

Package ‘probstats4econ’

May 23, 2024

Title Companion Package to Probability and Statistics for Economics and Business

Version 0.2.0

Description Utilities for multiple hypothesis testing, companion datasets from “Probability and Statistics for Economics and Business: An Introduction Using R” by Jason Abrevaya (MIT Press, under contract).

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.1

LazyData true

Depends R (>= 3.6.0)

Suggests testthat (>= 3.0.0), estimatr (>= 1.0.0)

Config/testthat/edition 3

URL <https://www.probstats4econ.com/package.html>,
<https://www.probstats4econ.com/>

Contact abrevaya@austin.utexas.edu

NeedsCompilation no

Author Jason Abrevaya [aut, cph],
Nathan Gardner Hattersley [aut, cre]

Maintainer Nathan Gardner Hattersley <nhattersley@utexas.edu>

Repository CRAN

Date/Publication 2024-05-23 15:40:08 UTC

R topics documented:

auctions	2
baseball	3
births	4
bitcoin	5
brands	5

cigdata	6
congress	6
cps	7
dictator	8
exams	9
houseprices	9
hrs	10
inflation	11
inflation_expectations	12
linear_combination	12
married	13
metricsgrades	14
mutualfunds	14
names2022	15
premier2020	16
resume	16
sp500	17
strikes	18
test_linear_restrictions	18
var_mean_indep	19
wald_test	20
website	21
widgets	22

Index 23

auctions	<i>Auction data</i>
----------	---------------------

Description

Data on eBay auctions, based upon the paper "Econometrics of Auctions by Least Squares" by Leonardo Rezende, Journal of Applied Econometrics, 2008, 23:925-948. The dataset consists of eBay auctions for Apple iPod mini devices in June and July 2006, limited to only auctions for the 4GB models.

Usage

auctions

Format

auctions:

A data frame with 684 rows and 14 columns:

ebay_auction_id eBay auction ID number

bidders Number of bidders

finalprice Final sales price

seller_feedback_pct Seller's positive feedback percentage (e.g., 90 = 90%)
seller_feedback_score Seller's feedback score (number of feedbacks received)
reserveprice Reserve price set by seller (value of 0.01 if no reserve price)
color_pink 1 if iPod is pink, 0 otherwise
color_blue 1 if iPod is blue, 0 otherwise
color_silver 1 if iPod is silver, 0 otherwise
color_green 1 if iPod is green, 0 otherwise
color_other 1 if iPod is another color, 0 otherwise
new 1 if condition listed is new, 0 otherwise
used 1 if condition listed is used, 0 otherwise
refurb 1 if condition listed is refurbished, 0 otherwise

Source

<https://journaldata.zbw.eu/dataset/econometrics-of-auctions-by-least-squares>

baseball	<i>Baseball attendance data</i>
----------	---------------------------------

Description

Data on 2022 attendance for Major League Baseball teams

Usage

baseball

Format

baseball:
 A data frame with 30 rows and 9 columns:
team Team name
attend_home Average home game attendance
attend_road Average road game attendance
winpct_22 Team winning percentage in 2022
winpct_21 Team winning percentage in 2021
playoff_21 1 if team made playoffs in 2021, 0 otherwise
capacity Capacity of home stadium
popul Population of team's metropolitan area (2020)
payroll Total team payroll in 2022 (in millions of dollars)

Source

various

births

*Birth outcome data***Description**

Data on birth outcomes in the United States for December 2021 births where mother's age is between 25 and 35 (inclusive), limited to singleton births, mother's first child, and having non-missing values for relevant variables

Usage

births

Format

births:

