Package: IPAG (via r-universe)

May 17, 2026

Title Tools for IPAG Courses

Version 0.1.0

Description Provides a collection of intuitive and user-friendly functions for computing confidence intervals for common statistical tasks, including means, differences in means, proportions, and odds ratios. The package also includes tools for linear regression analysis and several real-world datasets intended for teaching and applied statistical inference.

URL https://github.com/gpiaser/IPAG

BugReports https://github.com/gpiaser/IPAG/issues

Imports stats,

Suggests knitr, rmarkdown

VignetteBuilder knitr

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

Repository https://gpiaser.r-universe.dev

Date/Publication 2026-01-11 19:14:07 UTC

RemoteUrl https://github.com/gpiaser/ipag

RemoteRef HEAD

RemoteSha fac402e2fe0e413e12bae1c746d89667f117c4bd

Contents

`Beauty` *Beauty and teaching evaluations*

Description

Dataset from Hamermesh, D. S., & Parker, A. (2005), "Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity", Economics of Education Review, 24(4), 369–376.

Usage

```
data(Beauty)
```

Format

A data frame with the following variables:

n The professor's identification number.

score Average professor evaluation score, ranging from 1 (very unsatisfactory) to 5 (excellent).

rank Rank of professor: teaching, tenure track, or tenured.

ethnicity Ethnicity of professor: not minority or minority.

gender Gender of professor: female or male.

language Language of the school where the professor received education: English or non-English.

age Age of the professor.

cls_perc_eval Percentage of students in the class who completed the evaluation.

cls_did_eval Number of students in the class who completed the evaluation.

cls_students Total number of students enrolled in the class.

cls_level Class level: lower or upper.

cls_profs Number of professors teaching sections of the course in the sample: single or multiple.

cls_credits Number of credits of the class: one credit (e.g. lab, PE) or multi credit.

bty_f1lower Beauty rating of professor from lower-level female students (1 = lowest, 10 = highest).

bty_f1upper Beauty rating of professor from upper-level female students (1 = lowest, 10 = highest).

bty_f2upper Beauty rating of professor from second upper-level female students (1 = lowest, 10 = highest).

bty_m1lower Beauty rating of professor from lower-level male students (1 = lowest, 10 = highest).

bty_m1upper Beauty rating of professor from upper-level male students (1 = lowest, 10 = highest).

bty_m2upper Beauty rating of professor from second upper-level male students (1 = lowest, 10 = highest).

bty_avg Average beauty rating of the professor.

pic_outfit Outfit of professor in picture: not formal or formal.

pic_color Color of professor's picture: color or black and white.

Details

The dataset examines the relationship between instructors' physical attractiveness and student evaluation scores, controlling for demographic and class characteristics.

Source

Hamermesh, D. S., & Parker, A. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. Economics of Education Review, 24(4), 369–376. doi:10.1016/j.econedurev.2004.07.013

`Bosnia` *My dataset from CSV*

Description

This dataset was imported from a CSV file and included in the IPAG package for demonstration. Data are taken from the article by Augsburg, B., De Haas, R., Harmgart, H., & Meghir, C. (2015). The impacts of microcredit: Evidence from Bosnia and Herzegovina. American Economic Journal: Applied Economics, 7(1), 183-203.

Usage

```
data(Bosnia)
```

Format

A data frame with the following variables:

Income_0B Household income for the control group before the experiment

Income_1B Household income for the treatment group before the experiment

Income_0F Household income for the control group after the experiment

Income_1F Household income for the treatment group after the experiment

Details

doi: 10.1257/app.20130272

`ContentMarketing` *Content Marketing Dataset*

Description

Dataset from Koob (2021), "Determinants of content marketing effectiveness: Conceptual framework and empirical findings from a managerial perspective." PloS ONE, 16(4), e0249457.

Usage

```
data(ContentMarketing)
```

Format

A data frame with the following variables:

Firm The company's identification number.

CMEFFECT Effectiveness of the content marketing strategy. Marketing and communications executives rated the degree of effectiveness on a scale from 1 to 5 based on their perception and expertise.

CMSTRAT Content marketing strategy context. Four-item scale measuring whether the organization had a defined, comprehensible, and long-term content marketing strategy. Rated from 1 ("totally disagree") to 5 ("totally agree").

