

# Package ‘Crosstabs.Loglinear’

December 12, 2022

**Title** Cross Tabulation and Loglinear Analyses of Categorical Data

**Version** 0.1.1

**Date** 2022-12-11

**Description** Provides 'SPSS'- and 'SAS'-like output for cross tabulations of two categorical variables (CROSSTABS) and for hierarchical loglinear analyses of two or more categorical variables (LOGLINEAR). The methods are described in Agresti (2013, ISBN:978-0-470-46363-5), Ajzen & Walker (2021, ISBN:9780429330308), Field (2018, ISBN:9781526440273), Norusis (2012, ISBN:978-0-321-74843-0), Nussbaum (2015, ISBN:978-1-84872-603-1), Stevens (2009, ISBN:978-0-8058-5903-4), Tabachnik & Fidell (2019, ISBN:9780134790541), and von Eye & Mun (2013, ISBN:978-1-118-14640-8).

**License** GPL (>= 2)

**Imports** stats, utils

**LazyLoad** yes

**LazyData** yes

**NeedsCompilation** no

**Author** Brian O'Connor [aut, cre]

**Maintainer** Brian O'Connor <brian.oconnor@ubc.ca>

**Repository** CRAN

**Date/Publication** 2022-12-12 12:00:02 UTC

## R topics documented:

Crosstabs.Loglinear-package . . . . .	2
CROSSTABS . . . . .	2
datasets . . . . .	4
LOGLINEAR . . . . .	7

<b>Index</b>	<b>11</b>
--------------	-----------

Crosstabs.Loglinear-package

*Crosstabs.Loglinear*

---

### Description

Provides 'SPSS'- and 'SAS'-like output for cross tabulations of two categorical variables (CROSSTABS) and for hierarchical loglinear analyses of two or more categorical variables (LOGLINEAR).

### References

- Agresti, A. (2013). *Categorical data analysis (3rd ed)*. Hobokon, NJ: John Wiley & Sons.
- Ajzen, R., & Walker, C. M. (2021). *Categorical data analysis for the behavioral and social sciences (2nd ed.)*. New York, NY: Routledge.
- Field, A. (2018). Chapter 18: Categorical data. *Discovering statistics using SPSS (5th ed.)*. Los Angeles, CA: Sage.
- Noursis, M. J. (2012). Chapter 1: Model selection loglinear analysis. *IBM SPSS statistics 19: Advanced statistical procedures Companion*. Upper Saddle River, NJ: Prentice Hall.
- Nussbaum, E. M. (2015). *Categorical and nonparametric data analysis choosing the best statistical technique*. New York, NY: Routledge.
- Stevens, J. P. (2009). Chapter 14: Categorical data analysis: The log linear model. *Applied multivariate statistics for the social sciences (5th ed.)*. New York, NY: Routledge.
- Tabachnick, B. G., & Fidell, L. S. (2019). Chapter 16: Multiway frequency analysis. *Using multivariate statistics*. New York, NY: Pearson.
- von Eye, A., & Mun, E. Y. (2013). *Log-Linear modeling concepts, interpretation, and application*. Hoboken, NJ: Wiley.

---

CROSSTABS

*Cross tabulations of two categorical variables*

---

### Description

Provides 'SPSS'- and 'SAS'-like output for cross tabulations of two categorical variables. The input can be raw data, a contingency table, or a dataframe with cell frequency counts. The output includes the contingency table, expected frequencies, Pearson's Chi-Square, Yates's Chi-Square (continuity correction), the Likelihood Ratio, Fisher's Exact p, the Linear-by-Linear Association., the McNemar Test, the contingency coefficient C, phi, Cramer's V, Cohen's W, the residuals, standardized residuals, and adjusted residuals. Additional output for 2-by-2 tables includes the risk difference, the risk ratio, the odds ratio, and Yule's Q.

