

# Introduction to the Tetrachoric and Polychoric Correlation Coefficients

John S. Uebersax

Rev. 8 Sept. 2015

---

- [Introduction](#)
- [Summary](#)
- [Pros and Cons](#)
- [Intuitive Explanation](#)
- [Detailed Description](#)
- [Using the Polychoric Correlation to Measure Agreement](#)
- [Extensions and generalizations](#)
  - [Skewed distributions](#)
  - [Nonparametric distributions](#)
- [Examples](#)
  - [Tetrachoric correlation](#)
  - [Polychoric correlation](#)
  - [Polychoric correlation with relaxed assumptions](#)
- [Factor analysis and SEM](#)
- [Software](#) **Updated!**
  - [Tetrachoric correlation](#)
  - [Polychoric correlation](#)
  - [Tetrachoric/polychoric correlation coefficient in R](#) **New!**
  - [Generalized latent correlation](#)
- [References](#)

Go to [Agreement Statistics](#) main page

Go to [Latent Structure Analysis](#) main page

---

## Introduction

This page describes the tetrachoric and polychoric correlation coefficients, explains their meaning and uses, gives examples and references, provides programs for their estimation, and discusses other available software. While discussion is primarily oriented to rater agreement problems, it is general enough to apply to most other uses of these statistics.

A clear, concise description of the tetrachoric and polychoric correlation coefficients, including issues relating to their estimation, is found in Drasgow (1988). Olsson (1979) is also helpful.

What distinguishes the present discussion is the view that the tetrachoric and polychoric correlation models are special cases of latent trait modeling. (This is not a new observation, but it is sometimes overlooked). Recognizing this opens up important new possibilities. In particular, it allows one to relax the distributional assumptions which are the most limiting feature of the "classical" tetrachoric and polychoric correlation models.

Note in any case that the terms *tetrachoric correlation* and *polychoric correlation* are obsolete and arguably inaccurate. They refer to the tetrachoric series and polychoric series, numerical methods previously (before modern computers) used to facilitate calculations. Now these correlations are

estimated by maximum likelihood or other means. Hence a term like *latent correlation* (or latent continuous correlation) is more appropriate.

---

## Summary

The tetrachoric correlation (Pearson, 1901), for binary data, and the polychoric correlation, for ordered-category data, are excellent ways to measure rater agreement. They estimate what the correlation between raters would be if ratings were made on a continuous scale; they are, theoretically, invariant over changes in the number or "width" of rating categories. The tetrachoric and polychoric correlations also provide a framework that allows testing of marginal homogeneity between raters. Thus, these statistics let one separately assess both components of rater agreement: agreement on trait definition and agreement on definitions of specific categories.

These statistics make certain assumptions, however. With the polychoric correlation, the assumptions can be tested. The assumptions cannot be tested with the tetrachoric correlation if there are only two raters; in some applications, though, theoretical considerations may justify the use of the tetrachoric correlation without a test of model fit.

---

## Pros and Cons: Tetrachoric and Polychoric Correlation Coefficients

### Pros:

- These statistics express rater association in a familiar form--a correlation coefficient.
- They provide a way to separately quantify association and similarity of category definitions.
- They do not depend on number of rating levels; results can be compared for studies where the number of rating levels is different.
- They can be used even if different raters have different numbers of rating levels.
- The assumptions can be easily tested for the polychoric correlation.
- Estimation software is routinely available (e.g., SAS PROC FREQ, and PRELIS).

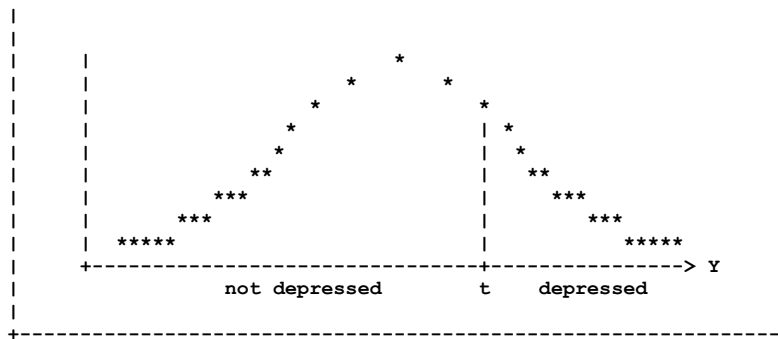
### Cons:

- Model assumptions not always appropriate--for example, if the latent trait is truly discrete.
  - For only two raters, there is no way to test the assumptions of the tetrachoric correlation.
- 

## Intuitive Explanation

Consider the example of two psychiatrists (Raters 1 and 2) making a diagnosis for presence/absence of Major Depression. Though the diagnosis is dichotomous, we allow that depression as a trait is continuously distributed in the population.

