Package 'wooldridge'

July 12, 2026

Type Package

Title 115 Data Sets from ``Introductory Econometrics: A Modern Approach, 8e" by Jeffrey M. Wooldridge

Version 1.4-7

Description Students learning both econometrics and R may find the introduction to both challenging. The wooldridge data package aims to lighten the task by efficiently loading any data set found in the text with a single command. Data sets have been compressed to a fraction of their original size. Documentation files contain page numbers, the original source, time of publication, and notes from the author suggesting avenues for further analysis and research. If one needs an introduction to R model syntax, a vignette contains solutions to examples from chapters of the text.
Data sets are from the 7th edition (Wooldridge 2020, ISBN-13 978-1-337-55886-0), and are backwards compatible with all previous versions of the text.

Depends R (>= 3.5.0)

License GPL-3

Encoding UTF-8

LazyData true

Suggests knitr, rmarkdown, stargazer, tinytest, xts, zoo

VignetteBuilder knitr

LazyDataCompression xz

URL https://justinmshea.github.io/wooldridge/

BugReports https://github.com/JustinMShea/wooldridge/issues/

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Justin M. Shea [aut, cre],
Kennth H. Brown [ctb]

Maintainer Justin M. Shea <jshea01@uic.edu>

Repository CRAN

Date/Publication 2026-07-12 07:00:03 UTC

Contents

admnrev . . . 4
affairs . . . 5
airfare . . . 6
alcohol . . . 7
apple . . . 9
approval . . . 10
athlet1 . . . 12
athlet2 . . . 13
attend . . . 14
audit . . . 15
barium . . . 16
beauty . . . 17
benefits . . . 19
beveridge . . . 20
big9salary . . . 20
bwght . . . 22
bwght2 . . . 23
campus . . . 25
card . . . 26
catholic . . . 27
cement . . . 29
census2000 . . . 30
ceosal1 . . . 31
ceosal2 . . . 32
charity . . . 33
consump . . . 34
corn . . . 35
countymurders . . . 36
cps78_85 . . . 38
cps91 . . . 39
crime1 . . . 40
crime2 . . . 41
crime3 . . . 43
crime4 . . . 44
discrim . . . 46
driving . . . 48
earns . . . 50
econmath . . . 51
elem94_95 . . . 52
engin . . . 53
expendshares . . . 54
ezanders . . . 55
ezunem . . . 57
fair . . . 58
fertil1 . . . 60
fertil2 . . . 61

fertil3 . . . 63
fish . . . 64
fringe . . . 65
gpa1 . . . 67
gpa2 . . . 69
gpa3 . . . 70
happiness . . . 71
hprice1 . . . 72
hprice2 . . . 73
hprice3 . . . 75
hseinv . . . 76
htv . . . 77
infmrt . . . 78
injury . . . 79
intdef . . . 81
intqrt . . . 82
inven . . . 83
jtrain . . . 84
jtrain2 . . . 85
jtrain3 . . . 87
jtrain98 . . . 88
k401k . . . 89
k401ksubs . . . 90
kielmc . . . 91
labsup . . . 92
lawsch85 . . . 94
loanapp . . . 95
lowbrth . . . 97
mathpnl . . . 99
meap00_01 . . . 101
meap01 . . . 102
meap93 . . . 103
meapsingle . . . 104
minwage . . . 105
mlb1 . . . 107
mroz . . . 109
murder . . . 111
nbasal . . . 112
ncaa_rpi . . . 113
nyse . . . 114
okun . . . 115
openness . . . 116
pension . . . 117
phillips . . . 118
pntsprd . . . 119
prison . . . 120
prminwge . . . 122
rdchem . . . 123

rdtelec 124
recid 125
rental 126
return 127
saving 128
school93_98 129
sleep75 130
slp75_81 132
smoke 133
traffic1 134
traffic2 135
twoyear 137
volat 138
vote1 140
vote2 141
voucher 142
wage1 143
wage2 145
wagepan 146
wageprc 148
wine 149

Index **150**

`admnrev` *admnrev*

Description

Wooldridge Source: Data from the National Highway Traffic Safety Administration: "A Digest of State Alcohol-Highway Safety Related Legislation," U.S. Department of Transportation, NHTSA. I used the third (1985), eighth (1990), and 13th (1995) editions. Data loads lazily.

Usage

```
data('admnrev')
```

Format

A data.frame with 153 observations on 5 variables:

- **state:** state postal code
- **year:** 85, 90, or 95
- **admnrev:** =1 if admin. revoc. law
- **daysfrst:** days suspended, first offense
- **daysscnd:** days suspended, second offense

Notes

This is not so much a data set as a summary of so-called "administrative per se" laws atthe state level, for three different years. It could be supplemented with drunk-driving fatalities for a nice econometric analysis. In addition, the data for 2000 or later years can be added, forming the basis for a term project. Many other explanatory variables could be included. Unemployment rates, state-level tax rates on alcohol, and membership in MADD are just a few possibilities.

Used in Text: not used

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(admnrev)
```

`affairs` *affairs*

Description

Wooldridge Source: R.C. Fair (1978), "A Theory of Extramarital Affairs," Journal of Political Economy 86, 45-61, 1978. I collected the data from Professor Fair's web cite at the Yale University Department of Economics. He originally obtained the data from a survey by Psychology Today. Data loads lazily.

Usage

```
data('affairs')
```

Format

A data.frame with 601 observations on 19 variables:

- **id:** identifier
- **male:** =1 if male
- **age:** in years
- **yrsmarr:** years married
- **kids:** =1 if have kids
- **relig:** 5 = very relig., 4 = somewhat, 3 = slightly, 2 = not at all, 1 = anti
- **educ:** years schooling
- **occup:** occupation, reverse Hollingshead scale
- **ratemarr:** 5 = vry hap marr, 4 = hap than avg, 3 = avg, 2 = smewht unhap, 1 = vry unhap

- **naffairs:** number of affairs within last year
- **affair:** =1 if had at least one affair
- **vryhap:** ratemarr == 5
- **hapavg:** ratemarr == 4
- **avgmarr:** ratemarr == 3
- **unhap:** ratemarr == 2
- **vryrel:** relig == 5
- **smerel:** relig == 4
- **slghtrel:** relig == 3
- **notrel:** relig == 2

Notes

This is an interesting data set for problem sets starting in Chapter 7. Even though naffairs (number of extramarital affairs a woman reports) is a count variable, a linear model can be used to model its conditional mean as an approximation. Or, you could ask the students to estimate a linear probability model for the binary indicator affair, equal to one of the woman reports having any extramarital affairs. One possibility is to test whether putting the single marriage rating variable, ratemarr, is enough, against the alternative that a full set of dummy variables is needed; see pages 239-240 for a similar example. This is also a good data set to illustrate Poisson regression (using naffairs) in Section 17.3 or probit and logit (using affair) in Section 17.1.

Used in Text: not used

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(affairs)
```

`airfare` *airfare*

Description

Wooldridge Source: Jiyoung Kwon, a former doctoral student in economics at MSU, kindly provided these data, which she obtained from the Domestic Airline Fares Consumer Report by the U.S. Department of Transportation. Data loads lazily.

Usage

```
data('airfare')
```

Format

A data.frame with 4596 observations on 14 variables:

- **year:** 1997, 1998, 1999, 2000
- **id:** route identifier
- **dist:** distance, in miles
- **passen:** avg. passengers per day
- **fare:** avg. one-way fare, $
- **bmktshr:** fraction market, biggest carrier
- **ldist:** log(distance)
- **y98:** =1 if year == 1998
- **y99:** =1 if year == 1999
- **y00:** =1 if year == 2000
- **lfare:** log(fare)
- **ldistsq:** ldist^2
- **concen:** = bmktshr
- **lpassen:** log(passen)

Notes

This data set nicely illustrates the different estimates obtained when applying pooled OLS, random effects, and fixed effects.

Used in Text: pages 506-507, 581

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(airfare)
```

---

alcohol *alcohol*

Description

Wooldridge Source: Terza, J.V. (2002), "Alcohol Abuse and Employment: A Second Look," Journal of Applied Econometrics 17, 393-404. I obtained these data from the Journal of Applied Econometrics data archive at http://qed.econ.queensu.ca/jae/. Data loads lazily.

Usage

```
data('alcohol')
```

Format

A data.frame with 9822 observations on 33 variables:

- **abuse:** =1 if abuse alcohol
- **status:** out of workforce = 1; unemployed = 2, employed = 3
- **unemrate:** state unemployment rate
- **age:** age in years
- **educ:** years of schooling
- **married:** =1 if married
- **famsize:** family size
- **white:** =1 if white
- **exhealth:** =1 if in excellent health
- **vghealth:** =1 if in very good health
- **goodhealth:** =1 if in good health
- **fairhealth:** =1 if in fair health
- **northeast:** =1 if live in northeast
- **midwest:** =1 if live in midwest
- **south:** =1 if live in south
- **centcity:** =1 if live in central city of MSA
- **outercity:** =1 if in outer city of MSA
- **qrt1:** =1 if interviewed in first quarter
- **qrt2:** =1 if interviewed in second quarter
- **qrt3:** =1 if interviewed in third quarter
- **beertax:** state excise tax, $ per gallon
- **cigtax:** state cigarette tax, cents per pack
- **ethanol:** state per-capita ethanol consumption
- **mothalc:** =1 if mother an alcoholic
- **fathalc:** =1 if father an alcoholic
- **livealc:** =1 if lived with alcoholic
- **inwf:** =1 if status > 1
- **employ:** =1 if employed
- **agesq:** age^2
- **beertaxsq:** beertax^2
- **cigtaxsq:** cigtax^2
- **ethanolsq:** ethanol^2
- **educsq:** educ^2

Used in Text

page 629

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(alcohol)
```

`apple` *apple*

Description

Wooldridge Source: These data were used in the doctoral dissertation of Jeffrey Blend, Department of Agricultural Economics, Michigan State University, 1998. The thesis was supervised by Professor Eileen van Ravensway. Drs. Blend and van Ravensway kindly provided the data, which were obtained from a telephone survey conducted by the Institute for Public Policy and Social Research at MSU. Data loads lazily.

Usage

```
data('apple')
```

Format

A data.frame with 660 observations on 17 variables:

- **id:** respondent identifier
- **educ:** years schooling
- **date:** date: month/day/year
- **state:** home state
- **regprc:** price of regular apples
- **ecoprc:** price of ecolabeled apples
- **inseason:** =1 if interviewed in Nov.
- **hhsize:** household size
- **male:** =1 if male
- **faminc:** family income, thousands
- **age:** in years
- **reglbs:** quantity regular apples, pounds
- **ecolbs:** quantity ecolabeled apples, lbs

- **numlt5:** # in household younger than 5
- **num5_17:** # in household 5 to 17
- **num18_64:** # in household 18 to 64
- **numgt64:** # in household older than 64

Notes

This data set is close to a true experimental data set because the price pairs facing a family were randomly determined. In other words, the family head was presented with prices for the eco-labeled and regular apples, and then asked how much of each kind of apple the family would buy at the given prices. As predicted by basic economics, the own price effect is negative (and strong) and the cross price effect is positive (and strong). While the main dependent variable, ecolbs, piles up at zero, estimating a linear model is still worthwhile. Interestingly, because the survey design induces a strong positive correlation between the prices of eco-labeled and regular apples, there is an omitted variable problem if either of the price variables is dropped from the demand equation. A good exam question is to show a simple regression of ecolbs on ecoprc and then a multiple regression on both prices, and ask students to decide whether the price variables must be positively or negatively correlated.

Used in Text: pages 201, 223, 266, 626-627

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(apple)
```

`approval` *approval*

Description

Wooldridge Source: Harbridge, L., J. Krosnick, and J.M. Wooldridge (forthcoming), "Presidential Approval and Gas Prices: Sociotropic or Pocketbook Influence?" in New Explorations in Political Psychology, ed. J. Krosnick. New York: Psychology Press (Taylor and Francis Group). Professor Harbridge kindly provided the data, of which I have used a subset. Data loads lazily.

Usage

```
data('approval')
```

Format

A data.frame with 78 observations on 16 variables:

- **id:** id
- **month:** month
- **year:** year
- **sp500:** S&P 500 index
- **cpi:** Consumer Price Index
- **cpifood:** CPI for food
- **approve:** Gallup approval rate, percent
- **gasprice:** average gas price, cents
- **unemploy:** unemployment rate, percent
- **katrina:** =1 for three months after Hurricane Katrina
- **rgasprice:** real gas price, 100*(gasprice/cpi)
- **lrgasprice:** log(rgasprice)
- **sep11:** =1 for 09/2001 and two months following
- **iraqinvade:** =1 for three months after Iraq invasion
- **lsp500:** log(sp500)
- **lcpifood:** log(cpifood)

Used in Text

343, 371, 400

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781305270107

Examples

```
str(approval)
```

athlet1 *athlet1*

Description

Wooldridge Sources: Peterson's Guide to Four Year Colleges, 1994 and 1995 (24th and 25th editions). Princeton University Press. Princeton, NJ. The Official 1995 College Basketball Records Book, 1994, NCAA. 1995 Information Please Sports Almanac (6th edition). Houghton Mifflin. New York, NY. Data loads lazily.

Usage

```
data('athlet1')
```

Format

A data.frame with 118 observations on 23 variables:

- **year:** 1992 or 1993
- **apps:** # applics for admission
- **top25:** perc frsh class in 25 hs perc
- **ver500:** perc frsh >= 500 on verbal SAT
- **mth500:** perc frsh >= 500 on math SAT
- **stufac:** student-faculty ratio
- **bowl:** = 1 if bowl game in prev yr
- **btitle:** = 1 if men's cnf chmps prv yr
- **finfour:** = 1 if men's final 4 prv yr
- **lapps:** log(apps)
- **d93:** =1 if year = 1993
- **avg500:** (ver500+mth500)/2
- **cfinfour:** change in finfour
- **clapps:** change in lapps
- **cstufac:** change in stufac
- **cbowl:** change in bowl
- **cavg500:** change in avg500
- **cbtitle:** change in btitle
- **lapps_1:** lapps lagged
- **school:** name of university
- **ctop25:** change in top25
- **bball:** =1 if btitle or finfour
- **cbball:** change in bball

Notes

These data were collected by Patrick Tulloch, an MSU economics major, for a term project. The "athletic success" variables are for the year prior to the enrollment and academic data. Updating these data to get a longer stretch of years, and including appearances in the "Sweet 16" NCAA basketball tournaments, would make for a more convincing analysis. With the growing popularity of women's sports, especially basketball, an analysis that includes success in women's athletics would be interesting.

Used in Text: page 697

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(athlet1)
```

`athlet2` *athlet2*

Description

Wooldridge Sources: Peterson's Guide to Four Year Colleges, 1995 (25th edition). Princeton University Press. 1995 Information Please Sports Almanac (6th edition). Houghton Mifflin. New York, NY Data loads lazily.

Usage

```
data('athlet2')
```

Format

A data.frame with 30 observations on 10 variables:

- **dscore:** home scr. - vist. scr., 1993
- **dinstt:** diff. in-state tuit., 1994
- **doutstt:** diff. out-state tuit., 1994
- **htpriv:** =1 if home team priv. sch.
- **vtpriv:** =1 if vist. team priv. sch.
- **dapps:** diff. in applications, 1994
- **htwrd:** =1 if home win. record, 1993
- **vtwrd:** =1 if vist. win. record, 1993
- **dwinrec:** htwrd - vtwrd
- **dpriv:** htpriv - vtpriv

Notes

These data were collected by Paul Anderson, an MSU economics major, for a term project. The score from football outcomes for natural rivals (Michigan-Michigan State, California-Stanford, Florida-Florida State, to name a few) is matched with application and academic data. The application and tuition data are for Fall 1994. Football records and scores are from 1993 football season. Extended these data to obtain a long stretch of panel data and other "natural" rivals could be very interesting.

Used in Text: page 697

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(athlet2)
```

`attend` *attend*

Description

Wooldridge Source: These data were collected by Professors Ronald Fisher and Carl Liedholm during a term in which they both taught principles of microeconomics at Michigan State University. Professors Fisher and Liedholm kindly gave me permission to use a random subset of their data, and their research assistant at the time, Jeffrey Guilfoyle, who completed his Ph.D. in economics at MSU, provided helpful hints. Data loads lazily.

Usage

```
data('attend')
```

Format

A data.frame with 680 observations on 11 variables:

- **attend:** classes attended out of 32
- **termGPA:** GPA for term
- **priGPA:** cumulative GPA prior to term
- **ACT:** ACT score
- **final:** final exam score
- **atndrte:** percent classes attended
- **hwrte:** percent homework turned in
- **frosh:** =1 if freshman

- **soph:** =1 if sophomore
- **missed:** number of classes missed
- **stndfnl:** (final - mean)/sd

Notes

The attendance figures were obtained by requiring students to slide their ID cards through a magnetic card reader, under the supervision of a teaching assistant. You might have the students use final, rather than the standardized variable, so that they can see the statistical significance of each variable remains exactly the same. The standardized variable is used only so that the coefficients measure effects in terms of standard deviations from the average score.

Used in Text: pages 111, 152, 199-200, 222

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(attend)
```

`audit` *audit*

Description

Wooldridge Source: These data come from a 1988 Urban Institute audit study in the Washington, D.C. area. I obtained them from the article "The Urban Institute Audit Studies: Their Methods and Findings," by James J. Heckman and Peter Siegelman. In Fix, M. and Struyk, R., eds., Clear and Convincing Evidence: Measurement of Discrimination in America. Washington, D.C.: Urban Institute Press, 1993, 187-258. Data loads lazily.

Usage

```
data('audit')
```

Format

A data.frame with 241 observations on 3 variables:

- **w:** =1 if white app. got job offer
- **b:** =1 if black app. got job offer
- **y:** b - w

Used in Text

pages 776-777, 784, 787

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(audit)
```

---

barium *barium*

Description

Wooldridge Source: C.M. Krupp and P.S. Pollard (1999), Market Responses to Antidumpting Laws: Some Evidence from the U.S. Chemical Industry, Canadian Journal of Economics 29, 199-227. Dr. Krupp kindly provided the data. They are monthly data covering February 1978 through December 1988. Data loads lazily.

Usage

```
data('barium')
```

Format

A data.frame with 131 observations on 31 variables:

- **chnimp:** Chinese imports, bar. chl.
- **bchlimp:** total imports bar. chl.
- **befile6:** =1 for all 6 mos before filing
- **affile6:** =1 for all 6 mos after filing
- **afdec6:** =1 for all 6 mos after decision
- **befile12:** =1 all 12 mos before filing
- **affile12:** =1 all 12 mos after filing
- **afdec12:** =1 all 12 mos after decision
- **chempi:** chemical production index
- **gas:** gasoline production
- **rtwex:** exchange rate index
- **spr:** =1 for spring months
- **sum:** =1 for summer months
- **fall:** =1 for fall months
- **lchnimp:** log(chnimp)
- **lgas:** log(gas)

- **lrtwex:** log(rtwex)
- **lchempi:** log(chempi)
- **t:** time trend
- **feb:** =1 if month is feb
- **mar:** =1 if month is march
- **apr:**
- **may:**
- **jun:**
- **jul:**
- **aug:**
- **sep:**
- **oct:**
- **nov:**
- **dec:**
- **percchn:** percent imports from china

Note

Rather than just having intercept shifts for the different regimes, one could conduct a full Chow test across the different regimes.

Used in Text: pages 361-362, 372, 377, 426, 442-443, 445, 663, 665, 672

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(barium)
```

---

beauty *beauty*

Description

Wooldridge Source: Hamermesh, D.S. and J.E. Biddle (1994), "Beauty and the Labor Market," American Economic Review 84, 1174-1194. Professor Hamermesh kindly provided me with the data. For manageability, I have included only a subset of the variables, which results in somewhat larger sample sizes than reported for the regressions in the Hamermesh and Biddle paper. Data loads lazily.

Usage

```
data('beauty')
```

Format

A data.frame with 1260 observations on 17 variables:

- **wage:** hourly wage
- **lwage:** log(wage)
- **belavg:** =1 if looks <= 2
- **abvavg:** =1 if looks >=4
- **exper:** years of workforce experience
- **looks:** from 1 to 5
- **union:** =1 if union member
- **goodhlth:** =1 if good health
- **black:** =1 if black
- **female:** =1 if female
- **married:** =1 if married
- **south:** =1 if live in south
- **bigcity:** =1 if live in big city
- **smllcity:** =1 if live in small city
- **service:** =1 if service industry
- **expersq:** exper^2
- **educ:** years of schooling

Used in Text

pages 238-239, 265-266

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(beauty)
```

benefits *benefits*

Description

Wooldridge Data loads lazily.

Usage

```
data('benefits')
```

Format

A data.frame with 1848 observations on 18 variables:

- **distid:** district identifier
- **schid:** school identifier
- **lunch:** percent eligible, free lunch
- **enroll:** school enrollment
- **staff:** staff per 1000 students
- **exppp:** expenditures per pupil
- **avgsal:** average teacher salary, $
- **avgben:** average teacher non-salary benefits, $
- **math4:** percent passing 4th grade math test
- **story4:** percent passing 4th grade reading test
- **bs:** avgben/avgsal
- **lavgsal:** log(avgsal)
- **lenroll:** log(enroll)
- **lstaff:** log(staff)
- **bsbar:** within-district avg of bs
- **lunchbar:** within-district avg of lunch
- **lenrollbar:** within-district avg of lenroll
- **lstaffbar:** within-district avg of lstaff

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(benefits)
```

beveridge *beveridge*

Description

Wooldridge Data loads lazily.

