On Consistency and Limitation of Parametric and Non-parametric Paired Sample Tests

Khairul Islam and Tanweer J Shapla

Department of Mathematics and Statistics
Eastern Michigan University, Ypsilanti, MI 48197

Abstract

A paired t-test is a very common test for before-after measurements under the assumption that the difference of the before-after measurements form a sample form a normal distribution. In the violation of normality assumption, the popular alternatives are non-parametric sign test or signed rank test. How do these tests behave in the presence of varying skewness or varying correlation in the paired population? In this paper, we carry out a simulation study to evaluate the consistency and limitations of various tests of paired samples due to varying skewness and correlation in the paired populations from beta, exponential and gamma distributions.

Key words: Tests of paired samples, Skewness, Correlation, Power.

1. Introduction

In many real-life situations, we are interested in two populations whose measurements are paired. For example, let μ_1 and μ_2 be the means of two populations that represent before and after measurements. By letting $\mu_d = \mu_1 - \mu_2$, we wish to test $H_0: \mu_d = 0$ against $H_a: \mu_d \neq 0$. Given a sample of n paired measurements $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, let us define $d_i = x_i - y_i$; $i = 1,2, \dots, n$; $\bar{d} = \bar{x} - \bar{y}$ and $s_d^2 = \frac{\sum(d_i - \bar{d})^2}{n-1}$. If d_i's form a sample from a normal distribution, the test statistic for testing $H_0: \mu_d = 0$ against $H_a: \mu_d \neq 0$ is given by

$$T = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$$

Under the null hypothesis, T follows a t-distribution with $(n-1)$ degrees of freedom. In the violation of normality assumption, the common practice is to implement nonparametric Sign test or Singed Rank [1-11]. Two identical distributions have identical means, medians or other location parameters. Therefore, many researchers compare t-test with Sign and Singed Rank test [7-9], which require identical distribution assumption. However, the Sign Test or Signed Rank test are transformed tests in that they use signs and ranks assigned to original data values. Given these facts, it makes sense to compare performance of these tests with other transformed tests, e.g., log-transformed paired t-test.

In this paper, we investigate consistency and limitations of four paired sample tests – paired t-test, Sign test, Signed Rank test and log-transformed paired t-test due to varying skewness and correlation in the paired populations via a Monte Carlo simulation.

2. Methodology

We implement the four tests of paired samples by generating paired samples from beta, exponential and gamma distributions with specified correlations. For evaluating consistency and limitation, we consider gamma distribution with varying skewness. For the completeness of the presentation, we briefly discuss paired-sample generation procedure and the four underlying tests to be compared.

2.1 Generating paired-sample with specified characteristics

To implement any paired-sample test, we ensure that a paired-sample gets generated from a bivariate population with a specified level of correlation so as to represent before and after measurements or measures of two controlled situations. To do so, we utilize an open source software R [12] and generate paired-sample with a specified variance-covariance or correlation matrix using the **mvrnorm** function available from package **MASS**.

For generating paired samples from paired beta, exponential and gamma distributions, we considered the variance covariance matrix between the paired populations (X, Y) to be

$$\Sigma_{(X,Y)} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

where ρ is the specified correlation between the paired populations (X, Y). We specify a set variance-covariance matrices to allow varying values of correlations ρ at 0.2, 0.4, 0.6 and 0.8, chosen arbitrarily, to assess any sensitiveness of underlying tests on estimated Type I error rate and power. For example, for generating from $\Sigma_{(X,Y)}$ with correlation $\rho = 0.2$, the following R code has been implemented

Sigma <- matrix(0.2, nrow=2, ncol=2) + diag(2)*0.8

Once Sigma has been specified, one could generate paired samples (x, y) of a given sample size n from a bi-variate normal distribution using **mvrnom**() function with a specified mean mu (μ) as follows:

mu <- rep(0,2)

rawvars <- mvrnorm(n,mu=mu, Sigma=Sigma)

pvars <- pnorm(rawvars)

Finally, a paired sample from a bivariate beta distribution with two shape parameters 2 and 5 (i.e., mean 0.29) can be obtained by implanting the following R code:

pairs<-qbeta(pvars, shape1=2, shape2=5)

x<-pairs[,1]+diff[dif]

y<-pairs[,2]

