

Package ‘PQA’

March 30, 2026

Title Perform the Pearson-Quetelet Analysis on Two-Way Contingency Tables

Version 1.0.0

Description Tools to perform Pearson-Quetelet analysis on two-way contingency tables. The package computes absolute and relative frequencies, Quetelet indices, Pearson-Quetelet decomposition, apex tables, and chi-square summaries for interpreting associations between categorical variables.

License LGPL-3

LazyData true

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 3.5)

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Boris Mirkin [aut],
Luca Coraggio [aut, cre],
Trevor Fenner [aut],
Zina Taran [aut]

Maintainer Luca Coraggio <luca.coraggio@unina.it>

Repository CRAN

Date/Publication 2026-03-30 17:10:08 UTC

Contents

bmi_mortality	2
bmi_mortality_all	3
pqa	4
print.pqa	5
print.pqa.chisq	6
print.pqa.subtable	7

summary.pqa	8
uk_crime_rubbish	9
usa_toddlers	10
usa_voting_prefs	11
us_covid_mortality	12
us_fall_accidents	13

Index	15
--------------	-----------

bmi_mortality	<i>BMI Category vs Mortality Outcome (Excluding First 5 Years)</i>
---------------	--

Description

Cross-classified counts of participants by BMI category at study entry and all-cause mortality outcome from the Leisure World Cohort Study (1981-2004), excluding the first 5 years of follow-up (Table 6 in the cited paper).

Usage

bmi_mortality

Format

A numeric matrix (also an array) with 4 rows and 2 columns:

Rows (BMI category): Underweight, Normal, Overweight, Obese.

Columns (mortality outcome): Died, Survived (Participants - Deaths).

Values Frequencies (counts).

Details

BMI thresholds at study entry:

- Underweight: BMI < 18.5
- Normal weight: BMI 18.5-24.9
- Overweight: BMI 25.0-29.9
- Obese: BMI >= 30

Category-level metadata (excluding first 5 years):

- Underweight: median BMI 17.8, range 12.4-18.5, participants 390, deaths 352
- Normal: median BMI 22.4, range 18.5-25.0, participants 7611, deaths 6091
- Overweight: median BMI 26.5, range 25.0-29.9, participants 2937, deaths 2345
- Obese: median BMI 31.6, range 30.0-54.1, participants 437, deaths 339

Totals in this dataset: 11,375 participants and 9,127 deaths.

Source

Corrada, Maria M., Kawas, Claudia H., Mozaffar, Farah, and Paganini-Hill, Annlia (2006). *Association of Body Mass Index and Weight Change with All-Cause Mortality in the Elderly*. *American Journal of Epidemiology*, 163(10), 938-949. Table 6, values excluding the first 5 years of follow-up. doi:10.1093/aje/kwj114

Examples

```
data(bmi_mortality)
bmi_mortality
rowSums(bmi_mortality)
colSums(bmi_mortality)
sum(bmi_mortality)
```

bmi_mortality_all	<i>BMI Category vs Mortality Outcome (Total Sample)</i>
-------------------	---

Description

Cross-classified counts of participants by BMI category at study entry and all-cause mortality outcome from the Leisure World Cohort Study (1981-2004), using the total sample values reported in Table 6 of the cited paper.

Usage

```
bmi_mortality_all
```

Format

A numeric matrix (also an array) with 4 rows and 2 columns:

Rows (BMI category): Underweight, Normal, Overweight, Obese.

Columns (mortality outcome): Died, Survived (Participants - Deaths).

Values Frequencies (counts).

Details

BMI thresholds at study entry:

- Underweight: BMI < 18.5
- Normal weight: BMI 18.5-24.9
- Overweight: BMI 25.0-29.9
- Obese: BMI >= 30

Category-level metadata (total sample):

- Underweight: median BMI 17.8, range 12.4-18.5, participants 556, deaths 518

- Normal: median BMI 22.4, range 18.5-25.0, participants 9021, deaths 7501
- Overweight: median BMI 26.5, range 25.0-29.9, participants 3376, deaths 2784
- Obese: median BMI 31.6, range 30.0-54.1, participants 498, deaths 400

Totals in this dataset: 13,451 participants and 11,203 deaths.