Usage

```
data('beveridge')
```

Format

A data.frame with 135 observations on 8 variables:

- **month:** dec 200 through feb 2012
- **urate:** unemployment rate, percent
- **vrate:** vacancy rate, percent
- **t:** linear time trend
- **urate_1:** urate_1
- **vrate_1:** vrate_1
- **curate:** urate - urate_1
- **cvrate:** urate - urate_1

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(beveridge)
```

big9salary *big9salary*

Description

Wooldridge Source: O. Baser and E. Pema (2003), "The Return of Publications for Economics Faculty," Economics Bulletin 1, 1-13. Professors Baser and Pema kindly provided the data. Data loads lazily.

Usage

```
data('big9salary')
```

Format

A data.frame with 786 observations on 30 variables:

- **id:** person identifier
- **year:** 92, 95, or 99
- **salary:** annual salary, $
- **pubindx:** publication index
- **totpge:** standardized total article pages
- **assist:** =1 if assistant professor
- **assoc:** =1 if associate professor
- **prof:** =1 if full professor
- **chair:** =1 if department chair
- **top20phd:** =1 if Ph.D. from top 20 dept.
- **yearphd:** year Ph.D. obtained
- **female:** =1 if female
- **osu:** =1 if Ohio State U.
- **iowa:** =1 if U. Iowa
- **indiana:** =1 if Indiana U.
- **purdue:** =1 if Purdue U.
- **msu:** =1 if Michigan State U.
- **minn:** =1 if U. Minnesota
- **mich:** =1 if U. Michigan
- **wisc:** =1 if U. Wisconsin
- **illinois:** =1 if U. Illinois
- **y92:** =1 if year == 92
- **y95:** =1 if year == 95
- **y99:** =1 if year == 99
- **lsalary:** log(salary)
- **exper:** years since first teaching job
- **expersq:** exper^2
- **pubindxsq:** pubindx^2
- **pubindx0:** =1 if pubindx == 0
- **lpubindx:** log(pubindx) if pubindx > 0

Notes

This is an unbalanced panel data set in the sense that as many as three years of data are available for each faculty member but where some have fewer than three years. It is not clear that something like a fixed effects or first differencing analysis makes sense: in effect, approaches that remove the heterogeneity control for too much by controlling for unobserved heterogeneity which, in this case, includes faculty intelligence, talent, and motivation. Presumably these factors enter into the publication index. It is hard to think we want to hold the main factors driving productivity fixed when trying to measure the effect of productivity on salary. Pooled OLS regression with "cluster robust" standard errors seems more natural. On the other hand, if we want to measure the return to having a degree from a top 20 Ph.D. program then we would want to control for factors that cause selection into a top 20 program. Unfortunately, this variable does not change over time, and so FD and FE are not applicable.

Used in Text: not used

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(big9salary)
```

`bwght` *bwght*

Description

Wooldridge Source: J. Mullahy (1997), "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," Review of Economics and Statistics 79, 596-593. Professor Mullahy kindly provided the data. He obtained them from the 1988 National Health Interview Survey. Data loads lazily.

Usage

```
data('bwght')
```

Format

A data.frame with 1388 observations on 14 variables:

- **faminc:** 1988 family income, $1000s
- **cigtax:** cig. tax in home state, 1988
- **cigprice:** cig. price in home state, 1988
- **bwght:** birth weight, ounces
- **fatheduc:** father's yrs of educ

- **motheduc:** mother's yrs of educ
- **parity:** birth order of child
- **male:** =1 if male child
- **white:** =1 if white
- **cigs:** cigs smked per day while preg
- **lbwght:** log of bwght
- **bwghtlbs:** birth weight, pounds
- **packs:** packs smked per day while preg
- **lfaminc:** log(faminc)

Used in Text

pages 18, 61, 110, 151, 165, 178, 184, 187-188, 258-259, 522-523

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(bwght)
```

---

bwght2 *bwght2*

Description

Wooldridge Source: Dr. Zhehui Luo, a recent MSU Ph.D. in economics and Visiting Research Associate in the Department of Epidemiology at MSU, kindly provided these data. She obtained them from state files linking birth and infant death certificates, and from the National Center for Health Statistics natality and mortality data. Data loads lazily.

Usage

```
data('bwght2')
```

Format

A data.frame with 1832 observations on 23 variables:

- **mage:** mother's age, years
- **meduc:** mother's educ, years
- **monpre:** month prenatal care began
- **npvis:** total number of prenatal visits

- **fage:** father's age, years
- **feduc:** father's educ, years
- **bwght:** birth weight, grams
- **omaps:** one minute apgar score
- **fmaps:** five minute apgar score
- **cigs:** avg cigarettes per day
- **drink:** avg drinks per week
- **lbw:** =1 if bwght <= 2000
- **vlbw:** =1 if bwght <= 1500
- **male:** =1 if baby male
- **mwhte:** =1 if mother white
- **mblck:** =1 if mother black
- **moth:** =1 if mother is other
- **fwhte:** =1 if father white
- **fblck:** =1 if father black
- **foth:** =1 if father is other
- **lbwght:** log(bwght)
- **magesq:** mage^2
- **npvissq:** npvis^2

Notes

There are many possibilities with this data set. In addition to number of prenatal visits, smoking and alcohol consumption (during pregnancy) are included as explanatory variables. These can be added to equations of the kind found in Exercise C6.10. In addition, the one- and five-minute APGAR scores are included. These are measures of the well being of infants just after birth. An interesting feature of the score is that it is bounded between zero and 10, making a linear model less than ideal. Still, a linear model would be informative, and you might ask students about predicted values less than zero or greater than 10.

Used in Text: pages 184, 223

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(bwght2)
```

campus *campus*

Description

Wooldridge Source: These data were collected by Daniel Martin, a former MSU undergraduate, for a final project. They come from the FBI Uniform Crime Reports and are for the year 1992. Data loads lazily.

Usage

```
data('campus')
```

Format

A data.frame with 97 observations on 7 variables:

- **enroll:** total enrollment
- **priv:** =1 if private college
- **police:** employed officers
- **crime:** total campus crimes
- **lcrime:** log(crime)
- **lenroll:** log(enroll)
- **lpolice:** log(police)

Notes

Colleges and universities are now required to provide much better, more detailed crime data. A very rich data set can now be obtained, even a panel data set for colleges across different years. Statistics on male/female ratios, fraction of men/women in fraternities or sororities, policy variables – such as a "safe house" for women on campus, as was started at MSU in 1994 – could be added as explanatory variables. The crime rate in the host town would be a good control.

Used in Text: pages 131-132

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(campus)
```

`card` *card*

Description

Wooldridge Source: D. Card (1995), Using Geographic Variation in College Proximity to Estimate the Return to Schooling, in Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp. Ed. L.N. Christophides, E.K. Grant, and R. Swidinsky, 201-222. Toronto: University of Toronto Press. Professor Card kindly provided these data. Data loads lazily.

Usage

```
data('card')
```

Format

A data.frame with 3010 observations on 34 variables:

- **id:** person identifier
- **nearc2:** =1 if near 2 yr college, 1966
- **nearc4:** =1 if near 4 yr college, 1966
- **educ:** years of schooling, 1976
- **age:** in years
- **fatheduc:** father's schooling
- **motheduc:** mother's schooling
- **weight:** NLS sampling weight, 1976
- **momdad14:** =1 if live with mom, dad at 14
- **sinmom14:** =1 if with single mom at 14
- **step14:** =1 if with step parent at 14
- **reg661:** =1 for region 1, 1966
- **reg662:** =1 for region 2, 1966
- **reg663:** =1 for region 3, 1966
- **reg664:** =1 for region 4, 1966
- **reg665:** =1 for region 5, 1966
- **reg666:** =1 for region 6, 1966
- **reg667:** =1 for region 7, 1966
- **reg668:** =1 for region 8, 1966
- **reg669:** =1 for region 9, 1966
- **south66:** =1 if in south in 1966
- **black:** =1 if black
- **smsa:** =1 in in SMSA, 1976

- **south:** =1 if in south, 1976
- **smsa66:** =1 if in SMSA, 1966
- **wage:** hourly wage in cents, 1976
- **enroll:** =1 if enrolled in school, 1976
- **KWW:** knowledge world of work score
- **IQ:** IQ score
- **married:** =1 if married, 1976
- **libcrd14:** =1 if lib. card in home at 14
- **exper:** age - educ - 6
- **lwage:** log(wage)
- **expersq:** exper^2

Notes

Computer Exercise C15.3 is important for analyzing these data. There, it is shown that the instrumental variable, 'nearc4', is actually correlated with 'IQ', at least for the subset of men for which an IQ score is reported. However, the correlation between 'nearc4" and 'IQ', once the other explanatory variables are netted out, is arguably zero. At least, it is not statistically different from zero. In other words, 'nearc4' fails the exogeneity requirement in a simple regression model but it passes, at least using the crude test described above, if controls are added to the wage equation. For a more advanced course, a nice extension of Card's analysis is to allow the return to education to differ by race. A relatively simple extension is to include black education (blackeduc) as an additional explanatory variable; its natural instrument is blacknearc4.

Used in Text: pages 526-527, 547

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(card)
```

`catholic` *catholic*

Description

Wooldridge Source: Altonji, J.G., T.E. Elder, and C.R. Taber (2005), "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling," Journal of Human Resources 40, 791-821. Professor Elder kindly provided a subset of the data, with some variables stripped away for confidentiality reasons. Data loads lazily.

Usage

```
data('catholic')
```

Format

A data.frame with 7430 observations on 13 variables:

- **id:** person identifier
- **read12:** reading standardized score
- **math12:** mathematics standardized score
- **female:** =1 if female
- **asian:** =1 if Asian
- **hispan:** =1 if Hispanic
- **black:** =1 if black
- **motheduc:** mother's years of education
- **fatheduc:** father's years of education
- **lfaminc:** log of family income
- **hsgrad:** =1 if graduated from high school by 1994
- **cathhs:** =1 if attended Catholic HS
- **parcath:** =1 if a parent reports being Catholic

Used in Text

pages 267, 551

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781305270107

Examples

```
str(catholic)
```

cement *cement*

Description

Wooldridge Source: J. Shea (1993), "The Input-Output Approach to Instrument Selection," Journal of Business and Economic Statistics 11, 145-156. Professor Shea kindly provided these data. Data loads lazily.

Usage

```
data('cement')
```

Format

A data.frame with 312 observations on 30 variables:

- **year:** 1964-1989
- **month:** 1-12
- **prccem:** BLS ppi for cement
- **ipcem:** industrial prod. index, cement
- **prcpet:** ppi for crude petroleum
- **rresc:** real residential construction
- **rnonc:** real nonres. construction
- **ip:** aggregate index of indus. prod.
- **rdefs:** real defense spending
- **milemp:** military employment
- **gprc:** log(prccem) - log(prccem[_n-1])
- **gcem:** log(ipcem) - log(ipcem[_n-1])
- **gprcpet:** log(prcpet) - log(prcpet[_n-1])
- **gres:** log(rresc) - log(rresc[_n-1])
- **gnon:** log(rnonc) - log(rnonc[_n-1])
- **gip:** log(ip) - log(ip[_n-1])
- **gdefs:** log(rdefs) - log(rdefs[_n-1])
- **gmilemp:** log(milemp) - log(milemp[_n-1])
- **jan:** =1 if month == 1
- **feb:** =1 if month == 2
- **mar:** =1 if month == 3
- **apr:** =1 if month == 4
- **may:** =1 if month == 5

- **jun:** =1 if month == 6
- **jul:** =1 if month == 7
- **aug:** =1 if month == 8
- **sep:** =1 if month == 9
- **oct:** =1 if month == 10
- **nov:** =1 if month == 11
- **dec:** =1 if month == 12

Notes

Compared with Shea's analysis, the producer price index (PPI) for fuels and power has been replaced with the PPI for petroleum. The data are monthly and have not been seasonally adjusted.

Used in Text: pages 579

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(cement)
```

`census2000` *census2000*

Description

Wooldridge Source: Obtained from the United States Census Bureau by Professor Alberto Abadie of the Harvard Kennedy School of Government. Professor Abadie kindly provided the data. Data loads lazily.

Usage

```
data('census2000')
```

Format

A data.frame with 29501 observations on 6 variables:

- **state:** State (ICPSR code)
- **puma:** Public Use Microdata Area
- **educ:** educational attainment
- **lweekinc:** log(weekly income)
- **exper:** years workforce experience
- **expersq:** exper^2

Used in Text

pages 452-453

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781305270107

Examples

```
str(census2000)
```

`ceosal1` *ceosal1*

Description

Wooldridge Source: I took a random sample of data reported in the May 6, 1991 issue of Businessweek. Data loads lazily.

Usage

```
data('ceosal1')
```

Format

A data.frame with 209 observations on 12 variables:

- **salary:** 1990 salary, thousands $
- **pcsalary:** percent change salary, 89-90
- **sales:** 1990 firm sales, millions $
- **roe:** return on equity, 88-90 avg
- **pcroe:** percent change roe, 88-90
- **ros:** return on firm's stock, 88-90
- **indus:** =1 if industrial firm
- **finance:** =1 if financial firm
- **consprod:** =1 if consumer product firm
- **utility:** =1 if transport. or utilties
- **lsalary:** natural log of salary
- **lsales:** natural log of sales

Notes

This kind of data collection is relatively easy for students just learning data analysis, and the findings can be interesting. A good term project is to have students collect a similar data set using a more recent issue of Businessweek, and to find additional variables that might explain differences in CEO compensation. My impression is that the public is still interested in CEO compensation. An interesting question is whether the list of explanatory variables included in this data set now explain less of the variation in log(salary) than they used to.

Used in Text: pages 32, 35-36, 39, 159-160, 218-219, 260-261, 263, 685, 692-693

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(ceosal1)
```

`ceosal2` *ceosal2*

Description

Wooldridge Source: See CEOSAL1.RAW Data loads lazily.

Usage

```
data('ceosal2')
```

Format

A data.frame with 177 observations on 15 variables:

- **salary:** 1990 compensation, $1000s
- **age:** in years
- **college:** =1 if attended college
- **grad:** =1 if attended graduate school
- **comten:** years with company
- **ceoten:** years as ceo with company
- **sales:** 1990 firm sales, millions
- **profits:** 1990 profits, millions
- **mktval:** market value, end 1990, mills.
- **lsalary:** log(salary)
- **lsales:** log(sales)

- **lmktval:** log(mktval)
- **comtensq:** comten^2
- **ceotensq:** ceoten^2
- **profmarg:** profits as percent of sales

Notes

Compared with CEOSAL1.RAW, in this CEO data set more information about the CEO, rather than about the company, is included.

Used in Text: pages 64, 111, 163, 214, 335, 699

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(ceosal2)
```

`charity` *charity*

Description

Wooldridge Source: P.H. Franses and R. Paap (2001), Quantitative Models in Marketing Research. Cambridge: Cambridge University Press. Professor Franses kindly provided the data. Data loads lazily.

Usage

```
data('charity')
```

Format

A data.frame with 4268 observations on 8 variables:

- **respond:** =1 if responded with gift
- **gift:** amount of gift, Dutch guilders
- **resplast:** =1 if responded to most recent mailing
- **weekslast:** number of weeks since last response
- **propresp:** response rate to mailings
- **mailsyear:** number of mailings per year
- **giftlast:** amount of most recent gift
- **avggift:** average of past gifts

Notes

This data set can be used to illustrate probit and Tobit models, and to study the linear approximations to them.

Used in Text: pages 65, 112-113, 266-267, 628

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(charity)
```

---

consump *consump*

Description

Wooldridge Source: I collected these data from the 1997 Economic Report of the President. Specifically, the data come from Tables B-71, 15, 29, and 32. Data loads lazily.

Usage

```
data('consump')
```

Format

A data.frame with 37 observations on 24 variables:

- **year:** 1959-1995
- **i3:** 3 mo. T-bill rate
- **inf:** inflation rate; CPI
- **rdisp:** disp. inc., 1992 $, bils.
- **rnondc:** nondur. cons., 1992 $, bils.
- **rserv:** services, 1992 $, bils.
- **pop:** population, 1000s
- **y:** per capita real disp. inc.
- **rcons:** rnondc + rserv
- **c:** per capita real cons.
- **r3:** i3 - inf; real ex post int.
- **lc:** log(c)
- **ly:** log(y)

- **gc:** lc - lc[_n-1]
- **gy:** ly - ly[_n-1]
- **gc_1:** gc[_n-1]
- **gy_1:** gy[_n-1]
- **r3_1:** r3[_n-1]
- **lc_ly:** lc - ly
- **lc_ly_1:** lc_ly[_n-1]
- **gc_2:** gc[_n-2]
- **gy_2:** gy[_n-2]
- **r3_2:** r3[_n-2]
- **lc_ly_2:** lc_ly[_n-2]

Notes

For a student interested in time series methods, updating this data set and using it in a manner similar to that in the text could be acceptable as a final project.

Used in Text: pages 377-378, 408-409, 442, 570-571, 579, 673

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(consump)
```

---

corn *corn*

Description

Wooldridge Source: G.E. Battese, R.M. Harter, and W.A. Fuller (1988), "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data," Journal of the American Statistical Association 83, 28-36. This small data set is reported in the article. Data loads lazily.

Usage

```
data('corn')
```

Format

A data.frame with 37 observations on 5 variables:

- **county:** county number
- **cornhec:** corn per hectare
- **soyhec:** soybeans per hectare
- **cornpix:** corn pixels per hectare
- **soypix:** soy pixels per hectare

Notes

You could use these data to illustrate simple regression when the population intercept should be zero: no corn pixels should predict no corn planted. The same can be done with the soybean measures in the data set.

Used in Text: pages 791-792

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(corn)
```

`countymurders` *countymurders*

Description

Wooldridge Source: Compiled by J. Monroe Gamble for a Summer Research Opportunities Program (SROP) at Michigan State University, Summer 2014. Monroe obtained data from the U.S. Census Bureau, the FBI Uniform Crime Reports, and the Death Penalty Information Center. Data loads lazily.

Usage

```
data('countymurders')
```

Format

A data.frame with 37349 observations on 20 variables:

- **arrests:** # of murder arrests
- **countyid:** county identifier: 1000*statefips + countyfips
- **density:** population density; per square mile
- **popul:** county population
- **perc1019:** percent pop. age 10-19
- **perc2029:** percent pop. age 20-29
- **percblack:** percent population black
- **percmale:** percent population male
- **rpcincmaint:** real per capita income maintenance
- **rpcpersinc:** real per capita personal income
- **rpcunemins:** real per capita unem insurance payments
- **year:** 1980-1996
- **murders:** # of murders
- **murdrate:** murders per 10,000 people
- **arrestrate:** murder arrests per 10,000
- **statefips:** state FIPS code
- **countyfips:** county FIPS code
- **execs:** # of executions
- **lpopul:** log(popul)
- **execrate:** executions per 10,000

Used in Text

pages 16, 58, 431, 457

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781305270107

Examples

```
str(countymurders)
```

cps78_85 *cps78_85*

Description

Wooldridge Source: Professor Henry Farber, now at Princeton University, compiled these data from the 1978 and 1985 Current Population Surveys. Professor Farber kindly provided these data when we were colleagues at MIT. Data loads lazily.

Usage

```
data('cps78_85')
```

Format

A data.frame with 1084 observations on 15 variables:

- **educ:** years of schooling
- **south:** =1 if live in south
- **nonwhite:** =1 if nonwhite
- **female:** =1 if female
- **married:** =1 if married
- **exper:** age - educ - 6
- **expersq:** exper^2
- **union:** =1 if belong to union
- **lwage:** log hourly wage
- **age:** in years
- **year:** 78 or 85
- **y85:** =1 if year == 85
- **y85fem:** y85*female
- **y85educ:** y85*educ
- **y85union:** y85*union

Notes

Obtaining more recent data from the CPS allows one to track, over a long period of time, the changes in the return to education, the gender gap, black-white wage differentials, and the union wage premium.