Similarly, a paired sample from a bivariate exponential distribution with mean 2 (i.e., rate=0.5) can be obtained by implanting the following R code:

pairs<-qexp(pvars, rate=0.5)

x<-pairs[,1]

y<-pairs[,2]

While considering varying skewness, we utilize gamma distributions with mean 2 and skewness fixed at 1, 2, 3 and 4, arbitrarily, by choosing shape parameter θ_1 and scale parameter θ_2 as follows:

θ_1	θ_2	$\mu = \theta_1\theta_2$	Skewness, $\gamma = 2/\sqrt{\theta_1}$
4	1/2	2	1
1	2	2	2
4/9	9/2	2	3
1/4	8	2	4

For example, a paired sample (x, y) from a bivariate gamma distribution with mean=2, correlation=0.2 and skewness=1 can be obtained by implanting the following R code sequentially:

```
Sigma <- matrix(0.2, nrow=2, ncol=2) + diag(2)*0.8

mu <- rep(0,2)

rawvars <- mvrnorm(n,mu=mu, Sigma=Sigma)

pvars <- pnorm(rawvars)

pairs<-qgamma(pvars,shape=4, scale=1/2)

x<-pairs[,1]

y<-pairs[,2]
```

In this study, the paired-sample size n is fixed at 15, 20, 25, 30 and 35, taken arbitrarily, to evaluate the finite sample performance of underlying tests.

While simulating under the null models, we generate (x, y) with $\mu_1 = \mu_2 = 2$ for exponential and gamma distribution, and with $\mu_1 = \mu_2 = 0.29$ for beta distribution so that $\mu_d = \mu_1 - \mu_2 = 0$. Under alternative models, we generate (x, y) with $\mu_d = \mu_1 - \mu_2 = \delta$, with δ chosen arbitrarily. The values of δ are chosen to be 0.05, 0.1 for beta distribution, 0.2, 0.3, 0.4 for exponential distribution, and 0.1, 0.2, 0.3 for gamma distribution so that an adequate comparison can be made.

2.2 Paired t-test

The test defined in Section 1 (Introduction) is called a paired t-test. One can consult with any standard text [11] for this test. In this article, the paired t-test has been implemented by executing the following R code:

```
t.test(x,y,paired=TRUE)
```

In the violation of normality of paired differences, the conclusion of this test may be invalid or misleading. Following [8], we employ Sign test and Wilcoxon test, defined in sections 2.3 and 2.4, respectively, to overcome this limitation. Clearly, two identical distributions have identical means, medians or other location parameters. Motivated by these facts, researchers often compare t-test with Sign and Singed Rank test [7-9], which require identical distribution assumption. However, unlike [7-9], this article takes into account the varying correlation and skewness of the paired distributions in studying consistency and limitations of the underlying tests.

2.3 Sign-Test

The sign test [1-11] is a non-parametric equivalent to the one-sample or paired sample t-test. Sign test tests the null hypothesis that the paired samples have the same (identical) distribution. Note that two distributions with identical shape will have the same mean, median or other location parameter. Thus, researchers opt to compare t-test with non-parametric Sign or Sign ranked test. The Sign test takes into account the signs of paired differences of x-and y-values. Given a sample of n ordered pairs $\{(x_i,\ y_i): i = 1,2, \dots, n\}$, the Sign test for $H_0: \mu_d = 0$ against $H_0: \mu_d \neq 0$ can be implemented using the test statistics T^+, T^- or $minT$ defined as follows:

$T^+ = \#\ of\ (x_i\text{'s} > y_i\text{'s})$

$T^- = \#\text{ of } (x_i\text{'s} < y_i\text{'s})$ and

$minT = \min\{T^+,\ T^-\}$

It appears that $T^+$and T^- follow binomial distribution with parameters $n^*$and $\frac{1}{2}$, where $n^* = T^+ + T^- \leq n$ such that a pair with $x_i - y_i = 0$ is being ignored. For detail about Sign test one could consult [1-9]. In this article, we implement the Sign test by executing the SIGN.test function available from R package BSDA:

SIGN.test(x,y,md=0)

