The polycor Package

March 30, 2007

Title  Polychoric and Polyserial Correlations
Date  2007/3/29
Version  0.7-3
Depends  R (>= 1.9.0), mvtnorm
Author  John Fox
Description  Computes polychoric and polyserial correlations by quick “two-step” methods or ML, optionally with standard errors; tetrachoric and biserial correlations are special cases.
Maintainer  John Fox <jfox@mcmaster.ca>
License  GPL version 2 or newer

R topics documented:

hetcor .......................................................... 1
polychor ......................................................... 3
polyserial ....................................................... 5
print.polycor .................................................. 7

Index  8

hetcor  Heterogeneous Correlation Matrix

Description

Computes a heterogenous correlation matrix, consisting of Pearson product-moment correlations between numeric variables, polyserial correlations between numeric and ordinal variables, and polychoric correlations between ordinal variables.
hetcor

Usage

hetcor(data, ..., ML = FALSE, std.err = TRUE, bins=4)
## S3 method for class 'data.frame':
hetcor(data, ML = FALSE, std.err = TRUE,
       use = c("complete.obs", "pairwise.complete.obs"), bins=4, ...)
## Default S3 method:
hetcor(data, ..., ML = FALSE, std.err = TRUE, bins=4)
## S3 method for class 'hetcor':
print(x, digits = max(3, getOption("digits") - 3, ...)
## S3 method for class 'hetcor':
as.matrix(x, ...)

Arguments

data a data frame consisting of factors and/or numeric variables, or the first of several
variables.
...
variables and/or arguments to be passed down.
ML if TRUE, compute maximum-likelihood estimates; if FALSE, compute quick
two-step estimates.
std.err if TRUE, compute standard errors.
bins number of bins to use for continuous variables in testing bivariate normality; the
default is 4.
use if "complete.obs", remove observations with any missing data; if "pairwise.complete.obs",
compute each correlation using all observations with valid data for that pair of
variables.
x an object of class "hetcor" to be printed, or from which to extract the corre-
lation matrix.
digits number of significant digits.

Value

Returns an object of class "hetcor" with the following components:
correlations the correlation matrix.
type the type of each correlation: "Pearson", "Polychoric", or "Polyserial".
std.errors the standard errors of the correlations, if requested.
n the number (or numbers) of observations on which the correlations are based.
NA.method the method by which any missing data were handled: "complete.obs" or
"pairwise.complete.obs".
ML TRUE for ML estimates, FALSE for two-step estimates.

Note

Although the function reports standard errors for product-moment correlations, transformations (the
most well known is Fisher's z-transformation) are available that make the approach to asymptotic
normality much more rapid.
polychor

Author(s)
John Fox  ⟨jfox@mcmaster.ca⟩

References


See Also
polychor, polyserial

Examples

```r
set.seed(12345)
R <- matrix(0, 4, 4)
R[upper.tri(R)] <- runif(6)
diag(R) <- 1
R <- cov2cor(t(R) %*% R)
round(R, 4)  # population correlations
data <- rmvnorm(1000, rep(0, 4), R)
round(cor(data), 4)  # sample correlations
x1 <- data[,1]
x2 <- data[,2]
y1 <- cut(data[,3], c(-Inf, .75, Inf))
y2 <- cut(data[,4], c(-Inf, -1, .5, 1.5, Inf))
data <- data.frame(x1, x2, y1, y2)
hetcor(data)  # Pearson, polychoric, and polyserial correlations, 2-step est.
hetcor(x1, x2, y1, y2, ML=TRUE)  # Pearson, polychoric, polyserial correlations, ML est.
```

---

polychor  

**Polychoric Correlation**

Description

Computes the polychoric correlation (and its standard error) between two ordinal variables or from their contingency table, under the assumption that the ordinal variables dissect continuous latent variables that are bivariate normal. Either the maximum-likelihood estimator or a (possibly much) quicker “two-step” approximation is available. For the ML estimator, the estimates of the thresholds and the covariance matrix of the estimates are also available.
Usage

polychor(x, y, ML = FALSE, control = list(), std.err = FALSE, maxcor=.9999)

Arguments

x
a contingency table of counts or an ordered categorical variable; the latter can be numeric, a factor, or an ordered factor, but if a factor, its levels should be in proper order.

y
if x is a variable, a second ordered categorical variable.

ML
if TRUE, compute the maximum-likelihood estimate; if FALSE, the default, compute a quicker “two-step” approximation.

control
optional arguments to be passed to the optim function.

std.err
if TRUE, return the estimated variance of the correlation (for the two-step estimator) or the estimated covariance matrix (for the ML estimator) of the correlation and thresholds; the default is FALSE.

maxcor
maximum absolute correlation (to insure numerical stability).