Source

Corrada, Maria M., Kawas, Claudia H., Mozaffar, Farah, and Paganini-Hill, Annlia (2006). *Association of Body Mass Index and Weight Change with All-Cause Mortality in the Elderly*. *American Journal of Epidemiology*, 163(10), 938-949. Table 6, total sample values. doi:10.1093/aje/kwj114

Examples

```
data(bmi_mortality_all)
bmi_mortality_all
rowSums(bmi_mortality_all)
colSums(bmi_mortality_all)
sum(bmi_mortality_all)
```

pqa

Pearson-Quetelet Analysis for Two-Way Contingency Tables

Description

Performs Pearson-Quetelet analysis (PQA) to examine associations between categorical variables through the Quetelet index and its decomposition of the chi-square statistic.

Usage

```
pqa(x)
```

Arguments

x A two-way [table](#) of counts. Higher-dimensional tables are not supported.

Details

The Quetelet index is computed as $q_{ij} = p_{ij}/(p_i p_j) - 1$, so 0 indicates independence, positive values indicate higher-than-expected frequency, and negative values indicate lower-than-expected frequency. The decomposition pq equals $p_{ij} q_{ij}$ and sums to ϕ^2 ; apex rescales pq to percentage contributions. When $\phi^2 = 0$ (perfect independence), apex is returned as a zero table.

The function automatically handles missing factor/level names and assesses chi-square validity based on expected frequencies:

- flag = 0: Valid.
- flag = 1: Unreliable (min. expected frequency < 5).
- flag = 2: Cannot be computed (min. expected frequency < 1 or df = 0).

Value

An object of class pqa, which is a list containing:

abs Absolute frequencies (counts).

rel Relative frequencies (proportions).

q Quetelet index values, measuring relative change in probability.

pq Pearson-Quetelet decomposition of the chi-square statistic.

apex Percentage contributions of each cell to the chi-square statistic.

chisq A list of class pqa.chisq with test results (stat, df, pval) and a validity flag.

See Also

[table](#) for creating contingency tables, [chisq.test](#) for chi-square tests

Examples

```
# Example 1: Using the built-in usa_voting_prefs dataset
data(usa_voting_prefs)
result <- pqa(usa_voting_prefs)
print(result$abs) # View absolute frequencies
print(result$chisq) # View chi-square test results

# Example 2: Using a matrix (converted to table first)
data_matrix <- matrix(c(10, 20, 15, 25), nrow = 2, ncol = 2)
dimnames(data_matrix) <- list(Gender = c("Male", "Female"), Preference = c("A", "B"))
result <- pqa(as.table(data_matrix))
```

print.pqa

Print Pearson-Quetelet Analysis Object

Description

Displays a summary of the available components within a pqa object.

Usage

```
## S3 method for class 'pqa'
print(x, pp = NULL, ...)
```

Arguments

x A pqa object.

pp Logical; if TRUE, prints a formatted summary. If FALSE, prints the raw structure. Defaults to the "pqa.pretty_print" option.

... Further arguments passed to or from other methods.

Details

Components include absolute (abs) and relative (rel) frequencies, Quetelet indices (q), Pearson-Quetelet decomposition (pq), apex (apex), and chi-square results (chisq).

Value

Invisibly returns the input object.

See Also

[pqa](#), [summary.pqa](#), [print.pqa.subtable](#)

Examples

```
data(usa_voting_prefs)
qt <- pqa(usa_voting_prefs)
print(qt)
```

<code>print.pqa.chisq</code>	<i>Print Chi-Square Test Results</i>
------------------------------	--------------------------------------

Description

Formatted print method for `pqa.chisq` objects, showing test statistics and validity assessments.

