

RESEARCH ARTICLE

A robust Pearson correlation test for a general point null using a surrogate bootstrap distribution

Alan D. Hutson *

Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263 United States of America

* Alan.hutson@roswellpark.org



Abstract

In this note we present a robust bootstrap test with good Type I error control for testing the general hypothesis $H_0: \rho = \rho_0$. In order to carry out this test we use what is termed a *surrogate bootstrap distribution*. The test was inspired by the studentized permutation for testing $H_0: \rho = 0$, which was proven to be exact in certain scenarios and asymptotically correct over all. We show that the bootstrap based test is robust to a variety of distributional scenarios in terms of proper Type I error control.

OPEN ACCESS

Citation: Hutson AD (2019) A robust Pearson correlation test for a general point null using a surrogate bootstrap distribution. PLoS ONE 14(5): e0216287. <https://doi.org/10.1371/journal.pone.0216287>

Editor: Eugene Demidenko, Dartmouth College Geisel School of Medicine, UNITED STATES

Received: October 4, 2018

Accepted: April 17, 2019

Published: May 16, 2019

Copyright: © 2019 Alan D. Hutson. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data are provided in a table in the manuscript.

Funding: This work was supported by Roswell Park Cancer Institute and National Cancer Institute (NCI) grant P30CA016056 (Hutson, Co-Inv), National Science Foundation under grant NSF IIS-1514204 (Hutson, Co-Inv) and NRG Oncology Statistical and Data Management Center grant U10CA180822 (Hutson MPI) and IOTN Moonshot grant U24CA232979-01 (Hutson MPI). The funders had no role in study design, data collection and

Introduction

The term “correlation” was introduced by Galton as a synonym for regression and the term “coefficient of correlation” was first used by Edgeworth, who also derived the first sample estimator of linear association [1], [2]. Estimation and testing based on the correlation coefficient is one of the most used approaches towards examining the association between two continuous variables. Even though Edgeworth first derived the estimator of the sample correlation coefficient it was Karl Pearson who is credited with much of its development via his 1896 paper [3], and why we often refer to the “Pearson Correlation Coefficient” as given by the form

$$\rho = \frac{\mu_{XY}}{\sigma_X \sigma_Y}, \tag{1}$$

where $-1 < \rho < 1$, μ_{XY} is the covariance between two random variables X and Y , σ_X is the standard deviation of X and σ_Y is the standard deviation of Y .

Throughout this note let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be n paired observations from a non-degenerate joint distribution $F_{XY}(x, y)$. The well-known sample estimator for ρ at (1) is given as

$$\hat{\rho} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}, \tag{2}$$

analysis, decision to publish, or preparation of the manuscript.

Competing interests: The author has declared that no competing interests exist.

where $\bar{X} = \sum_{i=1}^n X_i/n$ and $\bar{Y} = \sum_{i=1}^n Y_i/n$. The estimator $\hat{\rho}$ is also well-known to be the maximum likelihood estimator for the bivariate normal correlation parameter.

The sampling distribution for $\hat{\rho}$ at (2) under bivariate normality was derived by R. A. Fisher [4]. For the case when $\rho = 0$, “Student” [5] surmised that $\sqrt{(n - 2)\hat{\rho}^2/(1 - \hat{\rho}^2)}$ has a *t*-distribution with $n - 2$ degrees-of-freedom. The *t*-distribution based test, assuming bivariate normality, for testing the hypothesis $H_0: \rho = 0$ versus $H_1: \rho \neq 0$ is generally what is found in most statistical software packages. For the more general bivariate normal case when $\rho \neq 0$ the exact distribution for $\hat{\rho}$ at (2) is a bit more unwieldy [1]. Remarkably, Fisher discovered a large sample asymptotic normal variance stabilized approximation based on what is now known as Fisher’s *z*-transformation, which is given as

$$z = \frac{1}{2} \log \frac{1 + \hat{\rho}}{1 - \hat{\rho}}. \tag{3}$$

This transformed variable tends to normality much faster than the classic distribution free based asymptotic normal approximation. The first order variance of *z* at (3) is $1/(n - 3)$. Various refinements of Fisher’s work have been carried out over the years in terms of both bias reduction and refined variance expressions [1]. For the more general test of $H_0: \rho = \rho_0$ (including $H_0: \rho = 0$ versus one-sided alternatives) Fisher’s *z*-transformation approach, again under bivariate normality assumptions, is what most current statistical software packages employ.

Alternatively, the “distribution free” large sample approximation for the sampling distribution for $\hat{\rho}$ may be derived using the multivariate delta method as outlined in Serfling [6] page 126 and given by first noting that we can express the sample estimator as $\hat{\rho} = g(\mathbf{V})$, where

$$\mathbf{V} = \left(\bar{X}, \bar{Y}, \frac{1}{n} \sum_{i=1}^n X_i^2, \frac{1}{n} \sum_{i=1}^n Y_i^2, \frac{1}{n} \sum_{i=1}^n X_i Y_i \right), \tag{4}$$

such that as per Serfling [6] we have

$$\hat{\rho} \sim AN(\rho, n^{-1} \mathbf{d} \Sigma \mathbf{d}'). \tag{5}$$

For expression (5) Σ is the 5×5 variance-covariance matrix for the vector of the components of the summands of \mathbf{V} given by $(X_1, Y_1, X_1^2, Y_1^2, X_1 Y_1)$. The elements of \mathbf{d} , which will play a key role in our application are given as

$$\begin{aligned} d_1 &= \frac{\rho \mu_X}{\sigma_X^2} - \frac{\mu_Y}{\sigma_X \sigma_Y}, d_2 = \frac{\rho \mu_Y}{\sigma_Y^2} - \frac{\mu_X}{\sigma_X \sigma_Y}, \\ d_3 &= -\frac{\rho}{2\sigma_X^2}, d_4 = -\frac{\rho}{2\sigma_Y^2}, d_5 = \frac{1}{\sigma_X \sigma_Y}. \end{aligned} \tag{6}$$

More refined approximations for the sampling distribution of $\hat{\rho}$ via Edgeworth techniques for non-normal populations [7], [8].