**Usage**

```
CROSSTABS(data, data_type = 'raw', variables=NULL, Freq = NULL, verbose=TRUE)
```

**Arguments**

data	The input data, which can be raw data, a contingency table, or a dataframe with cell frequency counts (see the Examples below).
data_type	The kind of input data. The options are 'raw' (for raw data), 'cont.table' (for a two-dimensional contingency table), or 'counts' (for a dataframe with the cell frequency counts).
variables	(optional) The two variable names, which is required if data_type = 'raw' or 'counts', e.g., variables=c('varA','varB'). Not required if data_type = 'cont.table'.
Freq	(optional) If data_type = 'counts', then Freq is the name of the column in data that has the frequency counts. If unspecified, it will be assumed that the column is named 'Freq'.
verbose	(optional) Should detailed results be displayed in console? The options are: TRUE (default) or FALSE.

**Value**

A list with the following possible elements:

obsFreqs	The observed frequencies.
expFreqs	The expected frequencies.
modEStab	Model test and effect size coefficients.
residuals	The residuals.
stdresiduals	The standardized residuals.
adjresiduals	The adjusted residuals.
EStab2x2	For a 2-by-2 contingency table, a list with the risk difference, the risk ratio, the odds ratio, and Yule's Q values.

**Author(s)**

Brian P. O'Connor

**References**

- Agresti, A. (2013). *Categorical data analysis (3rd ed.)*. Hobokon, NJ: John Wiley & Sons.
- Ajzen, R., & Walker, C. M. (2021). *Categorical data analysis for the behavioral and social sciences (2nd ed.)*. New York, NY: Routledge.
- Field, A. (2018). Chapter 18: Categorical data. *Discovering statistics using SPSS (5th ed.)*. Los Angeles, CA: Sage.
- Noursis, M. J. (2012). Chapter 1: Model selection loglinear analysis. *IBM SPSS statistics 19:*

*Advanced statistical procedures Companion*. Upper Saddle River, NJ: Prentice Hall.

Stevens, J. P. (2009). Chapter 14: Categorical data analysis: The log linear model. *Applied multivariate statistics for the social sciences (5th ed.)*. New York, NY: Routledge.

Tabachnick, B. G., & Fidell, L. S. (2019). Chapter 16: Multiway frequency analysis. *Using multivariate statistics*. New York, NY: Pearson.

## Examples

```
# when 'data' is a raw data file (rather than counts/frequencies)
# Field (2018). Chapter 18: Categorical data -- cats only
CROSSTABS(data = subset(datasets$Field_2018_raw, Animal=='Cat'),
          data_type = 'raw',
          variables=c('Training','Dance') )

# when 'data' is a file with the counts/frequencies (rather than raw data points)
# Field (2018). Chapter 18: Categorical data -- cats only
CROSSTABS(data = subset(datasets$Field_2018, Animal=='Cat'),
          data_type = 'counts',
          variables=c('Training','Dance') )

# create and enter a two-dimensional contingency table for 'data'
# Field (2018). Chapter 18: Categorical data -- cats only
food <- c(28, 10)
affection <- c(48, 114)
Field_2018_cats_conTable <- rbind(food, affection)
colnames(Field_2018_cats_conTable) <- c('danced', 'did not dance')
names(attributes(Field_2018_cats_conTable)$dimnames) <- c('Training','Dance')
CROSSTABS(data = Field_2018_cats_conTable, data_type = 'cont.table')

# another way of creating the same two-dimensional contingency table for 'data'
# Field (2018). Chapter 18: Categorical data -- cats only
Field_2018_cats_conTable_2 <- matrix( c(28, 48, 10, 114), nrow = 2, ncol = 2)
colnames(Field_2018_cats_conTable_2) <- c('danced', 'did not dance')
rownames(Field_2018_cats_conTable_2) <- c('food', 'affection')
CROSSTABS(data = Field_2018_cats_conTable_2, data_type = 'cont.table')

# go to this web page to see many more examples of the CROSSTABS function analyses:
# https://oconnor-psych.ok.ubc.ca/loglinear/CROSSTABS\_vignettes.html
```