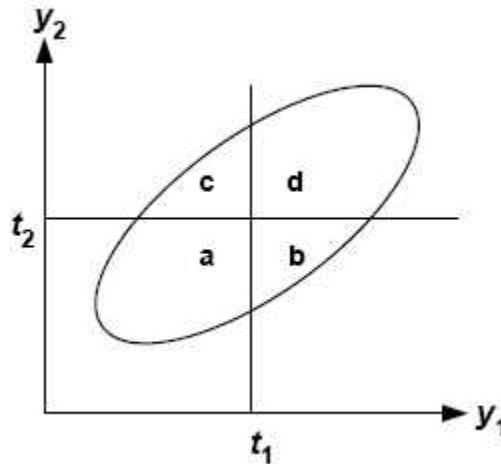
+-----+



**Figure 1 (draft). Latent continuous variable (depression severity,  $Y$ ); and discretizing threshold ( $t$ ).**

In diagnosing a given case, a rater considers the case's level of depression,  $Y$ , relative to some threshold,  $t$ : if the judged level is above the threshold, a positive diagnosis is made; otherwise the diagnosis is negative.

Figure 2 portrays the situation for two raters. It shows the distribution of cases in terms of depression level as judged by Rater 1 and Rater 2.



**Figure 2. Joint distribution (ellipse) of depression severity as judged by two raters ( $Y_1$  and  $Y_2$ ); and discretizing thresholds ( $t_1$  and  $t_2$ )**

$a$ ,  $b$ ,  $c$  and  $d$  denote the proportion of cases that fall in each region defined by the two raters' thresholds. For example,  $a$  is the proportion below both raters' thresholds and therefore diagnosed negative by both.

These proportions correspond to a summary of data as a  $2 \times 2$  cross-classification of the raters' ratings.

		Rater 1		
		-	+	
Rater 2	-	a	b	$a + b$
	+	c	d	$c + d$
		$a + c$	$b + d$	1

**Figure 3 (draft). Crossclassification proportions for binary ratings by two raters.**

Again, a, b, c and d in Figure 3 represent **proportions** (not frequencies).

Once we know the observed cross-classification proportions a, b, c and d for a study, it is a simple matter to estimate the model represented by Figure 2. Specifically, we estimate the location of the discretizing thresholds,  $t_1$  and  $t_2$ , and a third parameter,  $\rho$ , which determines the "fatness" of the ellipse.  $\rho$  is the tetrachoric correlation, or  $r^*$ . It can be interpreted here as the correlation between judged disease severity (before application of thresholds) as viewed by Rater 1 and Rater 2.

The principle of estimation is simple: basically, a computer program tries various combinations for  $t_1$ ,  $t_2$  and  $r^*$  until values are found for which the expected proportions for a, b, c and d in Figure 2 are as close as possible to the observed proportions in Figure 3. The parameter values that do so are regarded as (estimates of) the true, population values.

The polychoric correlation, used when there are more than two ordered rating levels is a straightforward extension of the model above. The difference is that there are more thresholds, more regions in Figure 2, and more cells in Figure 3. But again the idea is to find the values for thresholds and  $r^*$  that maximize similarity between model-expected and observed cross-classification proportions.

---

## Detailed Description

### Introduction

In many situations, even though a trait may be continuous, it may be convenient to divide it into ordered levels. For example, for research purposes, one may classify levels of headache pain into the categories none, mild, moderate and severe. Even for trait usually viewed as discrete, one might still consider continuous gradations--for example, people infected with the flu virus exhibit varying levels of symptom intensity.

The tetrachoric correlation and polychoric correlation coefficients are appropriate when the latent trait that forms the basis of ratings can be viewed as continuous. We will outline here the measurement model and assumptions for the tetrachoric correlation. The model and assumptions for the polychoric correlation are the same--the only difference is that there are more threshold parameters for the polychoric correlations, corresponding to the greater number ordered rating levels.

### Measurement Model

We begin with some notation and definitions. Let:

$X_1$  and  $X_2$  be the manifest (observed) ratings by Raters (or procedures, diagnostic tests, etc.) 1 and 2; these are discrete-valued variables;

$Y_1$ ,  $Y_2$  be latent continuous variables associated with  $X_1$  and  $X_2$ ; these are the pre-discretized, continuous "impressions" of the trait level, as judged by Raters 1 and 2;

$T$  be the true, latent trait level of a case.