There have been several investigations examining the robustness of the bivariate normality assumptions as it pertains to type I error control for testing $H_0: \rho = \rho_0$ using Fisher’s *z*-transformation or the more specific case of $H_0: \rho = 0$ using Student’s *t* distribution. One of the earliest of these investigations is due to Pearson [9] who concluded that “the results suggest that the normal bivariate surface can be mutilated and distorted to a remarkable degree without affecting the frequency distribution of *r*”, which has since been disproved [7]. Havlicek and Peterson [10] also incorrectly concluded “that the Pearson *r* is insensitive to extreme violations of the basic assumptions of normality.” There has been a substantial number of simulation based

studies, primarily in the psychology literature that note when tests about the correlation coefficient, assuming bivariate normality incorrectly, tend to fail [11]–[15].

We will repeat the simulation study of DiCiccio and Romano [16] and re-examine the Type I error control for testing $H_0: \rho = \rho_0$ based on Fisher's z -transformation, the classic large sample approximation and our new bootstrap approach. To the best of our knowledge there does not seem to be an extensive simulation study of the straightforward asymptotic approximation above given by using the corresponding moment estimators in place of their population counterparts and setting $\rho = \rho_0$ under the null. DiCiccio and Romano [16] do include results for the specific case for testing $H_0: \rho = 0$.

As an alternative to either assuming bivariate normality or using a large sample asymptotic approximation one may consider a permutation test approach for the specific test $H_0: \rho = 0$ [17]. The key sticking point that is often overlooked by practitioners is that the permutation test is only exact in terms of Type I error control when using the metric ρ to test $H_0: F_{XY} = F_X F_Y$. Only under specific distributional assumptions, e.g. bivariate normality or certain families of elliptical distributions, is the permutation test exact for testing $H_0: \rho = 0$. Heuristic investigations for permutation testing about the correlation coefficient go back several decades [18]. Even recently, Tuğran [19] make the same mistakes of the past when drawing conclusions regarding the permutation test's applicability under normality versus non-normality.

It was only until very recently that DiCiccio and Romano [16] provided a careful theoretical treatment of the subject summarizing the methodologies and providing guidance for the most appropriate permutation approach for testing $H_0: \rho = 0$. In their work, DiCiccio and Romano [16] provide a novel method of obtaining a permutation test that is exact when testing $H_0: \rho = 0$ is equivalent to testing $H_0: F_{XY} = F_X F_Y$ and asymptotically controls the Type I error when testing $H_0: \rho = 0$ is not equivalent to testing $H_0: F_{XY} = F_X F_Y$. The key to their approach is standardizing the test statistic using the large sample variance at (5) under $H_0: \rho = 0$ such that as $n \rightarrow \infty$ the quantiles of the studentized test statistic $\hat{\rho} / \sqrt{n^{-1} d \Sigma d'}$ for the permutation distribution and the true sampling distribution “converge almost surely to the corresponding quantiles of the standard normal distribution.” In fact, the specific form of the standardization they utilize for testing $H_0: \rho = 0$ can be traced back to Tschuprow [20].

As an overall strategy for testing $H_0: \rho = 0$ the permutation testing approach that appears in DiCiccio and Romano [16] appears to be the most robust strategy to date in terms of controlling Type I error rates across a variety of distributional assumptions. In particular it is clear that the over-reliance on the bivariate normality assumption, when testing $H_0: \rho = 0$, and utilized in many statistical software packages is not a particularly robust assumption. As noted by R. C. Geary [21]: “Normality is a myth; there never was, and never will be, a normal distribution.” One might extend this quote in an even stronger fashion relative to bivariate normality relative to its existence in real-life data analysis problems and testing about the correlation coefficient.

In this note we build upon the work of DiCiccio and Romano [16] by introducing a bootstrap-like resampling approach for testing the general null hypothesis $H_0: \rho = \rho_0$, which has virtually the same properties as the permutation approach for testing about $H_0: \rho = 0$. The general difference between our approach and the approach of DiCiccio and Romano [16] is that the permutation testing approach is essentially a sampling without replacement method under the null hypothesis while our approach is a sampling with replacement approach under the null hypothesis approach. This subtle difference will allow us to develop a more general testing approach for $H_0: \rho = \rho_0$. Both approaches rely on a properly standardized test statistic based on (5) under $H_0: \rho = 0$ such that the large sample Type I error control controlled exactly for certain scenarios and in general asymptotically. The test can also be inverted to provide precise

confidence interval for ρ as an alternative to the more standard bootstrap confidence interval methods [22].

Materials and methods

The main thrust of the bootstrap correlation test of $H_0: \rho = \rho_0$ is to generalize the approach of DiCiccio and Romano [16] by noting that we can approximate the distribution function of the sample correlation estimator $\hat{\rho}|H_0$ given $F_{XY}|H_0$ using a surrogate distribution function, which we will describe below. Bootstrap samples will be drawn from the surrogate distribution function $F_{ST}|H_0 \approx F_{XY}|H_0$, which is described in detail below.

In terms of the preliminaries let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be an i.i.d. bivariate sample from an absolutely continuous distribution F_{XY} with marginal distributions denoted F_X and F_Y . In terms of our application let the 2×2 positive definite correlation matrix for the standardized variables $(\frac{X_1 - \mu_X}{\sigma_X}, \frac{Y_1 - \mu_Y}{\sigma_Y})$ be given as as

$$\Gamma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{7}$$

Next, represent the Cholesky decomposition of the $p \times p$ matrix Γ as

$$\Gamma = \mathbf{A}'\mathbf{A}, \tag{8}$$

such that \mathbf{A}'^{-1} is defined. The Cholesky decomposition is a key component of the bootstrap test that we propose given below.

Denote the $n \times 2$ matrix of standardized observations as $(\mathbf{U}:\mathbf{V})$, where

$$\mathbf{U}' = \left(\frac{X_1 - \mu_X}{\sigma_X}, \frac{X_2 - \mu_X}{\sigma_X}, \dots, \frac{X_n - \mu_X}{\sigma_X} \right) \tag{9}$$

and

$$\mathbf{V}' = \left(\frac{Y_1 - \mu_Y}{\sigma_Y}, \frac{Y_2 - \mu_Y}{\sigma_Y}, \dots, \frac{Y_n - \mu_Y}{\sigma_Y} \right), \tag{10}$$

respectively. Apply the transformation

$$(\mathbf{S}:\mathbf{T})' = \mathbf{A}'^{-1}(\mathbf{U}:\mathbf{V})', \tag{11}$$

where \mathbf{A} is the decomposed matrix at (8), where more specifically stated

$$\mathbf{A}'^{-1} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1 - \rho^2} \end{pmatrix}, \tag{12}$$

such that we have transformed observations are given as $S_i = U_i$ and $T_i = \rho U_i + \sqrt{1 - \rho^2} V_i$, $i = 1, 2, \dots, n$.