Used in Text: pages 451, 476

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(cps78_85)
```

`cps91` *cps91*

Description

Wooldridge Source: Professor Daniel Hamermesh, at the University of Texas, compiled these data from the May 1991 Current Population Survey. Professor Hamermesh kindly provided these data. Data loads lazily.

Usage

```
data('cps91')
```

Format

A data.frame with 5634 observations on 24 variables:

- **husage:** husband's age
- **husunion:** =1 if hus. in union
- **husearns:** hus. weekly earns
- **huseduc:** husband's yrs schooling
- **husblck:** =1 if hus. black
- **hushisp:** =1 if hus. hispanic
- **hushrs:** hus. weekly hours
- **kidge6:** =1 if have child >= 6
- **earns:** wife's weekly earnings
- **age:** wife's age
- **black:** =1 if wife black
- **educ:** wife's yrs schooling
- **hispanic:** =1 if wife hispanic
- **union:** =1 if wife in union
- **faminc:** annual family income
- **husexp:** huseduc - husage - 6
- **exper:** age - educ - 6
- **kidlt6:** =1 if have child < 6
- **hours:** wife's weekly hours
- **expersq:** exper^2
- **nwifeinc:** non-wife inc, $1000s
- **inlf:** =1 if wife in labor force
- **hrwage:** earns/hours
- **lwage:** log(hrwage)

Notes

This is much bigger than the other CPS data sets even though the sample is restricted to married women. (CPS91.RAW contains many more observations than MROZ.RAW, too.) In addition to the usual human capital variables for the women in the sample, we have information on the husband. Therefore, we can estimate a labor supply function as in Chapter 16, although the validity of potential experience as an IV for log(wage) is questionable. (MROZ.RAW contains an actual experience variable.) Perhaps more convincing is to add hours to the wage offer equation, and instrument hours with indicators for young and old children. This data set also contains a union membership indicator. The web site for the National Bureau of Economic Research makes it very easy now to download CPS data files in a variety offormats. Go to http://www.nber.org/data/cps_basic.html.

Used in Text: page 627-628

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(cps91)
```

---

crime1 *crime1*

Description

Wooldridge Source: J. Grogger (1991), "Certainty vs. Severity of Punishment," Economic Inquiry 29, 297-309. Professor Grogger kindly provided a subset of the data he used in his article. Data loads lazily.

Usage

```
data('crime1')
```

Format

A data.frame with 2725 observations on 16 variables:

- **narr86:** # times arrested, 1986
- **nfarr86:** # felony arrests, 1986
- **nparr86:** # property crme arr., 1986
- **pcnv:** proportion of prior convictions
- **avgsen:** avg sentence length, mos.
- **tottime:** time in prison since 18 (mos.)
- **ptime86:** mos. in prison during 1986

- **qemp86:** # quarters employed, 1986
- **inc86:** legal income, 1986, $100s
- **durat:** recent unemp duration
- **black:** =1 if black
- **hispan:** =1 if Hispanic
- **born60:** =1 if born in 1960
- **pcnvsq:** pcnv^2
- **pt86sq:** ptime86^2
- **inc86sq:** inc86^2

Used in Text

pages 82-83, 173-174, 180, 252-253, 275, 299, 305-306, 607-608, 625

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(crime1)
```

---

crime2 *crime2*

Description

Wooldridge Source: These data were collected by David Dicicco, a former MSU undergraduate, for a final project. They came from various issues of the County and City Data Book, and are for the years 1982 and 1985. Unfortunately, I do not have the list of cities. Data loads lazily.

Usage

```
data('crime2')
```

Format

A data.frame with 92 observations on 34 variables:

- **pop:** population
- **crimes:** total number index crimes
- **unem:** unemployment rate
- **officers:** number police officers
- **pcinc:** per capita income

- **west:** =1 if city in west
- **nrtheast:** =1 if city in NE
- **south:** =1 if city in south
- **year:** 82 or 87
- **area:** land area, square miles
- **d87:** =1 if year = 87
- **popden:** people per sq mile
- **crmrte:** crimes per 1000 people
- **offarea:** officers per sq mile
- **lawexpc:** law enforce. expend. pc, $
- **polpc:** police per 1000 people
- **lpop:** log(pop)
- **loffic:** log(officers)
- **lpcinc:** log(pcinc)
- **llawexpc:** log(lawexpc)
- **lpopden:** log(popden)
- **lcrimes:** log(crimes)
- **larea:** log(area)
- **lcrmrte:** log(crmrte)
- **clcrimes:** change in lcrimes
- **clpop:** change in lpop
- **clcrmrte:** change in lcrmrte
- **lpolpc:** log(polpc)
- **clpolpc:** change in lpolpc
- **cllawexp:** change in llawexp
- **cunem:** change in unem
- **clpopden:** change in lpopden
- **lcrmrt_1:** lcrmrte lagged
- **ccrmrte:** change in crmrte

Notes

Very rich crime data sets, at the county, or even city, level, can be collected using the FBI's Uniform Crime Reports. These data can be matched up with demographic and economic data, at least for census years. The County and City Data Book contains a variety of statistics, but the years do not always match up. These data sets can be used investigate issues such as the effects of casinos on city or county crime rates.

Used in Text: pages 313-314, 459-460

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(crime2)
```

`crime3` *crime3*

Description

Wooldridge Source: E. Eide (1994), Economics of Crime: Deterrence of the Rational Offender. Amsterdam: North Holland. The data come from Tables A3 and A6. Data loads lazily.

Usage

```
data('crime3')
```

Format

A data.frame with 106 observations on 12 variables:

- **district:** district number
- **year:** 72 or 78
- **crime:** crimes per 1000 people
- **clrprc1:** clear-up perc, prior year
- **clrprc2:** clear-up perc, two-years prior
- **d78:** =1 if year = 78
- **avgclr:** (clrprc1 + clrprc2)/2
- **lcrime:** log(crime)
- **clcrime:** change in lcrime
- **cavgclr:** change in avgclr
- **cclrprc1:** change in clrprc1
- **cclrprc2:** change in clrprc2

Notes

These data are for the years 1972 and 1978 for 53 police districts in Norway. Much larger data sets for more years can be obtained for the United States, although a measure of the "clear-up" rate is needed.

Used in Text: pages 464-465, 477-478

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(crime3)
```

`crime4` *crime4*

Description

Wooldridge Source: From C. Cornwell and W. Trumball (1994), "Estimating the Economic Model of Crime with Panel Data," Review of Economics and Statistics 76, 360-366. Professor Cornwell kindly provided the data. Data loads lazily.

Usage

```
data('crime4')
```

Format

A data.frame with 630 observations on 59 variables:

- **county:** county identifier
- **year:** 81 to 87
- **crmrte:** crimes committed per person
- **prbarr:** 'probability' of arrest
- **prbconv:** 'probability' of conviction
- **prbpris:** 'probability' of prison sentenc
- **avgsen:** avg. sentence, days
- **polpc:** police per capita
- **density:** people per sq. mile
- **taxpc:** tax revenue per capita
- **west:** =1 if in western N.C.
- **central:** =1 if in central N.C.
- **urban:** =1 if in SMSA
- **pctmin80:** perc. minority, 1980
- **wcon:** weekly wage, construction
- **wtuc:** wkly wge, trns, util, commun
- **wtrd:** wkly wge, whlesle, retail trade

- **wfir:** wkly wge, fin, ins, real est
- **wser:** wkly wge, service industry
- **wmfg:** wkly wge, manufacturing
- **wfed:** wkly wge, fed employees
- **wsta:** wkly wge, state employees
- **wloc:** wkly wge, local gov emps
- **mix:** offense mix: face-to-face/other
- **pctymle:** percent young male
- **d82:** =1 if year == 82
- **d83:** =1 if year == 83
- **d84:** =1 if year == 84
- **d85:** =1 if year == 85
- **d86:** =1 if year == 86
- **d87:** =1 if year == 87
- **lcrmrte:** log(crmrte)
- **lprbarr:** log(prbarr)
- **lprbconv:** log(prbconv)
- **lprbpris:** log(prbpris)
- **lavgsen:** log(avgsen)
- **lpolpc:** log(polpc)
- **ldensity:** log(density)
- **ltaxpc:** log(taxpc)
- **lwcon:** log(wcon)
- **lwtuc:** log(wtuc)
- **lwtrd:** log(wtrd)
- **lwfir:** log(wfir)
- **lwser:** log(wser)
- **lwmfg:** log(wmfg)
- **lwfed:** log(wfed)
- **lwsta:** log(wsta)
- **lwloc:** log(wloc)
- **lmix:** log(mix)
- **lpctymle:** log(pctymle)
- **lpctmin:** log(pctmin)
- **clcrmrte:** lcrmrte - lcrmrte[_n-1]
- **clprbarr:** lprbarr - lprbarr[_n-1]
- **clprbcon:** lprbconv - lprbconv[_n-1]
- **clprbpri:** lprbpri - lprbpri[t-1]
- **clavgsen:** lavgsen - lavgsen[t-1]
- **clpolpc:** lpolpc - lpolpc[t-1]
- **cltaxpc:** ltaxpc - ltaxpc[t-1]
- **clmix:** lmix - lmix[t-1]

Notes

Computer Exercise C16.7 shows that variables that might seem to be good instrumental variable candidates are not always so good, especially after applying a transformation such as differencing across time. You could have the students do an IV analysis for just, say, 1987.

Used in Text: pages 471-472, 479, 504, 580

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(crime4)
```

---

discrim *discrim*

Description

Wooldridge Source: K. Graddy (1997), "Do Fast-Food Chains Price Discriminate on the Race and Income Characteristics of an Area?" Journal of Business and Economic Statistics 15, 391-401. Professor Graddy kindly provided the data set. Data loads lazily.

Usage

```
data('discrim')
```

Format

A data.frame with 410 observations on 37 variables:

- **psoda:** price of medium soda, 1st wave
- **pfries:** price of small fries, 1st wave
- **pentree:** price entree (burger or chicken), 1st wave
- **wagest:** starting wage, 1st wave
- **nmgrs:** number of managers, 1st wave
- **nregs:** number of registers, 1st wave
- **hrsopen:** hours open, 1st wave
- **emp:** number of employees, 1st wave
- **psoda2:** price of medium soday, 2nd wave
- **pfries2:** price of small fries, 2nd wave
- **pentree2:** price entree, 2nd wave
- **wagest2:** starting wage, 2nd wave

- **nmgrs2:** number of managers, 2nd wave
- **nregs2:** number of registers, 2nd wave
- **hrsopen2:** hours open, 2nd wave
- **emp2:** number of employees, 2nd wave
- **compown:** =1 if company owned
- **chain:** BK = 1, KFC = 2, Roy Rogers = 3, Wendy's = 4
- **density:** population density, town
- **crmrte:** crime rate, town
- **state:** NJ = 1, PA = 2
- **prpblck:** proportion black, zipcode
- **prppov:** proportion in poverty, zipcode
- **prpncar:** proportion no car, zipcode
- **hseval:** median housing value, zipcode
- **nstores:** number of stores, zipcode
- **income:** median family income, zipcode
- **county:** county label
- **lpsoda:** log(psoda)
- **lpfries:** log(pfries)
- **lhseval:** log(hseval)
- **lincome:** log(income)
- **ldensity:** log(density)
- **NJ:** =1 for New Jersey
- **BK:** =1 if Burger King
- **KFC:** =1 if Kentucky Fried Chicken
- **RR:** =1 if Roy Rogers

Notes

If you want to assign a common final project, this would be a good data set. There are many possible dependent variables, namely, prices of various fast-food items. The key variable is the fraction of the population that is black, along with controls for poverty, income, housing values, and so on. These data were also used in a famous study by David Card and Alan Krueger on estimation of minimum wage effects on employment. See the book by Card and Krueger, Myth and Measurement, 1997, Princeton University Press, for a detailed analysis.

Used in Text: pages 112, 166, 699-700

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(discrim)
```

`driving` *driving*

Description

Wooldridge Source: Freeman, D.G. (2007), "Drunk Driving Legislation and Traffic Fatalities: New Evidence on BAC 08 Laws," Contemporary Economic Policy 25, 293–308. Professor Freeman kindly provided the data. Data loads lazily.

Usage

```
data('driving')
```

Format

A data.frame with 1200 observations on 56 variables:

- **year:** 1980 through 2004
- **state:** 48 continental states, alphabetical
- **sl55:** speed limit == 55
- **sl65:** speed limit == 65
- **sl70:** speed limit == 70
- **sl75:** speed limit == 75
- **slnone:** no speed limit
- **seatbelt:** =0 if none, =1 if primary, =2 if secondary
- **minage:** minimum drinking age
- **zerotol:** zero tolerance law
- **gdl:** graduated drivers license law
- **bac10:** blood alcohol limit .10
- **bac08:** blood alcohol limit .08
- **perse:** administrative license revocation (per se law)
- **totfat:** total traffic fatalities
- **nghtfat:** total nighttime fatalities
- **wkndfat:** total weekend fatalities
- **totfatpvm:** total fatalities per 100 million miles
- **nghtfatpvm:** nighttime fatalities per 100 million miles
- **wkndfatpvm:** weekend fatalities per 100 million miles
- **statepop:** state population
- **totfatrte:** total fatalities per 100,000 population
- **nghtfatrte:** nighttime fatalities per 100,000 population

- **wkndfatrte:** weekend accidents per 100,000 population
- **vehicmiles:** vehicle miles traveled, billions
- **unem:** unemployment rate, percent
- **perc14_24:** percent population aged 14 through 24
- **sl70plus:** sl70 + sl75 + slnone
- **sbprim:** =1 if primary seatbelt law
- **sbsecon:** =1 if secondary seatbelt law
- **d80:** =1 if year == 1980
- **d81:**
- **d82:**
- **d83:**
- **d84:**
- **d85:**
- **d86:**
- **d87:**
- **d88:**
- **d89:**
- **d90:**
- **d91:**
- **d92:**
- **d93:**
- **d94:**
- **d95:**
- **d96:**
- **d97:**
- **d98:**
- **d99:**
- **d00:**
- **d01:**
- **d02:**
- **d03:**
- **d04:** =1 if year == 2004
- **vehicmilespc:**

Notes

Several more years of data are available and may further shed light on the effectiveness of several traffic laws.

Used in Text: not used, but see page 695

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(driving)
```

---

earns *earns*

Description

Wooldridge Source: Economic Report of the President, 1989, Table B-47. The data are for the non-farm business sector. Data loads lazily.

Usage

```
data('earns')
```

Format

A data.frame with 41 observations on 14 variables:

- **year:** 1947 to 1987
- **wkearns:** avg. real weekly earnings
- **wkhours:** avg. weekly hours
- **outphr:** output per labor hour
- **hrwage:** wkearns/wkhours
- **lhrwage:** log(hrwage)
- **loutphr:** log(outphr)
- **t:** time trend: t=1 to 47
- **ghrwage:** lhrwage - lhrwage[_n-1]
- **goutphr:** loutphr - loutphr[_n-1]
- **ghrwge_1:** ghrwage[_n-1]
- **goutph_1:** goutphr[_n-1]
- **goutph_2:** goutphr[_n-2]
- **lwkhours:** log(wkhours)

Notes

These data could be usefully updated, but changes in reporting conventions in more recent ERPs may make that difficult.

Used in Text: pages 363-364, 398, 407

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(earns)
```

---

econmath *econmath*

Description

Wooldridge Source: Compiled by Professor Charles Ballard, Michigan State University Department of Economics. Professor Ballard kindly provided the data. Data loads lazily.

Usage

```
data('econmath')
```

Format

A data.frame with 856 observations on 17 variables:

- **age:** age in years
- **work:** hours worked per week
- **study:** hours studying per week
- **econhs:** =1 if economics in high school
- **colgpa:** college GPA, beginning semester
- **hsgpa:** high school GPA
- **acteng:** ACT English score
- **actmth:** ACT math score
- **act:** ACT composite
- **mathscr:** math quiz score, 0-10
- **male:** =1 if male
- **calculus:** =1 if taken calculus course
- **attexc:** =1 if past attndce 'excellent'
- **attgood:** =1 if past attndce 'good'
- **fathcoll:** =1 if father has BA
- **mothcoll:** =1 if mother has BA
- **score:** course score, in percent

Used in Text

167, 185

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781305270107

Examples

```
str(econmath)
```

`elem94_95` *elem94_95*

Description

Wooldridge Source: Culled from a panel data set used by Leslie Papke in her paper "The Effects of Spending on Test Pass Rates: Evidence from Michigan" (2005), Journal of Public Economics 89, 821-839. Data loads lazily.

Usage

```
data('elem94_95')
```

Format

A data.frame with 1848 observations on 14 variables:

- **distid:** district identifier
- **schid:** school identifier
- **lunch:** percent eligible, free lunch
- **enrol:** enrollment
- **staff:** staff per 1000 students
- **exppp:** expenditures per pupil
- **avgsal:** average teacher salary, $
- **avgben:** average teacher non-salary benefits, $
- **math4:** percent passing 4th grade math test
- **story4:** percent passing 4th grade reading test
- **bs:** avgben/avgsal
- **lavgsal:** log(avgsal)
- **lenrol:** log(enrol)
- **lstaff:** log(staff)

Notes

Starting in 1995, the Michigan Department of Education stopped reporting average teacher benefits along with average salary. This data set includes both variables, at the school level, and can be used to study the salary-benefits tradeoff, as in Chapter 4. There are a few suspicious benefits/salary ratios, and so this data set makes a good illustration of the impact of outliers in Chapter 9.

Used in Text: pages 166-167, 341-342

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(elem94_95)
```

`engin` *engin*

Description

Wooldridge Source: Thada Chaisawangwong, a former graduate student at MSU, obtained these data for a term project in applied econometrics. They come from the Material Requirement Planning Survey carried out in Thailand during 1998. Data loads lazily.

Usage

```
data('engin')
```

Format

A data.frame with 403 observations on 17 variables:

- **male:** =1 if male
- **educ:** highest grade completed
- **wage:** monthly salary, Thai baht
- **swage:** starting wage
- **exper:** years on current job
- **pexper:** previous experience
- **lwage:** log(wage)
- **expersq:** exper^2
- **highgrad:** =1 if high school graduate
- **college:** =1 if college graduate
- **grad:** =1 if some graduate school

- **polytech:** =1 if a polytech
- **highdrop:** =1 if no high school degree
- **lswage:** log(swage)
- **pexpersq:** pexper^2
- **mleeduc:** male*educ
- **mleeduc0:** male*(educ - 14)

Notes

This is a nice change of pace from wage data sets for the United States. These data are for engineers in Thailand, and represents a more homogeneous group than data sets that consist of people across a variety of occupations. Plus, the starting salary is also provided in the data set, so factors affecting wage growth – and not just wage levels at a given point in time – can be studied. This is a good data set for a common term project that tests basic understanding of multiple regression and the interpretation of models with a logarithm for a dependent variable.

Used in Text: not used

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(engin)
```

`expendshares` *expendshares*

Description

Wooldridge Source: Blundell, R., A. Duncan, and K. Pendakur (1998), "Semiparametric Estimation and Consumer Demand," Journal of Applied Econometrics 13, 435-461. I obtained these data from the Journal of Applied Econometrics data archive at http://qed.econ.queensu.ca/jae/. Data loads lazily.

Usage

```
data('expendshares')
```

Format

A data.frame with 1519 observations on 13 variables:

- **sfood:** share of food expenditures (out of total)
- **sfuel:** share of fuel expenditures
- **sclothes:** share of clothing expenditures
- **salcohol:** share of alcohol expenditures
- **stransport:** share of transportation expenditures
- **sother:** share of other expenditures
- **totexpend:** total expenditure, British pounds per week
- **income:** family income, British pounds per week
- **age:** age of household head
- **kids:** number of children: 1 or 2
- **ltotexpend:** log(totexpend)
- **lincome:** log(income)
- **agesq:** age^2

Notes

The dependent variables in this data set – the expenditure shares – are necessarily bounded between zero and one. The linear model is at best an approximation, but the usual IV estimator likely gives good estimates of the average partial effects.

Used in Text: pages 581-582

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(expendshares)
```

---

ezanders *ezanders*

Description

Wooldridge Source: L.E. Papke (1994), "Tax Policy and Urban Development: Evidence from the Indiana Enterprise Zone Program," Journal of Public Economics 54, 37-49. Professor Papke kindly provided these data. Data loads lazily.

Usage

```
data('ezanders')
```

Format

A data.frame with 108 observations on 25 variables:

- **month:** name of month
- **uclms:** unemployment claims
- **ez:** =1 if enterprise zone
- **year:** 1980 through 1988
- **y81:** =1 if year == 1981
- **y82:**
- **y83:**
- **y84:**
- **y85:**
- **y86:**
- **y87:**
- **y88:**
- **luclms:** log(uclms)
- **jan:** =1 if month == JAN
- **feb:**
- **mar:**
- **apr:**
- **may:**
- **jun:**
- **jul:**
- **aug:**
- **sep:**
- **oct:**
- **nov:**
- **dec:**

Notes

These are actually monthly unemployment claims for the Anderson enterprise zone. Papke used annualized data, across many zones and non-zones, in her original analysis.

Used in Text: page 377

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(ezanders)
```

---

ezunem *ezunem*

Description

Wooldridge Source: See EZANDERS.RAW Data loads lazily.

Usage

```
data('ezunem')
```

Format

A data.frame with 198 observations on 37 variables:

- **year:** 1980 to 1988
- **uclms:** unemployment claims
- **ez:** =1 if have enterprise zone
- **d81:** =1 if year == 1981
- **d82:** =1 if year == 1982
- **d83:** =1 if year == 1983
- **d84:** =1 if year == 1984
- **d85:** =1 if year == 1985
- **d86:** =1 if year == 1986
- **d87:** =1 if year == 1987
- **d88:** =1 if year == 1988
- **c1:** =1 if city == 1
- **c2:** =1 if city == 2
- **c3:** =1 if city == 3
- **c4:**
- **c5:**
- **c6:**
- **c7:**
- **c8:**
- **c9:**
- **c10:**
- **c11:**
- **c12:**

- **c13:**
- **c14:**
- **c15:**
- **c16:**
- **c17:**
- **c18:**
- **c19:**
- **c20:**
- **c21:**
- **c22:** =1 if city == 22
- **luclms:** log(uclms)
- **guclms:** luclms - luclms[_n-1]
- **cez:** ez - ez[_n-1]
- **city:** city identifier, 1 through 22

Notes

A very good project is to have students analyze enterprise, empowerment, or renaissance zone policies in their home states. Many states now have such programs. A few years of panel data straddling periods of zone designation, at the city or zip code level, could make a nice study.

Used in Text: pages 470, 504

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(ezunem)
```

`fair` *fair*

Description

Wooldridge Source: R.C. Fair (1996), "Econometrics and Presidential Elections," Journal of Economic Perspectives 10, 89-102. The data set is provided in the article. Data loads lazily.

