

THE STATA JOURNAL

Editor

H. Joseph Newton
Department of Statistics
Texas A&M University
College Station, Texas 77843
979-845-8817; fax 979-845-6077
jnewton@stata-journal.com

Editor

Nicholas J. Cox
Department of Geography
Durham University
South Road
Durham City DH1 3LE UK
n.j.cox@stata-journal.com

Associate Editors

Christopher F. Baum
Boston College

Nathaniel Beck
New York University

Rino Bellocco
Karolinska Institutet, Sweden, and
Univ. degli Studi di Milano-Bicocca, Italy

Maarten L. Buis
Vrije Universiteit, Amsterdam

A. Colin Cameron
University of California–Davis

Mario A. Cleves
Univ. of Arkansas for Medical Sciences

William D. Dupont
Vanderbilt University

David Epstein
Columbia University

Allan Gregory
Queen's University

James Hardin
University of South Carolina

Ben Jann
ETH Zürich, Switzerland

Stephen Jenkins
University of Essex

Ulrich Kohler
WZB, Berlin

Frauke Kreuter
University of Maryland–College Park

Jens Lauritsen
Odense University Hospital

Stanley Lemeshow
Ohio State University

J. Scott Long
Indiana University

Thomas Lumley
University of Washington–Seattle

Roger Newson
Imperial College, London

Austin Nichols
Urban Institute, Washington DC

Marcello Pagano
Harvard School of Public Health

Sophia Rabe-Hesketh
University of California–Berkeley

J. Patrick Royston
MRC Clinical Trials Unit, London

Philip Ryan
University of Adelaide

Mark E. Schaffer
Heriot-Watt University, Edinburgh

Jeroen Weesie
Utrecht University

Nicholas J. G. Winter
University of Virginia

Jeffrey Wooldridge
Michigan State University

Stata Press Editorial Manager

Stata Press Copy Editors

Lisa Gilmore
Jennifer Neve and Deirdre Patterson

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the *Stata Journal*, including information for authors, see the web page

<http://www.stata-journal.com>

The *Stata Journal* is indexed and abstracted in the following:

- Science Citation Index Expanded (also known as SciSearch®)
- CompuMath Citation Index®

Copyright Statement: The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible web sites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press. Stata and Mata are registered trademarks of StataCorp LP.

Speaking Stata: Correlation with confidence, or Fisher's z revisited

Nicholas J. Cox
Department of Geography
Durham University
Durham City, UK
n.j.cox@durham.ac.uk

Abstract. Ronald Aylmer Fisher suggested transforming correlations by using the inverse hyperbolic tangent, or atanh function, a device often called Fisher's z transformation. This article reviews that function and its inverse, the hyperbolic tangent, or \tanh function, with discussions of their definitions and behavior, their use in statistical inference with correlations, and how to apply them in Stata. Examples show the use of Stata and Mata in calculator style. New commands `corrqi` and `corrqi` are also presented for correlation confidence intervals. The results of using bootstrapping to produce confidence intervals for correlations are also compared. Various historical comments are sprinkled throughout.

Keywords: pr0041, `corrqi`, `corrqi`, correlation, confidence intervals, Fisher's z , transformation, bootstrap, Mata

1 Introduction

A friend wanted to attach confidence intervals to his correlations. My first reaction was skeptical: why do you want to do that? Correlations, meaning Pearson or product-moment correlations, are not such a big deal in the first part of the twenty-first century as they were when Karl Pearson (1857–1936) introduced the modern way to estimate them. Good introductory texts now typically lead off on correlations as descriptive statistics, are firm in that you should always look at a scatterplot too, and point emphatically in the direction of regression as a way to quantify what the relationship is. No doubt many statisticians would want to go much further in downplaying correlations.

As an experiment, imagine a bizarre reform of physics teaching (say, Physics Lite or Physics for Dummies) in which basic laws such as Hooke's law or Ohm's law were introduced by statements like, "We find that there is a very close relationship between y and x ". The reaction would be immediate: these statements miss out on the key point, saying *what the relationship is*. Thus there is also a steadily increasing emphasis in statistical science on the importance of regression as trying to summarize what the relationship is.

Even granting all that, there can be good reason to put a magnifying glass on an estimated correlation. More frequently than regression texts imply, researchers may have a clutch of response variables all more or less on the same footing. Consider psychologists with bundles of measures of intelligence, biologists with measures of organism size, or

geomorphologists, such as my friend and me, with measures of landform size. When quantifying relationships between such variables, correlations are natural parameters. Yet again, correlations can seem to be the most interesting parameters whenever the units of measurement are arbitrary, but the strength of agreement between them is of central concern. Psychology (again), medicine, and ecology offer many examples. The point need not be labored strongly, because looking at correlation matrices is a standard starting point in multivariate analysis. An even stronger argument for correlations is that sometimes, as in genetics, they feature in theoretical treatments, so that comparing predicted and actual correlations is then required.

My second reaction was to review how to find such confidence intervals in Stata. Official Stata commands stop short. `correlate` does no inference, while `pwcorr` offers only a p -value from the standard t test of whether the population correlation ρ is really zero. See [R] `correlate` or the online help for more details. The latter t test cannot be converted into useful confidence intervals, because once the population correlation is really not zero, the sampling distribution of estimates r is substantially skewed, even for large sample sizes. That should not seem surprising, given that correlations are constrained to fall within $[-1, 1]$. Getting at the form of the distribution is a messy problem with, for the most part, only complicated exact solutions, but for practical data analysts an attractive way forward is offered by a suggestion of Ronald Aylmer Fisher (1890–1962): even though r usually has an awkward skewed distribution, the inverse hyperbolic tangent, or atanh , of r is much better behaved. This transform, within statistics often labeled Fisher's z , is said to be normally distributed to a good approximation. The basic tactic is now evident: apply standard normal-based technique on the z scale, and then back-transform using its inverse transform, the hyperbolic tangent, or \tanh .

Even though official Stata stops short, user programmers have been there before. Gleason (1996) and independently Seed (2001, 2004) have published Stata commands for correlation confidence intervals using Fisher's z transform. (See also Goldstein [1996] on a related problem not treated here.) Two by-products of writing this column are the new commands `corrcci` and `corrcti`, which are more up to date in Stata style and, in most respects, more versatile than their predecessors. The new commands will be discussed later in the column, but they are not its main theme. The problem of correlation confidence intervals is an example of a problem that Stata users frequently face: how far can you get with as little effort as possible in implementing something just beyond official Stata's functionality? This theme will itself be explored by showing how Stata, including Mata, can be used as a calculator for problems for which there are, so far, no canned commands.

My third reaction was, simply, to bootstrap. Correlations lend themselves well to bootstrapping and feature strongly as examples in the instant classic text of Efron and Tibshirani (1993). As always, bootstrapping requires care and caution even for a relatively simple statistic such as the Pearson correlation. A key question is clearly whether Fisher's z and bootstrapping yield similar answers.

2 Correlations from Pearson to Fisher and beyond

The idea of correlation is protean. [Rodgers and Nicewander \(1988\)](#) detailed 13 ways to look at the correlation coefficient, [Rovine and Von Eye \(1997\)](#) added another, and no doubt yet other perspectives could be added. Two of these perspectives appear especially important. Karl Pearson, his predecessors, and contemporaries almost all focused on the correlation as a key parameter in a bivariate normal distribution. Statistically minded people a century later are more likely to focus on the correlation as a measure of linear association, for which normality, whether joint or marginal, is not required. (In fact, George Udny Yule, Pearson's assistant at the time, took up the linear association viewpoint right away.)