Next denote $\sigma_{\hat{\rho}(X,Y)}^2 = n^{-1} \mathbf{d}_{XY}' \Sigma_{XY} \mathbf{d}_{XY}$, where Σ_{XY} is the 5×5 variance-covariance matrix for the vector $(X_1, Y_1, X_1^2, Y_1^2, X_1 Y_1)$ given earlier at (5) and \mathbf{d}_{XY} is the 1×5 vector defined at (6). For the transformed variables similarly denote $\sigma_{\hat{\rho}(S,T)}^2 = n^{-1} \mathbf{d}_{ST}' \Sigma_{ST} \mathbf{d}_{ST}$, where Σ_{ST} is the 5×5 variance-covariance matrix for the vector $(S_1, T_1, S_1^2, T_1^2, S_1 T_1)$ and the vector \mathbf{d}_{ST} is defined similar to \mathbf{d}_{XY} defined at (6) now using the moments based on S and T in place of X and Y .

Under $H_0: \rho = \rho_0$ the parameter ρ is “known”. The estimator for the variance of $\hat{\rho}(X, Y)|H_0 = \hat{\rho}_0(X, Y)$ will be denoted as

$$s_{\hat{\rho}_0(X,Y)}^2 = n^{-1} \hat{\mathbf{d}}_{XY,H_0} \hat{\Sigma}_{XY} \hat{\mathbf{d}}'_{XY,H_0}, \tag{13}$$

where population moments are replaced with their corresponding sample moment counterparts throughout \mathbf{d}_{XY,H_0} and Σ_{XY} , and $\rho = \rho_0$. In a similar fashion denote the estimator for the variance of $\hat{\rho}(S, T)|H_0 = \hat{\rho}_0(S, T)$ as

$$s_{\hat{\rho}_0(S,T)}^2 = n^{-1} \hat{\mathbf{d}}_{ST,H_0} \hat{\Sigma}_{ST} \hat{\mathbf{d}}'_{ST,H_0}. \tag{14}$$

Theorem 2.1 Under $H_0: \rho = \rho_0$ assume $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ are an i.i.d. bivariate sample from an absolutely continuous distribution F_{XY} . Also, assume $E(X_1^4) < \infty, E(Y_1^4) < \infty$ and $E(X_1^2 Y_1^2) < \infty$. Then as $n \rightarrow \infty$

$$\frac{\hat{\rho}_0(X, Y) - \rho_0}{s_{\hat{\rho}_0(X,Y)}} \rightarrow AN(0, 1). \tag{15}$$

The result follows straightforward noting that the sample moments used to calculate the estimator $s_{\hat{\rho}_0(X,Y)}$ are all averages and as $n \rightarrow \infty$ all converge to their constant population counterparts. Hence, under the conditions outlined above $s_{\hat{\rho}_0(X,Y)} \rightarrow \sigma_{\hat{\rho}_0(X,Y)}$ as $n \rightarrow \infty$. Noting that as per Serfling [6] page 126 that $\frac{\hat{\rho}_0(X,Y) - \rho_0}{\sigma_{\hat{\rho}_0(X,Y)}} \rightarrow AN(0, 1)$ as $n \rightarrow \infty$ completes the argument using Slutsky’s Theorem with the conditions $E(Y_1^4) < \infty$ and $E(X_1^2 Y_1^2) < \infty$.

Theorem 2.2 Under $H_0: \rho = \rho_0$ assume $(S_1, T_1), (S_2, T_2), \dots, (S_n, T_n)$ defined at (11) are an i.i.d. bivariate sample from an absolutely continuous distribution F_{ST} . Also, assume $E(S_1^4) < \infty, E(T_1^4) < \infty, E(S_1^2 T_1^2) < \infty$ and that μ_X, σ_X, μ_Y and σ_Y are known. Then as $n \rightarrow \infty$

$$\frac{\hat{\rho}_0(S, T) - \rho_0}{s_{\hat{\rho}_0(S,T)}} \rightarrow AN(0, 1). \tag{16}$$

The result follows using the same arguments as in Theorem 2.1 given the key feature that μ_X, σ_X, μ_Y and σ_Y are known quantities.

Comment 2.1 Under Theorem 2.1 and Theorem 2.2. we see that both estimators $\hat{\rho}_0(X, Y)$ and $\hat{\rho}_0(S, T)$ have the same large sample distributions and their expectations are given as $E(\hat{\rho}_0(X, Y)|H_0) = E(\hat{\rho}_0(S, T)|H_0) = \rho_0$.

Comment 2.2 Relative to Theorem 2.2 we can generate bootstrap samples denoted as the pairs $(S_1^*, T_1^*), (S_2^*, T_2^*), \dots, (S_n^*, T_n^*)$ under the null hypothesis $H_0: \rho = \rho_0$ such that $E(\hat{\rho}_0(S^*, T^*)|H_0) \rightarrow \rho_0$. This is done by first drawing independent bootstrap samples from the marginal distribution functions \hat{F}_X and \hat{F}_Y and applying an empirical version of the transformation at (11), i.e. replace μ_X, σ_X, μ_Y and σ_Y with their respective sample counterparts. Essentially we are drawing nonparametric bootstrap samples from $\hat{F}_{ST}|H_0$ to estimate the bootstrap distribution of the sample variance for $\hat{\rho}_0(S, T)|H_0$ with the scale based on the underlying properties of the distribution $F_{ST}|H_0$. We can then rescale based on the sample variance for $\hat{\rho}_0(X, Y)|H_0$ in order to have an asymptotically valid testing procedure. More specifically we can follow the the steps below to test $H_0: \rho = \rho_0$ versus $H_1: \rho > \rho_0$ as an extension of the rescaling ideas put forward by DiCiccio and Romano [16] for testing the specific hypothesis $H_0: \rho = 0$. Note also that the procedure we propose can be modified easily to handle the alternatives $H_1: \rho < \rho_0$ and $H_1: \rho \neq \rho_0$.