**Description**

A list with example data that were used in textbook presentations of categorical data analyses

**Usage**

```
data(datasets)
```

**Details**

A list with example data that were used in the following textbook presentations of categorical data analyses:

**datasets\$Agesti\_2019\_Tab9.3** is tabled data from Agresti (2019, p. 346).

**datasets\$Agesti\_2019\_Tab9.8** is tabled data from Agresti (2019, p. 351).

**datasets\$Ajzen\_2021\_Tab7.11** is tabled data from Ajzen and Walker (2021, p. 178).

**datasets\$Ajzen\_2021\_Tab7.16** is tabled data from Ajzen (2021, p. 180).

**datasets\$Field\_2018** is tabled data from Field (2018, Output 18.5 and Output 18.6).

**datasets\$Field\_2018\_raw** is raw data that simulates those from Field (2018, Output 18.5 and Output 18.6).

**datasets\$George\_2019\_26\_Hierarchical** is tabled data from George (2019, pp. 346-347).

**datasets\$Gray\_2012\_2wqy** is tabled data from Gray and Kinnear (2012, p. 538).

**datasets\$Gray\_2012\_3wqy** is tabled data from Gray and Kinnear (2012, p. 551).

**datasets\$Green\_2014** is tabled data from Green and Salkind (2014, p. 334).

**datasets\$Ho\_2014** is tabled data from Ho (2014, p. 513).

**datasets\$Howell\_2013** is tabled data from Howell (2013, p. 150).

**datasets\$Howell\_2017** is tabled data from Howell (2019, p. 512).

**datasets\$Meyers\_2013** is tabled data from Meyers (2013, p. 693).

**datasets\$Noursis\_2012\_marital** is tabled data from Noursis (2012a, p. 3).

**datasets\$Noursis\_2012\_voting\_degree** is tabled data from Noursis (2012b, p. 513).

**datasets\$Noursis\_2012\_voting\_degree\_sex** is tabled data from Noursis (2012b, p. 527).

**datasets\$Stevens\_2009\_HeadStart\_1** is tabled data from Stevens (2009, p. 472).

**datasets\$Stevens\_2009\_HeadStart\_2** is tabled data from Stevens (2009, p. 474).

**datasets\$Stevens\_2009\_Inf\_Survival** is tabled data from Stevens (2009, p. 481).

**datasets\$TabFid\_2019\_small** is tabled data from Tabachnick and Fidell (2019, p. 677).

**datasets\$Warner\_2020\_titanic** is tabled data from Warner (2020, p. 525).

**datasets\$Warner\_2020\_dog** is tabled data from Warner (2020, p. 530).

## References

- Agresti, A. (2013). *Categorical data analysis (3rd ed.)*. Hoboken, NJ: John Wiley & Sons.
- Ajzen, R., & Walker, C. M. (2021). *Categorical data analysis for the behavioral and social sciences (2nd ed.)*. New York, NY: Routledge.
- Field, A. (2018). Chapter 18: Categorical data. *Discovering statistics using SPSS (5th ed.)*. Los Angeles, CA: Sage.
- George, D., & Mallery, P. (2019). Chapter 26: Hierarchical log-linear models. *IBM SPSS statistics for Windows, version 25*. IBM Corp., Armonk, N.Y., USA.
- Gray, C. D., & Kinnear, P. R. (2012). Chapter 14: The analysis of multiway frequency tables. *IBM SPSS statistics 19 made simple*. Psychology Press.
- Green, S. B., Salkind, N. J. (2014). Chapter 41: Two-way contingency table analysis. *Using SPSS for Windows and Macintosh: Analyzing and understanding data*. New York, NY: Pearson.
- Ho, R. (2014). Chapter 19: Nonparametric tests. *Handbook of univariate and multivariate data analysis with IBM SPSS*. Boca Raton, FL: CRC Press.
- Howell, D. C. (2013). Chapter 6: Categorical data and chi-square. *Statistical methods for psychology (8th ed.)*. Belmont, CA: Wadsworth Cengage Learning.
