Usage**

samples

**Format**

Data frame with the following 10 variables:

**snum** sample number

**unit1** first unit in sample

**unit2** second unit in sample

**unit3** third unit in sample

**unit4** fourth unit in sample

**value1** value for first unit in sample

**value2** value for second unit in sample

**value3** value for third unit in sample

**value4** value for fourth unit in sample

**that** t hat, i.e. estimate of the population total based on the given sample

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. 26–27 and 444.

---

seals                      *Breathing Holes of Seals*

---

**Description**

Data on number of breathing holes found in sampled areas of Svalbard fjords, reconstructed from summary statistics given in Lydersen and Ryg (1991)

**Usage**

seals

**Format**

Data frame with the following 2 variables:

**zone** zone number for sampled area

**holes** number of breathing holes Imjak found in area

**Note**

The data are used in Chapter 4, Exercise 11.

**Source**

Lydersen, C. and Ryg, M. (1991). Evaluating breeding habitat and populations of ringed seals *Phoca hispida* in Svalbard fjords, *Polar Record*, 27: 223–228.

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 444.

---

selectrs                      *Steps used in Selecting an SRS*

---

**Description**

Steps used in selecting the simple random sample (SRS) in Example 2.4 of Lohr(1999).

**Usage**

selectrs

**Format**

Data frame with the following 5 variables:

- a** random number generated between 0 and 1
- b** ceiling(3048\*RN), with RN the random number in column a
- c** distinct values in column b
- d** new values generated to replace duplicates in b
- e** final set of distinct values to be used in sample

**Note**

the set of indices in column e was used to select observations from agpop into dataset agsrs.

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. 31–34 and 444.

---

 statepop

*Unequal-Probability Sample of Counties in the US*


---

**Description**

counties selected with probability proportional to 1992 population

**Usage**

statepop

**Format**

- state** state abbreviation
- county** county
- landarea** land area of county, 1990 (square miles)
- popn** population of county, 1992
- phys** number of physicians, 1990
- farmpop** farm population, 1990
- numfarm** number of farms, 1987
- farmacre** number of acres devoted to farming, 1987
- veterans** number of veterans, 1990
- percviet** percent of veterans from Vietnam era, 1990

**Source**

City and Counties Book, 1994

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. 190 – 192 and 444.

---

statepps	<i>Information on States</i>
----------	------------------------------

---

**Description**

Number of counties, land area, and population for the 50 states plus the District of Columbia

**Usage**

statepps

**Format**

Date frame with the following 7 variables:

**state** state name

**counties** number of counties in state

**cumcount** cumulative number of counties

**landarea** land area of state, 1990 (square miles)

**cumland** cumulative land area

**popn** population of state, 1992

**cumpopn** cumulative population

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 445.

---

syc	<i>Survey of Youth in Custody, 1987</i>
-----	---

---

**Description**

The 1987 Survey of Youth in Custody sampled juveniles and young adults in long-term, state-operated juvenile institutions. Residents of facilities at the end of 1987 were interviewed about family background, previous criminal history, and drug and alcohol use. Selected variables from the survey are contained in the syc data frame.

**Usage**

syc



**Format**

**stratum** stratum number

**psu** psu (facility) number

**psusize** number of eligible residents in psu

**initwt** initial weight

**finalwt** final weight

**randgrp** random group number

**age** age of resident

**race** race of resident: factor with levels 1 (white), 2 (black), 3 (Asian/Pacific Islander), 4 (American Indian, Aleut, Eskimo), 5 (other)

**ethnicity** ethnicity; factor with levels hispanic and notHispanic

**educ** highest grade before sent to correctional institution; factor with levels 0 (never attended), 1-12 (highest grade attended), 13 (GED), 14 (other)

**sex** factor with levels male and female

**livewith** factor with levels 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)

**famtime** Has anyone in your family, such as your mother, father, brother, sister, ever served time in jail or prison? factor with levels yes and no