Out testing algorithm now follows these steps:

1. Calculate the Pearson sample correlation $\hat{\rho}_0(X, Y)$ (the subscript is added for notational convenience), which has the same functional form as $\hat{\rho}$ ($= \hat{\rho}_0(X, Y)$) given at (2), and its corresponding variance estimate under H_0 , $s_{\hat{\rho}_0(X,Y)}^2$ given at (13).
2. Draw samples of size n independently and with replacement from each marginal distribution \hat{F}_X and \hat{F}_Y and denote the pairs of independently drawn observations as $(X_1^*, Y_1^*), (X_2^*, Y_2^*), \dots, (X_n^*, Y_n^*)$, where $\hat{F}_X(x) = \sum_{i=1}^n I_{(x_i \leq x)}/n$ and $\hat{F}_Y(y) = \sum_{i=1}^n I_{(y_i \leq y)}/n$, respectively.
3. Standardize each observation using the observed sample means and sample standard deviations as $\left(\frac{X_1^* - \bar{X}}{s_X}, \frac{Y_1^* - \bar{Y}}{s_Y}\right), \left(\frac{X_2^* - \bar{X}}{s_X}, \frac{Y_2^* - \bar{Y}}{s_Y}\right), \dots, \left(\frac{X_n^* - \bar{X}}{s_X}, \frac{Y_n^* - \bar{Y}}{s_Y}\right)$.
4. Denote the $n \times 2$ matrix of standardized bootstrap observations as $(\mathbf{U}^* : \mathbf{V}^*)$, where $\mathbf{U}^{*'} = \left(\frac{X_1^* - \bar{X}}{s_X}, \frac{X_2^* - \bar{X}}{s_X}, \dots, \frac{X_n^* - \bar{X}}{s_X}\right)$ and $\mathbf{V}^{*'} = \left(\frac{Y_1^* - \bar{Y}}{s_Y}, \frac{Y_2^* - \bar{Y}}{s_Y}, \dots, \frac{Y_n^* - \bar{Y}}{s_Y}\right)$, respectively.
5. Apply the transformation $A_0^{-1}(\mathbf{U}^* : \mathbf{V}^*)'$, where A_0 is the decomposed matrix at (8) replacing ρ with ρ_0 , where specifically stated

$$A_0^{-1} = \begin{pmatrix} 1 & 0 \\ \rho_0 & \sqrt{1 - \rho_0^2} \end{pmatrix}, \tag{17}$$

such that we have correlated bootstrap resampled paired values are given as $S_i^* = U_i^*$ and $T_i^* = \rho_0 U_i^* + \sqrt{1 - \rho_0^2} V_i^*, i = 1, 2, \dots, n$.

6. Calculate $\hat{\rho}_0^*(S^*, T^*)$ using the standard sample correlation formula (2) and the corresponding variance estimate $s_{\hat{\rho}_0^*(S,T)}^2$ at (14) under H_0 using the correlated values $(S_i^*, T_i^*), i = 1, 2, \dots, n$, to generate an empirical bootstrap null distribution of observations for a single bootstrap replicate.
7. Calculate the bootstrap rescaled and bias corrected bootstrap correlation estimate under H_0 as $\hat{\rho}^{**} = \rho_0 - \frac{s_{\hat{\rho}_0(X,Y)}}{s_{\hat{\rho}_0^*(S,T)}} (\hat{\rho}^*(S^*, T^*) - \rho_0)$.
8. Repeat steps 2-7 B times (usually $B = 500$ or $B = 1000$ is sufficient).
9. Calculate the bootstrap estimated p-value as $p = \frac{\sum_{j=1}^B I_{(\hat{\rho}_j^{**} > \hat{\rho}(X,Y))}}{B}$, where I denotes the standard indicator function.

An outline for the large sample validity of the test can be examined via Edgeworth expansion techniques [1], for an overview of Edgeworth expansion methods. Towards this end note that under $H_0: \rho = \rho_0$ we have

$$P\left(\frac{\hat{\rho}_0(S, T) - \rho_0}{s_{\hat{\rho}_0(S,T)}} \leq c | F_{ST}, H_0\right) = \Phi(c) + \frac{p_1(c | F_{ST}, H_0)}{\sqrt{n}} \phi(c) + \frac{p_2(c | F_{ST}, H_0)}{n} \phi(c) + o_p\left(\frac{1}{n}\right) \tag{18}$$

and

$$\begin{aligned}
 P\left(\frac{\hat{\rho}_0^*(S, T) - \rho_0}{s_{\hat{\rho}_0^*(S, T)}} \leq c | \hat{F}_{ST}, H_0\right) &= \Phi(c) + \frac{p_1(c | \hat{F}_{ST}, H_0)}{\sqrt{n}} \phi(c) \\
 &+ \frac{p_2(c | \hat{F}_{ST}, H_0)}{n} \phi(c) + o_p\left(\frac{1}{n}\right),
 \end{aligned}
 \tag{19}$$

where \hat{F}_{ST} is defined through the marginal distribution \hat{F}_X and \hat{F}_Y and the transformation at Step 5 in our algorithm and $\phi(c)$ and $\Phi(\cdot)$ are the standard normal density and cumulative distribution functions, respectively. Eqs (18) and (19) yields the expression for the Kolmogorov-Smirnov distance as

$$\begin{aligned}
 &\sup_c \left| P\left(\frac{\hat{\rho}_0(S, T) - \rho_0}{s_{\hat{\rho}_0(S, T)}} \leq c | F_{ST}, H_0\right) - P\left(\frac{\hat{\rho}_0^*(S, T) - \rho_0}{s_{\hat{\rho}_0^*(S, T)}} \leq c | \hat{F}_{ST}, H_0\right) \right|, \\
 &= \sup_c \left| \frac{p_1(c | F_{ST}, H_0) - p_1(c | \hat{F}_{ST}, H_0)}{\sqrt{n}} + \frac{p_2(c | F_{ST}, H_0) - p_2(c | \hat{F}_{ST}, H_0)}{n} \right| \phi(c) \\
 &+ o_p\left(\frac{1}{n}\right),
 \end{aligned}
 \tag{20}$$

