British Journal of Applied Science & Technology
4(12): 1808-1822, 2014

SCIENCEDOMAIN *international*
www.sciencedomain.org

Sample Size for Correlation Studies When Normality Assumption Violated

Azadeh Saki[1] and Hamed Tabesh[1*]

[1]Department of Biostatistics and Epidemiology, School of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

Authors' contributions

This work was carried out in collaboration between both authors. Author AS Conceptualized and participated in the design of the study and simulation study. Author HT participated in the design of the study and redaction. Both authors read and approved the final manuscript.

Original Research Article

Received 17th November 2013
Accepted 28th January 2014
Published 26th February 2014

ABSTRACT

Aims: Assessing the correlation between two variables is very important in observational and experimental researches. "How many sample size was required?" is one of the preliminary questions for correlation studies. Although achieving normality is rare the available techniques calculated the sample size based on fisher transformation statistics that supposed the bivariate normal distribution. This study conducted to find the sample size of correlation studies when the distribution of population is not bivriate normal.
Methodology: A Simulation study was used to compared the required sample size of the correlation test for ρ= 0.1, 0.2 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. Samples are drawn from bivariat normal, skewed, highly skewed, and heavy tailed distributions. The bias, variance, Mean Square Error (MSE), and the rejection rate of the fisher test for 10000 sample correlation coefficient were calculated. To achieving the nominal power the sample size was increased gradually.
Results: Both the mean Bias and Mean Square Error of Sample correlation increased when the bivariate distribution is not normal. The correlation test is robust against minor and major departures from bivariate normal assumption when the sample size of study was sufficiently large. To find the significance correlation between two variables with nominal power the required sample size depending on ρ and population distribution approximately 10 to 30 percent increased.
Conclusion: Departure from normality affected both accuracy and precision of sample correlation. Normality is not an ignorable assumption for correlation studies and it is

**Corresponding author: E-mail: tabesh-h@ajums.ac.ir;*

important to have information about population distribution to determine the sample size when designing a study.

Keywords: Correlation coefficient; fisher test; sample size; bias; simulation; normality.

1. INTRODUCTION

Experimental and observational studies are leading us to better results when they are carefully planned. Satisfied planning has many aspects, such as precise definitions and operationalization of the problem, selection of experimental or observational units from the appropriate population; correct randomization and the procedures must be followed carefully. Finally, the sample size of the study must be adequate, relative to the study goals. It must be "big enough" that an effect of such magnitude as to be of scientific significance will also be a waste of resources for not having the capability to produce useful results.

Assessing the correlation between variables is the main object of some medical, health and social researches. Achieving normality is rare, the available techniques calculate the sample size based on Fisher transformation statistics that assumed the joint distribution of variables is bivariate normal [1-5]. In most of these studies, however two variables are not followed the jointly normal distribution. For such studies, sample size formula which is based on Fisher transformation statistics is not adequate and its results may not be appropriate.

For example, to determine the correlation between frequency of breastfeeding and duration of each suckling during 24-hr period among exclusively breastfed infants [6], where they were not bivariate normal the researcher tries to determine big enough sample size but how it could do? In terms of statistical theory, this problem may be restated as follows.

We consider random bivariate samples $(x_i,\ y_i)$, i=1,2,…,n from continuous bivariate population e.g., suckling duration and frequency of breastfeeding during 24- hours. F(x,y) denote the joint distribution of (X, Y) and $\rho_{x,y}$ denote the correlation between X, Y. we intend to obtain adequate sample size to test:

$$H_0:\ \rho_{x,y}=0 \text{ against } H_1:\ \rho_{x,y}\neq 0. \qquad \text{(I)}$$

For the above mentioned problem, we know that for a bivariate normal population, sample size estimation which use Ronald Fisher's classic transformation to normalize the distribution of Karl Pearson's correlation coefficient (1-2) is common:

$$n=\left[\frac{Z_{\alpha/2}+Z_{\beta}}{\omega}\right]^2+3 \qquad \text{(II)}$$

Where:

$$\omega=\frac{1}{2}\ln\frac{1+r}{1-r}$$

and r is sample correlation between X, Y under hypothesis H_1 (3).

We are unaware of any studies to date that have focused on (e.g. It seems there are no explicit sample size formula for calculating adequate sample size to test (I) [7-9]. In this paper, simulation studies based on correlation strength and bivariate distribution of variables in the study are used to assess observed power of the study.

2. METHODOLOGY

First we calculated sample size for specified circumstances then simulations were run for bivariate normal, skewed, highly skewed and heavy tailed distributions, observed power for testing (I) with any of mentioned distributions calculated. Then sample size increased where nominal power achieved.

We illustrate here with examples of quick calculations for a study and then discuss a general approach of simulations in detail. For example let ρ=0.2, α=0.05, β=0.2 and suppose joint distribution of X, Y bivariate normal, where r, is Pearson correlation coefficient that is an estimation of correlation between X, Y and α, β are probability of type I and type II error respectively. Sample size, n=194, obtained by using formula (II). Simulations were run for bivariate normal with ρ=0.2, also simulations were run for skewed, heavy tailed and highly skewed distributions. Monte- Carlo rejection proportion for bivariate normal distribution was 80%, which was equal to specified nominal power, 1-β, but rejection proportion for skewed, highly skewed, and heavy tailed were less than 80%. So simulation were run more times for skewed, highly skewed, and heavy tailed distributions with greater sample sizes until rejection proportion approximately reach to 80%, specified nominal power.