Usage

```
## S3 method for class 'pqa.chisq'
print(x, pp = NULL, ...)
```

Arguments

<code>x</code>	A <code>pqa.chisq</code> object.
<code>pp</code>	Logical; if TRUE, prints formatted results. Defaults to the " <code>pqa.pretty_print</code> " option.
<code>...</code>	Further arguments passed to or from other methods.

Details

Displays the null hypothesis, chi-square statistic, degrees of freedom, and p-value. Includes warnings if the test is unreliable (expected frequencies < 5) or cannot be computed.

Value

Invisibly returns the input object.

See Also

[pqa](#), [print.pqa](#)

Examples

```
data(usa_voting_prefs)
qt <- pqa(usa_voting_prefs)
print(qt$chisq)
```

`print.pqa.subtable` *Print Pearson-Quetelet Analysis Subtables*

Description

Formatted print method for `pqa.subtable` components such as absolute frequencies, relative frequencies, Quetelet indices, decompositions, and apex.

Usage

```
## S3 method for class 'pqa.subtable'
print(x, pp = NULL, ...)
```

Arguments

<code>x</code>	A <code>pqa.subtable</code> object.
<code>pp</code>	Logical; if TRUE, prints a formatted contingency table. Defaults to the "pqa.pretty_print" option.
<code>...</code>	Further arguments passed to or from other methods.

Details

Formatting (rounding, scaling, and marginals) automatically adapts to the subtable type:

- **abs, rel, pq**: shown with 4 decimal places and marginals.
- **q**: shown as percentages with 2 decimal places and no marginals.
- **apex**: shown as percentages with 2 decimal places and marginals.

If `pp = FALSE`, the raw matrix-like object is printed via `print.AsIs()`.

Value

Invisibly returns the input object.

See Also

[pqa](#), [print.pqa](#), [print.pqa.chisq](#)

Examples

```
data(usa_voting_prefs)
qt <- pqa(usa_voting_prefs)
print(qt$abs)
print(qt$q)
```

summary.pqa

Summarize a Pearson-Quetelet Analysis

Description

Prints a textual summary of a pqa object, including absolute frequencies, chi-square test output, Quetelet signals of association/indifference, and apex-based contribution notes.

Usage

```
## S3 method for class 'pqa'
summary(object, ...)
```

Arguments

object A pqa object.
... Further arguments passed to or from other methods.

Details

The summary output includes:

- **Absolute Frequencies:** Contingency table with margins.
- **Chi-square Test:** Test statistics, flag, and significance messages (when the test is considered valid/reliable).
- **Association Analysis:** Cell-level signals for strong associations ($|q| > 30\%$) and row/column indifference patterns ($|q| < 10\%$ for all cells in a row/column).
- **Apex Notes:** "Odd" row/column contributions and the overall positive-vs-negative apex balance.

Value

Invisibly returns the input pqa object.

See Also

[pqa](#), [print.pqa](#), [print.pqa.subtable](#), [print.pqa.chisq](#)

Examples

```
# Create a pqa from the built-in usa_voting_prefs dataset and get summary
data(usa_voting_prefs)
qt <- pqa(usa_voting_prefs)

# Get comprehensive summary
summary(qt)
```

uk_crime_rubbish

UK Crime Survey: Rubbish on Street vs Crime Victimization

Description

A cross-classified data table from the British Crime Survey (2007-2008) showing the relationship between the perceived frequency of rubbish on streets and crime victimization status. This dataset is useful for illustrating contingency table analysis and chi-square tests of independence in statistical education and research.

Usage

```
uk_crime_rubbish
```

Format

An object of class `table` with 4 rows (Rubbish on street categories) and 2 columns (Crime victimization status):

Rows (Rubbish on street):

- **Very common:** Rubbish on street is very common

- **Fairly common:** Rubbish on street is fairly common
- **Not very common:** Rubbish on street is not very common
- **Not at all common:** Rubbish on street is not at all common

Columns (Crime victimization status):

- **Not a victim of crime:** Respondent was not a victim of crime

- **Victim of crime:** Respondent was a victim of crime

Values Frequencies or counts of survey respondents (integer numbers).