2.4 The Wilcoxon Singed Rank Test

The Wilcoxon Signed Rank test uses signs of $(x_i - y_i)$'s and their magnitudes as ranks. One can implement this test for $H_0: \mu_d = 0$ against $H_0: \mu_d \neq 0$ by using any of the following statistics:

$$T^+ = \sum_{i=1}^{n} u_i \times r_i$$

$$T^- = \sum_{i=1}^{n} (1 - u_i) \times r_i$$

where $u_i = 1$ if $x_i - y_i > 0$ and $r_i = rank\ of |x_i - y_i|$. It appears that $E[T^+] = E[T^-] = \frac{n(n+1)}{4}$ and $V[T^+] = V[T^-] = \frac{n(n+1)(2n+1)}{24}$. There is no explicit form of the exact probability distribution of T^+ and T^-, and thus one can use the statistic $Z = \frac{T^+ - \frac{n(n+1)}{4} - 0.5}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$ or $Z = \frac{T^- - \frac{n(n+1)}{4} - 0.5}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$, where $Z \sim \mathrm{N}(0,1)$ approximately. For detail about Signed Rank test one could consult with [1-11].

In this paper, we implement the Signed Rank test by executing the wilcox.test function available from the software R:

wilcox.test(x,y,mu=0, paired=TRUE)

2.5 Log-transformed paired *t*-test

The paired populations having identical distribution assumption should have identical means, medians or other location parameters. Given these facts, a Sing test or Signed Rank test is preferable to paired t-test for testing location parameter for non-normal paired sample. Recent studies [7-9] compare performance of Sing or Signed Rank test with paired *t*-test. The Sign test replaces the sample data values by the sign of (x-y) and Wilcoxon Singed Rank replaces the original sample data values by the signs of (x-y) and ranks assigned to |x-y|, and so they both are versions of transformed tests. Given these facts, it is inspiring to compare these tests (Sign test and Wilcoxon test) to a log-transformed paired t-test and report consistency and limitations, whatsoever, for better evaluation of their performances.

The log-transformation of ordered pair (x, y) can be made as follows:

- $u = log_{10}(x)$, where x is positively skewed.
- $u = log_{10}(x + c)$, where x is positively skewed with existence of zero values and c is a constant added to each data value so that log is defined.
- $u = log_{10}(K - x)$, where x is negatively skewed and $K = \max(x) + 1$.

A similar transformation v can be applied to y. Since exponential and gamma distributions are skewed, it is expected that the log-transformed paired t-test to ordered pair (u, v) would perform relatively better as compared to untransformed paired t-test and Sing or Signed Rank test.

2.6. Simulation and evaluation criteria

To evaluate performance of four underlying tests, we simulate paired sample from non-normal Beta, Exponential and Gamma distribution using the following steps:

(a) For beta and exponential distributions, fix paired sample size n and paired correlation ρ, and for gamma distribution, fix paired sample size n, paired correlation ρ and skewness γ. In simulations under null models, we fix means of the paired population at 0.29 for beta distribution and 2 for exponential and gamma distributions. Under alternative models, the mean difference for (i) beta distributions are fixed at 0.05, 0.10, (ii) exponential distributions are fixed at 0.2, 0.3, 0.4, and (iii) gamma distributions are fixed at 0.1, 0.2, 0.3. The paired correlation ρ is fixed at 0.2, 0.4, 0.6 and 0.8. In additions, the skewness of gamma distributions are fixed at 1, 2, 3 and 4.

(b) Generate a paired sample using any specification in (a) and implement paired t-test (t), Sign test (s), Signed Rank test (sr) and the log-transformed t-test (tt) and compute p-values of implemented four tests.

(c) Repeat steps (a)-(b) for 1000 times.

(d) Evaluate performance of underlying tests for 1000 repetitions in (c) via estimated powers and Type I error rates. Compute the estimated (i) power as the proportion of the rejection under the alternative models at 5% level of significance and (ii) Type I error rate as the proportion of rejection under the null models at 5% level of significance.

3. Analysis of simulation

The Type I error rates for beta, exponential and gamma distribution are presented in Tables 1.1, 1.2 and 1.3, respectively. The power of underlying tests for beta and exponential distributions are presented in Tables 2 and 3, respectively. Table 4 represents power of gamma distribution with varying mean difference δ and skewness γ when correlation ρ is 0.2. For better readability, the power of gamma distribution with varying mean difference (mdiff) and skewness (skew) are presented in Figures 4.1-4.4.