2.1 SIMULATION STUDY

Simulation were run for bivariate normal with absolute ρ= 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. Also simulations were run for skewed, heavy tailed, and highly skewed distribution generated using the g- and –h distribution, i.e. generating z_{ij} from a bivariate normal distribution and setting

$$x_{ij} = \frac{1}{g}\exp\{gz_{ij}\}^{-1}\exp\{\frac{1}{2}hz_{ij}^{2}\}\ .$$

For g=0 this expression is taken to be:

$$x_{ij} = z_{ij}\exp\{\frac{1}{2}hz_{ij}^{2}\}$$

As the g- and- h distribution provides a convenient method for considering a very wide range of situations corresponding to both symmetric and highly asymmetric distributions. The case g=h=0 corresponds a normal distribution. The case g=0 corresponds to a symmetric distribution, and, as g increases, the skewness increases as well. For example, with g=0.5 and h=0, the skewness is 1.75 which is great [10]. Similar to Tabesh et al. (2010), in the study, simulations were run with g=0.5, and g=0.5 to span the range of skewness values that seems to occur in practice [11].

The parameter h determines the heaviness of the tails. As h increase, the heaviness increases as well. With h=0.2 and g=0, the kurtosis equals 36. This might seem extreme, but even higher values were found by Wilcox, so our simulations were run for h=0.2 (10).

The simulation procedure for each ρ begins with a different sample sizes were desired at least 70% nominal power at α=0.05 for bivariate normal distribution. Then the correlation coefficients for bivariate samples extracted from normal, skewed, highly skewed, and heavy tailed distribution were estimated and then the rejection rate based on the Z fisher transformation was calculated for each distribution.

To evaluate the impact of departure from bivariate normal assumption on the accuracy, and precision of the estimated correlation coefficient; bias, variance, and mean square error of 10,000 sample correlations at different sample size for ρ=0.1,…, 0.9 were calculated as follow.

1- The Expected Bias of estimated correlation coefficient :

$$Bias\,(r,\rho) = E(r-\rho) = \frac{\sum_{i=1}^{I}(r_i-\rho)}{I}$$

2- The Variance of estimated correlation coefficient:

$$Var(r) = E(r-E(r))^2 = \frac{\sum_{i=1}^{I}(r_i-\bar{r})^2}{I}$$

3- The Mean Square Error of estimated correlation coefficient:

$$MSE\,(r) = E(r-\rho)^2 = \frac{\sum_{i=1}^{I}(r_i-\rho)^2}{I}$$

Where **I**=10000, is the number of iterations in simulation procedure.

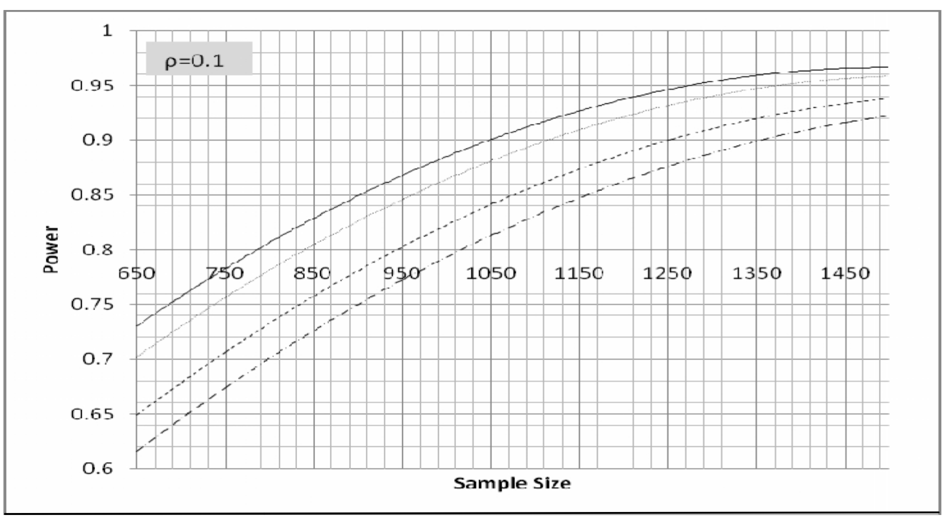
3. RESULTS AND DISCUSSION

Bias, variance, and MSE of 10,000 sample correlations at different sample size were calculated for ρ=0.1,…, 0.9 (Tables 1-9). As shown in Tables 1-9 the sample correlations are bias for all values of ρ and for all type of population distribution. As the sign of bias is uniformly negative we can conclude that the sample correlation is always an under estimation of population correlation. The minimum and maximum bias, variance and MSE are observed for normal and highly skewed bivariate distributions, respectively. The observed bias, variance and MSE of sample correlation tends to 0, when the sample size increases for all types of population distributions. Figs. 1-9 plot the power of correlation test, at the significance level α=0.05, as a function of the study sample size where the distribution of (X,Y) is bivariate normal, skewed, highly skewed, and heavy tailed for ρ=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, respectively. The lower and upper values of horizontal axis (sample size) at each Figure are according to achieve at least 0.60 to 0.95 powers for bivariate normal distribution with a given value of ρ. The power function of sample correlation from bivariate normal distribution at all Figures is upper than other distributions.

For ρ=0.1 the observed powers to test H0: ρ=0, when n=600 are 0.69, 0.66, 0.58, and 0.61 for bivariate normal, skewed, highly skewed distributions and according to Fig. 1 the power reached to 80% at n= 790, 845, 1020, 945 for these distributions, respectively. When the distributions are skewed, heavy tailed and highly skewed the required sample sizes for correlation test are respectively 7%, 20%, and 29% more than calculated sample size. The expected Bias is uniformly negative but the MSE tends to 0 as n increases, so the Pearson correlation coefficient is a bias but consistent estimator for population correlation for ρ=0.1.