A rating or diagnosis of a case begins with the case's true trait level,  $T$ . This information, along with "noise" (random error) and perhaps other information

unrelated to the true trait which a given rater may consider (unique variation), leads to each rater's impression of the case's trait level (Y1 and Y2). Each rater applies discretizing thresholds to this judged trait level to yield a dichotomous or ordered-category rating (X1 and X2).

Stated more formally, we have:

$$\begin{aligned} Y1 &= bT + u1 + e1, \\ Y2 &= bT + u2 + e2, \end{aligned}$$

where  $b$  is a regression coefficient,  $u1$  and  $u2$  are the unique components of the raters' impressions, and  $e1$  and  $e2$  represent random error or noise. It turns out that unique variation and error variation behave more or less the same in the model, and the former can be subsumed under the latter. Thus we may consider the simpler model:

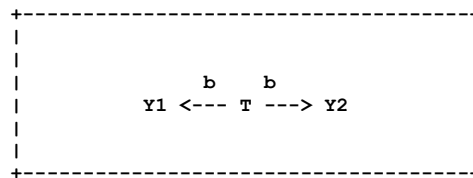
$$\begin{aligned} Y1 &= b1T + e1, \\ Y2 &= b2T + e2. \end{aligned}$$

The tetrachoric correlation assumes that the latent trait  $T$  is normally distributed. As scaling is arbitrary, we specify that  $T \sim N(0, 1)$ . Error is similarly assumed to be normally distributed (and independent both between raters and across cases). For reasons we need not pursue here, the model loses no generality by assuming that  $\text{var}(e1) = \text{var}(e2)$ . We therefore stipulate that  $e1, e2 \sim N(0, \sigma_{e_e})$ . A consequence of these assumptions is that  $Y1$  and  $Y2$  must also be normally distributed. To fix the scale, we specify that  $\text{var}(Y1) = \text{var}(Y2) = 1$ . It follows that  $b1 = b2 = b$  = the correlation of both  $Y1$  and  $Y2$  with the latent trait.

We define the tetrachoric correlation,  $r^*$ , as

$$r^* = b^2$$

A simple "path diagram" may clarify this:



**Figure 4 (draft). Path diagram.**

Here  $b$  is the path coefficient that reflects the influence of  $T$  on both  $Y1$  and  $Y2$ . Those familiar with the rules of path analysis will see that the correlation of  $Y1$  and  $Y2$  is simply the product of their degree of dependence on  $T$ --that is  $b^2$ .

As an aside, one might consider that the value of  $b$  is interesting in its own right, inasmuch as it offers a measure of the association of ratings with the true latent trait--i.e., a measure of rating validity or accuracy.

The tetrachoric correlation  $r^*$  is readily interpretable as a measure of the association between the ratings of Rater 1 and Rater 2. Because it estimates the correlation that exists between the pre-discretized judgements of the raters, it is, in theory, not affected by (1) the number of rating levels, or (2) the marginal proportions for rating levels (i.e., the 'base rates.'). The fact that this association is expressed in the familiar form of a correlation is also helpful.

The assumptions of the tetrachoric correlation coefficient may be expressed as follows:

1. The trait on which ratings are based is continuous.
2. The latent trait is normally distributed.
3. Rating errors are normally distributed.
4.  $\text{Var}(e)$  is homogeneous across levels of T.
5. Errors are independent between raters.
6. Errors are independent between cases.

Assumptions 1--4 can be alternatively expressed as the assumption that Y1 and Y2 follow a bivariate normal distribution.

We will assume that the one has sufficient theoretical understanding of the application to accept the assumption of latent continuity.

The second assumption--that of a normal distribution for T--is potentially more questionable. Absolute normality, however, is probably not necessary; a unimodal, roughly symmetrical distribution may be close enough. Also, the model implicitly allows for a monotonic transformation of the latent continuous variables. That is, a more exact way to express Assumptions 1-4 is that one can obtain a bivariate normal distribution by some monotonic transformation of Y1 and Y2.

The model assumptions can be tested for the polychoric correlation. This is done by comparing the observed numbers of cases for each combination of rating levels with those predicted by the model. This is done with the likelihood ratio chi-squared test,  $G^2$  (Bishop, Fienberg & Holland, 1975), which is similar the usual Pearson chi-squared test (the Pearson chi-square test can also be used; for more information on these tests, see the FAQ for testing model fit on the [Latent Class Analysis web site](#)).