Source

BMRB Social Research and Home Office, Research, Development and Statistics Directorate (2022). *British Crime Survey, 2007-2008* (data collection), 4th Edition. UK Data Service, SN: 6066. [doi:10.5255/UKDASN60662](https://doi.org/10.5255/UKDASN60662)

Examples

```
# Load the dataset into the workspace
data(uk_crime_rubbish)

# Display the entire table
print(uk_crime_rubbish)

# Calculate marginal totals (row sums and column sums)
rowSums(uk_crime_rubbish)
colSums(uk_crime_rubbish)

# Perform chi-square test of independence
chisq.test(uk_crime_rubbish)
```

usa_toddlers

USA School Readiness of Toddlers by Parent Education

Description

A cross-classified data table from the National Survey of Children's Health showing school readiness for children aged 3-5 years by the highest level of education of an adult in the household. The table reports nationwide counts for children who are on track versus those who need support.

Usage

```
usa_toddlers
```

Format

An object of class `table` with 4 rows (parent education categories) and 2 columns (school readiness):

Rows (parent education): • **Less than high school**

- **High School/GED**
- **College/Technical**
- **College or more**

Columns (school readiness): • **On track**

- **Need support**

Values Frequencies or counts of children (integer numbers).

Details

Totals in this dataset: 23,176 children overall, including 15,964 classified as On track and 7,212 as Need support.

The largest parent-education group is College or more (15,718 children), followed by College/Technical (4,388 children).

Source

Data Resource Center for Child & Adolescent Health (2024). *National Survey of Children's Health: School Readiness (Age 3-5 Years) by Parent Education*. Nationwide tabulation based on the highest level of education of an adult in the household. <https://www.childhealthdata.org/> Accessed 30 October 2024.

Examples

```
# Load the dataset into the workspace
data(usa_toddlers)

# Display the table
print(usa_toddlers)

# Calculate marginal totals
rowSums(usa_toddlers)
colSums(usa_toddlers)
sum(usa_toddlers)
```

usa_voting_prefs	<i>USA Residents Voting Preferences by Income Category</i>
------------------	--

Description

A cross-classified data table presenting the voting preferences of USA residents classified by their income category, according to a survey by the Pew Research Center (2014). These data are typically used to illustrate computations and contingency analyses in statistical scenarios.

Usage

```
usa_voting_prefs
```

Format

An object of class `table` with 4 rows (Income Categories) and 3 columns (Political Affiliations):

Rows (Income Categories):

- **I:** Less than \$30,000

- **II:** More than \$30,000 but less than \$50,000
- **III:** More than \$50,000 but less than \$100,000
- **IV:** \$100,000 or more

Columns (Political Affiliations):

- **R:** Republican or leaning toward Republican

- **U:** Undecided
- **D:** Democrat or leaning toward Democrat

Values Frequencies or counts of respondents (integer numbers).

Source

Pew Research Center (2014). *Religious Landscape Study: Compare Party Affiliation by Income Distribution*. <https://www.pewresearch.org/religion/religious-landscape-study/compare/party-affiliation/by/income-distribution/> Accessed 08 July 2022.

Examples

```
# Load the dataset into the workspace
data(usa_voting_prefs)

# Display the entire table
print(usa_voting_prefs)

# Calculate marginal totals (row sums and column sums)
rowSums(usa_voting_prefs)
colSums(usa_voting_prefs)
```

us_covid_mortality *US Mortality Data by Age and Gender (2020 vs 2015-2019 Average)*

Description

A dataset containing US mortality statistics by age group and gender, comparing 2020 deaths (including COVID-19 impact) with 2015-2019 averages. Includes all-cause deaths, non-COVID-19 deaths, and population data.