A data frame with 50,249 rows and 20 columns:

birthtime Birth time during day (in minutes, range is 0 to 2399)

birthwkday Day of week of birth (1=Sunday, 2=Monday, ..., 7=Saturday)

age Mother's age (in years)

nonhsgrad 1 if mother is not a HS graduate, 0 otherwise

hsgrad 1 if mother is HS graduate and has no add'l education, 0 otherwise

somecoll 1 if mother completed some college, 0 otherwise

collgrad 1 if mother is 4-year college graduate, 0 otherwise

married 1 if mother is married, 0 otherwise

smoke1 1 if mother smoked during first trimester, 0 otherwise

smoke2 1 if mother smoked during second trimester, 0 otherwise

smoke3 1 if mother smoked during third trimester, 0 otherwise

smokepre 1 if mother smoked before pregnancy, 0 otherwise

smoke 1 if mother smoked during pregnancy (any trimester), 0 otherwise

prenatal1 1 if first prenatal care during first trimester, 0 otherwise

prenatal2 1 if first prenatal care during second trimester, 0 otherwise

prenatal3 1 if first prenatal care during third trimester, 0 otherwise

nocare 1 if no prenatal care visit, 0 otherwise

male 1 if baby is a boy, 0 otherwise

bweight Birthweight (in grams)

bweight_lbs Birthweight (in pounds)

Source

<https://www.nber.org/research/data/vital-statistics-natality-birth-data>

bitcoin	<i>Bitcoin price and returns data</i>
---------	---------------------------------------

Description

Data on daily prices and returns for Bitcoin during 2020 and 2021

Usage

bitcoin

Format

bitcoin:

A data frame with 364 rows and 268 columns:

date Date

high Highest price (in dollars)

low Lowest price (in dollars)

close End-of-day price (in dollars)

return Daily return, based on end-of-day prices

Source

<https://finance.yahoo.com>

brands	<i>Brand data</i>
--------	-------------------

Description

Data on the purchase behavior of customers at a specific market. The dataset consists of customers who purchased one of five candy-bar brands in their previous visit to the market and records whether or not they make a purchase during this visit and, if so, which brand they purchase. The dataset is adapted from the full dataset that is referenced in the source citation.

Usage

brands

Format

brands:

A data frame with 14,560 rows and 3 columns:

purchase 1 if customer makes a purchase, 0 otherwise

brand Brand purchased (1 through 5), 0 if no purchase

last_brand Brand purchased (1 through 5) during last visit

Source

<https://medium.com/%40miradzji/purchase-probability-analysis-in-certain-market-segments-with-python>

cigdata	<i>State-level cigarette price and tax data</i>
---------	---

Description

Data on cigarette prices and taxes in 2019 for the 50 U.S. states plus the District of Columbia

Usage

cigdata

Format

cigdata:

A data frame with 51 rows and 9 columns:

state State abbreviation

statename State name

price_pack Average price per pack (in dollars)

pack_sales_per_capita Annual sales, packs per capita

totaltax_pct Total tax (federal plus state) per pack, as percent of pack price

totaltax_pack Total tax (federal plus state) per pack (in dollars)

cig_tax_revnu Total annual tax revenue (in dollars)

producer 1 if tobacco production > 20m pounds, 0 otherwise

Source

<https://healthdata.gov/dataset/The-Tax-Burden-on-Tobacco-1970-2019/etts-u9ii>

congress	<i>Congressional election data</i>
----------	------------------------------------

Description

Data on congressional election outcomes in the United States between 1948 and 1990, based upon the paper "Do Voters Affect or Elect Policies? Evidence from the U.S. House" by David S. Lee, Enrico Moretti, Matthew J. Butler, 2004, *Quarterly Journal of Economics*, 119: 807-859. This sample is restricted to elections where (i) the incumbent is running for re-election and (ii) are not running unopposed. There are 9,788 observations available, and demographic variables are available for 6,774 of the observations.