Usage

```
data('fair')
```

Format

A data.frame with 21 observations on 28 variables:

- **year:** 1916 to 1992, by 4
- **V:** prop. dem. vote
- **I:** =1 if demwh, -1 if repwh
- **DPER:** incumbent running
- **DUR:** duration
- **g3:** avg ann grwth rte, prev 3 qrts
- **p15:** avg ann inf rate, prev 15 qtrs
- **n:** quarters of good news
- **g2:** avg ann grwth rte, prev 2 qrts
- **gYR:** ann grwth rte, prev year
- **p8:** avg ann inf rate, prev 8 qtrs
- **p2YR:** inf rte over 2 yr period
- **Ig2:** I*g2
- **Ip8:** I*p8
- **demwins:** =1 if V > .5
- **In:** I*n
- **d:** =1 in 1920, 1944,1948
- **Id:** I*d
- **Ig3:** I*g3
- **Ip151md:** I*p15*(1-d)
- **In1md:** I*n*(1-d)

Notes

An updated version of this data set, through the 2004 election, is available at Professor Fair's web site at Yale University: http://fairmodel.econ.yale.edu/rayfair/pdf/2001b.htm. Students might want to try their own hands at predicting the most recent election outcome, but they should be restricted to no more than a handful of explanatory variables because of the small sample size.

Used in Text: pages 362-363, 440, 442

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(fair)
```

`fertil1` *fertil1*

Description

Wooldridge Source: W. Sander, "The Effect of Women's Schooling on Fertility," Economics Letters 40, 229-233.Professor Sander kindly provided the data, which are a subset of what he used in his article. He compiled the data from various years of the National Opinion Resource Center's General Social Survey. Data loads lazily.

Usage

```
data('fertil1')
```

Format

A data.frame with 1129 observations on 27 variables:

- **year:** 72 to 84, even
- **educ:** years of schooling
- **meduc:** mother's education
- **feduc:** father's education
- **age:** in years
- **kids:** # children ever born
- **black:** = 1 if black
- **east:** = 1 if lived in east at 16
- **northcen:** = 1 if lived in nc at 16
- **west:** = 1 if lived in west at 16
- **farm:** = 1 if on farm at 16
- **othrural:** = 1 if other rural at 16
- **town:** = 1 if lived in town at 16
- **smcity:** = 1 if in small city at 16
- **y74:** = 1 if year = 74
- **y76:**
- **y78:**
- **y80:**
- **y82:**
- **y84:**
- **agesq:** age^2
- **y74educ:**

- **y76educ:**
- **y78educ:**
- **y80educ:**
- **y82educ:**
- **y84educ:**

Notes

(1) Much more recent data can be obtained from the National Opinion Research Center website, http://www.norc.org/GSS+Website/Download/. Very rich pooled cross sections can be constructed to study a variety of issues – not just changes in fertility over time. It would be interesting to analyze a similar data set for a developing country, especially where efforts have been made to emphasize birth control. Some measure of access to birth control could be useful if it varied by region. Sometimes, one can find policy changes in the advertisement or availability of contraceptives.

Used in Text: pages 449-450, 476, 541, 625, 681

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(fertil1)
```

`fertil2` *fertil2*

Description

Wooldridge Source: These data were obtained by James Heakins, a former MSU undergraduate, for a term project. They come from Botswana's 1988 Demographic and Health Survey. Data loads lazily.

Usage

```
data('fertil2')
```

Format

A data.frame with 4361 observations on 27 variables:

- **mnthborn:** month woman born
- **yearborn:** year woman born
- **age:** age in years
- **electric:** =1 if has electricity

- **radio:** =1 if has radio
- **tv:** =1 if has tv
- **bicycle:** =1 if has bicycle
- **educ:** years of education
- **ceb:** children ever born
- **agefbrth:** age at first birth
- **children:** number of living children
- **knowmeth:** =1 if know about birth control
- **usemeth:** =1 if ever use birth control
- **monthfm:** month of first marriage
- **yearfm:** year of first marriage
- **agefm:** age at first marriage
- **idlnchld:** 'ideal' number of children
- **heduc:** husband's years of education
- **agesq:** age^2
- **urban:** =1 if live in urban area
- **urb_educ:** urban*educ
- **spirit:** =1 if religion == spirit
- **protest:** =1 if religion == protestant
- **catholic:** =1 if religion == catholic
- **frsthalf:** =1 if mnthborn <= 6
- **educ0:** =1 if educ == 0
- **evermarr:** =1 if ever married

Notes

Currently, this data set is used only in one computer exercise. Since the dependent variable of interest – number of living children or number of children every born – is a count variable, the Poisson regression model discussed in Chapter 17 can be used. However, some care is required to combine Poisson regression with an endogenous explanatory variable (educ). I refer you to Chapter 19 of my book Econometric Analysis of Cross Section and Panel Data. Even in the context of linear models, much can be done beyond Computer Exercise C15.2. At a minimum, the binary indicators for various religions can be added as controls. One might also interact the schooling variable, educ, with some of the exogenous explanatory variables.

Used in Text: page 547

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(fertil2)
```

`fertil3` *fertil3*

Description

Wooldridge Source: L.A. Whittington, J. Alm, and H.E. Peters (1990), "Fertility and the Personal Exemption: Implicit Pronatalist Policy in the United States," American Economic Review 80, 545-556. The data are given in the article. Data loads lazily.

Usage

```
data('fertil3')
```

Format

A data.frame with 72 observations on 24 variables:

- **gfr:** births per 1000 women 15-44
- **pe:** real value pers. exemption, $
- **year:** 1913 to 1984
- **t:** time trend, t=1,...,72
- **tsq:** t^2
- **pe_1:** pe[_n-1]
- **pe_2:** pe[_n-2]
- **pe_3:** pe[_n-3]
- **pe_4:** pe[_n-4]
- **pill:** =1 if year >= 1963
- **ww2:** =1, 1941 to 1945
- **tcu:** t^3
- **cgfr:** change in gfr: gfr - gfr_1
- **cpe:** pe - pe_1
- **cpe_1:** cpe[_n-1]
- **cpe_2:** cpe[_n-2]
- **cpe_3:** cpe[_n-3]
- **cpe_4:** cpe[_n-4]
- **gfr_1:** gfr[_n-1]
- **cgfr_1:** cgfr[_n-1]
- **cgfr_2:** cgfr[_n-2]
- **cgfr_3:** cgfr[_n-3]
- **cgfr_4:** cgfr[_n-4]
- **gfr_2:** gfr[_n-2]

Used in Text

pages 358, 377, 378, 397-398, 401, 408, 441, 649, 664-665, 673

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(fertil3)
```

`fish` *fish*

Description

Wooldridge Source: K Graddy (1995), "Testing for Imperfect Competition at the Fulton Fish Market," RAND Journal of Economics 26, 75-92. Professor Graddy's collaborator on a later paper, Professor Joshua Angrist at MIT, kindly provided me with these data. Data loads lazily.

Usage

```
data('fish')
```

Format

A data.frame with 97 observations on 20 variables:

- **prca:** price for Asian buyers
- **prcw:** price for white buyers
- **qtya:** quantity sold to Asians
- **qtyw:** quantity sold to whites
- **mon:** =1 if Monday
- **tues:** =1 if Tuesday
- **wed:** =1 if Wednesday
- **thurs:** =1 if Thursday
- **speed2:** min past 2 days wind speeds
- **wave2:** avg max last 2 days wave height
- **speed3:** 3 day lagged max windspeed
- **wave3:** avg max wave hghts of 3 & 4 day lagged hghts
- **avgprc:** ((prca*qtya) + (prcw*qtyw))/(qtya + qtyw)
- **totqty:** qtya + qtyw

- **lavgprc:** log(avgprc)
- **ltotqty:** log(totqty)
- **t:** time trend
- **lavgp_1:** lavgprc[_n-1]
- **gavgprc:** lavgprc - lavgp_1
- **gavgp_1:** gavgprc[_n-1]

Notes

This is a nice example of how to go about finding exogenous variables to use as instrumental variables. Often, weather conditions can be assumed to affect supply while having a negligible effect on demand. If so, the weather variables are valid instrumental variables for price in the demand equation. It is a simple matter to test whether prices vary with weather conditions by estimating the reduced form for price.

Used in Text: pages 443, 580

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(fish)
```

`fringe` *fringe*

Description

Wooldridge Source: F. Vella (1993), "A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors," International Economic Review 34, 441-457. Professor Vella kindly provided the data. Data loads lazily.

Usage

```
data('fringe')
```

Format

A data.frame with 616 observations on 39 variables:

- **annearn:** annual earnings, $
- **hrearn:** hourly earnings, $
- **exper:** years work experience

- **age:** age in years
- **depends:** number of dependents
- **married:** =1 if married
- **tenure:** years with current employer
- **educ:** years schooling
- **nrtheast:** =1 if live in northeast
- **nrthcen:** =1 if live in north central
- **south:** =1 if live in south
- **male:** =1 if male
- **white:** =1 if white
- **union:** =1 if union member
- **office:**
- **annhrs:** annual hours worked
- **ind1:** industry dummy
- **ind2:**
- **ind3:**
- **ind4:**
- **ind5:**
- **ind6:**
- **ind7:**
- **ind8:**
- **ind9:**
- **vacdays:** $ value of vac. days
- **sicklve:** $ value of sick leave
- **insur:** $ value of employee insur
- **pension:** $ value of employee pension
- **annbens:** vacdays+sicklve+insur+pension
- **hrbens:** hourly benefits, $
- **annhrssq:** annhrs^2
- **beratio:** annbens/annearn
- **lannhrs:** log(annhrs)
- **tenuresq:** tenure^2
- **expersq:** exper^2
- **lannearn:** log(annearn)
- **peratio:** pension/annearn
- **vserat:** (vacdays+sicklve)/annearn

Notes

Currently, this data set is used in only one Computer Exercise – to illustrate the Tobit model. It can be used much earlier. First, one could just ignore the pileup at zero and use a linear model where any of the hourly benefit measures is the dependent variable. Another possibility is to use this data set for a problem set in Chapter 4, after students have read Example 4.10. That example, which uses teacher salary/benefit data at the school level, finds the expected tradeoff, although it appears to less than one-to-one. By contrast, if you do a similar analysis with FRINGE.RAW, you will not find a tradeoff. A positive coefficient on the benefit/salary ratio is not too surprising because we probably cannot control for enough factors, especially when looking across different occupations. The Michigan school-level data is more aggregated than one would like, but it does restrict attention to a more homogeneous group: high school teachers in Michigan.

Used in Text: page 624-625

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(fringe)
```

---

gpa1 *gpa1*

Description

Wooldridge Source: Christopher Lemmon, a former MSU undergraduate, collected these data from a survey he took of MSU students in Fall 1994. Data loads lazily.

Usage

```
data('gpa1')
```

Format

A data.frame with 141 observations on 29 variables:

- **age:** in years
- **soph:** =1 if sophomore
- **junior:** =1 if junior
- **senior:** =1 if senior
- **senior5:** =1 if fifth year senior
- **male:** =1 if male
- **campus:** =1 if live on campus

- **business:** =1 if business major
- **engineer:** =1 if engineering major
- **colGPA:** MSU GPA
- **hsGPA:** high school GPA
- **ACT:** 'achievement' score
- **job19:** =1 if job <= 19 hours
- **job20:** =1 if job >= 20 hours
- **drive:** =1 if drive to campus
- **bike:** =1 if bicycle to campus
- **walk:** =1 if walk to campus
- **voluntr:** =1 if do volunteer work
- **PC:** =1 of pers computer at sch
- **greek:** =1 if fraternity or sorority
- **car:** =1 if own car
- **siblings:** =1 if have siblings
- **bgfriend:** =1 if boy- or girlfriend
- **clubs:** =1 if belong to MSU club
- **skipped:** avg lectures missed per week
- **alcohol:** avg # days per week drink alc.
- **gradMI:** =1 if Michigan high school
- **fathcoll:** =1 if father college grad
- **mothcoll:** =1 if mother college grad

Notes

This is a nice example of how students can obtain an original data set by focusing locally and carefully composing a survey.

Used in Text: pages 75, 77, 81, 129-130, 160, 232, 262, 295-296, 300-301

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(gpa1)
```

gpa2 *gpa2*

Description

Wooldridge Source: For confidentiality reasons, I cannot provide the source of these data. I can say that Data loads lazily.

Usage

```
data('gpa2')
```

Format

A data.frame with 4137 observations on 12 variables:

- **sat:** combined SAT score
- **tothrs:** total hours through fall semest
- **colgpa:** GPA after fall semester
- **athlete:** =1 if athlete
- **verbmath:** verbal/math SAT score
- **hsize:** size grad. class, 100s
- **hsrank:** rank in grad. class
- **hsperc:** high school percentile, from top
- **female:** =1 if female
- **white:** =1 if white
- **black:** =1 if black
- **hsizesq:** hsize^2

Used in Text

pages 106, 184, 208-209, 210-211, 221, 259, 262-263

they come from a midsize research university that also supports men's and women's athletics at the Division I level.

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(gpa2)
```

gpa3 *gpa3*

Description

Wooldridge Source: See GPA2.RAW Data loads lazily.

Usage

```
data('gpa3')
```

Format

A data.frame with 732 observations on 23 variables:

- **term:** fall = 1, spring = 2
- **sat:** SAT score
- **tothrs:** total hours prior to term
- **cumgpa:** cumulative GPA
- **season:** =1 if in season
- **frstsem:** =1 if student's 1st semester
- **crsgpa:** weighted course GPA
- **verbmath:** verbal SAT to math SAT ratio
- **trmgpa:** term GPA
- **hssize:** size h.s. grad. class
- **hsrank:** rank in h.s. class
- **id:** student identifier
- **spring:** =1 if spring term
- **female:** =1 if female
- **black:** =1 if black
- **white:** =1 if white
- **ctrmgpa:** change in trmgpa
- **ctothrs:** change in total hours
- **ccrsgpa:** change in crsgpa
- **ccrspop:** change in crspop
- **cseason:** change in season
- **hsperc:** percentile in h.s.
- **football:** =1 if football player

Used in Text

pages 246-248, 273, 297-298, 478

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(gpa3)
```

`happiness` *happiness*

Description

Wooldridge Data loads lazily.

Usage

```
data('happiness')
```

Format

A data.frame with 17137 observations on 33 variables:

- **year:** gss year for this respondent
- **workstat:** work force status
- **prestige:** occupational prestige score
- **divorce:** ever been divorced or separated
- **widowed:** ever been widowed
- **educ:** highest year of school completed
- **reg16:** region of residence, age 16
- **babies:** household members less than 6 yrs old
- **preteen:** household members 6 thru 12 yrs old
- **teens:** household members 13 thru 17 yrs old
- **income:** total family income
- **region:** region of interview
- **attend:** how often r attends religious services
- **happy:** general happiness
- **owngun:** =1 if own gun
- **tvhours:** hours per day watching tv
- **vhappy:** =1 if 'very happy'
- **mothfath16:** =1 if live with mother and father at 16
- **black:** =1 if black

- **gwbush04:** =1 if voted for G.W. Bush in 2004
- **female:** =1 if female
- **blackfemale:** black*female
- **gwbush00:** =1 if voted for G.W. Bush in 2000
- **occattend:** =1 if attend is 3, 4, or 5
- **regattend:** =1 if attend is 6, 7, or 8
- **y94:** =1 if year == 1994
- **y96:** =1 if year == 1996
- **y98:** =1 if year == 1998
- **y00:** =1 if year == 2000
- **y02:** =1 if year == 2002
- **y04:** =1 if year == 2004
- **y06:** =1 if year == 2006
- **unem10:** =1 if unemployed in last 10 years

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(happiness)
```

`hprice1` *hprice1*

Description

Wooldridge Source: Collected from the real estate pages of the Boston Globe during 1990. These are homes that sold in the Boston, MA area. Data loads lazily.

Usage

```
data('hprice1')
```

Format

A data.frame with 88 observations on 10 variables:

- **price:** house price, $1000s
- **assess:** assessed value, $1000s
- **bdrms:** number of bdrms
- **lotsize:** size of lot in square feet
- **sqrft:** size of house in square feet
- **colonial:** =1 if home is colonial style
- **lprice:** log(price)
- **lassess:** log(assess
- **llotsize:** log(lotsize)
- **lsqrft:** log(sqrft)

Notes

Typically, it is very easy to obtain data on selling prices and characteristics of homes, using publicly available data bases. It is interesting to match the information on houses with other information – such as local crime rates, quality of the local schools, pollution levels, and so on – and estimate the effects of such variables on housing prices.

Used in Text: pages 110, 153-154, 160-161, 165, 211-212, 221, 222, 234, 278, 280, 299, 307

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(hprice1)
```

`hprice2` *hprice2*

Description

Wooldridge Source: D. Harrison and D.L. Rubinfeld (1978), "Hedonic Housing Prices and the Demand for Clean Air," by Harrison, D. and D.L.Rubinfeld, Journal of Environmental Economics and Management 5, 81-102. Diego Garcia, a former Ph.D. student in economics at MIT, kindly provided these data, which he obtained from the book Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, by D.A. Belsey, E. Kuh, and R. Welsch, 1990. New York: Wiley. Data loads lazily.

Usage

```
data('hprice2')
```

Format

A data.frame with 506 observations on 12 variables:

- **price:** median housing price, $
- **crime:** crimes committed per capita
- **nox:** nit ox concen; parts per 100m
- **rooms:** avg number of rooms
- **dist:** wght dist to 5 employ centers
- **radial:** access. index to rad. hghwys
- **proptax:** property tax per $1000
- **stratio:** average student-teacher ratio
- **lowstat:** perc of people 'lower status'
- **lprice:** log(price)
- **lnox:** log(nox)
- **lproptax:** log(proptax)

Notes

The census contains rich information on variables such as median housing prices, median income levels, average family size, and so on, for fairly small geographical areas. If such data can be merged with pollution data, one can update the Harrison and Rubinfeld study. Presumably, this has been done in academic journals.

Used in Text: pages 108, 132-133, 190-191, 196-197.

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(hprice2)
```

hprice3 *hprice3*

Description

Wooldridge Data loads lazily.

Usage

```
data('hprice3')
```

Format

A data.frame with 321 observations on 19 variables:

- **year:** 1978, 1981
- **age:** age of house
- **agesq:** age^2
- **nbh:** neighborhood, 1-6
- **cbd:** dist. to cent. bus. dstrct, ft.
- **inst:** dist. to interstate, ft.
- **linst:** log(inst)
- **price:** selling price
- **rooms:** # rooms in house
- **area:** square footage of house
- **land:** square footage lot
- **baths:** # bathrooms
- **dist:** dist. from house to incin., ft.
- **ldist:** log(dist)
- **lprice:** log(price)
- **y81:** =1 if year = 1981
- **larea:** log(area)
- **lland:** log(land)
- **linstsq:** linst^2

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(hprice3)
```

hseinv *hseinv*

Description

Wooldridge Source: D. McFadden (1994), "Demographics, the Housing Market, and the Welfare of the Elderly," in D.A. Wise (ed.), Studies in the Economics of Aging. Chicago: University of Chicago Press, 225-285. The data are contained in the article. Data loads lazily.

Usage

```
data('hseinv')
```

Format

A data.frame with 42 observations on 14 variables:

- **year:** 1947-1988
- **inv:** real housing inv, millions $
- **pop:** population, 1000s
- **price:** housing price index; 1982 = 1
- **linv:** log(inv)
- **lpop:** log(pop)
- **lprice:** log(price)
- **t:** time trend: t=1,...,42
- **invpc:** per capita inv: inv/pop
- **linvpc:** log(invpc)
- **lprice_1:** lprice[_n-1]
- **linvpc_1:** linvpc[_n-1]
- **gprice:** lprice - lprice_1
- **ginvpc:** linvpc - linvpc_1

Used in Text

pages 367, 370, 407, 638-639, 822?

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(hseinv)
```

htv *htv*

Description

Wooldridge Source: J.J. Heckman, J.L. Tobias, and E. Vytlacil (2003), "Simple Estimators for Treatment Parameters in a Latent-Variable Framework," Review of Economics and Statistics 85, 748-755. Professor Tobias kindly provided the data, which were obtained from the 1991 National Longitudinal Survey of Youth. All people in the sample are males age 26 to 34. For confidentiality reasons, I have included only a subset of the variables used by the authors. Data loads lazily.