Table 1. Mean bias, Variance, MSE, and Monte Carlo rejection proportion for ρ=0.1

N		Normal	skewed	Highly skewed	Heavy tailed
600	Rejection rate	0.6871	0.6590	0.5753	0.6099
	Bias	-0.0002	-0.0029	-0.0107	-0.0086
	Variance	0.0017	0.0017	0.0018	0.0017
	MSE	0.0033	0.0034	0.0038	0.0035
800	Rejection rate	0.8057	0.7831	0.7027	0.7333
	Bias	-0.0004	-0.0031	-0.0110	-0.0085
	Variance	0.0010	0.0010	0.0011	0.0010
	MSE	0.0025	0.0025	0.0028	0.0026
1000	Rejection rate	0.8870	0.8673	0.7992	0.8275
	Bias	-0.0007	-0.0020	-0.010	-0.008
	Variance	0.0010	0.0010	0.0011	0.0010
	MSE	0.0020	0.0020	0.0023	0.0021
1200	Rejection rate	0.9333	0.9175	0.8577	0.8844
	Bias	-0.0006	-0.0029	-0.0110	-0.0087
	Variance	0.0008	0.0008	0.0009	0.0008
	MSE	0.0016	0.0017	0.0019	0.0018
1400	Rejection rate	0.9643	0.9546	0.9084	0.9286
	Bias	0.0002	-0.0025	-0.0104	0.0083
	Variance	0.0007	0.0007	0.0008	0.0007
	MSE	0.0014	0.0015	0.0017	0.0015



Normal .... Skewed ----- Heavy tailed -.-. Highly Skewed

Fig. 1. Power curves of correlation test for ρ= 0.1 when the bivriate distribution of population is normal, skewed, heavy tailed and highly skewed

For ρ=0.2 the observed powers to test (I), when n=150 are 0.69, 0.66, 0.59, and 0.62 for bivariate normal, skewed, highly skewed distributions and according to Fig. 2 the power reached to 80% at n= 195, 210, 250, 230 for these distributions, respectively. So the required sample sizes are 8%, 18%, and 28% more than normal distribution for skewed, heavy tailed and highly skewed distributions, respectively. So, the Pearson correlation coefficient is not an unbiased but it is a consistent estimator for population correlation where ρ=0.2 (Table 2)

Table 2. Mean bias, Variance, MSE, and Monte Carlo rejection proportion for ρ=0.2

N		Normal	Skewed	Highly Skewed	Heavy tailed
150	Rejection rate	0.6896	0.6641	0.5866	0.6203
	Bias	-0.0021	-.0066	-.0197	-.0155
	Variance	0.0062	0.0064	0.0074	.0066
	MSE	0.0123	0.0130	0.0151	0.0134
200	Rejection rate	0.8152	0.7852	0.7048	0.7512
	Bias	-0.0001	-.0049	-0.0187	-0.0144
	Variance	0.0045	0.0048	0.0055	0.0049
	MSE	0.0091	0.0096	0.0114	0.0010
250	Rejection rate	0.8913	0.8702	0.8034	0.8322
	Bias	-0.0002	-0.0047	-0.0181	-0.0147
	Variance	0.0037	0.0039	0.0045	0.0041
	MSE	0.0075	0.0079	0.0094	0.0084
300	Rejection rate	0.9344	0.9179	0.8601	0.8910
	Bias	-0.0003	-.0051	-.0191	-0.0151
	Variance	0.0032	0.0033	0.0038	0.0034
	MSE	0.0063	0.0066	0.0080	0.0070
350	Rejection rate	0.9670	0.9569	0.9111	0.9373
	Bias	-0.0004	-0.0052	-0.0192	-0.0156
	Variance	0.0026	0.0028	0.0032	0.0028
	MSE	0.0052	0.0056	0.0068	0.0059

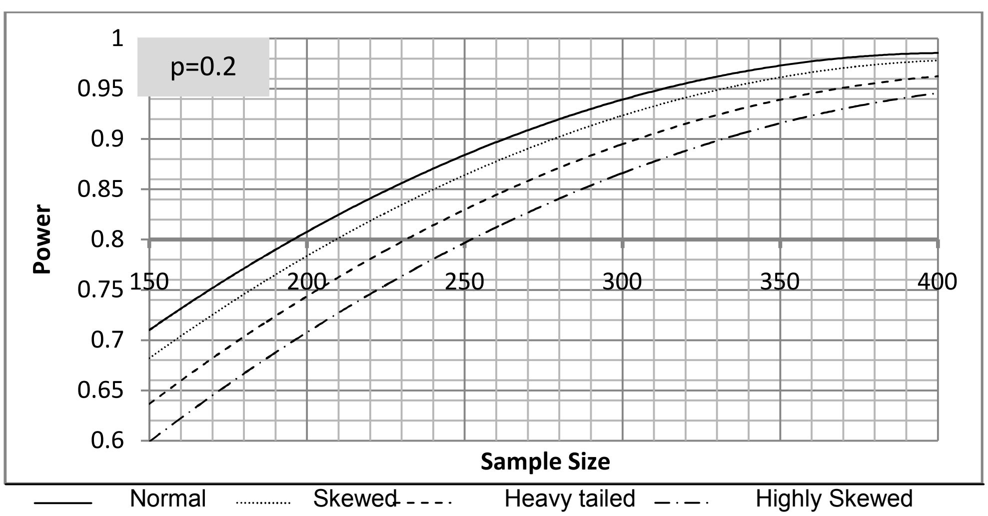


Fig. 2. Power curves of correlation test for ρ= 0.2 when the bivriate distribution of population is normal, skewed, heavy tailed and highly skewed

Fig. 3 shows that the required sample sizes for correlation test when ρ=0.3, at α=0.05 with 1-β=0.8 are 85, 92, 100, 108 for normal, skewed, heavy tailed and highly skewed distributions, respectively. So, the required sample size for correlation test must be increased between 8 to 27 percent according to population bivariate distribution.