Usage

congress

Format

congress:

A data frame with 9,788 rows and 15 columns:

state State code (ICPSR coding)

district District code

demvote Number of votes for Democrat candidate

repvote Number of votes for Republican candidate

year Year of election

demvoteshare Percentage of vote for Democrat candidate

lagdemvoteshare Percentage of vote for Democrat candidate in last election

totpop Population of Congressional district

medianincome Median (nominal) income of Congressional district

pcturban Percentage of Congressional district that is urban

pctblack Percentage of Congressional district that is black

pcthighschl Percentage of Congressional district that is HS graduates

votingpop Voting population of Congressional district

democrat 1 if Democrat wins election ($\text{demvoteshare} > 0.5$), 0 otherwise

lagdemocrat 1 if Democrat won last election ($\text{lagdemvoteshare} > 0.5$), 0 otherwise

Source

<https://eml.berkeley.edu/%7Emoretti/data3.html>

cps

Current Population Survey (CPS) data

Description

A subsample of the 2019 Current Population Survey (CPS) consisting of data on individuals aged 30 to 59 (inclusive)

Usage

cps

Format

cps:

A data frame with 4,013 rows and 17 columns:

statefips Two-character state code, including DC

gender Gender (Male, Female)

metro Metropolitan-area (Metro, Non-Metro)

race Race category (Black, White, Other)

hispanic Hispanic (Hispanic, Non-hispanic)

marstatus Marital status (Married, Divorced, Widowed, Never married)

lfstatus Labor-force status (Employed, Unemployed, Not in LF)

ottipcomm Earnings include overtime, tips, and/or commissions (Yes, No)

hourly Hourly-worker status (Hourly, Non-hourly)

unionstatus Union status (Union, Non-union)

age Age (in years)

hrslastwk Hours worked last week

unempwks Number of weeks unemployed

wagehr Hourly wage (in dollars); only for hourly employees

earnwk Earnings last week (in dollars)

ownchild Number of children in household

educ Highest education level attained (in years)

Source

<https://www.census.gov/programs-surveys/cps/data/datasets.html>

dictator

Dictator-game data

Description

Data on the results from "dictator games" played in an experimental study, based on the paper "Giving and taking in dictator games – differences by gender? A replication study of Chowdhury et al.", Journal of Comments and Replications in Economics, 2023. Each observation corresponds to one play of the game. Earnings are for the dictator. Two game variants are the "giving game" (dictator starts with endowment) and "taking game" (recipient starts with endowment).

Usage

dictator

Format

dictator:

A data frame with 137 rows and 5 columns:

earnings Earnings of the dictator (between 0 and 10)

giving 1 if giving game, 0 otherwise

taking 1 if taking game, 0 otherwise

female 1 if dictator is female, 0 otherwise

female_opp 1 if recipient is female, 0 otherwise

Source

<https://journaldata.zbw.eu/dataset/giving-and-taking-in-dictator-games-replication>

exams

Exam data

Description

Data on two exam scores for 77 university students

Usage

exams

Format

exams:

A data frame with 77 rows and 2 columns:

exam1 Score (out of 100) on the first exam

exam2 Score (out of 100) on the second exam

houseprices

Housing price data

Description

Data on house sales in Ames, Iowa between 2006 and 2010. The dataset is limited to one-family homes with public utilities and excludes new home sales.

Usage

houseprices

Format

houseprices:

A data frame with 973 rows and 16 columns:

lotarea Area of lot (in square feet)

overallqual Overall home quality (scale 1-10, 10 best)

yearbuilt Year house was built

yearremodadd Year house was remodeled (equal to yearbuilt if never)

bsmtfinsf Area of finished basement (in square feet, 0 if no finished basement)

grlivarea Total non-basement living area (in square feet)

fullbath Number of full bathrooms

halfbath Number of half bathrooms

bedroomabvgr Number of non-basement bedrooms

totrmsabvgrd Number of non-basement rooms (not including bathrooms)

fireplaces Number of fireplaces

garagecars Size of garage (0 if no garage)

mosold Month house sold (1=Jan,...,12=Dec)

yrsold Year house sold

saleprice Sales price of house (in dollars)

centralair 1 if house has central air, 0 otherwise

Source

<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>

hrs

Health-expenditure data

Description

Data on healthcare utilization and expenditures for adults 50 years and older in the United States, taken from the Health and Retirement Study (HRS) and Asset and Health Dynamics Among the Oldest Old (AHEAD). Data was originally used in the paper "On the distribution and dynamics of health care costs" by Eric French and John Bailey Jones, 2004, Journal of Applied Econometrics, 19: 705-721. This dataset is restricted to non-married individuals in the year 2000.