Table 1.1. Type I error rate of beta distribution with mean 0.29 for varying correlation ρ

n	t	s	sr	tt	t	s	sr	tt
	$\rho = 0.2$				$\rho = 0.4$			
15	0.03	0.03	0.03	0.04	0.04	0.03	0.04	0.04
20	0.04	0.04	0.04	0.04	0.05	0.04	0.04	0.04
25	0.04	0.03	0.05	0.04	0.04	0.04	0.04	0.04
30	0.06	0.05	0.06	0.05	0.04	0.04	0.05	0.05
35	0.06	0.05	0.06	0.05	0.05	0.04	0.05	0.04
	$\rho = 0.6$				$\rho = 0.8$			
15	0.04	0.04	0.04	0.05	0.04	0.04	0.05	0.04
20	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04
25	0.04	0.03	0.04	0.05	0.05	0.03	0.04	0.05
30	0.05	0.04	0.05	0.05	0.05	0.05	0.06	0.05
35	0.06	0.05	0.05	0.05	0.04	0.04	0.05	0.04

Table 1.2. Type I error rate of exponential distribution with mean 2 for varying correlation ρ

n	t	s	sr	tt	t	s	sr	tt
	$\rho = 0.2$				$\rho = 0.4$			
15	0.05	0.04	0.05	0.05	0.04	0.03	0.05	0.05
20	0.04	0.04	0.04	0.04	0.05	0.06	0.06	0.06
25	0.05	0.03	0.05	0.05	0.05	0.05	0.05	0.06
30	0.06	0.04	0.06	0.06	0.05	0.04	0.05	0.05
35	0.04	0.04	0.04	0.04	0.05	0.04	0.04	0.04
	$\rho = 0.6$				$\rho = 0.8$			
15	0.05	0.04	0.06	0.05	0.04	0.03	0.04	0.04
20	0.05	0.05	0.06	0.06	0.04	0.04	0.05	0.05
25	0.04	0.05	0.06	0.05	0.05	0.04	0.04	0.04
30	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.04
35	0.04	0.03	0.05	0.06	0.04	0.04	0.04	0.05

Table 1.3. Type I error rate of gamma distribution with mean 2 for varying correlation ρ and skewness γ.

	$\rho = 0.20$				$\rho = 0.40$				$\rho = 0.60$			
n	t	s	sr	tt	t	s	sr	tt	t	s	sr	tt
	γ =1											
15	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.04	0.03	0.03	0.04	0.04
20	0.05	0.04	0.05	0.05	0.05	0.04	0.05	0.06	0.06	0.06	0.06	0.06
25	0.05	0.03	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06
30	0.05	0.04	0.05	0.05	0.05	0.04	0.05	0.06	0.06	0.04	0.06	0.06
35	0.05	0.05	0.05	0.06	0.04	0.04	0.05	0.04	0.04	0.04	0.04	0.05
	γ =2											
15	0.04	0.04	0.04	0.04	0.05	0.04	0.06	0.05	0.05	0.03	0.05	0.04
20	0.04	0.04	0.04	0.04	0.05	0.04	0.06	0.05	0.04	0.04	0.05	0.05
25	0.04	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.04	0.06	0.06	0.06
30	0.05	0.04	0.05	0.06	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.04

35	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.05	0.05
	γ =3											
15	0.04	0.04	0.05	0.05	0.04	0.03	0.05	0.04	0.04	0.04	0.06	0.05
20	0.04	0.03	0.04	0.04	0.04	0.04	0.05	0.04	0.05	0.04	0.06	0.05
25	0.04	0.04	0.05	0.04	0.05	0.04	0.05	0.05	0.05	0.04	0.05	0.04
30	0.04	0.04	0.05	0.05	0.04	0.03	0.04	0.03	0.04	0.03	0.05	0.04
35	0.04	0.04	0.06	0.05	0.04	0.04	0.05	0.04	0.05	0.05	0.06	0.06
	γ =4											
15	0.04	0.04	0.05	0.04	0.03	0.04	0.04	0.05	0.03	0.03	0.05	0.03
20	0.04	0.04	0.05	0.04	0.03	0.04	0.04	0.03	0.03	0.04	0.04	0.04
25	0.05	0.04	0.06	0.05	0.06	0.05	0.06	0.05	0.04	0.04	0.05	0.04
30	0.04	0.04	0.05	0.05	0.04	0.06	0.05	0.05	0.04	0.04	0.05	0.04
35	0.04	0.04	0.05	0.04	0.04	0.05	0.04	0.06	0.05	0.05	0.05	0.05

Table 2. Power of beta distribution with varying mean difference δ and correlation ρ.