where the polynomials p_i depend on F_{ST} in a smooth way. If the distance between between $\hat{F}_{ST} - F_{ST}$ is $O_p(\sqrt{n})$ then expression (20) holds. Again, recall that \hat{F}_{ST} is defined through the marginal distribution \hat{F}_X and \hat{F}_Y and the transformation at Step 5 in our algorithm, where using standard techniques it is straightforward to show $\hat{F}_X - F_X = O_p(\sqrt{n})$ and $\hat{F}_Y - F_Y = O_p(\sqrt{n})$ and that $\bar{X} \rightarrow \mu_X, s_X \rightarrow \sigma_X, \bar{Y} \rightarrow \mu_Y$ and $s_Y \rightarrow \sigma_Y$, where μ_X, σ_X, μ_Y and σ_Y are constants. In other words, the distribution of $\frac{\hat{\rho}_0^*(S, T) - \rho_0}{s_{\hat{\rho}_0^*(S, T)}} \rightarrow AN(0, 1)$ under the assumptions provided and consistent with the original distribution of $\frac{\hat{\rho}_0(X, Y) - \rho_0}{s_{\hat{\rho}_0(X, Y)}}$ as per Theorem 2.1.

Comment 2.3. For the special and important case $H_0: \rho = 0$ the only practical difference between our approach and the permutation based method of DiCiccio and Romano [16] is that $s_{\hat{\rho}_0^*(S, T)}$ is recalculated per each bootstrap replicate due to sampling with replacement mechanism whereas the permutation value is fixed across all permutations under the null. We do show however in our simulation study that the bootstrap test for $H_0: \rho = 0$ does appear to have some cases where it works slightly better than the permutation test, likely due to a finer grid of values in the resampling scheme of with replacement as compared to without replacement.

Results

For our simulation study we tested $H_0: \rho = \rho_0$ versus $H_1: \rho > \rho_0$ for $\rho_0 = 0.0, 0.3$ and 0.6 with sample sizes $n = 10, 25, 50, 100, 200$. Each simulation utilized 10,000 Monte Carlo replications and the bootstrap resampling algorithm used $B = 500$ resamples. We compared Type I error control between the studentized permutation test of DiCiccio and Romano [16], the new bootstrap test, Fisher’s z-transform and the straight large sample approximation given Eq (5) using sample moment estimators for the variance estimation. The tests examined the Type I error control $\alpha = 0.05$ as was performed by DiCiccio and Romano [16].

We utilized the same marginal distributions for F_X and F_Y from DiCiccio and Romano [16] page 1218 for the case $H_0: \rho = 0$ for our simulation study:

1. Multivariate normal (MVN) with mean zero and identity covariance.

2. Multivariate t-distribution (Multivariate t_5) with 5 degrees of freedom.
3. Exponential given as $(X, Y) = rS'u$ where $S = \text{diag}(\sqrt{2}, 1)$, u is uniformly distributed on the two dimensional unit circle.
4. Circular given as the uniform distribution on a two dimensional unit circle.
5. $t_{4,1}$ where $X = W + Z$ and $Y = W - Z$, where W and Z are i.i.d. $t_{4,1}$ random variables.

One can see DiCiccio and Romano [16] for more detail regarding these specific distributions.

For the general case $H_0: \rho = \rho_0$ we utilized the marginal distributions above, standardized the marginal variables X and Y to mean = 0 and variance = 1 and applied the transformation at (11) under H_0 . The results are contained in Tables 1–3.

The first striking point to make is how inflated the Type I error can be for the test based on Fisher’s z-transformation. For the $t_{4,1}$ at $H_0: \rho = 0.3$ we see an estimated Type I error rate of 0.1847 at $n = 200$. In the other direction the test based on Fisher’s z-transformation can also have much lower than anticipated Type error rate, e.g. see $H_0: \rho = 0.3, n = 100$ for the circular distribution.

In terms of the straight large sample approximation [6] we see that it performs fairly well across all distributions for samples of size $n \geq 25$, but has inflated Type I error rates for $n = 10$ usually twice the desired α .

For the specific case $H_0: \rho_0 = 0$ we see that the bootstrap test and permutation test are comparable with the bootstrap test actually subtly outperforming the permutation test, even for large sample sizes, for certain scenarios. The bootstrap test and asymptotic tests also have

Table 1. Rejection Probabilities at $\alpha = 0.05$ for bootstrap, asymptotic and Fisher’s z tests for $H_0: \rho = \rho_0$ versus $H_1: \rho > \rho_0$ using the sample correlation coefficient.

Distribution	ρ_0	Test	n					
			10	25	50	100	200	
MVN	0.0	Bootstrap	0.0533	0.0514	0.0493	0.0464	0.0516	
		Asymptotic	0.1011	0.0698	0.0586	0.0523	0.0543	
		Fisher’s z	0.0497	0.0513	0.0510	0.0482	0.0519	
		Stud Perm*	0.0470	0.0525	0.0525	0.0511	0.0561	
	0.3	Bootstrap	0.0571	0.0538	0.0526	0.0533	0.0496	
		Asymptotic	0.1076	0.0687	0.0599	0.0551	0.0511	
		Fisher’s z	0.0492	0.0477	0.0491	0.0493	0.0483	
	0.6	Bootstrap	0.0641	0.0619	0.0549	0.0571	0.0543	
		Asymptotic	0.1136	0.0752	0.0601	0.0591	0.0545	
		Fisher’s z	0.0492	0.0474	0.0486	0.0509	0.0494	
	Exponential	0.0	Bootstrap	0.0524	0.0461	0.0511	0.0474	0.0504
			Asymptotic	0.0959	0.0643	0.0603	0.0512	0.0525
Fisher’s z			0.0279	0.0214	0.0237	0.0231	0.0234	
Stud Perm*			0.0679	0.0508	0.0476	0.0502	0.0485	
0.3		Bootstrap	0.0599	0.0545	0.0515	0.0560	0.0542	
		Asymptotic	0.1070	0.0725	0.0599	0.0597	0.0565	
		Fisher’s z	0.0272	0.0245	0.0223	0.0236	0.0243	
0.6		Bootstrap	0.0603	0.0604	0.0570	0.0558	0.0529	
		Asymptotic	0.1072	0.0761	0.0627	0.0591	0.0533	
		Fisher’s z	0.0244	0.0250	0.0222	0.0240	0.0213	

* Numbers abstracted from DiCiccio and Romano (2017)

<https://doi.org/10.1371/journal.pone.0216287.t001>

Table 2. Rejection Probabilities at $\alpha = 0.05$ for bootstrap, asymptotic and Fisher's z tests for $H_0: \rho = \rho_0$ versus $H_1: \rho > \rho_0$ using the sample correlation coefficient.