Either way, the boundedness of correlations within $[-1, 1]$ eases interpretations but complicates formal inference. We learn early that correlations of 1 and -1 arise from exact linear relationships, and we can use that fact in thinking about observed correlations in between. But the bounds impart skewness to the sampling distribution of correlations whenever the population correlation is not zero. (Even when it is zero, the bounds must apply to the sampling distribution as well, but that is, in practice, of little concern.)

Common terminology (Pearson's r , Fisher's z) credits Pearson with the method for calculating correlation that is now standard and Fisher with the transformation commonly used to produce confidence intervals. The actual history was, not surprisingly, more complicated. See [Stigler \(1986, 1999\)](#) and [Hald \(1998, 2007\)](#) for much more detail. The idea of a bivariate normal distribution goes back to Lagrange, Laplace, and Gauss. What Pearson did ([Pearson \[1896\]](#) and many other articles) included putting the calculation of correlation on a systematic basis and promoting its application and extension, but the underlying idea of correlation was very much the earlier child of Francis Galton. Fisher's work on inference with correlation was the culmination of a series of articles wrestling with the problem, including other contributions by Edgeworth, Yule, Filon, Sheppard, Student, and Soper. The assumption of bivariate normality was central in that thread. What happens when some other distribution applies received little attention until more recently. Rank correlation methods solve that problem by replacing it with another. It deserves more attention than we will give it directly, except that bootstrapping takes most of the sting out of the question.

The first mention of the atanh transformation in Fisher's work was as a closing aside in his first article on correlation ([Fisher 1915](#)). Then he made use of the transformation in an article on the genetics of twins ([Fisher 1919](#)). The formal development of the idea came later in a longer statistical article ([Fisher 1921](#)). More important than any of these for the spread of the transformation was its discussion within the text *Statistical Methods for Research Workers* ([Fisher 1925](#)) and later editions ([Fisher 1970, 1990](#)). Fisher had the chutzpah not only to invent new methods but promptly to include them as standard within his own textbook. That text was the ancestor of many other textbook accounts (e.g., [Snedecor and Cochran \[1989\]](#); [Altman \[1991\]](#); [Zar \[1999\]](#); [Bland \[2000\]](#); van Belle et al. [[2004](#)]).

3 The art of hyperbole

If atanh is an old friend to you, skim your way through to the next section, unless one of your concerns is how to explain all this to colleagues or students.

atanh is the notation used here for the inverse hyperbolic tangent function. A good reason for that notation is that Stata (from Stata 8) and Mata (from Stata 9) both have `atanh()` functions, as indeed do many other mathematical and statistical software packages, including R and Microsoft Excel. More pedantic comments on names and notation will follow in due course, but first I will explain how the atanh function arises from the \tanh function. That, in turn, requires an introduction of the \sinh and \cosh functions.

3.1 Hyperbolic \sinh and \cosh functions

To understand atanh , it helps to know a little about hyperbolic functions in general. Good accounts in varying styles are available within many introductions to calculus (e.g., [Abbott \[1940\]](#); [Hardy \[1952\]](#); [Gullberg \[1997\]](#); [Banner \[2007\]](#)). The account here presumes familiarity with basic ideas on logarithmic, exponential, and trigonometric functions. For more under those headings, the books just referenced are fine possibilities. I set aside discussion of these functions for complex arguments and instead focus on the special case of real arguments, all that is needed for our purpose.

Although they appeared earlier in various forms, the hyperbolic functions were first examined systematically by Johann Heinrich Lambert (1728–1777) and Vincenzo Riccati (1707–1775). [Barnett \(2004\)](#) gives a splendid historical survey. Among several other contributions to mathematics and science, Lambert proved that π is irrational and conjectured that e is transcendental. More statistically, he was an early proponent of what we now call maximum likelihood ([Hald 1998](#)). Riccati came from a talented mathematical family: Riccati equations are named for his father Jacopo or Giacomo.

The three main hyperbolic functions are called the hyperbolic sine, cosine, and tangent, or \sinh , \cosh , and \tanh . The names and notations should suggest a strong analogy with the trigonometric \sin , \cos , and \tan functions: the “h” naturally stands for hyperbolic, to be explained in a moment.

Stata itself has no `sinh()` or `cosh()` function, although Mata does, but this matters not at all. \sinh and \cosh can be calculated in Stata through their definitions:

$$\sinh x = \frac{\exp(x) - \exp(-x)}{2}$$

and

$$\cosh x = \frac{\exp(x) + \exp(-x)}{2}$$

Here and usually elsewhere, I spell out the exponential function e^x as $\exp(x)$, partly to underline subliminally that `exp()` is the Stata function name to use in computations. The choice between, say, $\sinh x$ or $\sinh(x)$ as notation is purely a matter of convenience; no nuance is implied by the difference. See what those definitions imply by considering the exponential functions in terms of infinite series:

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

and

$$\exp(-x) = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots$$

Then \sinh is the sum of terms in odd powers of x and \cosh is the sum of terms in even powers of x :

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$$

and

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$$

Thus \sinh is an odd function, such that $\sinh x = -\sinh(-x)$, and \cosh is an even function, such that $\cosh x = \cosh(-x)$. It is easy to draw the function graphs on the fly in Stata (Cox 2004). Plotting for a short interval around $x = 0$ suffices to give a good idea of overall behavior, because both functions explode toward one or the other infinity as x becomes arbitrarily large in magnitude; see figure 1.

```
. twoway function (exp(x) - exp(-x)) / 2, ra(-3 3) || function (exp(x) +  
> exp(-x)) / 2, ra(3 -3) lp(dash) legend(order(1 "sinh" 2 "cosh")) xla(-3(1)3)  
> ytitle("") yla(, ang(h))
```

(Continued on next page)

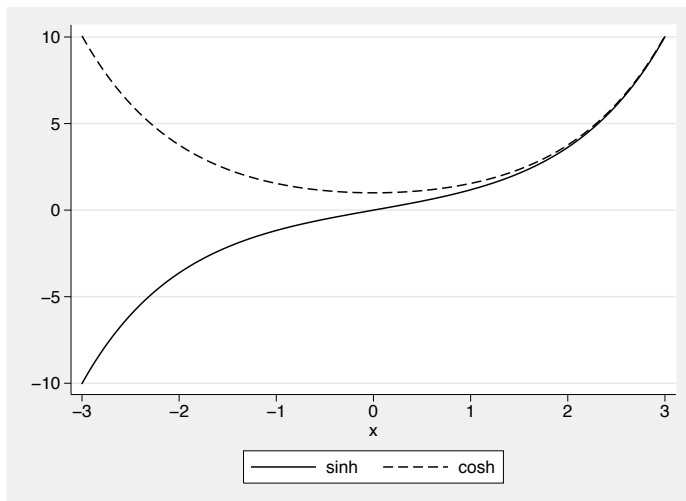


Figure 1. $\sinh x$ and $\cosh x$ for x near 0

The similarity—and equally crucial, the difference—of sine and cosine can be seen by comparing their series representations:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

and

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

The term hyperbolic is explained by the geometry implied by the hyperbolic functions. Among many beautiful identities satisfied by \sinh and \cosh is

$$\cosh^2 x - \sinh^2 x = 1$$

which defines a hyperbola with the locus $(\cosh x, \sinh x)$. Compare and contrast the key identity for trigonometric functions,

$$\cos^2 x + \sin^2 x = 1$$

which defines a circle. [Barnett \(2004\)](#) expands on the geometric motivation for the original work on hyperbolic functions. [Maor \(1994\)](#) gives further historical and geometrical orientation.