	$\delta = 0.05$				$\delta = 0.10$			
n	t	s	sr	tt	t	s	sr	tt
	$\rho = 0.2$							
15	0.14	0.09	0.14	0.16	0.42	0.29	0.41	0.50
20	0.18	0.14	0.17	0.22	0.54	0.39	0.52	0.63
25	0.22	0.15	0.21	0.27	0.65	0.50	0.65	0.81
30	0.27	0.20	0.27	0.36	0.76	0.57	0.75	0.86
35	0.31	0.23	0.31	0.41	0.80	0.61	0.79	0.91
	$\rho = 0.4$							
15	0.18	0.11	0.17	0.21	0.56	0.38	0.54	0.63
20	0.25	0.17	0.24	0.31	0.66	0.51	0.65	0.76
25	0.28	0.21	0.26	0.33	0.79	0.62	0.78	0.89
30	0.31	0.21	0.30	0.42	0.85	0.72	0.85	0.93
35	0.38	0.28	0.38	0.51	0.89	0.77	0.88	0.97
	$\rho = 0.6$							
15	0.23	0.15	0.22	0.28	0.69	0.53	0.69	0.77
20	0.34	0.25	0.33	0.40	0.84	0.68	0.83	0.90
25	0.39	0.32	0.39	0.48	0.91	0.81	0.91	0.96
30	0.44	0.36	0.45	0.57	0.96	0.88	0.96	0.98
35	0.53	0.40	0.53	0.67	0.98	0.91	0.98	0.99
	$\rho = 0.8$							
15	0.42	0.30	0.42	0.49	0.94	0.82	0.94	0.96
20	0.53	0.43	0.54	0.65	0.98	0.92	0.98	0.99
25	0.64	0.52	0.65	0.76	1.00	0.97	1.00	1.00
30	0.71	0.60	0.72	0.85	1.00	0.99	1.00	1.00
35	0.80	0.68	0.80	0.90	1.00	1.00	1.00	1.00

Table 3. Power of exponential distribution with varying mean difference δ and correlation ρ.

	$\delta = 0.2$				$\delta = 0.3$				$\delta = 0.4$			
n	t	s	sr	tt	t	s	sr	tt	t	s	sr	tt
	$\rho = 0.2$											
15	0.05	0.05	0.06	0.08	0.07	0.08	0.08	0.15	0.09	0.11	0.11	0.21
20	0.06	0.06	0.07	0.11	0.09	0.11	0.10	0.20	0.10	0.11	0.12	0.27
25	0.09	0.09	0.10	0.16	0.10	0.11	0.12	0.25	0.12	0.20	0.16	0.37
30	0.07	0.08	0.09	0.17	0.12	0.14	0.14	0.30	0.15	0.20	0.18	0.43
35	0.08	0.08	0.09	0.20	0.11	0.13	0.13	0.33	0.15	0.22	0.21	0.51
	$\rho = 0.4$											
15	0.06	0.06	0.07	0.11	0.08	0.09	0.10	0.18	0.11	0.12	0.14	0.25
20	0.08	0.08	0.08	0.15	0.09	0.14	0.13	0.25	0.10	0.15	0.14	0.32
25	0.07	0.10	0.10	0.17	0.09	0.14	0.12	0.28	0.16	0.21	0.18	0.43
30	0.08	0.11	0.10	0.20	0.13	0.17	0.16	0.36	0.19	0.28	0.25	0.54
35	0.10	0.11	0.11	0.25	0.13	0.19	0.18	0.42	0.20	0.28	0.26	0.62
	$\rho = 0.6$											
15	0.06	0.07	0.08	0.14	0.10	0.12	0.12	0.22	0.15	0.16	0.18	0.33
20	0.08	0.11	0.10	0.19	0.11	0.15	0.15	0.32	0.18	0.24	0.21	0.45
25	0.10	0.13	0.11	0.23	0.12	0.20	0.16	0.38	0.23	0.30	0.29	0.58
30	0.09	0.13	0.11	0.29	0.16	0.25	0.23	0.51	0.26	0.37	0.34	0.67
35	0.11	0.14	0.13	0.33	0.17	0.27	0.24	0.54	0.25	0.42	0.36	0.75
	$\rho = 0.8$											
15	0.08	0.12	0.12	0.20	0.15	0.21	0.20	0.34	0.22	0.28	0.28	0.51
20	0.10	0.16	0.13	0.29	0.19	0.27	0.23	0.52	0.28	0.40	0.37	0.67
25	0.12	0.21	0.17	0.38	0.23	0.36	0.31	0.61	0.34	0.49	0.44	0.79
30	0.15	0.23	0.20	0.46	0.23	0.40	0.34	0.72	0.41	0.57	0.53	0.88
35	0.16	0.28	0.24	0.53	0.25	0.44	0.40	0.79	0.45	0.61	0.59	0.92