Table 3. Mean bias, Variance, MSE, and Monte Carlo rejection proportion for ρ=0.3

N		Normal	Skewed	Highly Skewed	Heavy tailed
60	Rejection rate	0.6516	0.6279	0.5646	0.5902
	Bias	-0.0026	-0.0084	-0.0247	-0.0190
	Variance	0.0143	0.0153	0.0181	0.0154
	MSE	0.0286	0.0307	0.0367	0.0318
80	Rejection rate	0.7865	0.7660	0.6922	0.7264
	Bias	0.0009	-0.0051	-0.0291	-0.0165
	Variance	0.0107	0.0113	0.0134	0.0116
	MSE	0.0213	0.0227	0.0273	0.0235
100	Rejection rate	0.8663	0.8432	0.7727	0.8122
	Bias	-0.0013	-0.0074	-0.0247	-0.0189
	Variance	0.0084	0.0089	0.0107	0.0092
	MSE	0.0168	0.0179	0.0219	0.0188
120	Rejection rate	0.9194	0.9008	0.8391	0.8784
	Bias	-0.0016	-0.0078	-0.0253	-0.0098
	Variance	0.0070	0.0076	0.0092	0.0079
	MSE	0.0141	0.0153	0.0191	0.0163
140	Rejection rate	0.9509	0.9370	0.8866	0.9173
	Bias	-0.0018	-0.0080	-0.0259	-0.0214
	Variance	0.0061	0.0065	0.0079	0.0069
	MSE	0.0122	0.0131	0.0164	0.0143

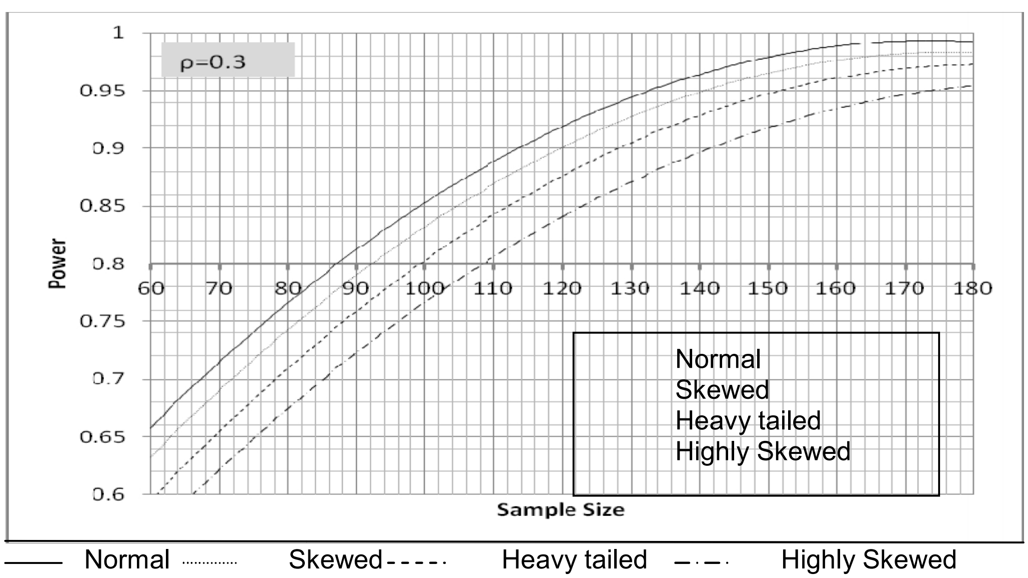


Fig. 3. Power curves of correlation test for ρ= 0.3 when the bivriate distribution of population is normal, skewed, heavy tailed and highly skewed

When the population correlation is ρ=0.4, the required sample size for correlation test increased 8, 15, and 26 percent for skewed, heavy tailed and highly skewed distribution (Table 4 & Fig. 4).

Table 4. Mean bias, Variance, MSE, and Monte Carlo rejection proportion for ρ=0.4

N		Normal	Skewed	Highly Skewed	Heavy tailed
35	Rejection rate	0.6877	0.6605	0.5975	0.6346
	Bias	-0.0035	-0.0108	-0.0291	-0.0197
	Variance	0.0210	0.0224	0.0267	0.0228
	MSE	0.0420	0.0449	0.0543	0.0460
45	Rejection rate	0.7963	0.7697	0.6998	0.7356
	Bias	-0.0028	-0.0096	-0.0282	-0.0209
	Variance	0.0164	0.0178	0.0218	0.0184
	MSE	0.0328	0.0358	0.0444	0.0372
60	Rejection rate	0.8970	0.8805	0.8228	0.8581
	Bias	-0.0015	-0.0085	-0.0277	-0.0212
	Variance	0.0122	0.0132	0.0162	0.0136
	MSE	0.0244	0.0256	0.0332	0.0277
70	Rejection rate	0.9405	0.9276	0.8762	0.9031
	Bias	-0.0016	-0.0083	-0.0301	-0.0240
	Variance	0.0103	0.0112	0.0140	0.0120
	MSE	0.0205	0.0224	0.0287	0.0244
80	Rejection rate	0.9604	0.9473	0.9097	0.9377
	Bias	-0.0039	-0.0094	-0.0292	-0.0242
	Variance	0.0091	0.0099	0.0126	0.0106
	MSE	0.0182	0.0201	0.0260	0.0217