3.2 tanh and atanh

Neither \sinh nor \cosh is needed directly for our purpose, but we do need their ratio, \tanh , that is, $\tanh = \sinh / \cosh$. That definition should remind you of a definition of tangent as the ratio of sine and cosine. From the definitions, various forms are possible:

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} = \frac{\exp(2x) - 1}{\exp(2x) + 1}$$

From figure 1, we could guess that \tanh is bounded by 1 and -1 , and this guess is supported by a direct graph; see figure 2.

```
. twoway function tanh(x), ra(-3 3) xla(-3(1)3) ytitle(tanh x) yla(, ang(h))
```

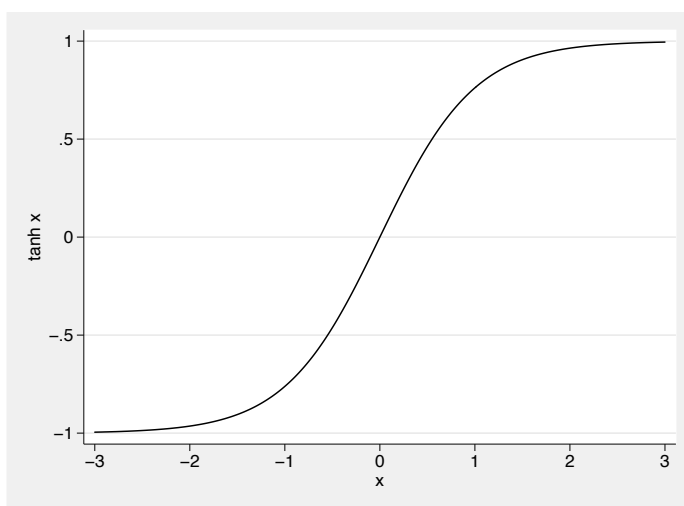


Figure 2. $\tanh x$ for x near 0

As positive x becomes very large, $\exp(-x)$ approaches 0 and \tanh approaches 1. Similarly, as negative x becomes very large, $\exp(x)$ approaches 0 and \tanh approaches -1 . So \tanh maps any finite argument to the interval $(-1, 1)$.

\tanh is single valued, and so its inverse function, atanh , is easy to define. If $r = \tanh x$, then $x = \operatorname{atanh} r$. Here the notation r , although not customary in calculus, honors the focus in this column on correlations. atanh is important enough in the story to deserve a separate graph; see figure 3.

```
. twoway function atanh(x), ra(-1 1) || function x, lp(dash) ra(-1 1)
> ytitle(atanh r) yla(, ang(h)) xtitle(r) legend(order(1 "atanh" 2 "equality"))
```

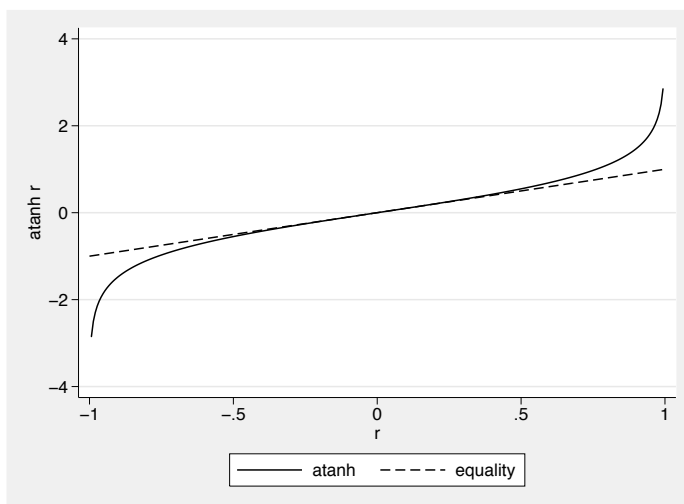


Figure 3. $\operatorname{atanh} r$; note how $\operatorname{atanh} r \approx r$ for r near 0

Thus \tanh is an odd function that maps the interval $(-1, 1)$ to $(-\infty, \infty)$. The reference line shows how atanh leaves values in the middle little changed but bites strongly in the tails near ± 1 . An explicit definition is given by

$$x = \operatorname{atanh} r = \frac{1}{2} \ln \left\{ \frac{(1+r)}{(1-r)} \right\} = \frac{1}{2} \{ \ln(1+r) - \ln(1-r) \}$$

If you want to see that derived, set $r = (e^{2x} - 1)/(e^{2x} + 1)$, so $1+r = 2e^{2x}/(e^{2x} + 1)$ and $1-r = 2/(e^{2x} + 1)$. Then $e^{2x} = (1+r)/(1-r)$ and $2x = \ln\{(1+r)/(1-r)\}$.

atanh also has a very neat series expansion:

$$\operatorname{atanh} r = r + \frac{r^3}{3} + \frac{r^5}{5} + \cdots$$

From that we can tell that for $r \approx 0$, $\tanh r \approx r$, which matches what is seen in figure 3.

3.3 An aside on names and notations

Other names and notations for the inverse hyperbolic tangent are commonly used. Arc-hyperbolic tangent is a common name. $\operatorname{arctanh}$, $\operatorname{argtanh}$, artanh , arth , and \tanh^{-1} are other notations I have seen, with or without extra spacing (e.g., $\operatorname{argtanh}$).

An objection to the name arc-hyperbolic tangent—or the notation $\operatorname{arctanh}$ —is that use of the word “arc” is based on an improper analogy with terminology for inverse

trigonometric functions, such as arctangent. In the latter case, “arc” means the arc of a circle, or equivalently angle, as is proper for circular functions. But the inverse of a hyperbolic function is geometrically interpreted not as an arc but as an area. This point is argued in detail by [Saler \(1985\)](#). Various minor misunderstandings in this territory may arise partly from the linguistic coincidence that arc, area, and argument have the same initial characters.

Some people object to notations such as \tanh^{-1} because of possible ambiguity between inverse and reciprocal.

It is fortunate that using \tanh and atanh as mathematical function names, partly because they are used as Stata function names, also avoids using a name and some notations that are widely regarded as objectionable.

3.4 Mapping back and forth

The key point for us is simple: atanh and \tanh offer ways of mapping back and forth between the interval $(-1, 1)$ and the entire real line. There are other pairs of functions that do that, but it turns out that this pair is especially suitable for dealing with correlations, which usually fall in that interval. We can work as we wish in the unconfined, but equivalent, space of the real line after using that transformation and then transform back again.

Notice the small print, and the *usually*. The limiting values -1 and 1 are excluded. This exclusion does not bite in practice. If you have a correlation with a magnitude of 1 , then you have made a mistake over which variables to correlate, or you are testing a program, or you do have an exact relationship and are presumably unconcerned with attaching a confidence interval.

Some statistical accounts of the atanh transformation cut out the background of hyperbolic functions and present it directly as $(1/2)\ln\{(1+r)/(1-r)\}$. That is charitable to any readership with limited calculus but wrenches the function out of its proper context. It was helpful to practitioners over several decades in which logarithmic tables were more widely accessible than hyperbolic function tables. Either way, we can see by looking at the definition that the function has the behavior of stretching the tails of correlation distributions indefinitely. As r approaches 1 , the result becomes arbitrarily large and positive, while as r approaches -1 , it becomes arbitrarily large and negative.

Here is yet another benefit in understanding. Put $r = 2x - 1$, or $x = (r + 1)/2$. Hence as r varies between -1 and 1 , so x varies between 0 and 1 , and

$$\operatorname{atanh} r = \frac{1}{2} \ln \left\{ \frac{2x}{2(1-x)} \right\} = \frac{1}{2} \operatorname{logit} x$$

So we see that atanh and logit are just sisters under the skin.

All that said, once you know that you can use `tanh()` and `atanh()` in Stata or Mata, you need never type out any definitions in terms of `exp()` or `ln()`.