The  $G^2$  test is assessed by considering the associated p value, with the appropriate degrees of freedom (df). The df are given by:

$$df = RC - R - C$$

where R is the number of levels used by the first rater and C is the number of levels used by the second rater. As this is a "goodness-of-fit" test, it is standard practice to set the alpha level fairly high (e.g., .10). A p value lower than the alpha level is evidence of model fit.

For the tetrachoric correlation  $R = C = 2$ , and there are no df with which to test the model. It is possible to test the model, though, when there are more than two raters.

---

## Using the Polychoric Correlation to Measure Agreement

Here are the steps one might follow to use the tetrachoric or polychoric correlation to assess agreement in a study. For convenience, we will mainly refer to the polychoric correlation, which includes the tetrachoric correlation as a special case.

## 1. Calculate the value of the polychoric correlation.

For this a computer program, such as those described in the [software section](#), is required.

## 2. Evaluate model fit.

The next step is to determine if the assumptions of the polychoric correlation are empirically valid. This is done with the goodness-of-fit test that compares observed crossclassification frequencies to model-predicted frequencies described previously. As noted, this test cannot be done for the tetrachoric correlation.

PRELIS includes a test of model fit when estimating the polychoric correlation. It is unknown whether SAS PROC FREQ includes such a test.

## 3. Assess magnitude and significance of correlation.

Assuming that model fit is acceptable, the next step is to note is the magnitude of the polychoric correlation. Its value is interpreted in the same way as a Pearson correlation. As the value approaches 1.0, more agreement on the trait definition is indicated. Values near 0 indicate little agreement on the trait definition.

One may wish to test the null hypothesis of no correlation between raters. There are at least two ways to do this. The first makes use of the estimated standard error of the polychoric correlation under the null hypothesis of  $r^* = 0$ . At least for the tetrachoric correlation, there is a simple closed-form expression for this standard error (Brown, 1977). Knowing this value, one may calculate a z value as:

$$z = \frac{r^*}{\text{sigma}_{r^*(0)}}$$

where the denominator is the standard error of  $r^*$  where  $r^* = 0$ . One may then assess statistical significance by evaluating the z value in terms of the associated tail probabilities of the standard normal curve.

The second method is via a chi-squared test. If  $r^* = 0$ , the polychoric correlation model is the same as the model of statistical independence. It therefore seems reasonable to test the null hypothesis of  $r^* = 0$  by testing the statistical independence model. Either the Pearson ( $X^2$ ) or likelihood-ratio ( $G^2$ ) chi-squared statistics can be used to test the independence model. The df for either test is  $(R - 1)(C - 1)$ . A significant chi-squared value implies that  $r^*$  is not equal to 0.

*[I now question whether the above is correct. For the polychoric correlation, data may fail the test of independence even with when  $r^* = 0$  (i.e., there may be some other kind of 'structure' to the data). If so, a better alternative would be to calculate a difference  $G^2$  statistic as:*

$$G^2_{H0} - G^2_{H1},$$

*where  $G^2_{H0}$  is the likelihood-ratio chi-squared for the independence model and  $G^2_{H1}$  is the likelihood-ratio chi-squared for the polychoric correlation model. The difference  $G^2$  can be evaluated as a chi-squared value with 1 df.*  
-- JSU, 27 Jul 00]

## 4. Testing equality of thresholds.

Equality of thresholds between raters can be tested by estimating what may be termed a **threshold-constrained polychoric correlation**. That is, one estimates the polychoric correlation with the added constraint(s) that the threshold(s) of Rater 1 is/are the same Rater 2's threshold(s). A difference  $G^2$  test is then made comparing the  $G^2$  statistic for this constrained model with the  $G^2$  for the unconstrained polychoric correlation model. The difference  $G^2$  statistic is evaluated as a chi-squared value with  $df = R - 1$ , where  $R$  is the number of rating levels (this test only applies when both raters use the same number of rating levels).

---

## Extensions and Generalizations

Here we briefly note some extensions and generalizations of the tetrachoric/polychoric correlation approach to analyzing rater agreement:

- **Modifying latent distribution assumptions.** When the assumption of latent bivariate normality is empirically or theoretically implausible, other distributional assumptions can be made. One may exploit the fact that the tetrachoric/polychoric correlation model is isomorphic with a [latent trait model](#). Within the latter framework one can have non-normal latent trait distributions, including skewed and nonparametric distributions.

**Skewed distributions.** A new page describing what might be colloquially called "skewed tetrachoric or polychoric correlation," but would be more accurately termed the [latent correlation with a skewed latent distribution](#) has been added. This page also describes a simple [computer program to implement the model](#) for binary ratings.