Usage

hrs

Format

hrs:

A data frame with 6,052 rows and 14 columns:

age Age (in years)**assets** Total assets (in dollars); bottom-coded at \$20,000**doctor_visits** Number of doctor visits**drug_costs** Drug costs (in dollars)**income** Income (in dollars); bottom-coded at \$5,000**hosp_nights** Number of nights spent in hospital**ins_private** 1 if insurance is private or employee-provided, 0 otherwise**ins_medicare** 1 if insurance is Medicare, 0 otherwise**ins_medicaid** 1 if insurance is Medicaid, 0 otherwise**ins_none** 1 if no health insurance, 0 otherwise**male** 1 if male, 0 otherwise**medical_costs** Total medical costs (in dollars)**nodrug_financial** 1 if did not take prescription drugs for financial reasons, 0 otherwise**outofpocket_costs** Total out-of-pocket medical costs (in dollars)**Source**<https://journaldata.zbw.eu/dataset/on-the-distribution-and-dynamics-of-health-care-costs>

inflation

*Inflation data***Description**

Data on inflation rates for 45 countries for a ten-year period (2010-2019).

Usage

inflation

Format

inflation:

A data frame with 450 rows and 3 columns:

country Country abbreviation**year** Year**inflation** Annual inflation rate (change in CPI)**Source**<https://data.oecd.org/price/inflation-cpi.htm>

 inflation_expectations

Inflation expectations data

Description

Data on individual inflation expectations, based on the paper: "Measuring consumer uncertainty about future inflation," by Wandu Bruine de Bruin, Charles F. Manski, Giorgio Topa, Wilbert van der Klaauw, 2011, Journal of Applied Econometrics, 26: 454-478. This dataset has only the observations with point estimates of inflation for individuals between 30 and 70 years of age. The survey took place in 2007 and 2008. The actual inflation, for benchmark, was 3.2% in 2006, 2.9% in 2007, and 3.8% in 2008.

Usage

```
inflation_expectations
```

Format

```
inflation_expectations:
```

A data frame with 290 rows and 6 columns:

inflation_pred Individual prediction of inflation next year (integer; e.g. 10=10%)

age Age (in years)

finlit_score Financial literacy test score (out of 12 points)

male 1 if male, 0 otherwise

collgrad 1 if college graduate, 0 otherwise

famincome_hi 1 if family income > \$75,000, 0 otherwise

Source

<https://journaldata.zbw.eu/dataset/measuring-consumer-uncertainty-about-future-inflation>

 linear_combination

Test a single linear restriction of a model

Description

linear_combination takes a set of regression results and a vector representing a linear combination of the parameters and returns the estimate, standard error, and p-value for the null hypothesis that the linear combination is equal to zero.

Usage

```
linear_combination(regresults, R)
```

Arguments

- regresults** A list containing two items: `coefficients`, which is a vector of coefficient estimates, and `vcov`, which is the variance-covariance matrix of the coefficient estimates.
- R** A vector of length equal to the number of coefficients, representing weights on each of the parameters.

Value

List with the following values:

- `estimate`, the point estimate of the linear combination
- `se`, the standard error of the point estimate
- `p_value`, the p-value for the null hypothesis that the linear combination is equal to zero

Examples

```
# test that the returns to one year of education are equal to ten years of age
model <- estimatr::lm_robust(earnwk ~ age + educ, data = cps)
R <- c(0, -10, 1) # 0 * `intercept` - 10 * `age` + 1 * `education`
linear_combination(model, R)
```

married	<i>Married-couple data</i>
---------	----------------------------

Description

Data on married couples in the United States from the 2003 Community Tracking Study (CTS) Household Survey.