4 Precise recipes for Fisher's z

Suppose we have an estimate of the Pearson correlation, r , for a sample of size n and seek confidence intervals for the population correlation, ρ . Fisher recommended working with $\operatorname{atanh} r$. A first approximation from his work is that $z = \operatorname{atanh} r$ is normally distributed with a mean of $\operatorname{atanh} \rho$ and a variance of $1/(n-3)$. A closer approximation is that z is normal with a mean of $\operatorname{atanh} \rho + 2\rho/(n-1)$ and the same variance. So atanh is a variance-stabilizing and, indeed, normalizing transformation.

A direct derivation of atanh as a variance-stabilizing transformation follows from the observation, attributed to [Pearson and Filon \(1898\)](#), that the variance of the sampling distribution of r varies as $(1-\rho^2)^2$. See, for example, [Bartlett \(1947\)](#), [Hotelling \(1953\)](#), or [Hoyle \(1973\)](#). The pivot of the argument is that $\int dr/(1-r^2)$ is $\operatorname{atanh} r$, as tabulated in any substantial mathematical reference work. In short, there is a direct route from the Pearson and Filon result to atanh , although it does not feature in Fisher's writings.

Variance-stabilizing does not necessarily bring normalizing in its own wake, although roughly that is often true in practice and is being claimed here. However, [Winterbottom \(1979\)](#) showed that choosing the transformation that best produces approximate normality also yields atanh . Thus atanh exhibits both of these two distinct virtues, at least in principle.

The bias on the z scale, $2\rho/(n-1)$, clearly increases with ρ , which we naturally estimate by r , but decreases with n ; several authors rely on the latter fact, at least implicitly, and neglect it. It is easy to include a correction for this bias in calculation, say, as an option in a program. Fisher himself ignored the bias when he judged it unimportant, as will be seen by reading [Fisher \(1925, sec. 35 and 36\)](#) or any later edition. The effect of the bias correction is to impart caution by shrinking estimates of correlation toward zero.

An independent suggestion was made by [Jeffreys \(1939, 139–144; 1961, 174–179\)](#), who put forth the approximation that z is normal with a mean of $\operatorname{atanh} \rho - 5\rho/2n$ and a variance of $1/n$. [Box and Tiao \(1973, 465–470\)](#) give a detailed derivation in a different style. [Lee \(2004, 159–164\)](#) offers a different and more complicated approximation for the variance. These authors all write in a Bayesian spirit, but their results are comparable with those discussed earlier, so any difference of philosophy need not detain us.

Again it is easy to program [Jeffreys'](#) procedure, if only as a matter of curiosity to see how it compares. As with Fisher's correction, the effect is to shrink estimates slightly toward zero, except a little more so. [Jeffreys'](#) intervals are typically a little narrower than Fisher's, although it takes a very small sample size to make this really obvious.

Yet further possibilities can be found in the literature. [Hotelling \(1953\)](#) published a substantial survey of the correlation problem, including small corrections to Fisher's algebra and a detailed exploration of alternative approaches. None is especially simple and none seems to have had much impact on statistical practice (see [Sokal and Rohlf \[1995\]](#) for an exception).

All these procedures are based on 1) bivariate normality, 2) large sample sizes, and 3) independence of observations. One, or two, or even all three of these assumptions are left implicit by many authors. Equally, one, or two, or even all three may not be satisfied, even roughly, by actual data. The bootstrap offers salvation in respect to the first two, but emphatically not the last. Different authors give different signals on the consequences of nonnormality. [Duncan and Layard \(1973\)](#) reported simulation results for various distributions, showing that Fisher's z can give some very poor results. Evidently, caution is imperative.

We can be precise about very small sample sizes. Only for $n > 4$ is the sampling distribution even unimodal, so the problem for smaller sample sizes is not even qualitatively similar to that of normalizing (or even symmetrizing) a skewed but unimodal sampling distribution. See [Cramér \(1946, 398–399\)](#) or [Stuart and Ord \(1994, 564\)](#) for careful explanations.

5 Applying Fisher's z in Stata

New commands `corrcci` and `corrcti` have already been promised, but let us first look at how you could do it for yourself from what has already been explained. [Fisher \(1925, sec. 34 and 35\)](#) quotes an analysis of wheat yield and autumn rainfall for eastern England for the 20 years 1885–1904, which yielded a correlation of -0.629 . We will revisit that analysis. Our main tool, other than inbuilt functions, will be the `display` command. Even many experienced Stata users underestimate the scope for using `display` as a quick calculator ([Ryan 2004](#)). See [P] `display` or the online help for more details.

Two standard Stata features prove their worth for this kind of work. You can keep a log of a sequence of calculations that is completely explicit about what was done. Further, you can retrieve previous commands and use them as a template for similar calculations.

A 95% confidence interval based on a normal sampling distribution would use `invnormal(0.975)` as a multiplier. You may already know that 1.96 is very close:

```
. di invnormal(0.975)
1.959964
```

Without using a bias correction, we get $(-0.838, -0.258)$ as the interval:

```
. di tanh(atanh(-0.629) + invnormal(0.975)/sqrt(17))
-.25840515
. di tanh(atanh(-0.629) - invnormal(0.975)/sqrt(17))
-.83820901
```

The interval is quite broad, as is reasonable for a fairly small sample size. It does at least exclude zero. Note also how the interval is asymmetric around -0.629 . With Fisher's bias correction, the interval is shunted toward zero and a little changed in shape:

```
. di tanh(atanh(-0.629) - (2 * -0.629)/19 + invnormal(0.975)/sqrt(17))
-.19563344
. di tanh(atanh(-0.629) - (2 * -0.629)/19 - invnormal(0.975)/sqrt(17))
-.81739277
```

Jeffreys' procedure is most notable here for giving a narrower interval, because the difference between $1/n$ and $1/(n-3)$ registers as n becomes small.

```
. di tanh(atanh(-0.629) - (5 * -0.629)/(2 * 20) + invnormal(0.975)/sqrt(20))
-.21925514
. di tanh(atanh(-0.629) - (5 * -0.629)/(2 * 20) - invnormal(0.975)/sqrt(20))
-.80028197
```

We could also do this in Mata. That lets us get both limits at the same time:

```
. mata : tanh(atanh(-0.629) :+ invnormal(.975) * (-1,1)/sqrt(17))
               1               2
1  -0.8382090146  -0.258405146

. mata : tanh(atanh(-0.629) - ((5 * -0.629)/40) :+ invnormal(.975) *
> (-1,1)/sqrt(20))
               1               2
1  -0.8002819748  -0.2192551437
```

The little thing that bites most commonly, unless you have internalized Mata's conformability rules, is that you may need to spell out that you want an elementwise operator, indicated by a colon prefix (e.g., `:+`).

Let us look at some other standard problems by using Stata or Mata calculator-style. From [van Belle et al. \(2004, 295–297, 322\)](#), data on the adenosine triphosphate (ATP) levels of $n = 17$ pairs of oldest and youngest sons yield $r = 0.5974$. Genetics theory predicts $\rho = 0.5$.

A standard normal deviate for a hypothesis test is thus

```
. di (atanh(0.5974) - atanh(0.5)) * sqrt(14)
.52304027
```

Note the use of $n - 3 = 14$ in calculating the standard error. A formal calculation of a p -value is hardly needed, except for publication. If it were, the result just needs to be wrapped by `normal()`.

From [Snedecor and Cochran \(1989, 188–189\)](#), two methods are being compared. For the first method, two independent measurements have a correlation of $r_1 = 0.862$ for $n_1 = 60$. For the second, we have $r_2 = 0.720$, $n_2 = 49$. Is there a difference in reliability, as measured by the correlation? Once correlations are mapped to the `atanh` scale, this is a standard normal-theory test.

```
. di (atanh(0.862) - atanh(0.720))/sqrt((1/57) + (1/46))
1.9850285
```

The numerator is a difference of z values. The denominator is the root of the sum of the variances. We will get a p -value too:

```
. di 2 * normal(-(atanh(0.862) - atanh(0.720))/sqrt((1/57) + (1/46)))
.0471413
```

Note the trick of negating the positive argument of `normal()` to get a result in the left tail of the distribution. The result qualifies as significant at the 5% level in a two-tailed test. You can get the corresponding one-tailed test by omitting the factor of 2.