Usage

```
data('htv')
```

Format

A data.frame with 1230 observations on 23 variables:

- **wage:** hourly wage, 1991
- **abil:** abil. measure, not standardized
- **educ:** highest grade completed by 1991
- **ne:** =1 if in northeast, 1991
- **nc:** =1 if in nrthcntrl, 1991
- **west:** =1 if in west, 1991
- **south:** =1 if in south, 1991
- **exper:** potential experience
- **motheduc:** highest grade, mother
- **fatheduc:** highest grade, father
- **brkhme14:** =1 if broken home, age 14
- **sibs:** number of siblings
- **urban:** =1 if in urban area, 1991
- **ne18:** =1 if in NE, age 18
- **nc18:** =1 if in NC, age 18
- **south18:** =1 if in south, age 18
- **west18:** =1 if in west, age 18
- **urban18:** =1 if in urban area, age 18
- **tuit17:** college tuition, age 17
- **tuit18:** college tuition, age 18
- **lwage:** log(wage)
- **expersq:** exper^2
- **ctuit:** tuit18 - tuit17

Notes

Because an ability measure is included in this data set, it can be used as another illustration of including proxy variables in regression models. See Chapter 9. Also, one can try the IV procedure with the ability measure included as an exogenous explanatory variable.

Used in Text: pages 550, 628

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(htv)
```

`infmrt` *infmrt*

Description

Wooldridge Source: Statistical Abstract of the United States, 1990 and 1994. (For example, the infant mortality rates come from Table 113 in 1990 and Table 123 in 1994.) Data loads lazily.

Usage

```
data('infmrt')
```

Format

A data.frame with 102 observations on 12 variables:

- **year:** 1987 or 1990
- **infmort:** deaths per 1,000 live births
- **afdcprt:** afdc partic., 1000s
- **popul:** population, 1000s
- **pcinc:** per capita income
- **physic:** drs. per 100,000 civilian pop.
- **afdcper:** percent on AFDC
- **d90:** =1 if year == 1990
- **lpcinc:** log(pcinc)
- **lphysic:** log(physic)
- **DC:** =1 for Washington DC
- **lpopul:** log(popul)

Notes

An interesting exercise is to add the percentage of the population on AFDC (afdcper) to the infant mortality equation. Pooled OLS and first differencing can give very different estimates. Adding the years 1998 and 2002 and applying fixed effects seems natural. Intervening years can be added, too, although variation in the key variables from year to year might be minimal.

Used in Text: pages 330-331, 339

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(infmrt)
```

`injury` *injury*

Description

Wooldridge Source: B.D. Meyer, W.K. Viscusi, and D.L. Durbin (1995), "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment," American Economic Review 85, 322-340. Professor Meyer kindly provided the data. Data loads lazily.

Usage

```
data('injury')
```

Format

A data.frame with 7150 observations on 30 variables:

- **durat:** duration of benefits
- **afchnge:** =1 if after change in benefits
- **highearn:** =1 if high earner
- **male:** =1 if male
- **married:** =1 if married
- **hosp:** =1 if inj. required hosp. stay
- **indust:** industry
- **injtype:** type of injury
- **age:** age at time of injury
- **prewage:** previous weekly wage, 1982 $
- **totmed:** total med. costs, 1982 $

- **injdes:** 4 digit injury description
- **benefit:** real dollar value of benefit
- **ky:** =1 for kentucky
- **mi:** =1 for michigan
- **ldurat:** log(durat)
- **afhigh:** afchnge*highearn
- **lprewage:** log(wage)
- **lage:** log(age)
- **ltotmed:** log(totmed); = 0 if totmed < 1
- **head:** =1 if head injury
- **neck:** =1 if neck injury
- **upextr:** =1 if upper extremities injury
- **trunk:** =1 if trunk injury
- **lowback:** =1 if lower back injury
- **lowextr:** =1 if lower extremities injury
- **occdis:** =1 if occupational disease
- **manuf:** =1 if manufacturing industry
- **construc:** =1 if construction industry
- **highlpre:** highearn*lprewage

Notes

This data set also can be used to illustrate the Chow test in Chapter 7. In particular, students can test whether the regression functions differ between Kentucky and Michigan. Or, allowing for different intercepts for the two states, do the slopes differ? A good lesson from this example is that a small R-squared is compatible with the ability to estimate the effects of a policy. Of course, for the Michigan data, which has a smaller sample size, the estimated effect is much less precise (but of virtually identical magnitude).

Used in Text: pages 458-459, 475-476

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(injury)
```

intdef *intdef*

Description

Wooldridge Source: Economic Report of the President, 2004, Tables B-64, B-73, and B-79. Data loads lazily.

Usage

```
data('intdef')
```

Format

A data.frame with 56 observations on 13 variables:

- **year:** 1948 to 2003
- **i3:** 3 month T-bill rate
- **inf:** CPI inflation rate
- **rec:** federal receipts, percent GDP
- **out:** federal outlays, percent GDP
- **def:** out - rec
- **i3_1:** i3[_n-1]
- **inf_1:** inf[_n-1]
- **def_1:** def[_n-1]
- **ci3:** i3 - i3_1
- **cinf:** inf - inf_1
- **cdef:** def - def_1
- **y77:** =1 if year >= 1977; change in FY

Used in Text

pages 356, 377, 430, 547-548

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(intdef)
```

intqrt *intqrt*

Description

Wooldridge Source: From Salomon Brothers, Analytical Record of Yields and Yield Spreads, 1990. The folks at Salomon Brothers kindly provided the Record at no charge when I was an assistant professor at MIT. Data loads lazily.

Usage

```
data('intqrt')
```

Format

A data.frame with 124 observations on 23 variables:

- **r3:** bond equiv. yield, 3 mo T-bill
- **r6:** bond equiv. yield, 6 mo T-bill
- **r12:** yield on 1 yr. bond
- **p3:** price of 3 mo. T-bill
- **p6:** price of 6 mo. T-bill
- **hy6:** 100*(p3 - p6[_n-1])/p6[_n-1])
- **hy3:** r3*(91/365)
- **spr63:** r6 - r3
- **hy3_1:** hy3[_n-1]
- **hy6_1:** hy6[_n-1]
- **spr63_1:** spr63[_n-1]
- **hy6hy3_1:** hy6 - hy3_1
- **cr3:** r3 - r3_1
- **r3_1:** r3[_n-1]
- **chy6:** hy6 - hy6_1
- **chy3:** hy3 - hy3_1
- **chy6_1:** chy6[_n-1]
- **chy3_1:** chy3[_n-1]
- **cr6:** r6 - r6_1
- **cr6_1:** cr6[_n-1]
- **cr3_1:** cr3[_n-1]
- **r6_1:** r6[_n-1]
- **cspr63:** spr63 - spr63_1

Notes

A nice feature of the Salomon Brothers data is that the interest rates are not averaged over a month or quarter – they are end-of-month or end-of-quarter rates. Asset pricing theories apply to such "point-sampled" data, and not to averages over a period. Most other sources report monthly or quarterly averages. This is a good data set to update and test whether current data are more or less supportive of basic asset pricing theories.

Used in Text: pages 405-406, 641, 646-647, 650, 652, 672, 673

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(intqrt)
```

`inven` *inven*

Description

Wooldridge Source: Economic Report of the President, 1997, Tables B-4, B-20, B-61, and B-71. Data loads lazily.

Usage

```
data('inven')
```

Format

A data.frame with 37 observations on 13 variables:

- **year:** 1959-1995
- **i3:** 3 mo. T-bill rate
- **inf:** CPI inflation rate
- **inven:** inventories, billions '92 $
- **gdp:** GDP, billions '92 $
- **r3:** real interest: i3 - inf
- **cinven:** inven - inven[_n-1]
- **cgdp:** gdp - gdp[_n-1]
- **cr3:** r3 - r3[_n-1]
- **ci3:** i3 - i3[_n-1]
- **cinf:** inf - inf[_n-1]
- **ginven:** log(inven) - log(inven[_n-1])
- **ggdp:** log(gdp) - log(gdp[_n-1])

Used in Text

pages 408, 444, 643, 830

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(inven)
```

---

jtrain *jtrain*

Description

Wooldridge Source: H. Holzer, R. Block, M. Cheatham, and J. Knott (1993), "Are Training Subsidies Effective? The Michigan Experience," Industrial and Labor Relations Review 46, 625-636. The authors kindly provided the data. Data loads lazily.

Usage

```
data('jtrain')
```

Format

A data.frame with 471 observations on 30 variables:

- **year:** 1987, 1988, or 1989
- **fcode:** firm code number
- **employ:** # employees at plant
- **sales:** annual sales, $
- **avgsal:** average employee salary
- **scrap:** scrap rate (per 100 items)
- **rework:** rework rate (per 100 items)
- **tothrs:** total hours training
- **union:** =1 if unionized
- **grant:** = 1 if received grant
- **d89:** = 1 if year = 1989
- **d88:** = 1 if year = 1988
- **totrain:** total employees trained
- **hrsemp:** tothrs/totrain

- **lscrap:** log(scrap)
- **lemploy:** log(employ)
- **lsales:** log(sales)
- **lrework:** log(rework)
- **lhrsemp:** log(1 + hrsemp)
- **lscrap_1:** lagged lscrap; missing 1987
- **grant_1:** lagged grant; assumed 0 in 1987
- **clscrap:** lscrap - lscrap_1; year > 1987
- **cgrant:** grant - grant_1
- **clemploy:** lemploy - lemploy[_n-1]
- **clsales:** lavgsal - lavgsal[_n-1]
- **lavgsal:** log(avgsal)
- **clavgsal:** lavgsal - lavgsal[_n-1]
- **cgrant_1:** cgrant[_n-1]
- **chrsemp:** hrsemp - hrsemp[_n-1]
- **clhrsemp:** lhrsemp - lhrsemp[_n-1]

Used in Text

pages 137, 161, 233, 254, 339, 465-466, 479, 486-487, 492, 504, 541-542, 774-775, 786-787, 788, 819.

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(jtrain)
```

---

jtrain2 *jtrain2*

Description

Wooldridge Source: R.J. Lalonde (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," American Economic Review 76, 604-620. Professor Jeff Biddle, at MSU, kindly passed the data set along to me. He obtained it from Professor Lalonde. Data loads lazily.

Usage

```
data('jtrain2')
```

Format

A data.frame with 445 observations on 19 variables:

- **train:** =1 if assigned to job training
- **age:** age in 1977
- **educ:** years of education
- **black:** =1 if black
- **hisp:** =1 if Hispanic
- **married:** =1 if married
- **nodegree:** =1 if no high school degree
- **mosinex:** # mnths prior to 1/78 in expmnt
- **re74:** real earns., 1974, $1000s
- **re75:** real earns., 1975, $1000s
- **re78:** real earns., 1978, $1000s
- **unem74:** =1 if unem. all of 1974
- **unem75:** =1 if unem. all of 1975
- **unem78:** =1 if unem. all of 1978
- **lre74:** log(re74); zero if re74 == 0
- **lre75:** log(re75); zero if re75 == 0
- **lre78:** log(re78); zero if re78 == 0
- **agesq:** age^2
- **mostrn:** months in training

Notes

Professor Lalonde obtained the data from the National Supported Work Demonstration job-training program conducted by the Manpower Demonstration Research Corporation in the mid 1970s. Training status was randomly assigned, so this is essentially experimental data. Computer Exercise C17.8 looks only at the effects of training on subsequent unemployment probabilities. For illustrating the more advanced methods in Chapter 17, a good exercise would be to have the students estimate a Tobit of re78 on train, and obtain estimates of the expected values for those with and without training. These can be compared with the sample averages.

Used in Text: pages 18, 340-341, 626

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(jtrain2)
```

jtrain3 *jtrain3*

Description

Wooldridge Source: R.H. Dehejia and S. Wahba (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," Journal of the American Statistical Association 94, 1053-1062. Professor Sergio Firpo, at the University of British Columbia, has used this data set in his recent work, and he kindly provided it to me. This data set is a subset of that originally used by Lalonde in the study cited for JTRAIN2.RAW. Data loads lazily.

Usage

```
data('jtrain3')
```

Format

A data.frame with 2675 observations on 20 variables:

- **train:** =1 if in job training
- **age:** in years, 1977
- **educ:** years of schooling
- **black:** =1 if black
- **hisp:** =1 if Hispanic
- **married:** =1 if married
- **re74:** '74 earnings, $1000s '82
- **re75:** '75 earnings, $1000s '82
- **unem75:** =1 if unem. all of '75
- **unem74:** =1 if unem. all of '74
- **re78:** '78 earnings, $1000s '82
- **agesq:** age^2
- **trre74:** train*re74
- **trre75:** train*re75
- **trun74:** train*unem74
- **trun75:** train*unem75
- **avgre:** (re74 + re75)/2
- **travgre:** train*avgre
- **unem78:** =1 if unem. all of '78
- **em78:** 1 - unem78

Used in Text

pages 340-341, 480-481

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(jtrain3)
```

`jtrain98` *jtrain98*

Description

Wooldridge Source: This is a data set I created many years ago intended as an update to the files JTRAIN2 and JTRAIN3. While the data were partly generated by me, the data attributes are similar to data sets used to evaluate job training programs. Data loads lazily.

Usage

```
data('jtrain98')
```

Format

A data.frame with 1130 observations on 10 variables:

- **train:** =1 if in job training
- **age:** in years
- **educ:** years of schooling
- **black:** =1 if black
- **hisp:** =1 if Hispanic
- **married:** =1 if married
- **earn96:** earnings in 1996, $1000s
- **unem96:** =1 if unemployed all of 1995
- **earn98:** earnings in 1998, $1000s
- **unem98:** =1 if unemployed all of 1998

Notes

The response variables, earn98 and unem98, both have discreteness: the former is a corner solutions (takes on the value zero and then a range of strictly positive values) and the latter is binary. One could use these in an exercise using methods in Chapter 17. unem98 can be used in a probit or logit model, earn98 in a Tobit model, or in Poisson regression (without assuming, of course, that the Poisson distribution is correct).

Used in Text: 101-102, 248, 601

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781337558860

Examples

```
str(jtrain98)
```

`k401k` *k401k*

Description

Wooldridge Source: L.E. Papke (1995), "Participation in and Contributions to 401(k) Pension Plans:Evidence from Plan Data," Journal of Human Resources 30, 311-325. Professor Papke kindly provided these data. She gathered them from the Internal Revenue Service's Form 5500 tapes. Data loads lazily.

Usage

```
data('k401k')
```

Format

A data.frame with 1534 observations on 8 variables:

- **prate:** participation rate, percent
- **mrate:** 401k plan match rate
- **totpart:** total 401k participants
- **totelg:** total eligible for 401k plan
- **age:** age of 401k plan
- **totemp:** total number of firm employees
- **sole:** = 1 if 401k is firm's sole plan
- **ltotemp:** log of totemp

Notes

This data set is used in a variety of ways in the text. One additional possibility is to investigate whether the coefficients from the regression of prate on mrate, log(totemp) differ by whether the plan is a sole plan. The Chow test (see Section 7.4), and the less restrictive version that allows different intercepts, can be used.

Used in Text: pages 63, 79, 136, 174, 219, 692

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(k401k)
```

`k401ksubs` *k401ksubs*

Description

Wooldridge Source: A. Abadie (2003), "Semiparametric Instrumental Variable Estimation of Treatment Response Models," Journal of Econometrics 113, 231-263. Professor Abadie kindly provided these data. He obtained them from the 1991 Survey of Income and Program Participation (SIPP). Data loads lazily.

Usage

```
data('k401ksubs')
```

Format

A data.frame with 9275 observations on 11 variables:

- **e401k:** =1 if eligble for 401(k)
- **inc:** annual income, $1000s
- **marr:** =1 if married
- **male:** =1 if male respondent
- **age:** in years
- **fsize:** family size
- **nettfa:** net total fin. assets, $1000
- **p401k:** =1 if participate in 401(k)
- **pira:** =1 if have IRA
- **incsq:** inc^2
- **agesq:** age^2

Notes

This data set can also be used to illustrate the binary response models, probit and logit, in Chapter 17, where, say, pira (an indicator for having an individual retirement account) is the dependent variable, and e401k [the 401(k) eligibility indicator] is the key explanatory variable.

Used in Text: pages 166, 174, 223, 264, 283, 301-302, 340, 549

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(k401ksubs)
```

`kielmc` *kielmc*

Description

Wooldridge Source: K.A. Kiel and K.T. McClain (1995), "House Prices During Siting Decision Stages: The Case of an Incinerator from Rumor Through Operation," Journal of Environmental Economics and Management 28, 241-255. Professor McClain kindly provided the data, of which I used only a subset. Data loads lazily.

Usage

```
data('kielmc')
```

Format

A data.frame with 321 observations on 25 variables:

- **year:** 1978 or 1981
- **age:** age of house
- **agesq:** age^2
- **nbh:** neighborhood, 1-6
- **cbd:** dist. to cent. bus. dstrct, ft.
- **intst:** dist. to interstate, ft.
- **lintst:** log(intst)
- **price:** selling price
- **rooms:** # rooms in house
- **area:** square footage of house
- **land:** square footage lot

- **baths:** # bathrooms
- **dist:** dist. from house to incin., ft.
- **ldist:** log(dist)
- **wind:** prc. time wind incin. to house
- **lprice:** log(price)
- **y81:** =1 if year == 1981
- **larea:** log(area)
- **lland:** log(land)
- **y81ldist:** y81*ldist
- **lintstsq:** lintst^2
- **nearinc:** =1 if dist <= 15840
- **y81nrinc:** y81*nearinc
- **rprice:** price, 1978 dollars
- **lrprice:** log(rprice)

Used in Text

pages 220, 454-457, 475, 477

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(kielmc)
```

---

labsup *labsup*

Description

Wooldridge Source: The subset of data for black or Hispanic women used in J.A. Angrist and W.E. Evans (1998) Data loads lazily.

Usage

```
data('labsup')
```

Format

A data.frame with 31857 observations on 20 variables:

- **kids:** number of kids
- **morekids:** had more than 2 kids
- **boys2:** first two births boys
- **girls2:** first two births girls
- **boy1st:** first birth boy
- **boy2nd:** second birth boy
- **samesex:** first two kids are of same sex
- **multi2nd:** =1 if 2nd birth is twin
- **age:** age of mom
- **agefstm:** age of mom at first birth
- **black:** =1 of black
- **hispan:** =1 if hispanic
- **worked:** mom worked last year
- **weeks:** weeks worked mom
- **hours:** hours of work per week, mom
- **labinc:** mom's labor income, $1000s
- **faminc:** family income, $1000s
- **nonmomi:** 'non-mom' income, $1000s
- **educ:** mom's years of education
- **agesq:**

Notes

This example can promote an interesting discussion of instrument validity, and in particular, how a variable that is beyond our control – for example, whether the first two children have the same gender – can, nevertheless, affect subsequent economic choices. Students are asked to think about such issues in Computer Exercise C13 in Chapter 15. A more egregious version of this mistake would be to treat a variable such as age as a suitable instrument because it is beyond our control: clearly age has a direct effect on many economic outcomes that would play the role of the dependent variable.

Used in Text: pages 530-531

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781337558860

Examples

```
str(labsup)
```

`lawsch85` *lawsch85*

Description

Wooldridge Source: Collected by Kelly Barnett, an MSU economics student, for use in a term project. The data come from two sources: The Official Guide to U.S. Law Schools, 1986, Law School Admission Services, and The Gourman Report: A Ranking of Graduate and Professional Programs in American and International Universities, 1995, Washington, D.C. Data loads lazily.

Usage

```
data('lawsch85')
```

Format

A data.frame with 156 observations on 21 variables:

- **rank:** law school ranking
- **salary:** median starting salary
- **cost:** law school cost
- **LSAT:** median LSAT score
- **GPA:** median college GPA
- **libvol:** no. volumes in lib., 1000s
- **faculty:** no. of faculty
- **age:** age of law sch., years
- **clsize:** size of entering class
- **north:** =1 if law sch in north
- **south:** =1 if law sch in south
- **east:** =1 if law sch in east
- **west:** =1 if law sch in west
- **lsalary:** log(salary)
- **studfac:** student-faculty ratio
- **top10:** =1 if ranked in top 10
- **r11_25:** =1 if ranked 11-25
- **r26_40:** =1 if ranked 26-40
- **r41_60:** =1 if ranked 41-60
- **llibvol:** log(libvol)
- **lcost:** log(cost)

Notes

More recent versions of both cited documents are available. One could try a similar analysis for, say, MBA programs or Ph.D. programs in economics. Quality of placements may be a good dependent variable, and measures of business school or graduate program quality could be included among the explanatory variables. Of course, one would want to control for factors describing the incoming class so as to isolate the effect of the program itself.