**Nonparametric distributions.** Example 3 below describes an alternative approach based on a nonparametric latent trait distribution.

- **Modifying measurement error assumptions.** One can easily relax the assumptions concerning measurement error. Hutchinson (2000) described models where the variance of measurement error differs according to the latent trait level of the object being rated. In theory, one could also consider non-Gaussian distributions for measurement error.
  - **More than two raters.** When there are more than two raters, the tetrachoric/polychoric correlation model generalizes to a latent trait model with normal-ogive (Gaussian cdf) response functions. Latent trait models can be used to (a) estimate the tetrachoric/polychoric correlation among all rater pairs; (b) simultaneously test whether all raters have the same definition of the latent trait; and (c) simultaneously test for equivalence of thresholds among all raters.
- 

## Examples



## Example 1. Tetrachoric Correlation

Table 1 summarizes hypothetical ratings by two raters on presence (+) or absence (-) of schizophrenia.

Rater 1	Rater 2		Total
	-	+	
-	40	10	50
+	20	30	50
Total	60	40	100

Table 1 (draft)

For these data, the tetrachoric correlation (std. error) is:

rho            0.6071    (0.1152)

which is much larger than the Pearson correlation of 0.4082 calculated for the same data.

The thresholds (std. errors) for the two raters are estimated as:

Rater 1    0.0000    (0.1253)  
Rater 2    0.2533    (0.1268)

## Example 2. Polychoric Correlation

Table 2 summarizes number of lambs born to 227 ewes over two years. These data were previously analyzed by Tallis (1962) and Drasgow (1988).

Tallis suggested that the number of lambs born is a manifestation of the ewe's fertility--a continuous and potentially normally distributed variable. Clearly the situation is more complex than the simple "continuous normal variable plus discretizing thresholds" assumptions allow for. We consider the data simply for the sake of a computational example.

Lambs born in 1953	Lambs born in 1952			Total
	None	1	2	
None	58	52	1	111
1	26	58	3	87
2	8	12	9	29
Total	92	122	13	227

Table 2 (draft)

Drasgow (1988; see also Olsson, 1979) described two different ways to calculate the polychoric correlation. The first method, the *joint maximum likelihood (ML) approach*, estimates all model parameters--i.e., rho and the thresholds--at the same time.

The second method, *two-step ML estimation*, first estimates the thresholds from the one-way marginal frequencies, then estimates rho, conditional on these thresholds, via maximum likelihood. For the tetrachoric correlation, both methods produce the same results; for the polychoric correlation, they may produce slightly different results.

The data in Table 2 are analyzed with the POLYCORR program (Uebersax, 2000). Application of the joint ML approach produces the following estimates (standard errors):

```

rho                0.4192  (0.0761)
threshold 2, 1952  -0.2421  (0.0836)
threshold 3, 1952   1.5938  (0.1372)
threshold 2, 1953  -0.0297  (0.0830)
threshold 3, 1953   1.1331  (0.1063)

```

With two-step estimation the results are:

```

rho                0.4199  (0.0747)
threshold 2, 1952  -0.2397
threshold 3, 1952   1.5781
threshold 2, 1953  -0.0276
threshold 3, 1953   1.1371

```

However the  $G^2$  statistic testing model fit for the joint ML and two-step estimates are 11.54 and 11.55, respectively, each with 3 df. The corresponding p-values, less than .01, suggest poor model fit and implausibility of the polychoric model assumptions. Acceptable fit could possibly be obtained by considering a skewed latent trait distribution.

### Example 3. Polychoric Correlation with Relaxed Distributional Assumptions

The data in Table 3, previously analyzed by Hutchinson (2000), summarize ratings on the health of 460 trees and shrubs by two raters. Rating levels denote increasing levels of plant health; i.e., 1 indicates the lowest level, and 6 the highest level.

Rating of Rater	Rating of Rater 1						Total
	1	2	3	4	5	6	
1	30	1	0	0	0	0	31
2	0	10	2	0	0	0	12
3	0	4	8	3	1	0	16
4	0	3	3	37	9	0	52
5	0	0	1	25	71	49	146
6	0	0	0	2	20	181	203
Total	30	18	14	67	101	230	460

Table 3 (draft).  
Ratings of plant health by two judges

The polychoric correlation (std. error) for these data is .954 using joint estimation. However there is reason to doubt the assumptions of the standard polychoric correlation model; the  $G^2$  model fit statistic is 57.33 on 24 df ( $p < .001$ ).

