How can I compute tetrachoric correlations in Stata?

Authored by

stats writer

July 1, 2024

Tetrachoric correlations are a statistical method used to measure the relationship between two dichotomous variables, or variables with only two possible values. In Stata, this can be computed using the "tetrachoric" command. This command takes in the two variables of interest and outputs the tetrachoric correlation coefficient, which ranges from -1 to 1 and indicates the strength and direction of the relationship between the two variables. This method is useful for analyzing data with binary outcomes, such as yes or no responses, and can provide valuable insights into the underlying relationship between these variables. To compute tetrachoric correlations in Stata, the data must be organized in a specific format, and the "polychoric" command can also be used to adjust for any underlying continuous variables. Overall, using the tetrachoric command in Stata can provide researchers with a powerful tool for understanding the associations between dichotomous variables in their data.

How can I compute tetrachoric correlations in Stata? | Stata FAQ

Stata has added a maximum likelihood tetrachoric command to Stata 9.2. The matrix of tetrachoric correlations is saved in r(Rho) for use pic pcamat or factormat.

If you need polychoric or polyserial correlations in addition to tetrachoric then the polychoric command by Stas Kolenikov is meant for you. The correlation matrix is displayed using the matrix list r(R) command.

You can download polychoric over the internet by typing search polychoric (see How can I used the search command to

search for
programs and get additional
help? for more information about using search).

Now, let's read in an example dataset,
tetra, which contains data several dichotomous variables
derived from the High School and Beyond study.

```
use https://stats.idre.ucla.edu/stat/stata/faq/tetra, clear

summarize female schtyp ses hon sci

Variable | Obs Mean Std. Dev. Min Max
-------------+--------------------------------------------------------
female | 200 .545 .4992205 0 1
schtyp | 200 .84 .367526 0 1
ses | 200 .765 .4250628 0 1
hon | 200 .265 .4424407 0 1
sci | 200 .335 .4731749 0 1
```

Let's compute a standard correlation matrix followed by a tetrachoric
correlation matrix.

```
corr female schtyp ses hon sci
(obs=200)

| female schtyp ses hon sci
-------------+---------------------------------------------
female | 1.0000
schtyp | -0.0153 1.0000
ses | -0.1512 -0.1776 1.0000
hon | 0.1391 0.0148 0.0389 1.0000
sci | -0.1599 -0.0659 0.1435 0.3179 1.0000

tetrachoric female schtyp ses hon sci

(obs=200)

| female schtyp ses hon sci
-------------+---------------------------------------------
female | 1.0000
schtyp | -0.0291 1.0000
ses | -0.2636 -0.4475 1.0000
hon | 0.2362 0.0304 0.0735 1.0000
sci | -0.2562 -0.1253 0.2668 0.5037 1.0000

mat list r(Rho) /* note the upper-case R */

symmetric r(Rho)
```

female schtyp ses hon sci
female 1
schtyp -.02907834 1
ses -.26359952 -.44745088 1
hon .23621208 .03040347 .07348643 1
sci -.25621193 -.12525624 .26676862 .50370596 1

And now for the polychoric version.

polychoric female schtyp ses hon sci

matrix list r(R)

symmetric r(R)
female schtyp ses hon sci
female 1
schtyp -.02893308 1
ses -.26297321 -.44527627 1
hon .23643947 .03044205 .07352325 1
sci -.25634248 -.12494757 .2666472 .504248 1