Used in Text: pages 107, 164-165, 239

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(lawsch85)
```

---

loanapp *loanapp*

Description

Wooldridge Source: W.C. Hunter and M.B. Walker (1996), "The Cultural Affinity Hypothesis and Mortgage Lending Decisions," Journal of Real Estate Finance and Economics 13, 57-70. Professor Walker kindly provided the data. Data loads lazily.

Usage

```
data('loanapp')
```

Format

A data.frame with 1989 observations on 59 variables:

- **occ:** occupancy
- **loanamt:** loan amt in thousands
- **action:** type of action taken
- **msa:** msa number of property
- **suffolk:** =1 if property in suffolk co.
- **appinc:** applicant income, $1000s
- **typur:** type of purchaser of loan
- **unit:** number of units in property
- **married:** =1 if applicant married
- **dep:** number of dependents

- **emp:** years employed in line of work
- **yjob:** years at this job
- **self:** =1 if self employed
- **atotinc:** total monthly income
- **cototinc:** coapp total monthly income
- **hexp:** propose housing expense
- **price:** purchase price
- **other:** other financing, $1000s
- **liq:** liquid assets
- **rep:** no. of credit reports
- **gdlin:** credit history meets guidelines
- **lines:** no. of credit lines on reports
- **mortg:** credit history on mortgage paym
- **cons:** credit history on consumer stuf
- **pubrec:** =1 if filed bankruptcy
- **hrat:** housing exp, percent total inc
- **obrat:** other oblgs, percent total inc
- **fixadj:** fixed or adjustable rate?
- **term:** term of loan in months
- **apr:** appraised value
- **prop:** type of property
- **inss:** PMI sought
- **inson:** PMI approved
- **gift:** gift as down payment
- **cosign:** is there a cosigner
- **unver:** unverifiable info
- **review:** number of times reviewed
- **netw:** net worth
- **unem:** unemployment rate by industry
- **min30:** =1 if minority pop. > 30percent
- **bd:** =1 if boarded-up val > MSA med
- **mi:** =1 if tract inc > MSA median
- **old:** =1 if applic age > MSA median
- **vr:** =1 if tract vac rte > MSA med
- **sch:** =1 if > 12 years schooling
- **black:** =1 if applicant black
- **hispan:** =1 if applicant Hispanic

- **male:** =1 if applicant male
- **reject:** =1 if action == 3
- **approve:** =1 if action == 1 or 2
- **mortno:** no mortgage history
- **mortperf:** no late mort. payments
- **mortlat1:** one or two late payments
- **mortlat2:** > 2 late payments
- **chist:** =0 if accnts deliq. >= 60 days
- **multi:** =1 if two or more units
- **loanprc:** amt/price
- **thick:** =1 if rep > 2
- **white:** =1 if applicant white

Notes

These data were originally used in a famous study by researchers at the Boston Federal Reserve Bank. See A. Munnell, G.M.B. Tootell, L.E. Browne, and J. McEneaney (1996), "Mortgage Lending in Boston: Interpreting HMDA Data," American Economic Review 86, 25-53.

Used in Text: pages 263-264, 300, 339-340, 624

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(loanapp)
```

`lowbrth` *lowbrth*

Description

Wooldridge Source: Source: Statistical Abstract of the United States, 1990, 1993, and 1994. Data loads lazily.

Usage

```
data('lowbrth')
```

Format

A data.frame with 100 observations on 36 variables:

- **year:** 1987 or 1990
- **lowbrth:** perc births low weight
- **infmort:** infant mortality rate
- **afdcprt:** # participants in AFDC, 1000s
- **popul:** population, 1000s
- **pcinc:** per capita income
- **physic:** # physicians, 1000s
- **afdcprc:** percent of pop in AFDC
- **d90:** =1 if year == 1990
- **lpcinc:** log of pcinc
- **cafdcprc:** change in afdcprc
- **clpcinc:** change in lpcinc
- **lphysic:** log of physic
- **clphysic:** change in lphysic
- **clowbrth:** change in lowbrth
- **cinfmort:** change in infmort
- **afdcpay:** avg monthly AFDC payment
- **afdcinc:** afdcpay as percent pcinc
- **lafdcpay:** log of afdcpay
- **clafdcpy:** change in lafdcpay
- **cafdcinc:** change in afdcinc
- **stateabb:** state postal code
- **state:** name of state
- **beds:** # hospital beds, 1000s
- **bedspc:** beds per capita
- **lbedspc:** log(bedspc)
- **clbedspc:** change in lbedspc
- **povrate:** percent people below poverty line
- **cpovrate:** change in povrate
- **afdcpsq:** afdcper^2
- **cafdcpsq:** change in afdcpsq
- **physicpc:** physicians per capita
- **lphypc:** log(physicpc)
- **clphypc:** change in lphypc
- **lpopul:** log(popul)
- **clpopul:** change in lpopul

Notes

This data set can be used very much like INFMRT.RAW. It contains two years of state-level panel data. In fact, it is a superset of INFMRT.RAW. The key is that it contains information on low birth weights, as well as infant mortality. It also contains state identifies, so that several years of more recent data could be added for a term project. Putting in the variable afcdprc and its square leads to some interesting findings for pooled OLS and fixed effects (first differencing). After differencing, you can even try using the change in the AFDC payments variable as an instrumental variable for the change in afdcprc.

Used in Text: not used

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(lowbrth)
```

`mathpnl` *mathpnl*

Description

Wooldridge Source: Leslie Papke, an economics professor at MSU, collected these data from Michigan Department of Education web site, www.michigan.gov/mde. These are district-level data, which Professor Papke kindly provided. She has used building-level data in "The Effects of Spending on Test Pass Rates: Evidence from Michigan" (2005), Journal of Public Economics 89, 821-839. Data loads lazily.

Usage

```
data('mathpnl')
```

Format

A data.frame with 3850 observations on 52 variables:

- **distid:** district identifier
- **intid:** intermediate school district
- **lunch:** percent eligible for free lunch
- **enrol:** school enrollment
- **ptr:** pupil/teacher: 1995-98
- **found:** foundation grant, $: 1995-98
- **expp:** expenditure per pupil

- **revpp:** revenue per pupil
- **avgsal:** average teacher salary
- **drop:** high school dropout rate, percent
- **grad:** high school grad. rate, percent
- **math4:** percent satisfactory, 4th grade math
- **math7:** percent satisfactory, 7th grade math
- **choice:** number choice students
- **psa:** # public school academy studs.
- **year:** 1992-1998
- **staff:** staff per 1000 students
- **avgben:** avg teacher fringe benefits
- **y92:** =1 if year == 1992
- **y93:** =1 if year == 1993
- **y94:** =1 if year == 1994
- **y95:** =1 if year == 1995
- **y96:** =1 if year == 1996
- **y97:** =1 if year == 1997
- **y98:** =1 if year == 1998
- **lexpp:** log(expp)
- **lfound:** log(found)
- **lexpp_1:** lexpp[_n-1]
- **lfnd_1:** lfnd[_n-1]
- **lenrol:** log(enrol)
- **lenrolsq:** lenrol^2
- **lunchsq:** lunch^2
- **lfndsq:** lfnd^2
- **math4_1:** math4[_n-1]
- **cmath4:** math4 - math4_1
- **gexpp:** lexpp - lexpp_1
- **gexpp_1:** gexpp[_n-1
- **gfound:** lfound - lfnd_1
- **gfnd_1:** gfound[_n-1]
- **clunch:** lunch - lunch[_n-1]
- **clnchsq:** lunchsq - lunchsq[_n-1]
- **genrol:** lenrol - lenrol[_n-1]
- **genrolsq:** genrol^2
- **expp92:** expp in 1992

- **lexpp92:** log(expp92)
- **math4_92:** math4 in 1992
- **cpi:** consumer price index
- **rexpp:** real spending per pupil, 1997$
- **lrexpp:** log(rexpp)
- **lrexpp_1:** lrexpp[_n-1]
- **grexpp:** lrexpp - lrexpp_1
- **grexpp_1:** grexpp[_n-1]

Used in Text

pages 479-480, 505-506

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(mathpnl)
```

---

meap00_01 *meap00_01*

Description

Wooldridge Source: Michigan Department of Education, www.michigan.gov/mde Data loads lazily.

Usage

```
data('meap00_01')
```

Format

A data.frame with 1692 observations on 9 variables:

- **dcode:** district code
- **bcode:** building code
- **math4:** percent students satisfactory, 4th grade math
- **read4:** percent students satisfactory, 4th grade reading
- **lunch:** percent students eligible for free or reduced lunch
- **enroll:** school enrollment
- **exppp:** expenditures per pupil: expend/enroll
- **lenroll:** log(enroll)
- **lexppp:** log(exppp)

Used in Text

pages 224, 302

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(meap00_01)
```

`meap01` *meap01*

Description

Wooldridge Source: Michigan Department of Education, www.michigan.gov/mde Data loads lazily.

Usage

```
data('meap01')
```

Format

A data.frame with 1823 observations on 11 variables:

- **dcode:** district code
- **bcode:** building code
- **math4:** percent students satisfactory, 4th grade math
- **read4:** percent students satisfactory, 4th grade reading
- **lunch:** percent students eligible for free or reduced lunch
- **enroll:** school enrollment
- **expend:** total spending, $
- **exppp:** expenditures per pupil: expend/enroll
- **lenroll:** log(enroll)
- **lexpend:** log(expend)
- **lexppp:** log(exppp)

Notes

This is another good data set to compare simple and multiple regression estimates. The expenditure variable (in logs, say) and the poverty measure (lunch) are negatively correlated in this data set. A simple regression of math4 on lexppp gives a negative coefficient. Controlling for lunch makes the spending coefficient positive and significant.

Used in Text: page 18

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(meap01)
```

---

meap93 *meap93*

Description

Wooldridge Source: I collected these data from the old Michigan Department of Education web site. See MATHPNL.RAW for the current web site. I used data on most high schools in the state of Michigan for 1993. I dropped some high schools that had suspicious-looking data. Data loads lazily.

Usage

```
data('meap93')
```

Format

A data.frame with 408 observations on 17 variables:

- **lnchprg:** perc of studs in sch lnch prog
- **enroll:** school enrollment
- **staff:** staff per 1000 students
- **expend:** expend. per stud, $
- **salary:** avg. teacher salary, $
- **benefits:** avg. teacher benefits, $
- **droprate:** school dropout rate, perc
- **gradrate:** school graduation rate, perc
- **math10:** perc studs passing MEAP math
- **sci11:** perc studs passing MEAP science
- **totcomp:** salary + benefits
- **ltotcomp:** log(totcomp)
- **lexpend:** log of expend
- **lenroll:** log(enroll)
- **lstaff:** log(staff)
- **bensal:** benefits/salary
- **lsalary:** log(salary)

Notes

Many states have data, at either the district or building level, on student performance and spending. A good exercise in data collection and cleaning is to have students find such data for a particular state, and to put it into a form that can be used for econometric analysis.

Used in Text: pages 50, 65, 111-112, 127-128, 155-156, 219, 336, 339, 696-697

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(meap93)
```

`meapsingle` *meapsingle*

Description

Wooldridge Source: Collected by Professor Leslie Papke, an economics professor at MSU, from the Michigan Department of Education web site, www.michigan.gov/mde, and the U.S. Census Bureau. Professor Papke kindly provided the data. Data loads lazily.

Usage

```
data('meapsingle')
```

Format

A data.frame with 229 observations on 18 variables:

- **dcode:** district code
- **bcode:** building code
- **math4:** percent satisfactory, 4th grade math
- **read4:** percent satisfactory, 4th grade reading
- **enroll:** school enrollment
- **exppp:** expenditures per pupil, $
- **free:** percent eligible, free lunch
- **reduced:** percent eligible, reduced lunch
- **lunch:** free + reduced
- **medinc:** zipcode median family, $ (1999)
- **totchild:** # of children (in zipcode)
- **married:** # of children in married-couple families

- **single:** # of children not in married-couple families
- **pctsgle:** percent of children not in married-couple families
- **zipcode:** school zipcode
- **lenroll:** log(enroll)
- **lexppp:** log(exppp)
- **lmedinc:** log(medinc)

Used in Text

100, 145-146, 198

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781305270107

Examples

```
str(meapsingle)
```

---

minwage *minwage*

Description

Wooldridge Source: P. Wolfson and D. Belman (2004), "The Minimum Wage: Consequences for Prices and Quantities in Low-Wage Labor Markets," Journal of Business & Economic Statistics 22, 296-311. Professor Belman kindly provided the data. Data loads lazily.

Usage

```
data('minwage')
```

Format

A data.frame with 612 observations on 58 variables:

- **emp232:** employment, sector 232, 1000s
- **wage232:** hourly wage, sector 232, $
- **emp236:**
- **wage236:**
- **emp234:**
- **wage234:**
- **emp314:**
- **wage314:**

- **emp228:**
- **wage228:**
- **emp233:**
- **wage233:**
- **emp394:**
- **wage394:**
- **emp231:**
- **wage231:**
- **emp226:**
- **wage226:**
- **emp387:**
- **wage387:**
- **emp056:**
- **wage056:**
- **unem:** civilian unemployment rate, percent
- **cpi:** Consumer Price Index (urban), 1982-1984 = 100
- **minwage:** Federal minimum wage, $/hour
- **lemp232:** log(emp232)
- **lwage232:** log(wage232)
- **gemp232:** lemp232 - lemp232[_n-1]
- **gwage232:** lwage232 - lwage232[_n-1]
- **lminwage:** log(minwage)
- **gmwage:** lminwage - lminwage[_n-1]
- **gmwage_1:** gmwage[_n-1]
- **gmwage_2:**
- **gmwage_3:**
- **gmwage_4:**
- **gmwage_5:**
- **gmwage_6:**
- **gmwage_7:**
- **gmwage_8:**
- **gmwage_9:**
- **gmwage_10:**
- **gmwage_11:**
- **gmwage_12:**
- **lemp236:**
- **gcpi:** lcpi - lcpi[_n-1]

- **lcpi:** log(cpi)
- **lwage236:**
- **gemp236:**
- **gwage236:**
- **lemp234:**
- **lwage234:**
- **gemp234:**
- **gwage234:**
- **lemp314:**
- **lwage314:**
- **gemp314:**
- **gwage314:**
- **t:** linear time trend, 1 to 612

Notes

The sectors corresponding to the different numbers in the data file are provided in the Wolfson and Bellman and article.

Used in Text: pages 379, 410, 444-445, 674-675

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(minwage)
```

`mlb1` *mlb1*

Description

Wooldridge Source: Collected by G. Mark Holmes, a former MSU undergraduate, for a term project. The salary data were obtained from the New York Times, April 11, 1993. The baseball statistics are from The Baseball Encyclopedia, 9th edition, and the city population figures are from the Statistical Abstract of the United States. Data loads lazily.

Usage

```
data('mlb1')
```

Format

A data.frame with 353 observations on 47 variables:

- **salary:** 1993 season salary
- **teamsal:** team payroll
- **nl:** =1 if national league
- **years:** years in major leagues
- **games:** career games played
- **atbats:** career at bats
- **runs:** career runs scored
- **hits:** career hits
- **doubles:** career doubles
- **triples:** career triples
- **hruns:** career home runs
- **rbis:** career runs batted in
- **bavg:** career batting average
- **bb:** career walks
- **so:** career strike outs
- **sbases:** career stolen bases
- **fldperc:** career fielding perc
- **frstbase:** = 1 if first base
- **scndbase:** =1 if second base
- **shrtstop:** =1 if shortstop
- **thrdbase:** =1 if third base
- **outfield:** =1 if outfield
- **catcher:** =1 if catcher
- **yrsallst:** years as all-star
- **hispan:** =1 if hispanic
- **black:** =1 if black
- **whitepop:** white pop. in city
- **blackpop:** black pop. in city
- **hisppop:** hispanic pop. in city
- **pcinc:** city per capita income
- **gamesyr:** games per year in league
- **hrunsyr:** home runs per year
- **atbatsyr:** at bats per year
- **allstar:** perc. of years an all-star
- **slugavg:** career slugging average

- **rbisyr:** rbis per year
- **sbasesyr:** stolen bases per year
- **runsyr:** runs scored per year
- **percwhte:** percent white in city
- **percblck:** percent black in city
- **perchisp:** percent hispanic in city
- **blckpb:** black*percblck
- **hispph:** hispan*perchisp
- **whtepw:** white*percwhte
- **blckph:** black*perchisp
- **hisppb:** hispan*percblck
- **lsalary:** log(salary)

Notes

The baseball statistics are career statistics through the 1992 season. Players whose race or ethnicity could not be easily determined were not included. It should not be too difficult to obtain the city population and racial composition numbers for Montreal and Toronto for 1993. Of course, the data can be pretty easily obtained for more recent players.

Used in Text: pages 143-149, 165, 244-245, 262

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(mlb1)
```

`mroz` *mroz*

Description

Wooldridge Source: T.A. Mroz (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," Econometrica 55, 765-799. Professor Ernst R. Berndt, of MIT, kindly provided the data, which he obtained from Professor Mroz. Data loads lazily.

Usage

```
data('mroz')
```

Format

A data.frame with 753 observations on 22 variables:

- **inlf:** =1 if in lab frce, 1975
- **hours:** hours worked, 1975
- **kidslt6:** # kids < 6 years
- **kidsge6:** # kids 6-18
- **age:** woman's age in yrs
- **educ:** years of schooling
- **wage:** est. wage from earn, hrs
- **repwage:** rep. wage at interview in 1976
- **hushrs:** hours worked by husband, 1975
- **husage:** husband's age
- **huseduc:** husband's years of schooling
- **huswage:** husband's hourly wage, 1975
- **faminc:** family income, 1975
- **mtr:** fed. marg. tax rte facing woman
- **motheduc:** mother's years of schooling
- **fatheduc:** father's years of schooling
- **unem:** unem. rate in county of resid.
- **city:** =1 if live in SMSA
- **exper:** actual labor mkt exper
- **nwifeinc:** (faminc - wage*hours)/1000
- **lwage:** log(wage)
- **expersq:** exper^2

Used in Text

pages 249-251, 260, 294, 519-520, 530, 535, 535-536, 565-566, 578-579, 593- 595, 601-603, 619-620, 625

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(mroz)
```

`murder` *murder*

Description

Wooldridge Source: From the Statistical Abstract of the United States, 1995 (Tables 310 and 357), 1992 (Table 289). The execution data originally come from the U.S. Bureau of Justice Statistics, Capital Punishment Annual. Data loads lazily.

Usage

```
data('murder')
```

Format

A data.frame with 153 observations on 13 variables:

- **id:** state identifier
- **state:** postal code
- **year:** 87, 90, or 93
- **mrdrte:** murders per 100,000 people
- **exec:** total executions, past 3 years
- **unem:** annual unem. rate
- **d90:** =1 if year == 90
- **d93:** =1 if year == 93
- **cmrdrte:** mrdrte - mrdrte[_n-1]
- **cexec:** exec - exec[_n-1]
- **cunem:** unem - unem[_n-1]
- **cexec_1:** cexec[_n-1]
- **cunem_1:** cunem[_n-1]

Notes

Prosecutors in different counties might pursue the death penalty with different intensities, so it makes sense to collect murder and execution data at the county level. This could be combined with better demographic information at the county level, along with better economic data (say, on wages for various kinds of employment).

Used in Text: pages 480, 505, 548

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(murder)
```

`nbasal` *nbasal*

Description

Wooldridge Source: Collected by Christopher Torrente, a former MSU undergraduate, for a term project. He obtained the salary data and the career statistics from The Complete Handbook of Pro Basketball, 1995, edited by Zander Hollander. New York: Signet. The demographic information (marital status, number of children, and so on) was obtained from the teams' 1994-1995 media guides. Data loads lazily.

Usage

```
data('nbasal')
```

Format

A data.frame with 269 observations on 22 variables:

- **marr:** =1 if married
- **wage:** annual salary, thousands $
- **exper:** years as professional player
- **age:** age in years
- **coll:** years played in college
- **games:** average games per year
- **minutes:** average minutes per year
- **guard:** =1 if guard
- **forward:** =1 if forward
- **center:** =1 if center
- **points:** points per game
- **rebounds:** rebounds per game
- **assists:** assists per game
- **draft:** draft number
- **allstar:** =1 if ever all star
- **avgmin:** minutes per game
- **lwage:** log(wage)
- **black:** =1 if black
- **children:** =1 if has children
- **expersq:** exper^2
- **agesq:** age^2
- **marrblck:** marr*black

Notes

A panel version of this data set could be useful for further isolating productivity effects of marital status. One would need to obtain information on enough different players in at least two years, where some players who were not married in the initial year are married in later years. Fixed effects (or first differencing, for two years) is the natural estimation method.

Used in Text: pages 222-223, 264-265

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(nbasal)
```

---

ncaa_rpi *ncaa_rpi*

Description

Wooldridge Source: Data on NCAA men's basketball teams, collected by Weizhao Sun for a senior seminar project in sports economics at Michigan State University, Spring 2017. He used various sources, including www.espn.com and www.teamrankings.com/ncaa-basketball/rpi-ranking/rpi-rating-by-team. Data loads lazily.

Usage

```
data('ncaa_rpi')
```

Format

A data.frame with 336 observations on 14 variables:

- **team:** Name
- **year:** Year
- **conference:** Conference
- **postrpi:** Post Rank
- **prerpi:** Preseason Rank
- **postrpi_1:** Post Rank 1 yr ago
- **postrpi_2:** Post Rank 2 yrs ago
- **recruitrank:** Recruits Rank
- **wins:** Number of games won
- **losses:** Number of games lost

- **winperc:** Winning Percentage
- **tourney:** Tournament dummy
- **coachexper:** Coach Experience
- **power5:** PowerFive Dummy

Notes

This is a nice example of how multiple regression analysis can be used to determine whether rankings compiled by experts – the so-called pre-season RPI in this case – provide additional information beyond what we can obtain from widely available data bases. A simple and interesting question is whether, once the previous year's post-season RPI is controlled for, does the pre-season RPI – which is supposed to add information on recruiting and player development – help to predict performance (such as win percentage or making it to the NCAA men's basketball tournament). For the binary outcome that indicates making it to the NCAA tournament, a probit or logit model can be used for courses that introduce more advanced methods. There are some other interesting variables, such as coaching experience, that can be included, too.