Usage

```
married
```

Format

`married`:

A data frame with 4,126 rows and 11 columns:

age_w Age of wife (in years)

age_h Age of husband (in years)

educ_w Education of wife (in years)

educ_h Education of husband (in years)

bmi_w Body mass index of wife (bottom-coded at 18, top-coded at 40)

bmi_h Body mass index of husband (bottom-coded at 18, top-coded at 40)

smoke_w 1 if wife smokes, 0 otherwise
smoke_h 1 if husband smokes, 0 otherwise
employed_w 1 if wife employed, 0 otherwise
employed_h 1 if husband employed, 0 otherwise
famincome Annual family income (in dollars, top-coded at \$150,000)

Source

<https://www.icpsr.umich.edu/web/HMCA/studies/4216>

metricsgrades	<i>Econometrics course data</i>
---------------	---------------------------------

Description

Data on performance in a graduate econometrics course, with GRE test information and domestic/international status available.

Usage

metricsgrades

Format

metricsgrades:
 A data frame with 68 rows and 4 columns:
gre_quant Score on GRE quantitative test (out of 170)
gre_verbal Score on GRE verbal test (out of 170)
domestic 1 if domestic student, 0 if international student
total Overall composite course grade (out of 100 points)

mutualfunds	<i>Mutual-fund performance data</i>
-------------	-------------------------------------

Description

Data on mutual funds categorized as "Large Blend Equity" funds by Morningstar, limited to funds in existence for more than 10 years. Data captured 2/28/2023.

Usage

mutualfunds

Format

mutualfunds:

A data frame with 208 rows and 11 columns:

name Name of mutual fund

fund_age Age of fund (in years)

expense_ratio Expense ratio (net)

aum Assets under management (in millions of dollars)

min_investment Minimum investment level (in dollars)

load Y if fund has a load (sales charge or fee), N if not

manager_tenure Tenure of current fund manager (in years)

return_1yr One-year annualized return

return_3yr Three-year annualized return

return_5yr Five-year annualized return

return_10yr Ten-year annualized return

Source

<https://www.fidelity.com>

names2022

Popular names data

Description

Data on the names of all babies born in the United States in 2022, as provided by the Social Security Administration. Each observation corresponds to a specific name and gender, with a count of that name provided. For confidentiality reasons, the minimum count for any name is 5. All other names (with fewer than 5 occurrences in the U.S.) are included within the observation having "OTHER" as the name. There are two "OTHER" observations, one for female babies and one for male babies.

Usage

names2022

Format

names2022:

A data frame with 31917 rows and 3 columns:

name Baby's name

gender F if female, M if male

count Number of babies with name and gender

Source

<https://www.ssa.gov/oact/babynames/limits.html>

`premier2020`*Premier League soccer data*

Description

Data on all game results for the 2020 Premier League soccer season. The Premier League consists of 20 teams. Each team plays every other team twice (home and away) during the season, so there are a total of 38 rounds in the season and 380 total games.

Usage`premier2020`**Format**`premier2020:`

A data frame with 380 rows and 5 columns:

round Round (values 1 to 38)

hometeam Home team

awayteam Away team

homegoals Number of goals by the home team

awaygoals Number of goals by the away team

Source

https://en.wikipedia.org/wiki/2020%E2%80%9321_Premier_League

`resume`*Resume response data*

Description

Data on responses to hypothetical resumes that were created for an experimental study, based upon "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment" by Amanda Agan and Sonja Starr, 2018, Quarterly Journal of Economics, 133: 191-235. This dataset considers only the subsample from before the ban-the-box initiative.