Hutchinson (2000) showed that the data can be fit by allowing measurement error variance to differ from low to high levels of the latent trait. Instead, we relax the assumption of a normally distributed latent trait. Using the [LLCA program](#) (Uebersax, 1993a) a latent trait model with a nonparametric latent trait distribution was fit to the data. The distribution was represented as six equally-spaced locations (located latent classes) along a unidimensional continuum, the density at each location (latent class prevalence) being estimated.

Model fit, assessed by the  $G^2$  statistic was 15.65 on 19 df ( $p = .660$ ). The LLCA program gave the correlation of each variable with the latent trait as .963. This value squared, .927, estimates what the correlation of the raters would be if they made their ratings on a continuous scale. This is a

generalization of the polychoric correlation, though perhaps we should reserve that term for the latent bivariate normal case. Instead, we simply term this the *latent correlation* between the raters.

(To see the input file for the LLCA program, [click here](#).)

The distribution of the latent trait estimated by the model is follows:

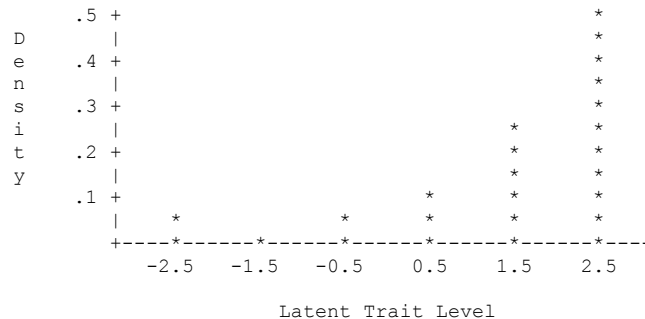


Figure 5 (draft).  
Estimated latent trait distribution

The shape suggests that the latent trait could be economically modeled with an asymmetric parametric distribution, such as a beta or exponential distribution.

---

## Factor analysis and SEM

A new, separate web page has been added on the topic of [factor analysis and SEM](#) with tetrachoric and polychoric correlations.

---

## Software

### Programs for tetrachoric correlation

[TetMat](#) is my free program to estimate a matrix of tetrachoric correlations. It also supplies other useful information such as one- and two-way marginal frequencies and rates, asymptotic standard errors of rho, p-values, confidence ranges, and thresholds. Provisions are made to smooth a potentially improper correlation matrix by the method of Knol and ten Berge (1989).

[Tcorr](#) is a simple utility for estimating a single tetrachoric correlation coefficient and its standard error. Just enter the frequencies of a fourfold table and get the answer. Also supplies threshold estimates.

[Dirk Enzmann](#) has written an SPSS macro to estimate a matrix of tetrachoric correlations. He also has a standalone version.

[Jim Fleming](#) also has a program to estimate a matrix of tetrachoric correlations and optionally smooth a poorly conditioned matrix.

Brown's (1977) algorithm [AS 116](#), a Fortran subroutine to calculate the tetrachoric correlation and its standard error, can be found at StatLib. Alternatively, you can download my program, Tcorr, above, which includes simple source code with an actual working version of Brown's subroutine.

[TESTFACT](#) is a very sophisticated program for item analysis using both classical and modern psychometric (IRT) methods. It includes provisions for calculating tetrachoric correlations.

## Programs for polychoric and tetrachoric correlation

[POLYCORR](#) is a program I've written to estimate the polychoric correlation and its standard error using either joint ML or two-step estimation. Goodness-of-fit and a lot of other information are also provided. **Note:** this program is just for a single pair of variables, or a few considered two at a time. It does not estimate a matrix of tetra- or polychoric correlations.

- [Basic version](#). This handles square tables only (i.e., models where both items have the same number of levels).
- [Advanced version](#). This allows non-square tables and has other advanced technical features, such as the ability to combine cells during estimation.

Note. At present the basic version doesn't run on 64-bit Windows operating systems. This will be addressed later in September 2015.

### [R](#)

(See next section.)

### [SAS](#)

A single polychoric or tetrachoric correlation can be calculated with the PLCORR option of SAS PROC FREQ. Example:

```
proc freq;
  tables var1*var2 / plcorr maxiter=100;
run;
```

Joint estimation is used. The standard error is supplied, but not thresholds. No goodness-of-fit test is performed.

### **Update!**

As of Base SAS® version 9.3 and SAS/STAT® version 13.1 there are two new ways to compute polychoric/tetrachoric correlation coefficients: with the new POLYCHORIC option of PROC CORR, and with the PROC IRT. If you have more than two variables, both of these procedures (especially the latter) can be used to supply a matrix of tetrachoric/polychoric correlation coefficients. Details and examples will follow here shortly. Meanwhile consult the SAS documentation for PROC CORR and PROC IRT.