Used in Text: not used

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781337558860

Examples

```
str(ncaa_rpi)
```

---

nyse *nyse*

Description

Wooldridge Source: These are Wednesday closing prices of value-weighted NYSE average, available in many publications. I do not recall the particular source I used when I collected these data at MIT. Probably the easiest way to get similar data is to go to the NYSE web site, www.nyse.com. Data loads lazily.

Usage

```
data('nyse')
```

Format

A data.frame with 691 observations on 8 variables:

- **price:** NYSE stock price index
- **return:** 100*(p - p(-1))/p(-1))
- **return_1:** lagged return
- **t:**
- **price_1:**
- **price_2:**
- **cprice:** price - price_1
- **cprice_1:** lagged cprice

Used in Text

pages 388-389, 407, 436, 438, 440-441, 442, 663-664

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(nyse)
```

---

okun *okun*

Description

Wooldridge Source: Economic Report of the President, 2007, Tables B-4 and B-42. Data loads lazily.

Usage

```
data('okun')
```

Format

A data.frame with 47 observations on 4 variables:

- **year:** 1959 through 2005
- **pcrgdp:** percentage change in real GDP
- **unem:** civilian unemployment rate
- **cunem:** unem - unem[_n-1]

Used in Text

410, 444

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(okun)
```

---

openness *openness*

Description

Wooldridge Source: D. Romer (1993), "Openness and Inflation: Theory and Evidence," Quarterly Journal of Economics 108, 869-903. The data are included in the article. Data loads lazily.

Usage

```
data('openness')
```

Format

A data.frame with 114 observations on 12 variables:

- **open:** imports as percent GDP, '73-
- **inf:** avg. annual inflation, '73-
- **pcinc:** 1980 per capita inc., U.S. $
- **land:** land area, square miles
- **oil:** =1 if major oil producer
- **good:** =1 if 'good' data
- **lpcinc:** log(pcinc)
- **lland:** log(land)
- **lopen:** log(open)
- **linf:** log(inf)
- **opendec:** open/100
- **linfdec:** log(inf/100)

Used in Text

pages 566, 579

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(openness)
```

`pension` *pension*

Description

Wooldridge Source: L.E. Papke (2004), "Individual Financial Decisions in Retirement Saving: The Role of Participant-Direction," Journal of Public Economics 88, 39-61. Professor Papke kindly provided the data. She collected them from the National Longitudinal Survey of Mature Women, 1991. Data loads lazily.

Usage

```
data('pension')
```

Format

A data.frame with 194 observations on 19 variables:

- **id:** family identifier
- **pyears:** years in pension plan
- **prftshr:** =1 if profit sharing plan
- **choice:** =1 if can choose method invest
- **female:** =1 if female
- **married:** =1 if married
- **age:** age in years
- **educ:** highest grade completed
- **finc25:** $15,000 < faminc92 <= $25,000
- **finc35:** $25,000 < faminc92 <= $35,000
- **finc50:** $35,000 < faminc92 <= $50,000
- **finc75:** $50,000 < faminc92 <= $75,000
- **finc100:** $75,000 < faminc92 <= $100,000
- **finc101:** $100,000 < faminc92
- **wealth89:** net worth, 1989, $1000
- **black:** =1 if black
- **stckin89:** =1 if owned stock in 1989
- **irain89:** =1 if had IRA in 1989
- **pctstck:** 0=mstbnds,50=mixed,100=mststcks

Used in Text

page 506

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(pension)
```

---

phillips *phillips*

Description

Wooldridge Source: Economic Report of the President, 2004, Tables B-42 and B-64. Data loads lazily.

Usage

```
data('phillips')
```

Format

A data.frame with 56 observations on 7 variables:

- **year:** 1948 through 2003
- **unem:** civilian unemployment rate, percent
- **inf:** percentage change in CPI
- **inf_1:** inf[_n-1]
- **unem_1:** unem[_n-1]
- **cinf:** inf - inf_1
- **cunem:** unem - unem_1

Used in Text

pages 355-356, 379, 390-391, 408, 409, 409, 418, 428, 443, 548-549, 642, 656, 659, 662, 672, 817.

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(phillips)
```

pntsprd *pntsprd*

Description

Wooldridge Source: Collected by Scott Resnick, a former MSU undergraduate, from various newspaper sources. Data loads lazily.

Usage

```
data('pntsprd')
```

Format

A data.frame with 553 observations on 12 variables:

- **favscr:** favored team's score
- **undscr:** underdog's score
- **spread:** las vegas spread
- **favhome:** =1 if favored team at home
- **neutral:** =1 if neutral site
- **fav25:** =1 if favored team in top 25
- **und25:** =1 if underdog in top 25
- **fregion:** favorite's region of country
- **uregion:** underdog's region of country
- **scrdiff:** favscr - undscr
- **sprdcvr:** =1 if spread covered
- **favwin:** =1 if favored team wins

Notes

The data are for the 1994-1995 men's college basketball seasons. The spread is for the day before the game was played. One might collect more recent data and determine whether the spread has become a less accurate predictor of the actual outcome in more recent years. In other words, in the simple regression of the actual score differential on the spread, is the variance larger in more recent years. (We should fully expect the slope coefficient not to be statistically different from one.)

Used in Text: pages 300, 624, 697

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(pntsprd)
```

prison *prison*

Description

Wooldridge Source: S.D. Levitt (1996), "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Legislation," Quarterly Journal of Economics 111, 319-351. Professor Levitt kindly provided me with the data, of which I used a subset. Data loads lazily.

Usage

```
data('prison')
```

Format

A data.frame with 714 observations on 45 variables:

- **state:** alphabetical; DC = 9
- **year:** 80 to 93
- **govelec:** =1 if gubernatorial election
- **black:** proportion black
- **metro:** proportion in metro. areas
- **unem:** proportion unemployed
- **criv:** viol. crimes per 100,000
- **crip:** prop. crimes per 100,000
- **lcriv:** log(criv)
- **lcrip:** log(crip)
- **gcriv:** lcriv - lcriv_1
- **gcrip:** lcrip - lcrip_1
- **y81:** =1 if year == 81
- **y82:**
- **y83:**
- **y84:**
- **y85:**
- **y86:**
- **y87:**
- **y88:**
- **y89:**
- **y90:**
- **y91:**

- **y92:**
- **y93:**
- **ag0_14:** prop. pop. 0 to 14 yrs
- **ag15_17:** prop. pop. 15 to 17 yrs
- **ag18_24:** prop. pop. 18 to 24 yrs
- **ag25_34:** prop. pop. 25 to 34 yrs
- **incpc:** per capita income, nominal
- **polpc:** police per 100,000 residents
- **gincpc:** log(incpc) - log(incpc_1)
- **gpolpc:** lpolpc - lpolpc_1
- **cag0_14:** change in ag0_14
- **cag15_17:** change in ag15_17
- **cag18_24:** change in ag18_24
- **cag25_34:** change in ag25_34
- **cunem:** change in unem
- **cblack:** change in black
- **cmetro:** change in metro
- **pris:** prison pop. per 100,000
- **lpris:** log(pris)
- **gpris:** lpris - lpris[_n-1]
- **final1:** =1 if fnl dec on litig, curr yr
- **final2:** =1 if dec on litig, prev 2 yrs

Used in Text

pages 573-574

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(prison)
```

prminwge *prminwge*

Description

Wooldridge Source: A.J. Castillo-Freeman and R.B. Freeman (1992), "When the Minimum Wage Really Bites: The Effect of the U.S.-Level Minimum Wage on Puerto Rico," in Immigration and the Work Force, edited by G.J. Borjas and R.B. Freeman, 177-211. Chicago: University of Chicago Press. The data are reported in the article. Data loads lazily.

Usage

```
data('prminwge')
```

Format

A data.frame with 38 observations on 25 variables:

- **year:** 1950-1987
- **avgmin:** weighted avg min wge, 44 indust
- **avgwage:** wghted avg hrly wge, 44 indust
- **kaitz:** Kaitz min wage index
- **avgcov:** wghted avg coverage, 8 indust
- **covt:** economy-wide coverage of min wg
- **mfgwage:** avg manuf. wage
- **prdef:** Puerto Rican price deflator
- **prepop:** PR employ/popul ratio
- **prepopf:** PR employ/popul ratio, alter.
- **prgnp:** PR GNP
- **prunemp:** PR unemployment rate
- **usgnp:** US GNP
- **t:** time trend: 1 to 38
- **post74:** time trend: starts in 1974
- **lprunemp:** log(prunemp)
- **lprgnp:** log(prgnp)
- **lusgnp:** log(usgnp)
- **lkaitz:** log(kaitz)
- **lprun_1:** lprunemp[_n-1]
- **lprepop:** log(prepop)
- **lprep_1:** lprepop[_n-1]
- **mincov:** (avgmin/avgwage)*avgcov
- **lmincov:** log(mincov)
- **lavgmin:** log(avgmin)

Notes

Given the ongoing debate on the employment effects of the minimum wage, this would be a great data set to try to update. The coverage rates are the most difficult variables to construct.

Used in Text: pages 356-357, 369-370, 420-421, 434

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(prminwge)
```

---

rdchem *rdchem*

Description

Wooldridge Source: From Businessweek R&D Scoreboard, October 25, 1991. Data loads lazily.

Usage

```
data('rdchem')
```

Format

A data.frame with 32 observations on 8 variables:

- **rd:** R&D spending, millions
- **sales:** firm sales, millions
- **profits:** profits, millions
- **rdintens:** rd as percent of sales
- **profmarg:** profits as percent of sales
- **salessq:** sales^2
- **lsales:** log(sales)
- **lrd:** log(rd)

Notes

It would be interesting to collect more recent data and see whether the R&D/firm size relationship has changed over time.

Used in Text: pages 64, 139-140, 159-160, 204, 218, 327-329, 339

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(rdchem)
```

---

rdtelec *rdtelec*

Description

Wooldridge Source: See RDCHEM.RAW Data loads lazily.

Usage

```
data('rdtelec')
```

Format

A data.frame with 29 observations on 6 variables:

- **rd:** R&D spending, millions $
- **sales:** firm sales, millions $
- **rdintens:** rd as percent of sales
- **lrd:** log(rd)
- **lsales:** log(sales)
- **salessq:** sales^2

Notes

According to these data, the R&D/firm size relationship is different in the telecommunications industry than in the chemical industry: there is pretty strong evidence that R&D intensity decreases with firm size in telecommunications. Of course, that was in 1991. The data could easily be updated, and a panel data set could be constructed.

Used in Text: not used

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(rdtelec)
```

recid *recid*

Description

Wooldridge Source: C.-F. Chung, P. Schmidt, and A.D. Witte (1991), "Survival Analysis: A Survey," Journal of Quantitative Criminology 7, 59-98. Professor Chung kindly provided the data. Data loads lazily.

Usage

```
data('recid')
```

Format

A data.frame with 1445 observations on 18 variables:

- **black:** =1 if black
- **alcohol:** =1 if alcohol problems
- **drugs:** =1 if drug history
- **super:** =1 if release supervised
- **married:** =1 if married when incarc.
- **felon:** =1 if felony sentence
- **workprg:** =1 if in N.C. pris. work prg.
- **property:** =1 if property crime
- **person:** =1 if crime against person
- **priors:** # prior convictions
- **educ:** years of schooling
- **rules:** # rules violations in prison
- **age:** in months
- **tserved:** time served, rounded to months
- **follow:** length follow period, months
- **durat:** min(time until return, follow)
- **cens:** =1 if duration right censored
- **ldurat:** log(durat)

Used in Text

pages 611-612, 625

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(recid)
```

`rental` *rental*

Description

Wooldridge Source: David Harvey, a former MSU undergraduate, collected the data for 64 "college towns" from the 1980 and 1990 United States censuses. Data loads lazily.

Usage

```
data('rental')
```

Format

A data.frame with 128 observations on 23 variables:

- **city:** city label, 1 to 64
- **year:** 80 or 90
- **pop:** city population
- **enroll:** # college students enrolled
- **rent:** average rent
- **rnthsg:** renter occupied units
- **tothsg:** occupied housing units
- **avginc:** per capita income
- **lenroll:** log(enroll)
- **lpop:** log(pop)
- **lrent:** log(rent)
- **ltothsg:** log(tothsg)
- **lrnthsg:** log(rnthsg)
- **lavginc:** log(avginc)
- **clenroll:** change in lrent from 80 to 90
- **clpop:** change in lpop
- **clrent:** change in lrent
- **cltothsg:** change in ltothsg
- **clrnthsg:** change in lrnthsg
- **clavginc:** change in lavginc
- **pctstu:** percent of population students
- **cpctstu:** change in pctstu
- **y90:** =1 if year == 90

Notes

These data can be used in a somewhat crude simultaneous equations analysis, either focusing on one year or pooling the two years. (In the latter case, in an advanced class, you might have students compute the standard errors robust to serial correlation across the two time periods.) The demand equation would have ltothsg as a function of lrent, lavginc, and lpop. The supply equation would have ltothsg as a function of lrent, pctst, and lpop. Thus, in estimating the demand function, pctstu is used as an IV for lrent. Clearly one can quibble with excluding pctstu from the demand equation, but the estimated demand function gives a negative price effect. Getting information for 2000, and adding many more college towns, would make for a much better analysis. Information on number of spaces in on-campus dormitories would be a big improvement, too.

Used in Text: pages 160, 477, 503-504

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(rental)
```

`return` *return*

Description

Wooldridge Source: Collected by Stephanie Balys, a former MSU undergraduate, from the New York Stock Exchange and Compustat. Data loads lazily.

Usage

```
data('return')
```

Format

A data.frame with 142 observations on 12 variables:

- **roe:** return on equity, 1990
- **rok:** return on capital, 1990
- **dkr:** debt/capital, 1990
- **eps:** earnings per share, 1990
- **netinc:** net income, 1990 (mills.)
- **sp90:** stock price, end 1990
- **sp94:** stock price, end 1994
- **salary:** CEO salary, 1990 (thous.)

- **return:** percent change s.p., 90-94
- **lsalary:** log(salary)
- **lsp90:** log(sp90)
- **lnetinc:** log(netinc)

Notes

More can be done with this data set. Recently, I discovered that lsp90 does appear to predict return (and the log of the 1990 stock price works better than sp90). I am a little suspicious, but you could use the negative coefficient on lsp90 to illustrate "reversion to the mean."

Used in Text: page 162-163

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(return)
```

`saving` *saving*

Description

Wooldridge Source: Unknown Data loads lazily.

Usage

```
data('saving')
```

Format

A data.frame with 100 observations on 7 variables:

- **sav:** annual savings, $
- **inc:** annual income, $
- **size:** family size
- **educ:** years educ, household head
- **age:** age of household head
- **black:** =1 if household head is black
- **cons:** annual consumption, $

Notes

I remember entering this data set in the late 1980s, and I am pretty sure it came directly from an introductory econometrics text. But so far my search has been fruitless. If anyone runs across this data set, I would appreciate knowing about it.

Used in Text: not used

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(saving)
```

---

school93_98 *school93_98*

Description

Wooldridge Source: L.E. Papke (2005), "The Effects of Spending on Test Pass Rates: Evidence from Michigan," Journal of Public Economics 89, 821-839. Data loads lazily.

Usage

```
data('school93_98')
```

Format

A data.frame with 10668 observations on 18 variables:

- **distid:**
- **schid:**
- **lunch:** percent eligible for free lunch
- **enrol:** number of students
- **exppp:** exp per pupil
- **math4:**
- **year:** 1993 = school year 1992-1993
- **y93:**
- **y94:**
- **y95:**
- **y96:**
- **y97:**
- **y98:**

- **rexpp:** (exppp/cpi)1.605: 1997 $
- **found:**
- **lenrol:** log(enrol)
- **lrexpp:** log(rexpp)
- **lavgrexpp:** log((rexpp + L.rexpp)/2)

Notes

This is closer to the data actually used in the Papke paper as it is at the school (building) level. It is unbalanced because data on scores and some of the spending and other variables is missing for some schools. While the usual RE and FE methods can be applied directly, obtaining the correlated random effects version of the Hausman test is more advance. Computer Exercise 17 in Chapter 14 walks the reader through it.

Used in Text: page 491

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781337558860

Examples

```
str(school93_98)
```

`sleep75` *sleep75*

Description

Wooldridge Source: J.E. Biddle and D.S. Hamermesh (1990), "Sleep and the Allocation of Time," Journal of Political Economy 98, 922-943. Professor Biddle kindly provided the data. Data loads lazily.

Usage

```
data('sleep75')
```

Format

A data.frame with 706 observations on 34 variables:

- **age:** in years
- **black:** =1 if black
- **case:** identifier
- **clerical:** =1 if clerical worker

- **construc:** =1 if construction worker
- **educ:** years of schooling
- **earns74:** total earnings, 1974
- **gdhlth:** =1 if in good or excel. health
- **inlf:** =1 if in labor force
- **leis1:** sleep - totwrk
- **leis2:** slpnaps - totwrk
- **leis3:** rlxall - totwrk
- **smsa:** =1 if live in smsa
- **lhrwage:** log hourly wage
- **lothinc:** log othinc, unless othinc < 0
- **male:** =1 if male
- **marr:** =1 if married
- **prot:** =1 if Protestant
- **rlxall:** slpnaps + personal activs
- **selfe:** =1 if self employed
- **sleep:** mins sleep at night, per wk
- **slpnaps:** minutes sleep, inc. naps
- **south:** =1 if live in south
- **spsepay:** spousal wage income
- **spwrk75:** =1 if spouse works
- **totwrk:** mins worked per week
- **union:** =1 if belong to union
- **worknrm:** mins work main job
- **workscnd:** mins work second job
- **exper:** age - educ - 6
- **yngkid:** =1 if children < 3 present
- **yrsmarr:** years married
- **hrwage:** hourly wage
- **agesq:** age^2

Notes

In their article, Biddle and Hamermesh include an hourly wage measure in the sleep equation. An econometric problem that arises is that the hourly wage is missing for those who do not work. Plus, the wage offer may be endogenous (even if it were always observed). Biddle and Hamermesh employ extensions of the sample selection methods in Section 17.5. See their article for details.

Used in Text: pages 64, 106-107, 162, 259, 263, 299

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(sleep75)
```

---

slp75_81 *slp75_81*

Description

Wooldridge Source: See SLEEP75.RAW Data loads lazily.

Usage

```
data('slp75_81')
```

Format

A data.frame with 239 observations on 20 variables:

- **age75:** age in 1975
- **educ75:** years educ in '75
- **educ81:** years educ in '81
- **gdhlth75:** = 1 if good hlth in '75
- **gdhlth81:** =1 if good hlth in '81
- **male:** =1 if male
- **marr75:** = 1 if married in '75
- **marr81:** =1 if married in '81
- **slpnap75:** mins slp wk, inc naps, '75
- **slpnap81:** mins slp wk, inc naps, '81
- **totwrk75:** minutes worked per week, '75
- **totwrk81:** minutes worked per week, '81
- **yngkid75:** = 1 if child < 3, '75
- **yngkid81:** =1 if child < 3, '81
- **ceduc:** change in educ
- **cgdhlth:** change in gdhlth
- **cmarr:** change in marr
- **cslpnap:** change in slpnap
- **ctotwrk:** change in totwrk
- **cyngkid:** change in yngkid

Used in Text

pages 463-464

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(slp75_81)
```

`smoke` *smoke*

Description

Wooldridge Source: J. Mullahy (1997), "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," Review of Economics and Statistics 79, 596-593. Professor Mullahy kindly provided the data. Data loads lazily.

Usage

```
data('smoke')
```

Format

A data.frame with 807 observations on 10 variables:

- **educ:** years of schooling
- **cigpric:** state cig. price, cents/pack
- **white:** =1 if white
- **age:** in years
- **income:** annual income, $
- **cigs:** cigs. smoked per day
- **restaurn:** =1 if rest. smk. restrictions
- **lincome:** log(income)
- **agesq:** age^2
- **lcigpric:** log(cigprice)

Notes

If you want to do a "fancy" IV version of Computer Exercise C16.1, you could estimate a reduced form count model for cigs using the Poisson regression methods in Section 17.3, and then use the fitted values as an IV for cigs. Presumably, this would be for a fairly advanced class.