Usage`resume`

Format

resume:

A data frame with 7,332 rows and 7 columns:

crime 1 if applicant has criminal record, 0 otherwise

drugcrime 1 if applicant has committed drug crime, 0 otherwise

propertycrime 1 if applicant has committed property crime, 0 otherwise

ged 1 if applicant has GED, 0 otherwise

empgap 1 if applicant has a gap in employment, 0 otherwise

black 1 if applicant is black, 0 otherwise

response 1 if applicant received positive response, 0 otherwise

Source

[doi:10.7910/DVN/VPHMNT](https://doi.org/10.7910/DVN/VPHMNT)

sp500

Monthly returns data for S&P 500 companies

Description

Data on monthly returns for S&P 500 companies between Jan 1991 and Apr 2021

Usage

sp500

Format

sp500:

A data frame with 364 rows and 268 columns:

Date Date, as a string, indicating the endpoint of the month

IDX Monthly return for the S&P 500 index

AAPL, ABMD, ..., ZION Monthly company returns, where variable name is the company stock ticker symbol

Source

<https://finance.yahoo.com>

strikes	<i>Strike duration data</i>
---------	-----------------------------

Description

Data on the length of worker contract strikes within U.S. manufacturing for the period 1968-1976, based upon "The Duration of Contract strikes in U.S. Manufacturing" by John Kennan, 1985, Journal of Econometrics, 28: 5-28.

Usage

```
strikes
```

Format

```
strikes:
```

A data frame with 566 rows and 1 column:

duration Strike duration (in weeks)

Source

<https://cameron.econ.ucdavis.edu/mmabook/mmadata.html>

test_linear_restrictions	<i>Test multiple linear restrictions simultaneously</i>
--------------------------	---

Description

test_linear_restrictions takes a set of regression results and tests multiple linear restrictions simultaneously.

Usage

```
test_linear_restrictions(regresults, R, r = default_test(R))
```

Arguments

regresults	A list containing two items: coefficients, which is a vector of coefficient estimates, and vcov, which is the variance-covariance matrix of the coefficient estimates.
R	A matrix of linear restrictions. Each row of R represents a different linear restriction. R should have the same number of columns as length(regresults\$coefficients).
r	A vector of constants, equal to the number of rows in R. This is what we are testing that each linear restriction is equal to.

Value

A list with the following items:

- W: The Wald (chi-square) statistic
- p_value: The p-value of the test

Examples

```
# test both that the returns to one year of education are
# equal to ten years of age, and that the intercept is zero
model <- estimatr::lm_robust(earnwk ~ age + educ, data = cps)
R <- matrix(c(0, -10, 1, 1, 0, 0), nrow = 2, byrow = TRUE)
test_linear_restrictions(model, R)
```

var_mean_indep	<i>Variance helper functions</i>
----------------	----------------------------------

Description

These functions help calculate the variance matrix of different kinds of samples. `var_mean_indep` creates an asymptotic covariance matrix for the sample means of a list of independent samples. `var_prop_indep` creates an asymptotic covariance matrix for the sample proportions of a list of independent samples. `var_mean_onesample` creates an asymptotic covariance matrix for the sample means of several variables from the same sample.

Usage

```
var_mean_indep(x_vectors)

var_mean_onesample(df, vars = names(df))

var_prop_indep(pi_hat, nobs)
```

Arguments

<code>x_vectors</code>	A list of vectors, representing the different independent samples.
<code>df</code>	A data.frame object
<code>vars</code>	A character vector of variable names in df.
<code>pi_hat</code>	A vector of sample proportions.
<code>nobs</code>	The sample size.

Value

A matrix, representing the asymptotic covariance matrix of the sample means.

Examples

```

# list of independent samples
x_vectors <- list(
  rnorm(1000, mean = 1, sd = 2),
  rnorm(10, mean = 4, sd = 0.5),
  rnorm(1000000, mean = 0, sd = 1)
)
var_mean_indep(x_vectors)

# sample proportions
pi_hat <- c(0.1, 0.6, 0.3)
nobs <- 1000
var_prop_indep(pi_hat, nobs)

# covariance of educ and age in cps dataset
var_mean_onesample(cps, c("educ", "age"))

```

wald_test

Wald test statistic and p-value

Description

Given the parameter estimates and their variance-covariance matrix, `wald_test` calculates the Wald test statistic and p-value for a set of linear constraints on the parameters.