From [Sokal and Rohlf \(1995, 580–582\)](#), the correlation between length of wing and width of band on wing in females of the butterfly *Heliconius charitonius* was measured in 10 populations (biological sense) from different places in the Caribbean. A quick Google search elicits a correction of name to *Heliconius charithonia*, with the English name zebra longwing, and some very pretty photographs. First, we test the null hypothesis of homogeneity, implying that the several values of r estimate the same ρ . Then, if we fail to reject that, we combine the estimates to get the corresponding overall estimate. (The presumption is that the original data are not available.)

One tactic is to put the sample correlations and sizes into vectors in Mata. Within Mata, we put the sample sizes in one vector and the correlations in another, which we promptly transform.

```
: n = (100, 46, 28, 74, 33, 27, 52, 26, 20, 17)
: z = atanh((0.29, 0.70, 0.58, 0.56, 0.55, 0.67, 0.65, 0.61, 0.64, 0.56))
```

Let j index samples. We need a weighted mean, \bar{z} , of the z_j using the reciprocals of the separate variances $1/(n_j - 3)$ as weights.

```
: zmean = sum((n :- 3) :* z)/sum(n :- 3)
```

Then the test statistic is a chi-squared statistic, $\sum_j \{(z_j - \bar{z})/\text{se}_j\}^2$, for the standard error $\text{se}_j = \sqrt{1/(n_j - 3)}$. There are $10 - 1 = 9$ degrees of freedom.

```
: chisq = sum((n :- 3) :* (z :- zmean):^2)
: chisq
15.26351671
: chi2tail(9, chisq)
.0839473192
```

With this p -value, 0.084, we fail to reject a null hypothesis of homogeneity at conventional levels, and so move on to get an overall estimate:

```
: tanh(zmean)
.547824694
```

The overall correlation we take as 0.548. Data snoopers might wonder about the deviant correlation of 0.29 and want to know if there is some story to tell. It is based on the largest sample size, which is worrying.

6 Comparison with the bootstrap

6.1 Fisher's (?) wheat yield and rainfall data

Fisher (1925, 1970) did not give the data for wheat yield and autumn rainfall for eastern England for the 20 years 1885–1904 for which he gave a correlation of -0.629 . Some detective work identifies the likely data source as Shaw (1905, 286). A data file, `shaw.dta`, is included in the electronic media for this column.

The first step, as always, is a scatterplot (figure 4).

```
. use shaw
. scatter yield rainfall, yla(, ang(h))
```

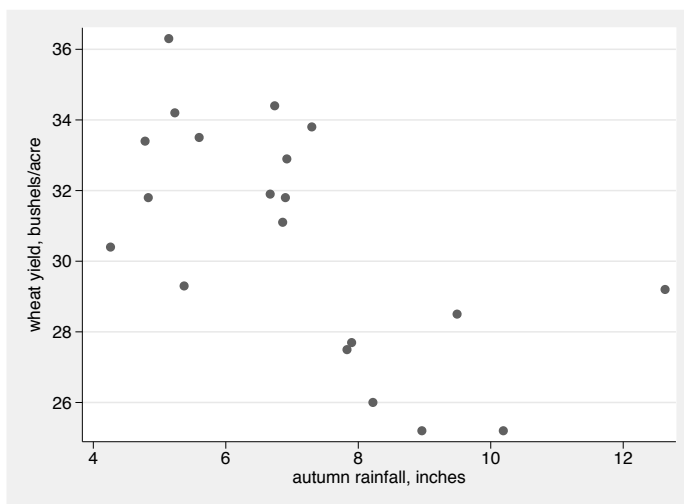


Figure 4. Scatterplot of wheat yield and autumn rainfall for eastern England 1885–1904; data from Shaw (1905) are presumably those used by Fisher (1925, 1970)

The units of measurement may well seem bizarre. They naturally wash out when we get a correlation, but for those curious, the bushel is a unit of volume here and about 36.37 liters; the acre is a unit of area and about 0.47 hectare; and the inch is a unit of length and is 25.4 mm.

Naturally, these data are time series, so you should be wondering about trend and serial correlation. Fisher ignored those questions, but they can be examined and indicate no strong patterns in either case. Successive years appear close to independent.

Why is high autumn rainfall bad for wheat yield? The main story appears to be that the more rainfall, the more nutrients are washed out of reach of the growing crop.

We move straight to a Pearson correlation with alternative confidence intervals using the new command `corrqi`. A formal specification follows in the next section.

```
. corrqi yield rainfall
(obs=20)

               correlation and 95% limits
yield    rainfall    -0.616   -0.832   -0.238
. corrqi yield rainfall, fisher
(obs=20)

               correlation and 95% limits
yield    rainfall    -0.616   -0.811   -0.176
. corrqi yield rainfall, jeffreys
(obs=20)

               correlation and 95% limits
yield    rainfall    -0.616   -0.793   -0.201
```

The first small surprise is that we get -0.616 . Either we are wrong somewhere or Fisher himself made a small mistake. (He would have achieved so much more had he used Stata.) The intervals are all quite wide, as befits a small sample size, but they do exclude zero.

Because we have the data, we can examine whether there is normality. Normal probability plots produced by using `qnorm` show a little skewness in the case of rainfall, but nothing that seems pathological (figure 5).

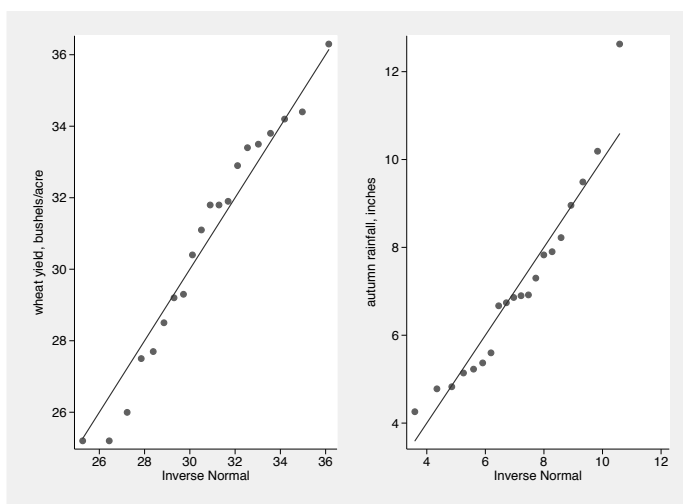


Figure 5. Normal probability plots for wheat yield (left) and rainfall (right)

`summarize` gives a moment-based skewness for rainfall of 0.901 and a kurtosis of 3.567. A test attributed to [Doornik and Hansen \(1994\)](#) for bivariate normality, implemented in the Stata command `omninorm` by Christopher F. Baum and myself and downloadable by using `ssc`, yields a marginal p -value of 0.0631. So, the data are not in perfect condition for Fisher's z , but they could be much worse.