Distribution	ρ_0	Test	n					
			10	25	50	100	200	
$t_{4,1}$	0.0	Bootstrap	0.0485	0.0432	0.0450	0.0428	0.0451	
		Asymptotic	0.0991	0.0608	0.0550	0.0467	0.0462	
		Fisher's z	0.1167	0.1438	0.1657	0.1769	0.1900	
		Stud Perm*	0.0444	0.0428	0.0426	0.0442	0.0391	
	0.3	Bootstrap	0.0533	0.0546	0.0512	0.0527	0.0527	
		Asymptotic	0.1006	0.0638	0.0556	0.0516	0.0494	
		Fisher's z	0.1036	0.1371	0.1604	0.1767	0.1847	
	0.6	Bootstrap	0.0576	0.0570	0.0587	0.0659	0.0570	
		Asymptotic	0.1071	0.0621	0.0552	0.0580	0.0469	
		Fisher's z	0.0991	0.1187	0.1471	0.1591	0.1619	
	Multivariate t_5	0.0	Bootstrap	0.0473	0.0477	0.0470	0.0523	0.0521
			Asymptotic	0.0947	0.0638	0.0545	0.0543	0.0543
Fisher's z			0.0497	0.0466	0.0471	0.0513	0.0516	
Stud Perm*			0.0507	0.0462	0.0460	0.0456	0.0471	
0.3		Bootstrap	0.0538	0.0530	0.0475	0.0440	0.0479	
		Asymptotic	0.1084	0.0689	0.0579	0.0505	0.0510	
		Fisher's z	0.0593	0.0612	0.0609	0.0636	0.0661	
0.6		Bootstrap	0.0516	0.0509	0.0459	0.0407	0.0412	
		Asymptotic	0.1069	0.0687	0.0570	0.0493	0.0489	
		Fisher's z	0.0763	0.0874	0.0878	0.0981	0.1027	

*Numbers abstracted from DiCiccio and Romano (2017)

<https://doi.org/10.1371/journal.pone.0216287.t002>

similar properties for large sample sizes thus heuristically validating our theoretical arguments from the previous section. The bootstrap test appears to work well across all scenarios in terms of Type I error control and thusly would appear to be a robust method for testing the general hypothesis $H_0: \rho = \rho_0$.

Table 3. Rejection Probabilities at $\alpha = 0.05$ for bootstrap, asymptotic and Fisher's z tests for $H_0: \rho = \rho_0$ versus $H_1: \rho > \rho_0$ using the sample correlation coefficient.

Distribution	ρ_0	Test	n				
			10	25	50	100	200
Circular	0.0	Bootstrap	0.0570	0.0480	0.0514	0.0540	0.0438
		Asymptotic	0.1010	0.0645	0.0605	0.0570	0.0455
		Fisher's z	0.0279	0.0240	0.0242	0.0234	0.0193
		Stud Perm*	0.0674	0.0468	0.0488	0.0484	0.0521
	0.3	Bootstrap	0.0612	0.0551	0.0511	0.0535	0.0503
		Asymptotic	0.1039	0.0701	0.0603	0.0552	0.0521
		Fisher's z	0.0273	0.0231	0.0233	0.0198	0.0211
	0.6	Bootstrap	0.0622	0.0595	0.0581	0.0533	0.0544
		Asymptotic	0.1097	0.0751	0.0643	0.0568	0.0550
		Fisher's z	0.0254	0.0227	0.0243	0.0224	0.0227

* Numbers abstracted from DiCiccio and Romano (2017)

<https://doi.org/10.1371/journal.pone.0216287.t003>

Table 4. Power comparisons at $\alpha = 0.05$ for bootstrap versus exact Student’s t-test for $H_0: \rho = 0$ versus $H_1: \rho > 0$ using the sample correlation coefficient.

Distribution	ρ	Test	n		
			10	25	50
MVN	0.3	Bootstrap	0.186	0.402	0.670
		t-test	0.201	0.415	0.677
	0.6	Bootstrap	0.484	0.917	0.999
		t-test	0.562	0.940	0.999

<https://doi.org/10.1371/journal.pone.0216287.t004>

In Table 4 we compared the power of the bootstrap test versus the exact Student’s t-test under multivariate normality assumptions. We see that the bootstrap test compares favorably to a test based on optimal assumptions and one that is limited by the form of the null hypothesis, i.e. the exact Student t-test for testing $H_0: \rho_0 = 0$ is specific to this form of the null hypothesis and does not generalize to other values of ρ_0 . We see from Table 4 that the exact test has a slight power advantage for small samples ($n = 10$), which dimensions rapidly as a function of larger sample sizes. A similar analogy would be found comparing the Wilcoxon rank-sum test versus the two-sample t-test in terms of efficiency where the two-sample t-test is only more efficient in general to the Wilcoxon rank-sum test under the strict and often unrealistic assumption of normality.

Example

As a straightforward example of our approach we tested $H_0: \rho = \rho_0$ versus $H_1: \rho > \rho_0$ for $\rho_0 = 0$ and $\rho_0 = 0.3$ using lactate levels measured in the blood and the cerebrospinal fluid (CSF) in 13 female subjects [20]. The data are provided in Table 5. A scatterplot of the paired data is given in Fig 1.