Value

If std.err is TRUE, returns an object of class "polycor" with the following components:

type
set to "polychoric".

rho
the polychoric correlation.

row.cuts
estimated thresholds for the row variable (x), for the ML estimate.

col.cuts
estimated thresholds for the column variable (y), for the ML estimate.

var
the estimated variance of the correlation, or, for the ML estimate, the estimated covariance matrix of the correlation and thresholds.

n
the number of observations on which the correlation is based.

chi sq
chi-square test for bivariate normality.

df
degrees of freedom for the test of bivariate normality.

ML
TRUE for the ML estimate, FALSE for the two-step estimate.

Otherwise, returns the polychoric correlation.

Author(s)

John Fox (jfox@mcmaster.ca)

References


Polyserial Correlation

Description

Computes the polyserial correlation (and its standard error) between a quantitative variable and an ordinal variables, based on the assumption that the joint distribution of the quantitative variable and a latent continuous variable underlying the ordinal variable is bivariate normal. Either the maximum-likelihood estimator or a quicker “two-step” approximation is available. For the ML estimator the estimates of the thresholds and the covariance matrix of the estimates are also available.

Usage

polyserial(x, y, ML = FALSE, control = list(), std.err = FALSE, maxcor = .9999, bins = 4)

Arguments

x

a numerical variable.

y

an ordered categorical variable; can be numeric, a factor, or an ordered factor, but if a factor, its levels should be in proper order.

ML

if TRUE, compute the maximum-likelihood estimate; if FALSE, the default, compute a quicker “two-step” approximation.

control

optional arguments to be passed to the optim function.

std.err

if TRUE, return the estimated variance of the correlation (for the two-step estimator) or the estimated covariance matrix of the correlation and thresholds (for the ML estimator); the default is FALSE.

maxcor

maximum absolute correlation (to insure numerical stability).

bins

the number of bins into which to dissect x for a test of bivariate normality; the default is 4.

Examples

set.seed(12345)
data <- rmvnorm(1000, c(0, 0), matrix(c(1, .5, .5, 1), 2, 2))x <- data[,1]y <- data[,2]cor(x, y) # sample correlationx <- cut(x, c(-Inf, .75, Inf))y <- cut(y, c(-Inf, -1, .5, 1.5, Inf))polychor(x, y) # 2-step estimatepolychor(x, y, ML=TRUE, std.err=TRUE) # ML estimate
Value

If `std.err` is `TRUE`, returns an object of class "polycor" with the following components:

- **type**: set to "polyserial".
- **rho**: the polyserial correlation.
- **cuts**: estimated thresholds for the ordinal variable \( y \), for the ML estimator.
- **var**: the estimated variance of the correlation, or, for the ML estimator, the estimated covariance matrix of the correlation and thresholds.
- **n**: the number of observations on which the correlation is based.
- **chisq**: chi-square test for bivariate normality.
- **df**: degrees of freedom for the test of bivariate normality.
- **ML**: `TRUE` for the ML estimate, `FALSE` for the two-step estimate.

Otherwise, returns the polyserial correlation.

Author(s)

John Fox ⟨jfox@mcmaster.ca⟩

References


See Also

`hetcor`, `polychor`, `print.polycor`, `optim`

Examples

```r
set.seed(12345)
data <- rmvnorm(1000, c(0, 0), matrix(c(1, .5, .5, 1), 2, 2))
x <- data[,1]
y <- data[,2]
cor(x, y)  # sample correlation
y <- cut(y, c(-Inf, -1, .5, 1.5, Inf))
polyserial(x, y)  # 2-step estimate
polyserial(x, y, ML=TRUE, std.err=TRUE)  # ML estimate
```
Description

print method for objects of class `polycor`, produced by `polychor` and `polyserial`.

Usage

```r
## S3 method for class 'polycor':
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

- `x`: an object of class `polycor`, as returned by `polychor` or `polyserial`.
- `digits`: number of significant digits to be printed.
- `...`: not used.

Value

Invisibly returns `x`: used for its side effect — i.e., printing.

Author(s)

John Fox (jfox@mcmaster.ca)

See Also

`polychor`, `polyserial`

Examples

```r
set.seed(12345)
data <- rmvnorm(1000, c(0, 0), matrix(c(1, .5, .5, 1), 2, 2))
x <- data[,1]
y <- data[,2]
cor(x, y) # sample correlation
x <- cut(x, c(-Inf, .75, Inf))
y <- cut(y, c(-Inf, -1, .5, 1.5, Inf))
polychor(x, y, ML=TRUE, std.err=TRUE) # polychoric correlation, ML estimate
```
Index

*Topic methods
  print.polycor, 7

*Topic models
  hetcor, 1
  polychor, 3
  polyserial, 5

*Topic print
  print.polycor, 7

as.matrix.hetcor(hetcor), 1

hetcor, 1, 4, 6

optim, 4, 6

polychor, 3, 3, 6, 7
polyserial, 3, 4, 5, 7
print.hetcor(hetcor), 1
print.polycor, 4, 6, 7