Let's bootstrap. Using `correlate` directly will always be faster than using `corrcci`, which carries interpretive overhead and does extra work.

```
. bootstrap correlation = r(rho), nodots nowarn reps(10000) seed(1921)
> saving(shaw_r, replace): correlate yield rainfall

Bootstrap results                                Number of obs      =          20
                                                Replications       =       10000
```

```
command: correlate yield rainfall
correlation: r(rho)
```

| | Observed Coef. | Bootstrap Std. Err. | z | P> z | Normal-based [95% Conf. Interval] | |
|-------------|-------------------|------------------------|-------|-------|--------------------------------------|-----------|
| correlation | -.6160026 | .1112054 | -5.54 | 0.000 | -.8339612 | -.3980439 |

```
. estat bootstrap, all
```

```
Bootstrap results                                Number of obs      =          20
                                                Replications       =       10000
```

```
command: correlate yield rainfall
correlation: r(rho)
```

| | Observed Coef. | Bias | Bootstrap Std. Err. | [95% Conf. Interval] | | |
|-------------|-------------------|-----------|------------------------|----------------------|-----------|------|
| correlation | -.61600258 | -.0130758 | .11120543 | -.8339612 | -.3980439 | (N) |
| | | | | -.8440725 | -.4089272 | (P) |
| | | | | -.8218457 | -.3796614 | (BC) |

```
(N) normal confidence interval
(P) percentile confidence interval
(BC) bias-corrected confidence interval
```

The agreement of Fisher-type intervals and bootstrap intervals is not good. The Fisher-type intervals are wider and closer to zero than the bootstrap intervals.

Let's look at the sampling distribution:

```
. use shaw_r, replace
. gen z = atanh(corr)
```

The density traces (figure 6) and normal probability plots (figure 7) indicate that the standard advertisement oversells Fisher's z here. The bootstrap sampling distribution of r is not expected to be close to normal, but that of z is not really better. So although it is now easy to calculate several different confidence intervals, it is not clear for this example whether any is really well based. Those who thought that this was predictable for $n = 20$ can be especially comfortable with the interpretation that they were right and R. A. Fisher wrong here.

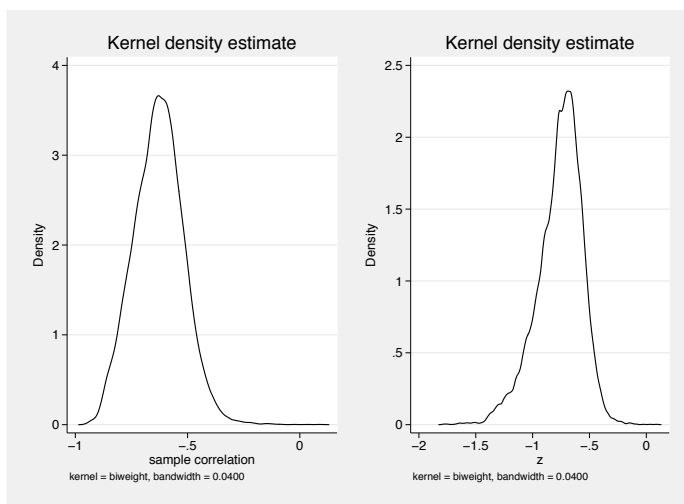


Figure 6. Kernel density estimate of the sampling distribution of the sample correlation (left) and its z transform (right)

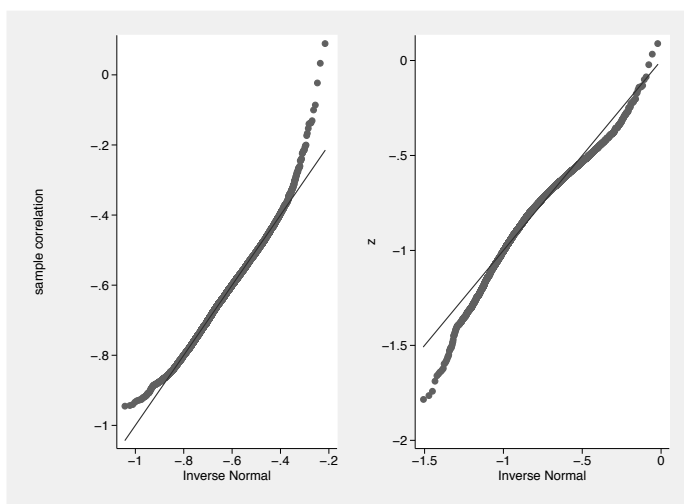


Figure 7. Normal probability plots of the sampling distribution of the sample correlation (left) and its z transform (right)

(Continued on next page)

6.2 Efron's law school data

Bradley Efron (born 1938) never met R. A. Fisher (Efron 1998, 96). But just as we can use Efron's bootstrap on Fisher's datasets, we can also apply Fisher-type intervals to one of Efron's datasets.

Efron and Tibshirani (1993, 21) give data for average LSAT scores (a national law test) and undergraduate grade-point averages (GPAs) for the entering classes of 82 American law schools in 1973. A data file, `law_school.dta`, is included in the electronic media for this column.

```
. use law_school, clear
```

A scatterplot shows a well-behaved relationship (figure 8). There is one moderate outlier, but it appears wholly consistent with the main trend.

```
. scatter GPA LSAT, yla(2.6(0.2)3.4, ang(h))
```

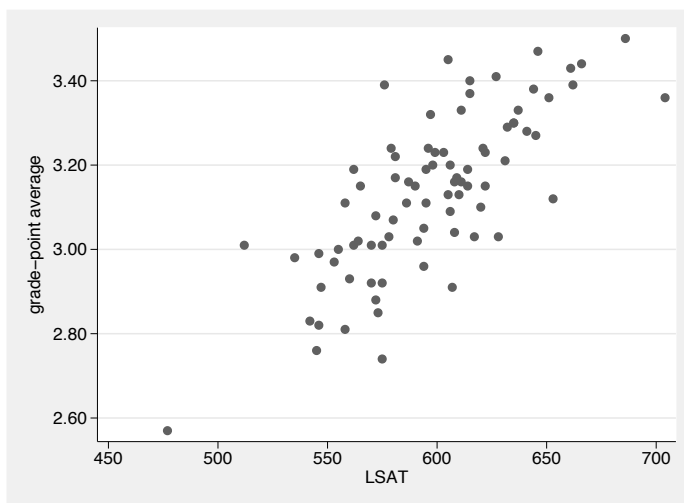


Figure 8. Relationship between GPA and LSAT for Efron and Tibshirani's law school data

Univariate and bivariate analyses not given here support the impression of good behavior.

```

. corrci GPA LSAT
(obs=82)
      correlation and 95% limits
GPA  LSAT      0.760    0.650    0.839
. corrci GPA LSAT, fisher
(obs=82)
      correlation and 95% limits
GPA  LSAT      0.760    0.639    0.833
. corrci GPA LSAT, jeffreys
(obs=82)
      correlation and 95% limits
GPA  LSAT      0.760    0.639    0.830
. bootstrap correlation = r(rho), nodots nowarn reps(10000) seed(1921)
> saving(law_school_r, replace): correlate GPA LSAT
Bootstrap results
                                     Number of obs      =      82
                                     Replications        =    10000

      command:  correlate GPA LSAT
      correlation:  r(rho)

```

| | Observed Coef. | Bootstrap Std. Err. | z | P> z | Normal-based [95% Conf. Interval] | |
|-------------|-------------------|------------------------|-------|-------|--------------------------------------|----------|
| correlation | .7599978 | .0517114 | 14.70 | 0.000 | .6586453 | .8613503 |

```

. estat bootstrap, all
Bootstrap results
                                     Number of obs      =      82
                                     Replications        =    10000

      command:  correlate GPA LSAT
      correlation:  r(rho)

```

| | Observed Coef. | Bias | Bootstrap Std. Err. | [95% Conf. Interval] | | |
|-------------|-------------------|----------|------------------------|----------------------|----------|------|
| correlation | .75999782 | -.003048 | .0517114 | .6586453 | .8613503 | (N) |
| | | | | .6465373 | .8484399 | (P) |
| | | | | .6456656 | .8478602 | (BC) |

```

(N)   normal confidence interval
(P)   percentile confidence interval
(BC)  bias-corrected confidence interval

```

This time, the intervals do appear to be marching almost to the same tune. The atanh-based intervals are closer to zero.