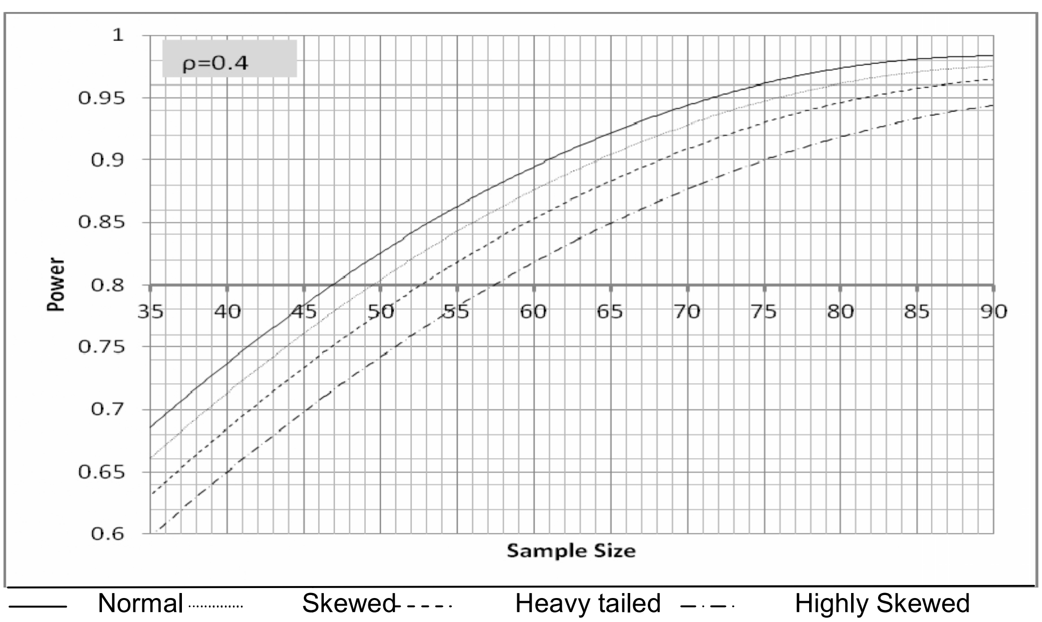


Fig. 4. Power curves of correlation test for ρ= 0.4 when the bivriate distribution of population is normal, skewed, heavy tailed and highly skewed

When ρ=0.5, and ρ=0.6, according to type of departure from bivriate normal distribution the required sample size for correlation test increased between 7% to 25%, and 5 to 21%, respectively (Figs. 5 & 6).

Table 5. Mean bias, Variance, MSE, and Monte Carlo rejection proportion for ρ=0.5

N		Normal	Skewed	Highly Skewed	Heavy tailed
22	Rejection rate	0.6913	0.6675	0.6113	0.6451
	Bias	-0.0082	-0.0148	-0.0325	-0.0240
	Variance	0.0210	0.0232	0.0285	0.0238
	MSE	0.0565	0.0609	0.0739	0.0621
30	Rejection rate	0.8267	0.8066	0.7496	0.7883
	Bias	-0.0066	-0.0114	-0.0304	-0.0227
	Variance	0.0204	0.0220	0.0270	0.0226
	MSE	0.0408	0.0442	0.0551	0.0457
36	Rejection rate	0.8869	0.8701	0.8167	0.8509
	Bias	-0.0054	-0.0123	-0.0312	-0.0239
	Variance	0.0171	0.0188	0.0234	0.0193
	MSE	0.0343	0.0377	0.0478	0.0393
42	Rejection rate	0.9379	0.9220	0.8734	0.9092
	Bias	-0.0048	-0.0126	-0.0326	-0.0244
	Variance	0.0142	0.0156	0.0197	0.0164
	MSE	0.0248	0.0313	0.0405	0.0334
48	Rejection rate	0.9600	0.9485	0.9127	0.9373
	Bias	-0.0043	-0.0114	-0.0313	-0.0256
	Variance	0.0125	0.0138	0.0364	0.0297
	MSE	0.0251	0.0278	0.0364	0.0297

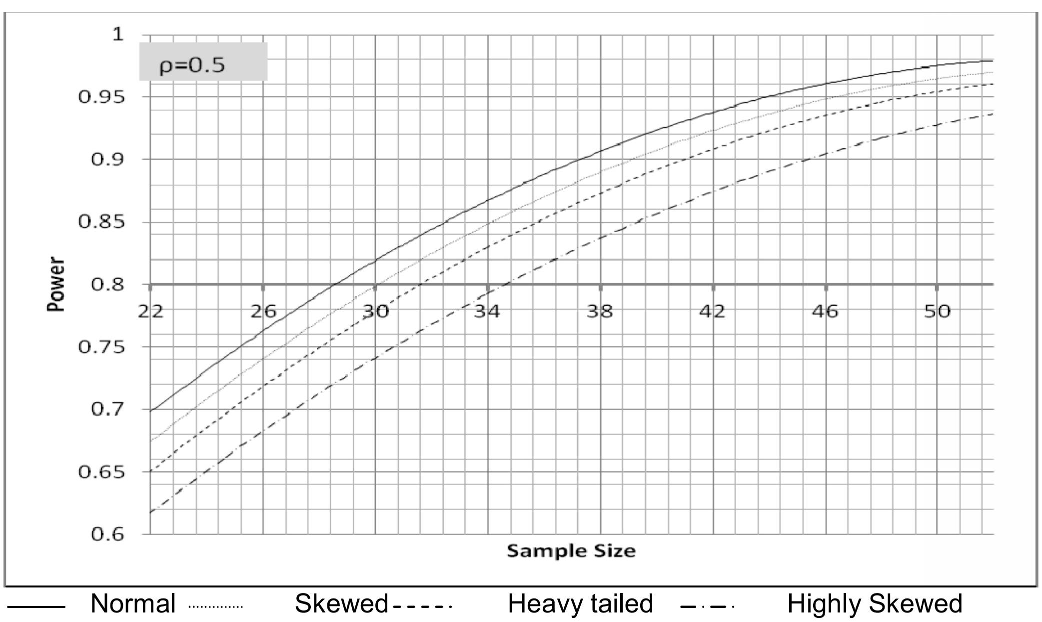


Fig. 5. Power curves of correlation test for ρ= 0.5 when the bivriate distribution of population is normal, skewed, heavy tailed and highly skewed

Table 6. Mean bias, Variance, MSE, and Monte Carlo rejection proportion for ρ=0.6