We used SAS 9.4 (SAS Institute Cary, NC) to test the marginal normality of the data using the Shapiro-Wilk test, the multivariate skewness and kurtosis coming from a bivariate normal distribution using the built in Mardia tests and the overall test of bivariate normality using the Henze-Zirkler T test. The results from SAS given in Table 6. We see that none of the tests

Table 5. Example blood and CSF lactate levels on $n = 13$ female subjects.

Subject	Blood	CSF
	Lactate (mM)	Lactate (mM)
1	3.5	7.800
2	2.7	3.400
3	1.7	5.900
4	2.9	6.400
5	0.6	2.400
6	1.1	2.000
7	3.5	4.400
8	1.9	4.300
9	1.5	5.700
10	1.6	3.900
11	2.2	3.400
12	1.5	4.528
13	1.6	4.600

<https://doi.org/10.1371/journal.pone.0216287.t005>

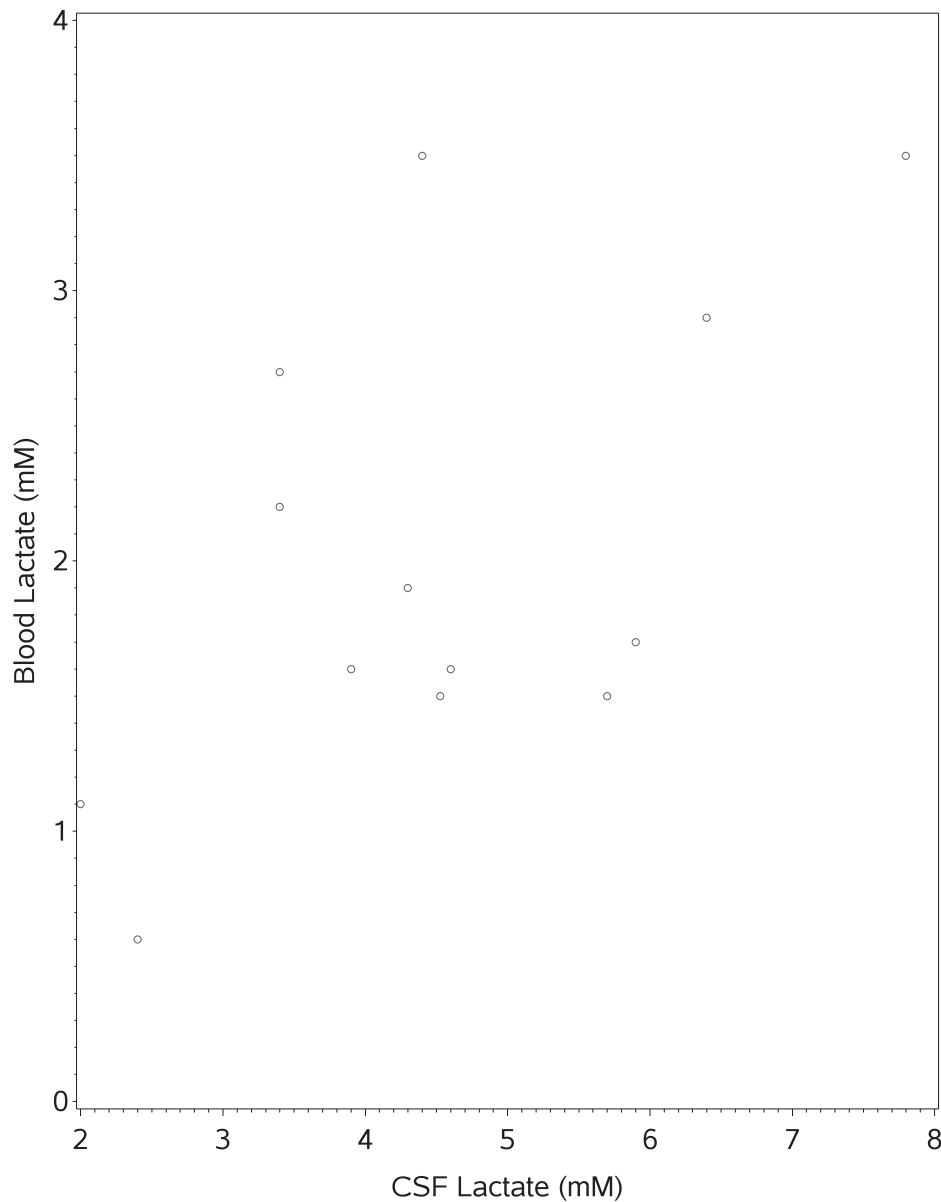


Fig 1. Scatterplot of blood lactate levels versus CSF lactate levels.

<https://doi.org/10.1371/journal.pone.0216287.g001>

Table 6. Results of normality tests for example data.

Normality Test			
Equation	Test Statistic	Value	Prob
blood	Shapiro-Wilk W	0.93	0.3457
csf	Shapiro-Wilk W	0.97	0.9115
	Mardia Skewness	1.81	0.7698
	Mardia Kurtosis	-0.77	0.4386
	Henze-Zirkler T	0.35	0.4613

<https://doi.org/10.1371/journal.pone.0216287.t006>

rejects a given aspect of normality at $\alpha = 0.05$. We do note that there is low power to do so given $n = 13$.

The estimated correlation between blood and CSF lactate levels was $\hat{\rho} = 0.57$. For the test $H_0: \rho = 0$ versus $H_1: \rho > 0$ the p-values were 0.0198, 0.0398, 0.0770 and 0.0628 for the Fisher's z transformation based test, the large sample test, the bootstrap test and the studentized permutation test of DiCiccio and Romano [16], respectively. The bootstrap and permutation tests used 5,000 resamples. If we were testing at level $\alpha = 0.05$ there is agreement between the permutation and bootstrap test of not rejecting H_0 while the test based on Fisher's z transformation and the large sample approximation indicate to reject H_0 . Given the small sample size and the results of our simulation study we would recommend that the permutation or bootstrap p-values are likely the more useful calculations in terms of drawing the correct conclusion, particularly given the closeness of the values between the Fisher's z transformation based p-value and the large sample approximation, which we demonstrated did not control the Type I error in small samples.

For the test $H_0: \rho = 0.3$ versus $H_1: \rho > 0.3$ the p-values were 0.149, 0.1399, 0.1792 and not applicable (doesn't generalize) for the Fisher's z based test, the large sample test, the bootstrap test and the studentized permutation test of DiCiccio and Romano [16], respectively. All three tests do not reject H_0 at $\alpha = 0.05$.