The results can be reproduced and analyzed in detail by any interested reader. (Note the practice of setting the seed, here 1921.) The final graph can convey a positive message. The atanh transformation does a nice job with the sampling distribution of correlations (figure 9).

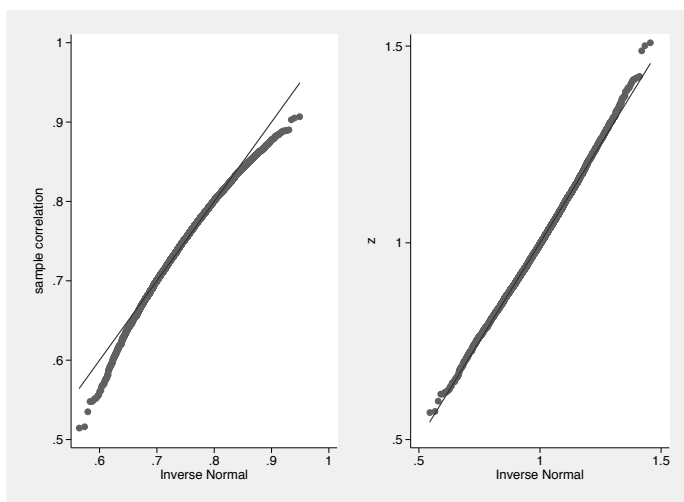


Figure 9. Normal probability plots of the sampling distribution of the sample correlation (left) and its z transform (right)

7 Commands `corrqi` and `corrqi`

A formal specification of the new commands `corrqi` and `corrqi` follows. Compared with the commands of [Gleason \(1996\)](#) and [Seed \(2001, 2004\)](#), the main differences are the `fisher` and `jeffreys` options for Fisher's bias correction and Jeffreys' procedure, and more flexibility in output as provided by the `format()`, `abbrev()`, and `saving()` options. Indeed, for some circumstances, `corrqi` might prove more flexible and amenable in output than `correlate` itself.

[Gleason \(1996\)](#) gives a command, `z_rplt`, for an original graphical display, which I have not tried to emulate. The fact that `corrqi` allows output to be saved to a file should greatly ease preparation of users' own alternative displays. Several pertinent ideas were covered in my last column ([Cox 2008](#)). [Seed \(2001, 2004\)](#) also addresses Spearman rank correlation, which I have not considered here.

7.1 corrci

Syntax

```
corrci varlist [if] [in] [weight] [, level(#) matrix format(%fmt)
      abbrev(#) [fisher|jeffreys] saving(filename [, postfile_options])]
```

varlist may contain time-series operators; see [U] 11.4.3 Time-series varlists.

aweights and *fweights* are allowed; see [U] 11.1.6 weight.

Description

corrci calculates Pearson correlations for two or more numeric variables specified in *varlist* together with confidence intervals calculated by using Fisher's z transform.

For sample size n , correlation r , and confidence level *level*, the default procedure sets $d = \text{invnormal}(0.5 + \text{level}/200)/\text{sqrt}(n - 3)$ and $z = \text{atanh}(r)$ and then calculates limits as $\tanh(z - d)$ and $\tanh(z + d)$.

corrci handles missing values through listwise deletion, meaning that the entire observation is omitted from the estimation sample if any of the variables in *varlist* is missing for that observation.

Remarks

Although *varlist* may contain time-series operators, no allowance is made for correlation structure or trend or any other time-series aspect in calculation, so interpretation is the user's responsibility.

corrci requires at least 5 nonmissing observations in *varlist*. With very small samples—or even very large ones—interpretation remains the user's responsibility.

Options

level(#) specifies the confidence level, as a percentage, for confidence intervals. The default is **level**(95) or whatever is set by **set level**; see [U] 20.7 Specifying the width of confidence intervals.

matrix specifies that correlations and confidence intervals be displayed as matrices. The default is to display them in a list. Whether **matrix** is specified has no effect on the saving of results in r-class matrices, which is automatic.

format(%*fmt*) specifies a numeric format, %*fmt*, for the display of correlations. The default is **format**(%9.3f). See [D] **format**.

abbrev(#) specifies that variable names be shown abbreviated to # characters. This option has no effect on output under the **matrix** option. It affects the saving of variable names under **saving()**.

fisher specifies that a bias correction attributed to Fisher (1921, 1925) be used. This sets $z' = z - 2r/(n-1)$ and then calculates limits as $\tanh(z' - d)$ and $\tanh(z' + d)$. There is no consequence for the display or saving of r itself.

jeffreys specifies that a procedure attributed to Jeffreys (1939, 1961) be used. This sets $d = \text{invnormal}(0.5 + \text{level}/200) / \text{sqrt}(n)$ and $z'' = z - 5r/2n$ and then calculates limits as $\tanh(z'' - d)$ and $\tanh(z'' + d)$. There is no consequence for the display or saving of r itself.

Only one of the **fisher** and **jeffreys** options can be specified.

saving(filename [, postfile_options]) specifies that results be saved to a Stata data file, *filename*, using **postfile**. *postfile_options* are options of **postfile**. The data file will contain the string variables **var1** and **var2** and the float variables **r**, **lower**, and **upper**. This is a rarely used option but could be useful if the correlations were to be displayed graphically or to be analyzed further.

Saved results

Matrices

| | |
|----------------|---------------------------------|
| r(corr) | matrix of correlation |
| r(lb) | matrix of lower limits |
| r(ub) | matrix of upper limits |
| r(z) | matrix of $z = \text{atanh } r$ |

7.2 corrcii

Syntax

```
corrcii #n #r [ , level(#) format(%fmt) [fisher|jeffreys] ]
```

Description

corrcii calculates confidence intervals for Pearson correlations using Fisher's z transform given a sample size of $\#n$ and a sample correlation of $\#r$.

For sample size n , correlation r , and confidence level *level*, the default procedure sets $d = \text{invnormal}(0.5 + \text{level}/200)/\text{sqrt}(n-3)$ and $z = \text{atanh}(r)$ and then calculates limits as $\tanh(z - d)$ and $\tanh(z + d)$.

corrcii is an immediate command. See [U] 19 Immediate commands for more detail.

Remarks

`corrcci` requires n to be at least 5. With very small samples—or even very large ones—interpretation remains the user’s responsibility.

Options

`level(#)`, `format(%fmt)`, `fisher`, and `jeffreys` options are the same as for `corrci`, specified above.

Saved results

Scalars

| | |
|----------------------|--|
| <code>r(corr)</code> | scalar with correlation |
| <code>r(lb)</code> | scalar with lower limit |
| <code>r(ub)</code> | scalar with upper limit |
| <code>r(z)</code> | scalar with $z = \operatorname{atanh} r$ |

8 Conclusion

The problem of estimating confidence intervals for correlation can be approached in two quite distinct ways. The classic method based on atanh transformation was originated by R. A. Fisher. The modern method based on bootstrapping was originated by Bradley Efron. As it happens, the latter has been better supported in Stata than the former, although user-written commands for the former have been available for several years. I have underlined how the availability of `atanh()` and `tanh()` functions in Stata makes calculator-style work much easier than might be supposed. That applies not just to confidence intervals but also to tests on the fly of a null hypothesis of nonzero correlation, or of whether two or more correlations are different, and to pooling correlations to get an overall estimate. I have also introduced two new confidence interval commands, `corrci` and `corrcci`, for separate use.