For those with older SAS version, a SAS macro, [%POLYCHOR](#), can construct a matrix of polychoric or tetrachoric correlations. The macro is relatively slow.

For tetrachoric correlations, if there is a single 0 frequency in the 2×2 crossclassification table for a pair of variables (see Figure 3 above), the PLCORR option of PROC FREQ and %POLYCHOR might unnecessarily supply a missing value result, at least if maxiter is left at the default value of 20. So far I have found this problem is avoided by setting maxiter higher, e.g., to 40, 50 or 100.

### [SPSS](#)

SPSS has no intrinsic procedure to estimate polychoric correlations.

As noted above, [Dirk Enzmann](#) has written an SPSS macro to estimate a matrix of tetrachoric correlations.

[PRELIS](#). A useful program for estimating a matrix of polychoric or tetrachoric correlations is PRELIS. It includes a goodness-of-fit test for each pair of variables. Standard errors can be requested. PRELIS uses two-step estimation. Because it is supplied with LISREL, PRELIS is widely available. Most university computation centers probably already have copies and/or site licenses.

[Mplus](#) can estimate a matrix of polychoric and tetrachoric correlations and estimate their standard errors. Two-step estimation is used. Features similar to PRELIS/LISREL.

#### [Stata](#)

Stata's internal function for tetrachoric correlations is a very rough approximation (e.g., actual tetrachoric correlation = .5172, Stata reports .6169!) based on Edwards and Edwards (1984) and is unsuitable for many or most applications. A more accurate external module has been written by [Stas Kolenikov](#) to estimate a matrix of polychoric or tetrachoric correlations and their standard errors.

[MicroFACT](#) will estimate polychoric and tetrachoric correlations and standard errors. Provisions for smoothing an improper correlation matrix are supplied. No goodness-of-fit tests. A free student version that handles up to thirty variables can be downloaded. Also does factor analysis.

#### [WinBUGS](#)

As Albert (1992) and others have shown, it is possible to estimate the polychoric correlation using Bayesian (MCMC) methods. The WinBUGS program can probably be used for this. (If anyone has code for this, I'll place it here.)

## Calculating the Tetrachoric/Polychoric Correlation Coefficient in R

**John Fox** has contributed an R library, [polycor](#), with functions to estimate the polychoric (and tetrachoric) correlation.

### How to Calculate the Polychoric Correlation Coefficient using R

1. Download R from the CRAN website (free) and install on your computer: <http://cran.r-project.org/>
2. Download and install RStudio (free): <https://www.rstudio.com>
3. Open RStudio.
4. At command prompt (">") type: **install.packages("polycor")**
5. When this action completes, type: **library("polycor")**
6. Supply frequencies for a two-way crossclassification table as a vector, e.g.,

```
x <- c(10, 20, 30, 40, 50, 60, 70, 80, 90)
```

7. Arrange column as a matrix, e.g.,

```
y <- matrix(x, 3, 3)
```

8. Estimate polychoric correlation using the 'polychor' (note spelling) function.

```
polychor(y, ML=T, std.err=T)
```

Results:

```
Polychoric Correlation, ML est. =  
-0.1183 (0.06098)  
Test of bivariate normality:  
Chisquare = 1.216, df = 3, p = 0.7491  
  
Row Thresholds  
Threshold Std.Err.  
1 -0.6227 0.06345  
2 0.2535 0.05976  
  
Column Thresholds  
Threshold Std.Err.  
1 -1.11100 0.07450  
2 -0.08275 0.05914
```

The estimated value of the polychoric correlation coefficient is -0.1183 with an estimated standard error of 0.06098.

If x has only two levels (rows and columns), the tetrachoric correlation is estimated.

It is also possible to supply raw data in the form of two vectors, x and y. For example, let x = c(1, 2, 1, 1, 2, 1, 1, 1, 2, 1), y = c(2, 1, 2, 2, 1, 1, 1, 2, 1, 2). Note that these are two variables measured on the same subjects/objects, so their lengths and orderings must match.

```
x <- c(1, 2, 1, 1, 2, 1, 1, 1, 2, 1)  
y <- c(2, 1, 2, 2, 1, 1, 1, 2, 1, 2)  
polychor(x, y, ML=T, std.err=T)
```

Results:

```
Polychoric Correlation, ML est. =  
0.3672 (0.3574)  
  
Row Threshold  
Threshold Std.Err.  
0.1573 0.3147  
  
Column Threshold  
Threshold Std.Err.  
0.1573 0.3147
```

(Note however that  $n=10$ ) is far too few observations from which to estimate the tetrachoric correlation coefficient; an  $n$  of at least 50 is recommended.)