N		Normal	Skewed	Highly Skewed	Heavy tailed
14	Rejection rate	0.6688	0.6527	0.6130	0.6359
	Bias	-0.0150	-0.0196	-0.0347	-0.0259
	Variance	0.0361	0.0388	0.0461	0.0384
	MSE	0.0725	0.0781	0.0934	0.0776
18	Rejection rate	0.7963	0.7754	0.7232	0.7607
	Bias	-0.0107	-0.0173	-0.0341	-0.0246
	Variance	0.0267	0.0292	0.0357	0.0292
	MSE	0.0535	0.0587	0.0726	0.0590
24	Rejection rate	0.9039	0.8885	0.8392	0.8719
	Bias	-0.0090	-0.0149	-0.0319	-0.0246
	Variance	0.0192	0.0210	0.0268	0.0224
	MSE	0.0384	0.0423	0.0545	0.0454
28	Rejection rate	0.9425	0.9308	0.8930	0.9208
	Bias	-0.0055	-0.0127	-0.0316	-0.0232
	Variance	0.0161	0.0174	0.0228	0.0189
	MSE	0.0322	0.0357	0.0465	0.0383
34	Rejection rate	0.9726	0.9671	0.9412	0.9591
	Bias	-0.0035	-0.0107	-0.0298	-0.0227
	Variance	0.0133	0.0148	0.0193	0.0160
	MSE	0.0265	0.0296	0.0394	0.0324

Fig 6. Power curves of correlation test forρ= 0.6 when the bivriate distribution of population is normal, skewed, heavy tailed and highly skewed

For ρ>0.6 the required sample size for highly skewed distribution is about 15% more than calculated sample size (Figs. 7-9). But when the departure is minor there is no significance difference in the power of correlation test between normal and non normal distribution.

Table 7. Mean bias, Variance, MSE, and Monte Carlo rejection proportion for ρ=0.7

N		Normal	Skewed	Highly Skewed	Heavy tailed
10	Rejection rate	0.6804	0.6652	0.6319	0.6544
	Bias	-0.0220	-0.0263	-0.0384	-0.0316
	Variance	0.0369	0.0389	0.0452	0.0392
	MSE	0.0743	0.0786	0.0919	0.0796
14	Rejection rate	0.8557	0.8380	0.7922	0.8185
	Bias	-0.0163	-0.0208	-0.0343	-0.0279
	Variance	0.0251	0.0269	0.0326	0.0275
	MSE	0.0505	0.0543	0.0663	0.0557
18	Rejection rate	0.9347	0.9223	0.8896	0.9180
	Bias	-0.0093	-0.0152	-0.0306	-0.0232
	Variance	0.0178	0.0194	0.0246	0.0203
	MSE	0.0357	0.0392	0.0501	0.0412
22	Rejection rate	0.9707	0.9664	0.9466	0.9606
	Bias	-0.0099	-0.0149	-0.0304	-0.0251
	Variance	0.0140	0.0154	0.0201	0.0164
	MSE	0.0281	0.0312	0.0413	0.0336
26	Rejection rate	0.9893	0.9860	0.9743	0.9838
	Bias	-0.0067	-0.0129	-0.0259	-0.0236
	Variance	0.0114	0.0126	0.0167	0.0138
	MSE	0.0228	0.0254	0.0342	0.0282

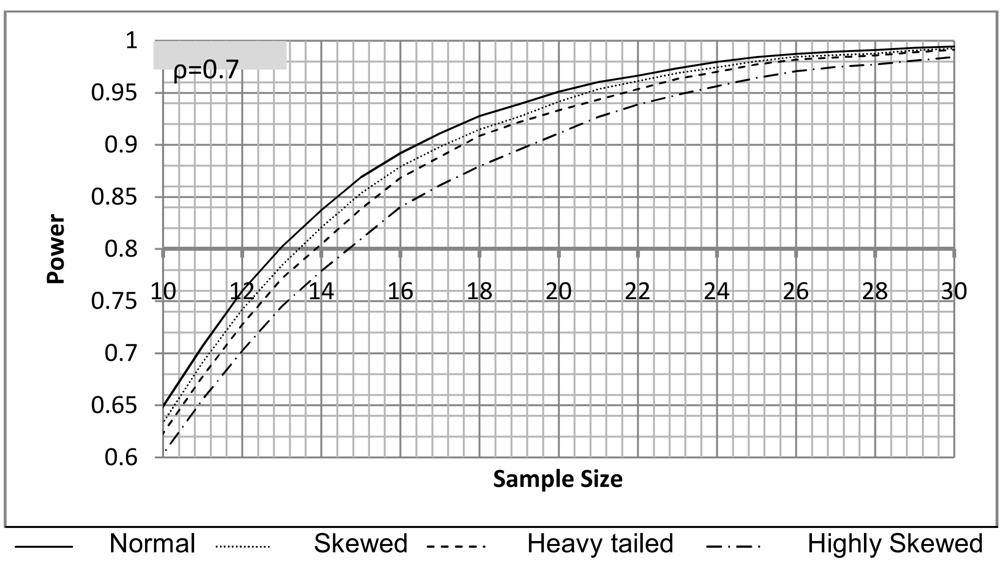


Fig. 7. Power curves of correlation test for ρ= 0.7 when the bivriate distribution of population is normal, skewed, heavy tailed and highly skewed

Table 8. Mean bias, Variance, MSE, and Monte Carlo rejection proportion for ρ = 0.8