As with any statistical method, this problem raises questions about how far the underlying assumptions are satisfied in practice and how well the methods work with real data. At a minimum, the existence of easy-to-use commands for both methods allows comparison between them. It would be foolish to generalize from just two substantial examples, but there is some mild irony in the thought that one of Fisher’s examples was probably based on too small a sample for any confidence interval procedure to work well, even by conveying uncertainty through its width. Similarly, one of Efron’s example datasets is sufficiently well behaved that the classical method works just about as well. However, the thoughts that sample size is important and that it is crucial to be careful about your analyses are much, much older than any of the work discussed here.

9 Acknowledgments

This column is dedicated to my friend, colleague, and PhD supervisor Ian S. Evans with best wishes on his notional retirement. Ian has been a Stata user since 1991 and a careful, thoughtful user of correlations and statistical methods throughout his career.

I thank Christopher F. Baum for his initial work on the program `omninorm` used here.

10 References

- Abbott, P. 1940. *Teach Yourself Calculus*. London: English University Press.
- Altman, D. G. 1991. *Practical Statistics for Medical Research*. London: Chapman & Hall.
- Banner, A. 2007. *The Calculus Lifesaver: All the Tools You Need to Excel at Calculus*. Princeton, NJ: Princeton University Press.
- Barnett, J. H. 2004. Enter, stage center: The early drama of the hyperbolic functions. *Mathematics Magazine* 77: 15–30.
- Bartlett, M. S. 1947. The use of transformations. *Biometrics* 3: 39–52.
- Bland, M. 2000. *An Introduction to Medical Statistics*. 3rd ed. Oxford: Oxford University Press.
- Box, G. E. P., and G. C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison–Wesley.
- Cox, N. J. 2004. Stata tip 15: Function graphs on the fly. *Stata Journal* 4: 488–489.
- . 2008. Speaking Stata: Between tables and graphs. *Stata Journal* 8: 269–289.
- Cramér, H. 1946. *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press.
- Doornik, J. A., and H. Hansen. 1994. An omnibus test for univariate and multivariate normality. Unpublished working paper.
<http://ideas.repec.org/p/nuf/econwp/9604.html> or
<http://www.doornik.com/research/normal2.pdf>.
- Duncan, G. T., and M. W. J. Layard. 1973. A Monte-Carlo study of asymptotically robust tests for correlation coefficients. *Biometrika* 60: 551–558.
- Efron, B. 1998. R. A. Fisher in the 21st century: Invited paper presented at the 1996 R. A. Fisher lecture. *Statistical Science* 13: 95–122.
- Efron, B., and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall.

- Fisher, R. A. 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10: 507–521. Reprinted in *Collected Papers of R. A. Fisher, Volume I: 1912–1924*, ed. J. H. Bennett, 84–98. Adelaide: University of Adelaide, 1971. Text accessible at <http://digital.library.adelaide.edu.au/coll/special/fisher/4.pdf>.
- . 1919. The genesis of twins. *Genetics* 4: 489–499. Reprinted in *Collected Papers of R. A. Fisher, Volume I: 1912–1924*, ed. J. H. Bennett, 177–187. Adelaide: University of Adelaide, 1971.
- . 1921. On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron* 1(4): 3–32. Reprinted in *Collected Papers of R. A. Fisher, Volume I: 1912–1924*, ed. J. H. Bennett, 205–235. Adelaide: University of Adelaide, 1971. Text accessible at <http://digital.library.adelaide.edu.au/coll/special/fisher/14.pdf>.
- . 1925. *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd. Text accessible at <http://psychclassics.yorku.ca/Fisher/Methods/>.
- . 1970. *Statistical Methods for Research Workers*. 14th ed. New York: Hafner.
- . 1990. *Statistical Methods, Experimental Design, and Scientific Inference*. Oxford: Oxford University Press.
- Gleason, J. R. 1996. sg51: Inference about correlations using the Fisher z-transform. *Stata Technical Bulletin* 32: 13–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 121–128. College Station, TX: Stata Press.
- Goldstein, R. 1996. sg52: Testing dependent correlation coefficients. *Stata Technical Bulletin* 32: 18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 128–129. College Station, TX: Stata Press.
- Gullberg, J. 1997. *Mathematics: From the Birth of Numbers*. New York: W. W. Norton.
- Hald, A. 1998. *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley.
- . 2007. *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713–1935*. New York: Springer.
- Hardy, G. H. 1952. *A Course of Pure Mathematics*. 10th ed. Cambridge: Cambridge University Press.
- Hotelling, H. 1953. New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society, Series B* 15: 193–232.
- Hoyle, M. H. 1973. Transformations: An introduction and a bibliography. *International Statistical Review* 41: 203–223.
- Jeffreys, H. 1939. *Theory of Probability*. Oxford: Oxford University Press.
- . 1961. *Theory of Probability*. 3rd ed. Oxford: Oxford University Press.

- Lee, P. M. 2004. *Bayesian Statistics: An Introduction*. 3rd ed. London: Arnold.
- Maor, E. 1994. e: *The Story of a Number*. Princeton, NJ: Princeton University Press.
- Pearson, K. 1896. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London, Series A* 187: 253–318. Reprinted in *Karl Pearson's Early Statistical Papers*, 113–178. Cambridge: Cambridge University Press, 1948.
- Pearson, K., and L. N. G. Filon. 1898. Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society of London, Series A* 191: 229–311. Reprinted in *Karl Pearson's Early Statistical Papers*, 179–261. Cambridge: Cambridge University Press, 1948.
- Rodgers, J. L., and W. A. Nicewander. 1988. Thirteen ways to look at the correlation coefficient. *American Statistician* 42: 59–66.
- Rovine, M. J., and A. Von Eye. 1997. A 14th way to look at the correlation coefficient: Correlation as the proportion of matches. *American Statistician* 51: 42–46.
- Ryan, P. 2004. Stata tip 4: Using display as an online calculator. *Stata Journal* 4: 93.
- Saler, B. M. 1985. Inverse hyperbolic functions as areas. *College Mathematics Journal* 16: 129–131.
- Seed, P. T. 2001. sg159: Confidence intervals for correlations. *Stata Technical Bulletin* 59: 27–28. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 267–269. College Station, TX: Stata Press.
- . 2004. Software update: Confidence intervals for correlations. *Stata Journal* 4: 490.
- Shaw, W. N. 1905. Seasons in the British Isles from 1878. *Journal of the Royal Statistical Society* 68: 247–319.
- Snedecor, G. W., and W. G. Cochran. 1989. *Statistical Methods*. 8th ed. Ames, IA: Iowa State University Press.
- Sokal, R. R., and F. J. Rohlf. 1995. *Biometry: The Principles and Practice of Statistics in Biological Research*. 3rd ed. New York: Freeman.
- Stigler, S. M. 1986. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, MA: Harvard University Press.
- . 1999. *Statistics on the Table: The History of Statistical Concepts and Methods*. Cambridge, MA: Harvard University Press.
- Stuart, A., and J. K. Ord. 1994. *Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory*. 6th ed. London: Arnold.

van Belle, G., L. D. Fisher, P. J. Heagerty, and T. Lumley. 2004. *Biostatistics: A Methodology for the Health Sciences*. 2nd ed. Hoboken, NJ: Wiley.

Winterbottom, A. 1979. A note on the derivation of Fisher's transformation of the correlation coefficient. *American Statistician* 33: 142–143.

Zar, J. H. 1999. *Biostatistical Analysis*. 4th ed. Upper Saddle River, NJ: Prentice Hall.

About the author

Nicholas Cox is a statistically minded geographer at Durham University. He contributes talks, postings, FAQs, and programs to the Stata user community. He has also coauthored 15 commands in official Stata. He wrote several inserts in the *Stata Technical Bulletin* and is an editor of the *Stata Journal*.