See also functions for the polychoric and polyserial correlation coefficients in the [psych](#) R library of **William Revelle** at Northwestern.

Another R function for the polychoric correlation coefficient has been written by [David Duffy](#).

## Generalized latent correlation

The [glc program](#) generalizes the tetrachoric correlation to estimate the latent correlation between binary variables assuming a skewed latent trait distribution.

The skewed distribution is modeled as a mixture of two Gaussian distributions, the parameters of which the user supplies; that is, one specifies in advance the shape of the latent trait distribution, based on prior beliefs/knowledge. This program is much simpler to use than those described below. Several sets of data (summarized as a series of 2× tables) can be analyzed in a single run.

The [LTMA](#) program can similarly be used to estimate a generalized polychoric correlation, based on a latent trait mixture model (Uebersax & Grove, 1993). This is basically a fancier version of the [glc program](#): (1) it handles ordered-categorical as well as dichotomous variables, and (2) it will estimate the shape of the latent trait distribution from the data (again, modeling it as a mixture of two component Gaussians).

The [LLCA](#) program can be used to estimate a polychoric correlation with nonparametric distributional assumptions. The latent trait is represented as a sequence of latent classes on a single continuum (Uebersax 1993b). That is, the latent trait distribution is modeled as a "histogram," where the densities at each point are estimated, rather than as a continuous parametric distribution.

## Helpful Links

- ["Polychoric correlation" - Wikipedia](#)
  - [Correlation page \(Statnotes\) - by G. David Garson](#)
- 

## References

- Albert, J.H. Bayesian estimation of the polychoric correlation coefficient. *Journal of Computation and Simulation*, 1992, 44, 47-61.
- Bishop YMM, Fienberg SE, Holland PW. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Massachusetts: MIT Press, 1975
- Brown MB. Algorithm AS 116: the tetrachoric correlation and its standard error. *Applied Statistics*, 1977, 26, 343-351.
- Drasgow F. Polychoric and polyserial correlations. In Kotz L, Johnson NL (Eds.), *Encyclopedia of Statistical Sciences*. Vol. 7 (pp. 69-74). New York: Wiley, 1988.
- Edwards JH, Edwards AWF. Approximating the tetrachoric correlation coefficient. *Biometrics*, 1984, 40, 563.
- Harris B. Tetrachoric correlation coefficient. In Kotz L, Johnson NL (Eds.), *Encyclopedia of Statistical sciences*. Vol. 9 (pp. 223-225). New York: Wiley, 1988.
- Hutchinson TP. Kappa muddles together two sources of disagreement: tetrachoric correlation is better. *Research in Nursing and Health*, 1993, 16, 313-315.

Hutchinson TP. Assessing the health of plants: Simulation helps us understand observer disagreements. *Environmetrics*, 2000, 11, 305-314.

Joreskog KG, Sorbom, D. PRELIS User's Manual, Version 2. Chicago: Scientific Software, Inc., 1996.

Knol DL, ten Berge JMF. Least-squares approximation of an improper correlation matrix by a proper one. *Psychometrika*, 1989, 54, 53-61.

Loehlin JC. *Latent Variable Models*, 3rd ed. Lawrence Erlbaum, 1999.

Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 1979, 44(4), 443-460.

Pearson K. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London*, Series A, 1900, vol. 195, pp. 1-47.

Tallis GM. The maximum likelihood estimation of correlation from contingency tables. *Biometrics*, 1962, 342-353.

Uebersax JS. LLCA: Located latent class analysis. Computer program documentation, 1993a.

Uebersax JS. Statistical modeling of expert ratings on medical treatment appropriateness. *Journal of the American Statistical Association*, 1993b, 88, 421-427.

Uebersax JS. POLYCORR: A program for estimation of the standard and extended polychoric correlation coefficient. Computer program documentation, 2000.

Uebersax JS, Grove WM. A latent trait finite mixture model for the analysis of rating agreement. *Biometrics*, 1993, 49, 823-835.

---

**To cite this article:**

Uebersax JS. The tetrachoric and polychoric correlation coefficients. Statistical Methods for Rater Agreement. Web, accessed <day-month-year>.

---

[\(Top of Page\)](#)

Go to [Agreement Statistics](#) site

Go to [Latent Structure Analysis](#) site

*Last updated: 8 Sep 2015* (Detailed R instructions.)

---

(c) 2010–2015 [John Uebersax PhD](#) [email](#)