N		Normal	Skewed	Highly Skewed	Heavy tailed
7	Rejection rate	0.6750	0.6638	0.6335	0.6557
	Bias	-0.0298	-0.0332	-0.0422	-0.0358
	Variance	0.0400	0.0415	0.0460	0.0417
	MSE	0.0809	0.0840	0.0938	0.0847
9	Rejection rate	0.8245	0.8126	0.7771	0.8028
	Bias	-0.0202	-0.0233	-0.0331	-0.0281
	Variance	0.0242	0.0259	0.0306	0.0261
	MSE	0.0489	0.0523	0.0622	0.0530
11	Rejection rate	0.9063	0.8946	0.8639	0.8880
	Bias	-0.0167	-0.0208	-0.0319	-0.0257
	Variance	0.0186	0.0200	0.0244	0.0205
	MSE	0.0374	0.0404	0.0498	0.0417
13	Rejection rate	0.9499	0.9443	0.9239	0.9408
	Bias	-0.0136	-0.0176	-0.0289	-0.0235
	Variance	0.0145	0.0157	0.0196	0.0163
	MSE	0.0292	0.0317	0.0400	0.0333
15	Rejection rate	0.9784	0.09742	0.9581	0.99727
	Bias	-0.0099	-0.0141	-0.0259	-0.0202
	Variance	0.0117	0.0129	0.0166	0.0134
	MSE	0.0236	0.0262	0.0339	0.0273

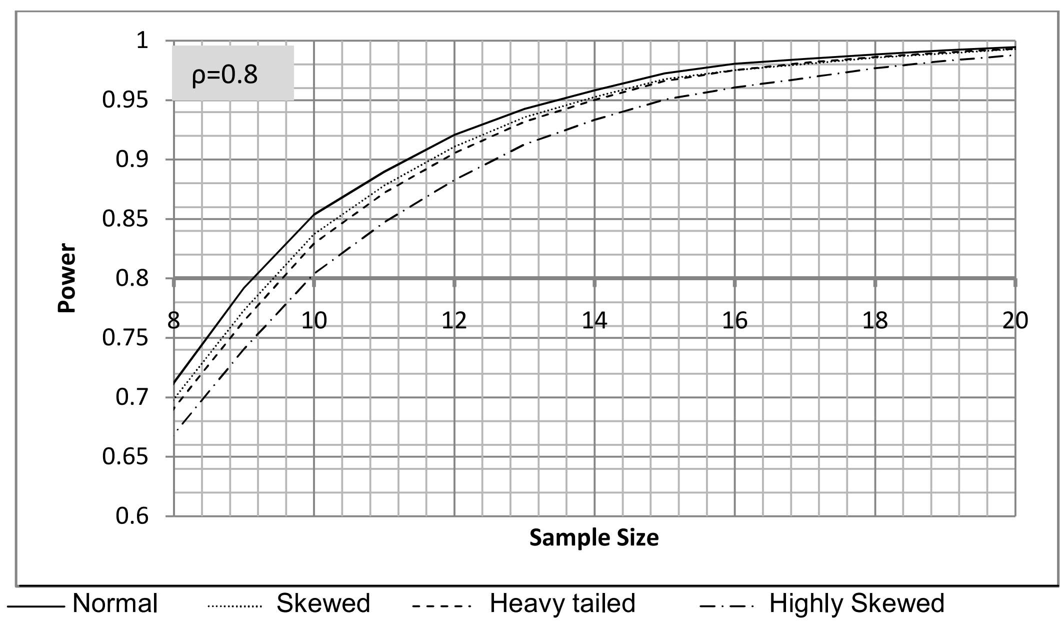


Fig. 8. Power curves of correlation test for ρ= 0.8 when the bivriate distribution of population is normal, skewed, heavy tailed and highly skewed

Table 9. Mean bias, Variance, MSE and Monte Carlo rejection proportion for ρ = 0.9

N		Normal	Skewed	Highly Skewed	Heavy tailed
5	Rejection rate	0.6565	0.6518	0.6375	0.6412
	Bias	-0.0342	-0.0354	-0.0397	-0.0367
	Variance	0.0321	0.0354	0.0349	0.0329
	MSE	0.0653	0.0669	0.0714	0.0671
6	Rejection rate	0.8193	0.8077	0.7808	0.7998
	Bias	-0.0228	-0.0248	-0.0303	-0.0268
	Variance	0.0193	0.0201	0.0228	0.0200
	MSE	0.0392	0.0408	0.0455	0.0408
7	Rejection rate	0.9045	0.8983	0.8767	0.8966
	Bias	-0.0179	-0.0203	-0.0263	-0.0209
	Variance	0.0146	0.0153	0.0173	0.0151
	MSE	0.0295	0.0310	0.0354	0.9327
8	Rejection rate	0.9406	0.9380	0.9218	0.9327
	Bias	-0.0168	-0.0184	-0.0238	-0.214
	Variance	0.0108	0.0113	0.0132	0.0116
	MSE	0.0219	0.0230	0.0269	0.0238
9	Rejection rate	0.9706	0.9679	0.9580	0.9668
	Bias	-0.0131	-0.0151	-0.0209	-0.0177
	Variance	0.0082	0.0088	0.0105	0.0089
	MSE	0.0166	0.0178	0.0215	0.0181