CPROD Content production context. Reflects the organization's efforts to optimize content value for customers, meet content quality standards, and plan and create content systematically.

CDIST1 Content distribution context / intermediate number of media platforms. Measures the number of media platforms used to distribute content.

CDIST2 Content distribution context / joint deployment of print and digital platforms. Measures the simultaneous use of print and digital media for content distribution.

CPROM Content Promotion Context. Measures the importance attached to content promotion. Respondents indicated the share of total content marketing investment devoted to promotion activities.

CMPERME Content Marketing Performance Measurement Context. Captures the frequency of content marketing performance measurement across print and digital platforms and the use of performance data to guide improvement.

CMORG Content Marketing Organization. Captures structural specialization, autonomy in content marketing, and processes and systems that enable specialization.

SIZE Organization size. Three dummy variables categorize organizations by number of employees: "Tiny" (250-499), "Small" (500-999), "Medium" (1,000-4,999), "Big" (>=5,000).

SECTOR Sector affiliation. Dummy variable distinguishing organizations in the "industrial" or "service" sector.

Source

Koob, C. (2021). Determinants of content marketing effectiveness: Conceptual framework and empirical findings from a managerial perspective. PloS ONE, 16(4), e0249457.

covid19 *My dataset from CSV*

Description

This dataset was imported from a CSV file and included in the IPAG package for demonstration. The reference article is Escobar, L. E., Molina-Cruz, A., & Barillas-Mury, C. (2020). BCG vaccine protection from severe coronavirus disease 2019 (COVID-19). Proceedings of the National Academy of Sciences, 117(30), 17720-17726.

Usage

```
data(covid19)
```

Format

A data frame with the following variables:

total_deaths_per_million Number of deaths per million inhabitants as of April 22, 2020.

country The name of the country.

Cal2013 Daily caloric intake.

ca2014 Per capita CO2 emissions in 2014.

BMI Body mass index in 2016 (male population).

Sras Number of people who died of SARS in 2004.

dtp3_2011 Proportion of children under one year of age vaccinated with the DTP vaccine (diphtheria, tetanus, poliomyelitis) in 2011.

BCG_policy BCG vaccination policy: `"current"`, `"never"` or `"interrupted"`.

lati Latitude of the country's capital.

longi Longitude of the country's capital.

Trade2018 Imported and exported goods as a percentage of GDP in 2018.

H2015 Health expenditure per capita in 2015.

Health2010 Percentage of the state budget allocated to health in 2010.

TB Number of tuberculosis cases per 100,000 inhabitants.

PIBhab GDP per capita.

Superf Area of the country.

Demo Democracy index of the country.

HDI_2018 Human Development Index in 2018.

Expectancy Life expectancy at birth.

Children Number of children per woman.

PopulationD Population density of the country.

Pop Total population of the country.number of children per woman

Gini Measure of income inequality (0 = perfect equality, 1 = perfect inequality).

AgeMed Median age of the population.

debut Number of days between the first confirmed Covid-19 case in China and the first confirmed case in the country.

Details

https://doi.org/10.1073/pnas.2008410117

Source

Various international public databases (WHO, World Bank, etc.)

`Housing` *Hedonic housing prices and environmental quality*

Description

Dataset from Harrison Jr, D., & Rubinfeld, D. L. (1978), "Hedonic housing prices and the demand for clean air", Journal of Environmental Economics and Management, 5(1), 81–102.

Usage

```
data(Housing)
```

Format

A data frame with the following variables:

CRIM Per capita crime rate by town.

ZN Proportion of residential land zoned for lots over 25,000 square feet.

INDUS Proportion of non-retail business acres per town.

CHAS Charles River dummy variable: 1 if the tract bounds the river, 0 otherwise.

NOX Nitric oxides concentration (parts per 10 million).

RM Average number of rooms per dwelling.

AGE Proportion of owner-occupied units built prior to 1940.

DIS Weighted distances to five Boston employment centres.

RAD Index of accessibility to radial highways.

TAX Full-value property tax rate per $10,000.

PTRATIO Pupil–teacher ratio by town.

B Computed as $1000(B_k - 0.63)^2$, where B_k is the proportion of Black residents by town.

LSTAT Percentage of lower-status population.

MEDV Median value of owner-occupied homes in thousands of US dollars.

Details

The dataset is a cross-section of housing values in Boston suburbs and is widely used to study hedonic pricing models and the demand for environmental quality.