- Howell, D. C. (2017). Chapter 19: Chi-square. *Fundamental statistics for the behavioral sciences*. Belmont, CA: Wadsworth Cengage Learning.
- Meyers, L. S., Gamst, G. C., & Guarino, A. J. (2013). Chapter 66: Hierarchical loglinear analysis. *Performing data analysis using IBM SPSS*. Hoboken, NJ: Wiley.
- Nouris, M. J. (2012a). Chapter 22: General loglinear analysis. *IBM SPSS statistics 19: Statistical procedures companion*. Upper Saddle River, NJ: Prentice Hall.
- Nouris, M. J. (2012b). Chapter 1: Model selection Loglinear analysis. *IBM SPSS Statistics 19: Advanced statistical procedures Companion*. Upper Saddle River, NJ: Prentice Hall.
- Stevens, J. P. (2009). Chapter 14: Categorical data analysis: The log linear model. *Applied multivariate statistics for the social sciences (5th ed.)*. New York, NY: Routledge.
- Tabachnick, B. G., & Fidell, L. S. (2019). Chapter 16: Multiway frequency analysis. *Using multivariate statistics*. New York, NY: Pearson.
- Warner, R. M. (2021). Chapter 17: Chi-square analysis of contingency tables. *Applied statistics: Basic bivariate techniques (3rd ed.)*. Thousand Oaks, CA: SAGE Publications.

## Examples

```
names(datasets)
```

```
datasets$Agresti_2019_Tab9.3
datasets$Agresti_2019_Tab9.8
datasets$Ajzen_2021_Tab7.11
datasets$Ajzen_2021_Tab7.16
datasets$Field_2018
head(datasets$Field_2018_raw)
datasets$George_2019_26_Hierarchical
datasets$George_2019_27_Nonhierarchical
datasets$Gray_2012_2way
datasets$Gray_2012_3way
datasets$Green_2014
datasets$Ho_2014
datasets$Howell_2013
datasets$Howell_2017
datasets$Meyers_2013
datasets$Nouris_2012_marital
datasets$Nouris_2012_voting_degree
datasets$Nouris_2012_voting_degree_sex
datasets$Stevens_2009_HeadStart_1
datasets$Stevens_2009_HeadStart_2
datasets$Stevens_2009_Inf_Survival
datasets$TabFid_2019_small
datasets$Warner_2020_titanic
datasets$Warner_2020_dog
```

### Description

Provides 'SPSS'- and 'SAS'-like output for hierarchical loglinear analyses of two or more categorical variables. The input can be raw data, a contingency table, or a dataframe with cell frequency counts. The output includes: (1) a table with the K-Way and higher-order effects; (2) a table with the K-Way effects; (3) a table with the the partial associations; (4) a table with the parameter estimates; (5) a table with the backward elimination statistics; (6) a table with the final model goodness of fit tests; and (7) a table with the final model observed and expected frequencies, standardized residuals, and adjusted residuals.

### Usage

```
LOGLINEAR(data, data_type = 'raw', variables=NULL, Freq = 'Freq', verbose=TRUE)
```

### Arguments

data	The input data, which can be raw data or a dataframe with cell frequency counts (see the Examples below).
data_type	The kind of input data. The options are 'raw' (for raw data), 'cont.table' (for a two-dimensional contingency table), or 'counts' (for a dataframe with the cell frequency counts).
variables	The variable names. Two or more variable names must be specified, as in, variables=c('varA', 'varB', 'varC').
Freq	(optional) If data_type = 'counts', then Freq is the name of the column in data that has the frequency counts. If unspecified, it will be assumed that the column is named 'Freq'.
verbose	(optional) Should detailed results be displayed in console? The options are: TRUE (default) or FALSE.