Used in Text: pages 183, 288-289, 298, 301, 578, 627

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(smoke)
```

---

traffic1 *traffic1*

Description

Wooldridge Source: I collected these data from two sources, the 1992 Statistical Abstract of the United States (Tables 1009, 1012) and A Digest of State Alcohol-Highway Safety Related Legislation, 1985 and 1990, published by the U.S. National Highway Traffic Safety Administration. Data loads lazily.

Usage

```
data('traffic1')
```

Format

A data.frame with 51 observations on 13 variables:

- **state:**
- **admn90:** =1 if admin. revoc., '90
- **admn85:** =1 if admin. revoc., '85
- **open90:** =1 if open cont. law, '90
- **open85:** =1 if open cont. law, '85
- **dthrte90:** deaths per 100 mill. miles, '90
- **dthrte85:** deaths per 100 mill. miles, '85
- **speed90:** =1 if 65 mph, 1990
- **speed85:** =0 always
- **cdthrte:** dthrte90 - dthrte85
- **cadmn:** admn90 - admn85
- **copen:** open90 - open85
- **cspeed:** speed90 - speed85

Notes

In addition to adding recent years, this data set could really use state-level tax rates on alcohol. Other important law changes include defining driving under the influence as having a blood alcohol level of .08 or more, which many states have adopted since the 1980s. The trend really picked up in the 1990s and continued through the 2000s.

Used in Text: pages 467-468, 688?

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(traffic1)
```

---

traffic2 *traffic2*

Description

Wooldridge Source: P.S. McCarthy (1994), "Relaxed Speed Limits and Highway Safety: New Evidence from California," Economics Letters 46, 173-179. Professor McCarthy kindly provided the data. Data loads lazily.

Usage

```
data('traffic2')
```

Format

A data.frame with 108 observations on 48 variables:

- **year:** 1981 to 1989
- **totacc:** statewide total accidents
- **fatacc:** statewide fatal accidents
- **injacc:** statewide injury accidents
- **pdoacc:** property damage only accidents
- **ntotacc:** noninterstate total acc.
- **nfatacc:** noninterstate fatal acc.
- **ninjacc:** noninterstate injur acc.
- **npdoacc:** noninterstate property acc.
- **rtotacc:** tot. acc. on rural 65 mph roads
- **rfatacc:** fat. acc. on rural 65 mph roads

- **rinjacc:** inj. acc. on rural 65 mph roads
- **rpdoacc:** prp. acc. on rural 65 mph roads
- **ushigh:** acc. on U.S. highways
- **cntyrds:** acc. on county roads
- **strtes:** acc. on state routes
- **t:** time trend
- **tsq:** t^2
- **unem:** state unemployment rate
- **spdlaw:** =1 after 65 mph in effect
- **beltlaw:** =1 after seatbelt law
- **wkends:** # weekends in month
- **feb:** =1 if month is Feb.
- **mar:**
- **apr:**
- **may:**
- **jun:**
- **jul:**
- **aug:**
- **sep:**
- **oct:**
- **nov:**
- **dec:**
- **ltotacc:** log(totacc)
- **lfatacc:** log(fatacc)
- **prcfat:** 100*(fatacc/totacc)
- **prcrfat:** 100*(rfatacc/rtotacc)
- **lrtotacc:** log(rtotacc)
- **lrfatacc:** log(rfatacc)
- **lntotacc:** log(ntotacc)
- **lnfatacc:** log(nfatacc)
- **prcnfat:** 100*(nfatacc/ntotacc)
- **lushigh:** log(ushigh)
- **lcntyrds:** log(cntyrds)
- **lstrtes:** log(strtes)
- **spdt:** spdlaw*t
- **beltt:** beltlaw*t
- **prcfat_1:** prcfat[_n-1]

Notes

Many states have changed maximum speed limits and imposed seat belt laws over the past 25 years. Data similar to those in TRAFFIC2.RAW should be fairly easy to obtain for a particular state. One should combine this information with changes in a state's blood alcohol limit and the passage of per se and open container laws.

Used in Text: pages 378-379, 409, 443, 674, 695-696

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(traffic2)
```

`twoyear` *twoyear*

Description

Wooldridge Source: T.J. Kane and C.E. Rouse (1995), Labor-Market Returns to Two- and Four-Year Colleges, American Economic Review 85, 600-614. With Professor Rouse's kind assistance, I obtained the data from her web site at Princeton University. Data loads lazily.

Usage

```
data('twoyear')
```

Format

A data.frame with 6763 observations on 23 variables:

- **female:** =1 if female
- **phsrank:** percent high school rank; 100 = best
- **BA:** =1 if Bachelor's degree
- **AA:** =1 if Associate's degree
- **black:** =1 if African-American
- **hispanic:** =1 if Hispanic
- **id:** ID Number
- **exper:** total (actual) work experience
- **jc:** total 2-year credits
- **univ:** total 4-year credits
- **lwage:** log hourly wage

- **stotal:** total standardized test score
- **smcity:** =1 if small city, 1972
- **medcity:** =1 if med. city, 1972
- **submed:** =1 if suburb med. city, 1972
- **lgcity:** =1 if large city, 1972
- **sublg:** =1 if suburb large city, 1972
- **vlgcity:** =1 if very large city, 1972
- **subvlg:** =1 if sub. very lge. city, 1972
- **ne:** =1 if northeast
- **nc:** =1 if north central
- **south:** =1 if south
- **totcoll:** jc + univ

Notes

As possible extensions, students can explore whether the returns to two-year or four-year colleges depend on race or gender. This is partly done in Problem 7.9 but where college is aggregated into one number. Also, should experience appear as a quadratic in the wage specification?

Used in Text: pages 140-143, 165, 261, 340

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(twoyear)
```

`volat` *volat*

Description

Wooldridge Source: J.D. Hamilton and L. Gang (1996), "Stock Market Volatility and the Business Cycle," Journal of Applied Econometrics 11, 573-593. I obtained these data from the Journal of Applied Econometrics data archive at http://qed.econ.queensu.ca/jae/ Data loads lazily.

Usage

```
data('volat')
```

Format

A data.frame with 558 observations on 17 variables:

- **date:** 1947.01 to 1993.06
- **sp500:** S&P 500 index
- **divyld:** div. yield annualized rate
- **i3:** 3 mo. T-bill annualized rate
- **ip:** index of industrial production
- **pcsp:** pct chg, sp500, ann rate
- **rsp500:** return on sp500: pcsp + divyld
- **pcip:** pct chg, IP, ann rate
- **ci3:** i3 - i3[_n-1]
- **ci3_1:** ci3[_n-1]
- **ci3_2:** ci3[_n-2]
- **pcip_1:** pcip[_n-1]
- **pcip_2:** pcip[_n-2]
- **pcip_3:** pcip[_n-3]
- **pcsp_1:** pcip[_n-1]
- **pcsp_2:** pcip[_n-2]
- **pcsp_3:** pcip[_n-3]

Used in Text

pages 378, 670, 671, 674

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(volat)
```

vote1 *vote1*

Description

Wooldridge Source: From M. Barone and G. Ujifusa, The Almanac of American Politics, 1992. Washington, DC: National Journal. Data loads lazily.

Usage

```
data('vote1')
```

Format

A data.frame with 173 observations on 10 variables:

- **state:** state postal code
- **district:** congressional district
- **democA:** =1 if A is democrat
- **voteA:** percent vote for A
- **expendA:** camp. expends. by A, $1000s
- **expendB:** camp. expends. by B, $1000s
- **prtystrA:** percent vote for president
- **lexpendA:** log(expendA)
- **lexpendB:** log(expendB)
- **shareA:** 100*(expendA/(expendA+expendB))

Used in Text

pages 34, 39, 164, 221-222, 299, 699

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(vote1)
```

vote2 *vote2*

Description

Wooldridge Source: See VOTE1.RAW Data loads lazily.

Usage

```
data('vote2')
```

Format

A data.frame with 186 observations on 26 variables:

- **state:** state postal code
- **district:** U.S. Congressional district
- **democ:** =1 if incumbent democrat
- **vote90:** inc. share two-party vote, 1990
- **vote88:** inc. share two-party vote, 1988
- **inexp90:** inc. camp. expends., 1990
- **chexp90:** chl. camp. expends., 1990
- **inexp88:** inc. camp. expends., 1988
- **chexp88:** chl. camp. expends., 1988
- **prtystr:** percent vote pres., same party, 1988
- **rptchall:** =1 if a repeat challenger
- **tenure:** years in H.R.
- **lawyer:** =1 if law degree
- **linexp90:** log(inexp90)
- **lchexp90:** log(chexp90)
- **linexp88:** log(inexp88)
- **lchexp88:** log(chexp88)
- **incshr90:** 100*(inexp90/(inexp90+chexp90))
- **incshr88:** 100*(inexp88/(inexp88+chexp88))
- **cvote:** vote90 - vote88
- **clinexp:** linexp90 - linexp88
- **clchexp:** lchexp90 - lchexp88
- **cincshr:** incshr90 - incshr88
- **win88:** =1 by definition
- **win90:** =1 if inc. wins, 1990
- **cwin:** win90 - win88

Notes

These are panel data, at the Congressional district level, collected for the 1988 and 1990 U.S. House of Representative elections. Of course, much more recent data are available, possibly even in electronic form.

Used in Text: pages 335-336, 478, 699

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(vote2)
```

---

voucher *voucher*

Description

Wooldridge Source: Rouse, C.E. (1998), "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," Quarterly Journal of Economics 113, 553-602. Professor Rouse kindly provided the original data set from her paper. Data loads lazily.

Usage

```
data('voucher')
```

Format

A data.frame with 990 observations on 19 variables:

- **studyid:** student identifier
- **black:** = 1 if African-American
- **hispanic:** = 1 if Hispanic
- **female:** = 1 if female
- **appyear:** year of first application: 90 to 93
- **mnce:** math NCE test score, 1994
- **select:** = 1 if ever selected to attend choice school
- **choice:** = 1 if attending choice school, 1994
- **selectyrs:** years selected to attend choice school
- **choiceyrs:** years attended choice school
- **mnce90:** mnce in 1990
- **selectyrs1:** = 1 if selectyrs == 1

- **selectyrs2:** = 1 if selectyrs == 2
- **selectyrs3:** = 1 if selectyrs == 3
- **selectyrs4:** = 1 if selectyrs == 4
- **choiceyrs1:** = 1 if choiceyrs == 1
- **choiceyrs2:** = 1 if choiceyrs == 2
- **choiceyrs3:** = 1 if choiceyrs == 3
- **choiceyrs4:** = 1 if choiceyrs == 4

Notes

This is a condensed version of the data set used by Professor Rouse. The original data set had missing information on many variables, including post-policy and pre-policy test scores. I did not impute any missing data and have dropped observations that were unusable without filling in missing data. There are 990 students in the current data set but pre-policy test scores are available for only 328 of them. This is a good example of where eligibility for a program is randomized but participation need not be. In addition, even if we look at just the effect of eligibility (captured in the variable selectyrs) on the math test score (mnce), we need to confront the fact that attrition (students leaving the district) can bias the results. Controlling for the pre-policy test score, mnce90, can help – but at the cost of losing two-thirds of the observations. A simple regression of mnce on selectyrs followed by a multiple regression that adds mnce90 as a control is informative. The selectyrs dummy variables can be used as instrumental variables for the choiceyrs variable to try to estimate the effect of actually participating in the program (rather than estimating the so- called intention-to-treat effect). Computer Exercise C15.11 steps through the details.

Used in Text: pages 550-551

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(voucher)
```

`wage1` *wage1*

Description

Wooldridge Source: These are data from the 1976 Current Population Survey, collected by Henry Farber when he and I were colleagues at MIT in 1988. Data loads lazily.

Usage

```
data('wage1')
```

Format

A data.frame with 526 observations on 24 variables:

- **wage:** average hourly earnings
- **educ:** years of education
- **exper:** years potential experience
- **tenure:** years with current employer
- **nonwhite:** =1 if nonwhite
- **female:** =1 if female
- **married:** =1 if married
- **numdep:** number of dependents
- **smsa:** =1 if live in SMSA
- **northcen:** =1 if live in north central U.S
- **south:** =1 if live in southern region
- **west:** =1 if live in western region
- **construc:** =1 if work in construc. indus.
- **ndurman:** =1 if in nondur. manuf. indus.
- **trcommpu:** =1 if in trans, commun, pub ut
- **trade:** =1 if in wholesale or retail
- **services:** =1 if in services indus.
- **profserv:** =1 if in prof. serv. indus.
- **profocc:** =1 if in profess. occupation
- **clerocc:** =1 if in clerical occupation
- **servocc:** =1 if in service occupation
- **lwage:** log(wage)
- **expersq:** exper^2
- **tenursq:** tenure^2

Notes

Barry Murphy, of the University of Portsmouth in the UK, has pointed out that for several observations the values for exper and tenure are in logical conflict. In particular, for some workers the number of years with current employer (tenure) is greater than overall work experience (exper). At least some of these conflicts are due to the definition of exper as "potential" work experience, but probably not all. Nevertheless, I am using the data set as it was supplied to me.

Used in Text: pages 7, 17, 33-34, 37, 76, 91, 125, 183, 194-195, 220, 231, 234, 235-236, 240-241, 243-244, 263, 272, 326, 678

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(wage1)
```

`wage2` *wage2*

Description

Wooldridge Source: M. Blackburn and D. Neumark (1992), "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," Quarterly Journal of Economics 107, 1421-1436. Professor Neumark kindly provided the data, of which I used just the data for 1980. Data loads lazily.

Usage

```
data('wage2')
```

Format

A data.frame with 935 observations on 17 variables:

- **wage:** monthly earnings
- **hours:** average weekly hours
- **IQ:** IQ score
- **KWW:** knowledge of world work score
- **educ:** years of education
- **exper:** years of work experience
- **tenure:** years with current employer
- **age:** age in years
- **married:** =1 if married
- **black:** =1 if black
- **south:** =1 if live in south
- **urban:** =1 if live in SMSA
- **sibs:** number of siblings
- **brthord:** birth order
- **meduc:** mother's education
- **feduc:** father's education
- **lwage:** natural log of wage

Notes

As with WAGE1.RAW, there are some clear inconsistencies among the variables tenure, exper, and age. I have not been able to track down the causes, and so any changes would be effectively arbitrary. Instead, I am using the data as provided by the authors of the above QJE article.

Used in Text: pages 64, 106, 111, 165, 218-219, 220-221, 262, 310-312, 338, 519-520, 534, 546-547, 549, 678

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(wage2)
```

---

wagepan *wagepan*

Description

Wooldridge Source: F. Vella and M. Verbeek (1998), "Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men," Journal of Applied Econometrics 13, 163-183. I obtained the data from the Journal of Applied Econometrics data archive at http://qed.econ.queensu.ca/jae/. This is generally a nice resource for undergraduates looking to replicate or extend a published study. Data loads lazily.

Usage

```
data('wagepan')
```

Format

A data.frame with 4360 observations on 44 variables:

- **nr:** person identifier
- **year:** 1980 to 1987
- **agric:** =1 if in agriculture
- **black:** =1 if black
- **bus:**
- **construc:** =1 if in construction
- **ent:**
- **exper:** labor mkt experience
- **fin:**
- **hisp:** =1 if Hispanic
- **poorhlth:** =1 if in poor health
- **hours:** annual hours worked
- **manuf:** =1 if in manufacturing
- **married:** =1 if married
- **min:**
- **nrthcen:** =1 if north central

- **nrtheast:** =1 if north east
- **occ1:**
- **occ2:**
- **occ3:**
- **occ4:**
- **occ5:**
- **occ6:**
- **occ7:**
- **occ8:**
- **occ9:**
- **per:**
- **pro:**
- **pub:**
- **rur:**
- **south:** =1 if south
- **educ:** years of schooling
- **tra:**
- **trad:**
- **union:** =1 if in union
- **lwage:** log(wage)
- **d81:** =1 if year == 1981
- **d82:**
- **d83:**
- **d84:**
- **d85:**
- **d86:**
- **d87:**
- **expersq:** exper^2

Used in Text

pages 480, 494-495, 505

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(wagepan)
```

wageprc *wageprc*

Description

Wooldridge Source: Economic Report of the President, various years. Data loads lazily.

Usage

```
data('wageprc')
```

Format

A data.frame with 286 observations on 20 variables:

- **price:** consumer price index
- **wage:** nominal hourly wage
- **t:** time trend = 1, 2 , 3, ...
- **lprice:** log(price)
- **lwage:** log(wage)
- **gprice:** lprice - lprice[_n-1]
- **gwage:** lwage - lwage[_n-1]
- **gwage_1:** gwage[_n-1]
- **gwage_2:** gwage[_n-2]
- **gwage_3:**
- **gwage_4:**
- **gwage_5:**
- **gwage_6:**
- **gwage_7:**
- **gwage_8:**
- **gwage_9:**
- **gwage_10:**
- **gwage_11:**
- **gwage_12:**
- **gprice_1:** gprice[_n-1]

Notes

These monthly data run from January 1964 through October 1987. The consumer price index averages to 100 in 1967.

Used in Text: pages 405, 444-445, 671.

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(wageprc)
```

---

wine *wine*

Description

Wooldridge Source: These data were reported in a New York Times article, December 28, 1994. Data loads lazily.

Usage

```
data('wine')
```

Format

A data.frame with 21 observations on 5 variables:

- **country:**
- **alcohol:** liters alcohol from wine, per capita
- **deaths:** deaths per 100,000
- **heart:** heart disease dths per 100,000
- **liver:** liver disease dths per 100,000

Notes

The dependent variables deaths, heart, and liver can be each regressed against alcohol as nice simple regression examples. The conventional wisdom is that wine is good for the heart but not for the liver, something that is apparent in the regressions. Because the number of observations is small, this can be a good data set to illustrate calculation of the OLS estimates and statistics.

Used in Text: not used

Source

https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Examples

```
str(wine)
```

Index

∗ **datasets**
admnrev, 4
affairs, 5
airfare, 6
alcohol, 7
apple, 9
approval, 10
athlet1, 12
athlet2, 13
attend, 14
audit, 15
barium, 16
beauty, 17
benefits, 19
beveridge, 20
big9salary, 20
bwght, 22
bwght2, 23
campus, 25
card, 26
catholic, 27
cement, 29
census2000, 30
ceosal1, 31
ceosal2, 32
charity, 33
consump, 34
corn, 35
countymurders, 36
cps78_85, 38
cps91, 39
crime1, 40
crime2, 41
crime3, 43
crime4, 44
discrim, 46
driving, 48
earns, 50
econmath, 51
elem94_95, 52
engin, 53
expendshares, 54
ezanders, 55
ezunem, 57
fair, 58
fertil1, 60
fertil2, 61
fertil3, 63
fish, 64
fringe, 65
gpa1, 67
gpa2, 69
gpa3, 70
happiness, 71
hprice1, 72
hprice2, 73
hprice3, 75
hseinv, 76
htv, 77
infmrt, 78
injury, 79
intdef, 81
intqrt, 82
inven, 83
jtrain, 84
jtrain2, 85
jtrain3, 87
jtrain98, 88
k401k, 89
k401ksubs, 90
kielmc, 91
labsup, 92
lawsch85, 94
loanapp, 95
lowbrth, 97
mathpnl, 99
meap00_01, 101
meap01, 102

meap93, 103
meapsingle, 104
minwage, 105
mlb1, 107
mroz, 109
murder, 111
nbasal, 112
ncaa_rpi, 113
nyse, 114
okun, 115
openness, 116
pension, 117
phillips, 118
pntsprd, 119
prison, 120
prminwge, 122
rdchem, 123
rdtelec, 124
recid, 125
rental, 126
return, 127
saving, 128
school93_98, 129
sleep75, 130
slp75_81, 132
smoke, 133
traffic1, 134
traffic2, 135
twoyear, 137
volat, 138
vote1, 140
vote2, 141
voucher, 142
wage1, 143
wage2, 145
wagepan, 146
wageprc, 148
wine, 149

admnrev, 4
affairs, 5
airfare, 6
alcohol, 7
apple, 9
approval, 10
athlet1, 12
athlet2, 13
attend, 14
audit, 15
barium, 16
beauty, 17
benefits, 19
beveridge, 20
big9salary, 20
bwght, 22
bwght2, 23

campus, 25
card, 26
catholic, 27
cement, 29
census2000, 30
ceosal1, 31
ceosal2, 32
charity, 33
consump, 34
corn, 35
countymurders, 36
cps78_85, 38
cps91, 39
crime1, 40
crime2, 41
crime3, 43
crime4, 44

discrim, 46
driving, 48

earns, 50
econmath, 51
elem94_95, 52
engin, 53
expendshares, 54
ezanders, 55
ezunem, 57

fair, 58
fertil1, 60
fertil2, 61
fertil3, 63
fish, 64
fringe, 65

gpa1, 67
gpa2, 69
gpa3, 70

happiness, 71
hprice1, 72

hprice2, 73
hprice3, 75
hseinv, 76
htv, 77

infmrt, 78
injury, 79
intdef, 81
intqrt, 82
inven, 83

jtrain, 84
jtrain2, 85
jtrain3, 87
jtrain98, 88

k401k, 89
k401ksubs, 90
kielmc, 91

labsup, 92
lawsch85, 94
loanapp, 95
lowbrth, 97

mathpnl, 99
meap00_01, 101
meap01, 102
meap93, 103
meapsingle, 104
minwage, 105
mlb1, 107
mroz, 109
murder, 111

nbasal, 112
ncaa_rpi, 113
nyse, 114

okun, 115
openness, 116

pension, 117
phillips, 118
pntsprd, 119
prison, 120
prminwge, 122

rdchem, 123
rdtelec, 124
recid, 125
rental, 126
return, 127

saving, 128
school93_98, 129
sleep75, 130
slp75_81, 132
smoke, 133

traffic1, 134
traffic2, 135
twoyear, 137

volat, 138
vote1, 140
vote2, 141
voucher, 142

wage1, 143
wage2, 145
wagepan, 146
wageprc, 148
wine, 149