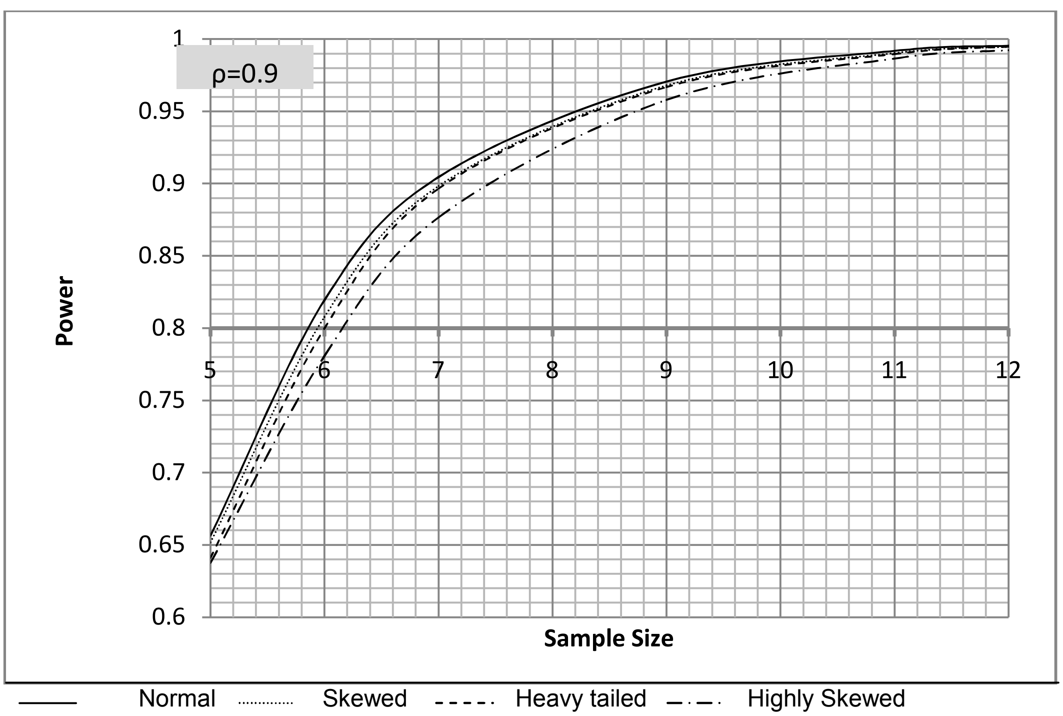


Fig. 9. Power curves of correlation test for ρ= 0.9 when the bivriate distribution of population is normal, skewed, heavy tailed and highly skewed

The maximum reduction in the power of correlation test is observed for the highly skewed populations, and then followed by heavy tailed and skewed distributions, respectively. The distances between power curves of four type of population distributions are similar for $\rho <= 0.6$, and for $\rho > 0.6$ power curves are closer together with increasing ρ.

Hence, the correlation test is a robust test against minor and major departures from bivariate normal assumption when the sample size of study was sufficiently large. Immediately this question arises: how many sample size is sufficiently large? The answer of this question is obtained from Figs. 1-9. Supposed that the calculated sample size for testing H0: $\rho = 0$ vs. H1: $\rho \neq 0$, at α level of type I Error and β level of type II Error base on normality assumption is n. when $\rho<=0.6$ the required sample size to achieving the desired power at α level for skewed, heavy tailed, and highly skewed distributions are approximately equal to 1.1n, 1.2n, and 1.3n, respectively. For $\rho > 0.6$ the percent of increasing the sample size is approximately 10 to 20 percent. Also, the results of our study show that despite the consistency of the sample correlation, it is a bias estimator for population correlation. This result is in agrees with previous studies (2,5,12-15). Gorsuch reported that the mean bias of sample correlation is an ignorable problem in study designing [12], our findings supported this claim. But the normality is not an ignorable assumption and the mean bias increased when the departure of normality is serious.

4. CONCLUSION

A direct conclusion of this study is that the power of correlation test with fisher transformation statistics reduced when the distribution of population is not bivariate normal. But, the desire power achieved by increasing the sample size. So, it is important to have information about population distribution to determine the sample size when designing a study.

Pearson correlation coefficient is a biased estimator of population correlation. This bias is not vigorous, but since the bivariate distribution is not normal, the absolute value of bias arises. This estimator is consistence and as sample size increases, the MSE goes to zero.

ACKNOWLEDGEMENTS

This work was financially supported by vice-chancellor for Research Affairs of Ahvaz Jundishapur University of Medical Sciences.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Fisher RA. The distribution of the partial correlation coefficient. Metron. 1924;3:329-32.
2. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika. 1915;10:507-21.
3. Desu MM, Raghavarao D. Sample size methodology. Boston: Academic Press; 1990.
4. Cohen J. Statistical power analysis for the behavioral sciences. Routledge;1988.
5. Bonett DG, Wright TA. Sample Size Requirements for Estimating Pearson, Kendall and Spearman Correlations. Psychometrik. 2000;65(1):23-8.

6. Saki A, Eshraghian MR, Tabesh H. Patterns of daily duration and frequency of breastfeeding among exclusively breastfed infants in Shiraz, Iran. A 6-month follow-up study using Baysian generalized linear mixed model. Global Journal of Health Sciences. 2013;5(2):123-34.
7. Stürmer GR, Cargnelutti Filho A, Guedes JVC, Stacke4 RF. Sample size for estimating the population of stink bugs in soybean crops. Revista Ciência Agronômica. 2014;45(1):155-67.
8. Cargnelutti Filho A, Toebe M, Burin C, Silveira TRd, Casarotto eG. Tamanho de amostra para estimação do coeficiente de correlação linear de Pearson entre caracteres de milho. Pesq agropec bras, Brasília,. 2010;45(12):1363-71.
9. Cargnelutti Filho A, Lopes SJ, Toebe M, Silveira TRd, Schwantes eIA. Tamanho de amostra para estimação do coeficiente de correlação de Pearson entre caracteres de Crambe abyssinica. Revista Ciência Agronômica. 2011;42(1):149-58.
10. Wilcox R. Simlation results on solutions to the multivariate Behrens-Fisher problem via trimmed means. The Statistician. 1995;44(2):213-25.
11. Tabesh H, Ayatollahi SMT, Towhidi M. A simple powerful bivariate test for two sample location problems in experimental and observational studies Theoretical Biology and Medical Modeling. 2010;7(13).
12. Gorsuch RL, Lehmann CS. Correlation Coefficients: Mean Bias and Confidence Interval Distortions. Journal of Methods and Measurement in the Social Sciences. 2010;1(2):52-65.
13. Fisher RA. On the "probable error" of a coefficient of correlation deduced from a small sample. Metron. 1921;1:3-32.
14. Filliben JJ. The Probability Plot Correlation Coefficient Test for Normality. Technometrics. 1975;17(1):111-7.
15. Donald W. Zimmerman, Bruno D. Zumbo, Williams RH. Bias in Estimation and Hypothesis Testing of Correlation. Psicológica. 2003;24:133-58.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=445&id=5&aid=3841