### Details

The purpose of hierarchical loglinear procedures is to find a model that best fits data given the model-fitting constraints, and to then provide the model parameters. The analyses begin with the saturated model, which includes all possible terms and for which there is a perfect fit to data. Terms are then tested for possible exclusion, which occurs when removal of a term does not result in a statistically significant reduction in fit *and* when a term is not involved in any higher order interactions. This function provides statistics for the saturated model, for the hierarchal removal of the model terms, for the backward elimination steps, and for the final model.

When data\_type = 'cont.table', the data must be a two-dimensional contingency table that has the names of the table dimensions/variables. See the Examples below.

### Value

A list with the following possible elements:

KwayHO	A table with the K-Way and higher-order effects.
Kway	A table with the K-Way effects.



PartialAssociations	A table with the partial associations.
paramests	A table with the parameter estimates.
StepSummTab	A table with the backward elimination statistics.
FinalModeltests	A table with the final model goodness of fit tests.
FinalModelcells	A table with the final model observed and expected frequencies and adjusted residuals.

**Author(s)**

Brian P. O'Connor

**References**

- Agresti, A. (2013). *Categorical data analysis (3rd ed.)*. Hobokon, NJ: John Wiley & Sons.
- Ajzen, R., & Walker, C. M. (2021). *Categorical data analysis for the behavioral and social sciences (2nd ed.)*. New York, NY: Routledge.
- Field, A. (2018). Chapter 18: Categorical data. *Discovering statistics using SPSS (5th ed.)*. Los Angeles, CA: Sage.
- Nouris, M. J. (2012). Chapter 1: Model selection loglinear analysis. *IBM SPSS statistics 19: Advanced statistical procedures Companion*. Upper Saddle River, NJ: Prentice Hall.
- Nussbaum, E. M. (2015). *Categorical and nonparametric data analysis choosing the best statistical technique*. New York, NY: Routledge.
- Stevens, J. P. (2009). Chapter 14: Categorical data analysis: The log linear model. *Applied multivariate statistics for the social sciences (5th ed.)*. New York, NY: Routledge.
- Tabachnick, B. G., & Fidell, L. S. (2019). Chapter 16: Multiway frequency analysis. *Using multivariate statistics*. New York, NY: Pearson.
- von Eye, A., & Mun, E. Y. (2013). *Log-Linear modeling concepts, interpretation, and application*. Hoboken, NJ: Wiley.

**Examples**

```
# Field (2018). Chapter 19: Categorical data -- cats & dogs, entering raw data
LOGLINEAR(data = datasets$Field_2018,
           data_type = 'counts',
           variables=c('Animal', 'Training', 'Dance'),
           Freq = 'Freq' )

# Field (2018). Chapter 19: Categorical data -- cats & dogs, entering raw counts
LOGLINEAR(data = datasets$Field_2018_raw,
```

```
data_type = 'raw',
variables=c('Animal', 'Training', 'Dance'),
Freq = NULL )

# Field (2018). Chapter 19: Categorical data -- cats & dogs, entering a table
# example of creating and entering a two-dimensional contingency table for 'data'
food <- c(28, 10)
affection <- c(48, 114)
Field_2018_cats_conTable <- as.table(rbind(food, affection))
colnames(Field_2018_cats_conTable) <- c('danced', 'did not dance')
names(attributes(Field_2018_cats_conTable)$dimnames) <- c('Training', 'Dance')
LOGLINEAR(data = Field_2018_cats_conTable,
           data_type = 'cont.table',
           variables=c('Training', 'Dance') )

# go to this web page to see many more examples of the LOGLINEAR function analyses:
# https://oconnor-psych.ok.ubc.ca/loglinear/LOGLINEAR\_vignettes.html
```

# Index

CROSSTABS, [2](#)

Crosstabs.Loglinear-package, [2](#)

datasets, [4](#)

LOGLINEAR, [7](#)