Usage

```
us_covid_mortality
```

Format

A data.frame with 22 rows (11 age groups × 2 genders) and 8 variables:

Age Character vector: age groups (<1, 1-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+)

Gender Character vector: "Male" or "Female"

Deaths_2020 Numeric: Total deaths in 2020

NonCOVID_Deaths_2020 Numeric: Non-COVID-19 deaths in 2020

COVID_Deaths_Percentage Numeric: Percentage of deaths attributed to COVID-19

Population_2020 Numeric: Population in 2020

Average_Deaths_2015_2019 Numeric: Average deaths for 2015-2019 period

Average_Population_2015_2019 Numeric: Average population for 2015-2019 period

Source

Jacobson, Sheldon H. and Jokela, Janet A. (2021). *Beyond COVID-19 Deaths during the COVID-19 Pandemic in the United States*. *Health Care Management Science*, 24, 661-665. doi:10.1007/s10729021095704

Examples

```
# Load the dataset
data(us_covid_mortality)

# View the structure
str(us_covid_mortality)

# Summary statistics by gender
aggregate(Deaths_2020 ~ Gender, data = us_covid_mortality, FUN = sum)

# COVID-19 impact analysis
us_covid_mortality$COVID_Impact <- with(
  us_covid_mortality,
  Deaths_2020 - Average_Deaths_2015_2019
)
summary(us_covid_mortality$COVID_Impact)
```

us_fall_accidents	<i>US Construction Fall Accidents by Occupation and Injury Degree</i>
-------------------	---

Description

Cross-classified counts of US construction fall accidents by occupation and injury degree, derived from the 2000-2020 data analysis reported by Halabi et al. (2022). The table summarizes how fall accidents are distributed across 17 occupation groups and 3 injury-severity categories.

Usage

```
us_fall_accidents
```

Format

An object of class table with 17 rows (occupation groups) and 3 columns (injury degree):

Rows (occupation groups, shown here with full names): Roofers, Construction laborers, Carpenters, Laborers, except construction, Supervisors and Engineers, Painters, plasterers, construction and maintenance, Installers and Repairers, Structural metal workers, Operators, Electricians, Truck drivers, heavy, Technicians, Mechanics, Janitors and cleaners, Helpers, construction trades, Installers (Drywalls, elevators), Plumbers, Sales engineers, workers.

Columns (injury degree): Fatality, Hospitalized, Non Hospitalized.

Values Frequencies (counts of fall accidents).

Details

Totals in this dataset: 15,495 accidents overall, including 5,701 fatal cases, 8,955 hospitalized cases, and 839 non-hospitalized cases.

The table stored in the package uses abbreviated row names for compact display; the full occupation labels are listed above for readability.

The largest occupation groups by total number of accidents are Roofers (2,967), Construction laborers (2,725), and Carpenters (1,665).

Source

Halabi, Y., Xu, H., Long, D., Chen, Y., Yu, Z., Alhaek, F., and Alhaddad, W. (2022). *Causal Factors and Risk Assessment of Fall Accidents in the US Construction Industry: A Comprehensive Data Analysis (2000-2020)*. *Safety Science*, 146, 105537. Table 6 (p. 8), "Fall accidents distributed by occupation and injury degree". doi:10.1016/j.ssci.2021.105537

Examples

```
data(us_fall_accidents)
us_fall_accidents
rowSums(us_fall_accidents)
colSums(us_fall_accidents)
sum(us_fall_accidents)
```

Index

* datasets

- bmi_mortality, 2
- bmi_mortality_all, 3
- uk_crime_rubbish, 9
- us_covid_mortality, 12
- us_fall_accidents, 13
- usa_toddlers, 10
- usa_voting_prefs, 11

- bmi_mortality, 2
- bmi_mortality_all, 3

- chisq.test, 5

- pqa, 4, 6–8
- print.pqa, 5, 7, 8
- print.pqa.chisq, 6, 7, 8
- print.pqa.subtable, 6, 7, 8

- summary.pqa, 6, 8

- table, 4, 5

- uk_crime_rubbish, 9
- us_covid_mortality, 12
- us_fall_accidents, 13
- usa_toddlers, 10
- usa_voting_prefs, 11