Source

Harrison Jr, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. Journal of Environmental Economics and Management, 5(1), 81–102. doi:10.1016/0095-0696(78)900062

`linear_regress` *Linear regression summary*

Description

This function performs a linear regression and returns a summary including:

- Adjusted R-squared
- Overall F-test p-value
- Table with parameter estimates, confidence intervals (default 99%), p-values, and significance stars (*, **, ***)

Usage

```
linear_regress(formula, data, level = 0.99)
```

Arguments

`formula`	A formula like Y ~ X1 + X2
`data`	A data frame
`level`	Confidence level (default 0.99)

Value

Object of class 'linear_regress'

Examples

```
data(Housing, package = "IPAG")
linear_regress(MEDV ~ RM + LSTAT, data = Housing)
```

`McKinsey` *McKinsey / OECD Education Dataset*

Description

Dataset combining information from:

- McKinsey, "Valuing the merit of teachers", Direction interministérielle de la transformation publique.
- OECD (2012), "Does Performance-Based Pay Improve Teaching?", PISA in Focus, No. 16, OECD Publishing, Paris.

Usage

```
data(McKinsey)
```

Format

A data frame with the following variables:

COUNTRIES The name of the country.

READING Teacher efficiency measured by PISA reading tests.

YSALARY Teacher salaries in relation to GDP per capita. 0 means salaries equal GDP per capita, 0.5 means 1.5 times higher than GDP per capita, 1 means 2 times higher than GDP per capita.

YGDP GDP per capita in USD 1,000.

EXPEND Cumulative expenditure by educational establishments in USD 1,000.

PERF Teacher merit pay (y = yes, n = no).

Details

The dataset contains teacher efficiency as measured by reading performance on PISA tests, along with explanatory variables related to salary, GDP, expenditures, and performance-based pay.

Source

- McKinsey, "Valuing the merit of teachers", Direction interministérielle de la transformation publique.
- OECD (2012), "Does Performance-Based Pay Improve Teaching?", PISA in Focus, No. 16, OECD Publishing, Paris, doi:10.1787/5k98q27r2stben

mean_ci *Confidence interval for a mean*

Description

Confidence interval for a mean

Usage

```
mean_ci(x, level = 0.99, na.rm = TRUE)
```

Arguments

x	Numeric vector
level	Confidence level (default 0.99)
na.rm	Remove NA values

Value

Object of class 'mean_ci'

Examples

```
x <- c(4.2, 5.1, 6.3, 5.8, 4.9)
mean_ci(x)
mean_ci(x, level = 0.95)
```

mean_diff_ci *Confidence interval for the difference of means*

Description

Confidence interval for the difference of means

Usage

```
mean_diff_ci(x, y, level = 0.99, paired = FALSE, na.rm = TRUE)
```

Arguments

x	Numeric vector
y	Numeric vector
level	Confidence level (default 0.99)
paired	Logical; are the samples paired?
na.rm	Remove NA values

Value

Object of class 'mean_diff_ci'

Examples

```
x <- c(5.1, 4.9, 6.2, 5.8, 5.4)
y <- c(4.8, 4.7, 5.9, 5.2, 5.0)
mean_diff_ci(x, y)
mean_diff_ci(x, y, paired = TRUE)
```

---

oddsratio_ci *Confidence interval for odds ratio from a 2x2 table*

Description

Confidence interval for odds ratio from a 2x2 table

Usage

```
oddsratio_ci(a, b, c, d, level = 0.99)
```

Arguments

a, b, c, d	Cell counts of the 2x2 contingency table
level	Confidence level (default 0.99)

Value

Object of class 'oddsratio_ci'

Examples

```
oddsratio_ci(a = 12, b = 5, c = 4, d = 15)
oddsratio_ci(a = 12, b = 5, c = 4, d = 15, level = 0.95)
```

prop_ci *Confidence interval for a proportion*

Description

Confidence interval for a proportion

Usage

```
prop_ci(trials, successes, level = 0.99)
```

Arguments

trials	Number of trials
successes	Number of successes
level	Confidence level (default 0.99)

Value

Object of class 'prop_ci'

Examples

```
# 45 successes out of 100 trials
prop_ci(trials = 100, successes = 45)
prop_ci(trials = 100, successes = 45, level = 0.95)
```

Index