**crimtype** most serious crime in current offense; one of violent (e.g. murder, rape, robbery, assault), property (e.g. burglary, larceny, arson, fraud, motor vehicle theft), drug (drug possession or trafficking), publicorder (weapons violation, perjury, failure to appear in court), juvenile (juvenile-status offense, e.g. truancy, running away, incorrigible behavior)

**everviol** Ever put on probation or sent to correctional institution for violent offense? factor with levels no and yes

**numarr** number of times arrested (integer)

**probtn** number of times on probation

**corrinst** number of times previously committed to correctional institution

**evertime** Prior to being sent here, did you ever serve time in a correctional institution? factor with levels yes and no

**prviol** previously arrested for violent offense; factor with levels no and yes

**prprop** previously arrested for property offense; factor with levels no and yes

**prdrug** previously arrested for drug offense; factor with levels no and yes

**prpub** previously arrested for public-order offense; factor with levels no and yes

**prjuv** previously arrested for juvenile-status offense; factor with levels no and yes

**agefirst** age first arrested (integer)

**usewepn** Did you use a weapon... for this incident? factor with levels yes and no

**alcuse** Did you drink alcohol at all during the year before being sent here this time? factor with levels yes, noduringyear, noatall

**everdrug** Ever used illegal drugs? factor with levels no, yes

**Source**

Inter-University Consortium on Political and Social Research, NCJ-130915, U.S. Department of Justice 1989.

**References**

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. 235–239 and 445.

---

teachers

*Elementary School Teacher Workload Data*

---

**Description**

Selected variables from a study on elementary school teacher workload in Maricopa County, Arizona.

**Usage**

teachers

**Format**

data frame with the following 6 variables:

**dist** school district size; factor with levels large and me/sm (medium/small)

**school** school identifier

**hrwork** number of hours required to work at school per week

**size** class size

**preprmin** minutes spent per week in school on preparation

**assist** minutes per week that a teacher's aide works with the teacher in the classroom

**Note**

The study is described in Exercise 16 of Chapter 15. The psu sizes are given in [teachmi](#). The large stratum had 245 schools; the small/medium stratum had 66 schools.

**Source**

Data courtesy of Rita Gnap (1995).

**References**

Gnap, R. (1995). Teacher load in Arizona elementary school districts in Maricopa County. Ph.D. diss., Arizona State University

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 446.

---

teachmi

*Cluster Sizes for Elementary School Teacher Workload Data*

---

**Description**

Cluster sizes for the study on elementary school teacher workload in Maricopa County, Arizona.

**Usage**

teachmi

**Format**

data frame with the following 6 variables:

**dist** school district size; factor with levels large and me/sm (medium/small)

**school** school identifier

**popteach** number of teachers in that school

**ssteach** number of surveys returned from that school

**Note**

The study is described in Exercise 16 of Chapter 15. The actual date are given in [teachers](#).

**Source**

Data courtesy of Rita Gnap (1995).

**References**

Gnap, R. (1995). Teacher load in Arizona elementary school districts in Maricopa County. Ph.D. diss., Arizona State University

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 446.

---

teachnr

*Follow-Up Study of Nonrespondents from Gnap (1995)*

---

**Description**

Follow-up study of nonrespondents from the Gnap (1995) study on the workload of elementary school teachers in Maricopa County, Arizona.

**Usage**

teachnr

**Format**

data frame with the following 6 variables:

**hrwork** number of hours required to work at school per week

**size** class size

**preprmin** minutes spent per week in school on preparation

**assist** minutes per week that a teacher's aide works with the teacher in the classroom

**Note**

The study is described in Exercise 16 of Chapter 15. The actual date are given in [teachers](#). Cluster size data for the original study are given in [teachmi](#).

**Source**

Data courtesy of Rita Gnap (1995).

**References**

Gnap, R. (1995). Teacher load in Arizona elementary school districts in Maricopa County. Ph.D. diss., Arizona State University

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 446.