Usage

```

wald_test(
  gamma_hat,
  var_gamma_hat,
  R = diag(length(gamma_hat)),
  r = default_test(R)
)

```

Arguments

<code>gamma_hat</code>	L x 1 vector of parameter estimates
<code>var_gamma_hat</code>	L x L variance-covariance matrix of parameter estimates
<code>R</code>	Q x L matrix of linear constraints to be tested. Defaults to identity matrix of size L
<code>r</code>	Q x 1 vector of test values for the linear constraints. Defaults to a vector of zeros of length Q to test that all the contrasts are equal to zero.

Value

A list with the following elements:

- **W**: Wald test statistic
- **p_value**: p-value for the Wald test (χ^2_Q distribution)

Examples

```
# test that union workers earn the same as non-union workers
cps$union <- as.numeric(cps$unionstatus == "Union")
model <- lm(earnwk ~ union, data = cps)
gamma_hat <- coef(model)
var_gamma_hat <- vcov(model)
wald_test(gamma_hat, var_gamma_hat, R = c(0, 1))

# test that non-union workers make 900/week
# *and* union workers make 1000/week
wald_test(
  gamma_hat,
  var_gamma_hat,
  R = matrix(c(0, 1, 1, 1), nrow = 2),
  r = c(900, 1000)
)
```

 website

Website visitor arrival data

Description

Data on the arrival time of website visitors during a specific hour for a hypothetical website.

Usage

```
website
```

Format

website:

A data frame with 748 rows and 2 columns:

arrival Arrival time during the hour (in minutes)

time_since_last Time since last visitor (in minutes)

widgets

Hypothetical data for widgets.com website

Description

Data on purchases for an e-mail experiment run by widgets.com

Usage

widgets

Format

widgets:

A data frame with 3,000 rows and 4 columns:

emailA 1 if customer receives e-mail A, 0 otherwise

emailB 1 if customer receives e-mail B, 0 otherwise

purchase 1 if customer makes a purchase, 0 otherwise

amount Total purchase (in dollars)

Index

* datasets

- auctions, [2](#)
 - baseball, [3](#)
 - births, [4](#)
 - bitcoin, [5](#)
 - brands, [5](#)
 - cigdata, [6](#)
 - congress, [6](#)
 - cps, [7](#)
 - dictator, [8](#)
 - exams, [9](#)
 - houseprices, [9](#)
 - hrs, [10](#)
 - inflation, [11](#)
 - inflation_expectations, [12](#)
 - married, [13](#)
 - metricsgrades, [14](#)
 - mutualfunds, [14](#)
 - names2022, [15](#)
 - premier2020, [16](#)
 - resume, [16](#)
 - sp500, [17](#)
 - strikes, [18](#)
 - website, [21](#)
 - widgets, [22](#)
-
- auctions, [2](#)
 - baseball, [3](#)
 - births, [4](#)
 - bitcoin, [5](#)
 - brands, [5](#)
 - cigdata, [6](#)
 - congress, [6](#)
 - cps, [7](#)
 - dictator, [8](#)
 - exams, [9](#)
 - houseprices, [9](#)
 - hrs, [10](#)
 - inflation, [11](#)
 - inflation_expectations, [12](#)
 - linear_combination, [12](#)
 - married, [13](#)
 - metricsgrades, [14](#)
 - mutualfunds, [14](#)
 - names2022, [15](#)
 - premier2020, [16](#)
 - resume, [16](#)
 - sp500, [17](#)
 - strikes, [18](#)
 - test_linear_restrictions, [18](#)
 - var_mean_indep, [19](#)
 - var_mean_onesample (var_mean_indep), [19](#)
 - var_prop_indep (var_mean_indep), [19](#)
 - wald_test, [20](#)
 - website, [21](#)
 - widgets, [22](#)