Discussion and conclusion

In this note we present a robust bootstrap test with good Type I error control for testing the general hypothesis $H_0: \rho = \rho_0$. The test was inspired by the studentized permutation given by DiCiccio and Romano [16] for testing $H_0: \rho = 0$, which was proven to be exact in certain scenarios and asymptotically correct overall. We believe that statistical software packages should employ both the bootstrap test and the studentized permutation given by DiCiccio and Romano [16] as the default tests over tests based on bivariate normal assumptions. In addition, it should be noted that the bootstrap test can be inverted to form a confidence interval about ρ . This will be examined in our future work.

Acknowledgments

This work was supported by Roswell Park Cancer Institute and National Cancer Institute (NCI) grant P30CA016056, National Science Foundation under grant NSF IIS-1514204 and NRG Oncology Statistical and Data Management Center grant U10CA180822 and IOTN Moonshot grant U24CA232979-01. We wish to thank the Academic Editor and two reviewers for their thoughtful critiques, which led to an improved version of this paper.

Author Contributions

Conceptualization: Alan D. Hutson.

Data curation: Alan D. Hutson.

Formal analysis: Alan D. Hutson.

Funding acquisition: Alan D. Hutson.

Investigation: Alan D. Hutson.

Methodology: Alan D. Hutson.

Project administration: Alan D. Hutson.

Resources: Alan D. Hutson.

Software: Alan D. Hutson.

Validation: Alan D. Hutson.

Visualization: Alan D. Hutson.

References

1. Kendall M. and Stuart A. (1979) *The Advanced Theory of Statistics*. Charles Griffin & Company Limited, Buck, England.
2. Stigler S. M. (1986) *The History of Statistics: The Measurement of Uncertainty before 1900*. The Belknap Press of Harvard University Press, Cambridge, MA.
3. Pearson K. (1896) Mathematical contributions to the theory of evolution, III: regression, heredity and panmixia. *Philosophical Transactions of the Royal Society of London (A)* 187 253–318. <https://doi.org/10.1098/rsta.1896.0007>
4. Fisher R. A. (1915) Frequency-distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10 507–521. <https://doi.org/10.2307/2331838>
5. “Student” (1908) On the probable error of a correlation coefficient. *Biometrika* 6 302–310. <https://doi.org/10.1093/biomet/6.2-3.302>
6. Serfling R. J. (1980) *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York, NY.
7. Gayen A. K. (1951) The Frequency Distribution of the Product-Moment Correlation Coefficient in Random Samples of Any Size Drawn from Non-Normal Universes. *Biometrika* 38 219–247. <https://doi.org/10.2307/2332329> PMID: 14848124
8. Ogasawara H. (2006) Asymptotic expansion of the sample correlation coefficient under nonnormality. *Computational Statistics & Data Analysis* 50 891–910. <https://doi.org/10.1016/j.csda.2004.11.001>
9. Pearson E. S. (1929). Some notes on sampling tests with two variables. *Biometrika* 21:337–360. <https://doi.org/10.2307/2332565>
10. Havlicek L. L. and Peterson N. L. (1977) Effect of the Violation of Assumptions Upon Significance Levels of the Pearson r . *Psychological Bulletin* 84 373–377.
11. Berry K. J. and Mielke P. W. Jr., (2000) A Monte Carlo investigation of the Fisher Z transformation for normal and nonnormal distributions. *Psychological Reports* 87 1101–1114. <https://doi.org/10.2466/pr0.2000.87.3f.1101> PMID: 11272750
12. Bishara A. J. and Hittner J. B. (2012) Testing the significance of a correlation with non-normal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods* 17: 399–417 PMID: 22563845
13. Bishara A. J. and Hittner J. B. (2015) Reducing Bias and Error in the Correlation Coefficient Due to Non-normality. *Educational and Psychological Measurement* 75 785–804. <https://doi.org/10.1177/0013164414557639> PMID: 29795841
14. Bishara A. J. and Hittner J. B. (2017) Confidence intervals for correlations when data are not normal. *Behavioral Research* 49 294–309. <https://doi.org/10.3758/s13428-016-0702-8>
15. Edgell S. E. and Noon S. M. (1984) Effect of Violation of Normality on the t Test of the Correlation Coefficient. *Psychological Bulletin* 95 576–583.
16. DiCiccio T. J. and Romano J. P. (2017) Robust Permutation Tests For Correlation And Regression Coefficients. *Journal of the American Statistical Association* 112 1211–1220. <https://doi.org/10.1080/01621459.2016.1202117>
17. Lehmann E. L. (1991) *Testing Statistical Hypotheses, 2nd Edition*. Wadsworth & Brooks/Cole, Belmont, CA.
18. Hayes A. F. (1996) Permutation Test Is Not Distribution-Free: Testing $H_0: \rho = 0$. *Psychological Methods* 1 184–198.
19. Tuğran E., Kocak M., Mirtağoğlu H., Yiğit S. and Mendes M. (2015) A Simulation Based Comparison of Correlation Coefficients with Regard to Type I Error Rate and Power. *Journal of Data Analysis and Information Processing* 3 87–101. <https://doi.org/10.4236/jdaip.2015.33010>
20. Tschuporow A. A. (1925). *Grundbegriffe und Grundprobleme der Korrelationstheorie*. Leipzig: Teubner (English translation as the *Mathematical Theory of Correlation*, William Hodge and Co. Ltd., 1939).
21. Geary R. C. (1947) Testing for normality. *Biometrika* 34 209–242. <https://doi.org/10.2307/2332434> PMID: 18918691

22. Lee W.-C. and Rogers J. L. (1998) Bootstrapping Correlation Coefficients Using Univariate and Bivariate Sampling. *Psychological Methods* 3 91–103.
23. Neiberger R. E., George J. C., Perkins L., Theriaque D. W., Hutson A. D., Stacpoole P. W. (2002) Renal Manifestations of Congenital Lactic Acidosis. *American Journal of Kidney Diseases* 39 12–23. <https://doi.org/10.1053/ajkd.2002.29872> PMID: 11774096