---

winter

*ASU Winter Closure Survey*

---

**Description**

Selected variables from the Arizona State University Winter Closure Survey, taken in January 1995. This survey was taken to investigate the attitudes and opinions of university employees toward the closing of the university between December 25 and January 1.

**Usage**

winter

**Format**

data frame with the following 6 variables:

**class** stratum number; factor with levels *faculty*, *classstaff* (classified staff), *admstaff* (administrative staff) and *acprof* (academic professional)

**yearasu** factor with levels 1 (1-2 years), 2 (3-4 years), 3 (5-9 years), 4 (10-14 years) and 5 (15 or more years)

**vacation** In the past, have you *usually* taken vacation days in the entire period between December 25 and January 1? factor with levels *no* and *yes*

- work** Did you work on campus during Winter Break Closure? factor with levels no and yes
- havediff** Did the Winter Break Closure cause you any difficulty/concerns? factor with levels no and yes
- negaeffe** Did the Winter Break Closure *negatively* affect your work productivity? factor with levels no and yes
- ownsupp** I was unable to obtain staff support in my department/office. factor with levels yes and no
- othersup** I was unable to obtain staff support in other departments/offices. factor with levels yes and no
- utility** I was unable to access computers, copy machine, etc. in my department/office. factor with levels yes and no
- environ** I was unable to endure environmental conditions - e.g., not properly climatized. factor with levels yes and no
- uniserve** I was unable to access university services necessary to my work; factor with levels yes and no
- workelse** I was unable to work on my assignments because I work in another department/office; factor with levels yes and no
- offclose** I was unable to work on my assignments because my office was closed; factor with levels yes and no
- treatsta** compared to other departments/offices, I feel staff in my department/office were treated fairly; factor with levels strongagr (strongly agree), agree, undecided, disagree, strdisagr (strongly disagree)
- treatme** compared to other people working in my department/office, I feel I was treated fairly; factor with levels strongagr (strongly agree), agree, undecided, disagree, strdisagr (strongly disagree)
- process** How satisfied are you with the process used to inform staff about Winter Closure? factor with levels verysat (very satisfied), satisfied, undecided, dissatisfied and verydissat (very dissatisfied)
- satbreak** How satisfied are you with the fact that ASU had a Winter Break Closure this year? factor with levels verysat (very satisfied), satisfied, undecided, dissatisfied and verydissat (very dissatisfied)
- breakaga** Would you want to have Winter Break Closure again? factor with levels no and yes

### Source

courtesy of the ASU Office of University Evaluation.

### References

Lohr (1999). Sampling: Design and Analysis, Duxbury, p. TODO and 447–448.

# Index

## \* datasets

agpop, 2  
agsrs, 3  
agstrat, 4  
anthrop, 5  
anthsrs, 6  
anthuneq, 7  
audit, 7  
books, 8  
certify, 9  
coots, 10  
counties, 11  
divorce, 12  
golfsrs, 13  
htpop, 14  
htsrs, 14  
htstrat, 15  
journal, 15  
measles, 16  
ncvs, 17  
nybight, 18  
otters, 19  
ozone, 20  
samples, 21  
seals, 22  
selectrs, 22  
statepop, 23  
statepps, 24  
syc, 24  
teachers, 26  
teachmi, 27  
teachnr, 27  
winter, 28

agpop, 2  
agsrs, 3  
agstrat, 4  
anthrop, 5  
anthsrs, 6  
anthuneq, 7

audit, 7  
books, 8  
certify, 9  
coots, 10  
counties, 11  
divorce, 12  
golfsrs, 13  
htpop, 14  
htsrs, 14  
htstrat, 15  
journal, 15  
lahiri.design, 16  
measles, 16  
ncvs, 17  
nybight, 18  
otters, 19  
ozone, 20  
samples, 21  
seals, 22  
selectrs, 22  
statepop, 23  
statepps, 24  
syc, 24  
teachers, 26, 27, 28  
teachmi, 26, 27, 28  
teachnr, 27  
winter, 28