Table 4. Power of gamma distribution with varying mean difference δ and skewness γ when correlation ρ is 0.2.

	$\delta = 0.1$				$\delta = 0.2$				$\delta = 0.3$			
n	t	s	sr	tt	t	s	sr	tt	t	s	sr	tt
	γ =1											
15	0.07	0.04	0.06	0.07	0.09	0.06	0.08	0.10	0.15	0.08	0.14	0.18
20	0.06	0.04	0.06	0.07	0.12	0.10	0.13	0.13	0.19	0.14	0.20	0.23
25	0.07	0.06	0.08	0.08	0.12	0.11	0.12	0.14	0.22	0.18	0.24	0.29
30	0.08	0.06	0.08	0.08	0.12	0.09	0.11	0.16	0.24	0.16	0.23	0.30
35	0.07	0.06	0.08	0.08	0.17	0.13	0.19	0.22	0.28	0.20	0.29	0.38
	γ =2											
15	0.05	0.03	0.06	0.06	0.06	0.06	0.07	0.08	0.09	0.07	0.09	0.14
20	0.05	0.05	0.05	0.07	0.06	0.06	0.07	0.12	0.08	0.09	0.08	0.18
25	0.05	0.06	0.06	0.08	0.07	0.08	0.07	0.15	0.08	0.10	0.10	0.24
30	0.05	0.05	0.05	0.08	0.05	0.08	0.06	0.15	0.09	0.11	0.11	0.27
35	0.05	0.04	0.06	0.08	0.08	0.09	0.09	0.21	0.11	0.16	0.15	0.35
	γ =3											

15	0.05	0.06	0.06	0.10	0.04	0.08	0.06	0.17	0.06	0.13	0.10	0.27
20	0.04	0.05	0.04	0.13	0.05	0.08	0.07	0.22	0.07	0.17	0.12	0.37
25	0.05	0.07	0.06	0.16	0.05	0.13	0.08	0.30	0.06	0.20	0.11	0.46
30	0.05	0.08	0.06	0.21	0.06	0.14	0.09	0.38	0.08	0.22	0.13	0.56
35	0.06	0.09	0.08	0.27	0.06	0.14	0.08	0.48	0.08	0.22	0.14	0.63
	γ =4											
15	0.04	0.08	0.06	0.23	0.03	0.14	0.09	0.38	0.06	0.20	0.13	0.48
20	0.05	0.14	0.08	0.33	0.04	0.21	0.10	0.53	0.06	0.27	0.14	0.64
25	0.06	0.17	0.08	0.46	0.05	0.25	0.11	0.63	0.05	0.33	0.14	0.77
30	0.03	0.18	0.06	0.59	0.05	0.31	0.14	0.74	0.06	0.37	0.16	0.85
35	0.05	0.18	0.08	0.62	0.04	0.29	0.12	0.84	0.07	0.44	0.20	0.92

Fig 4.1: Power of gamma distribution with varying mean difference (mdiff) and skewness (skew) when correlation ρ= 0.2.

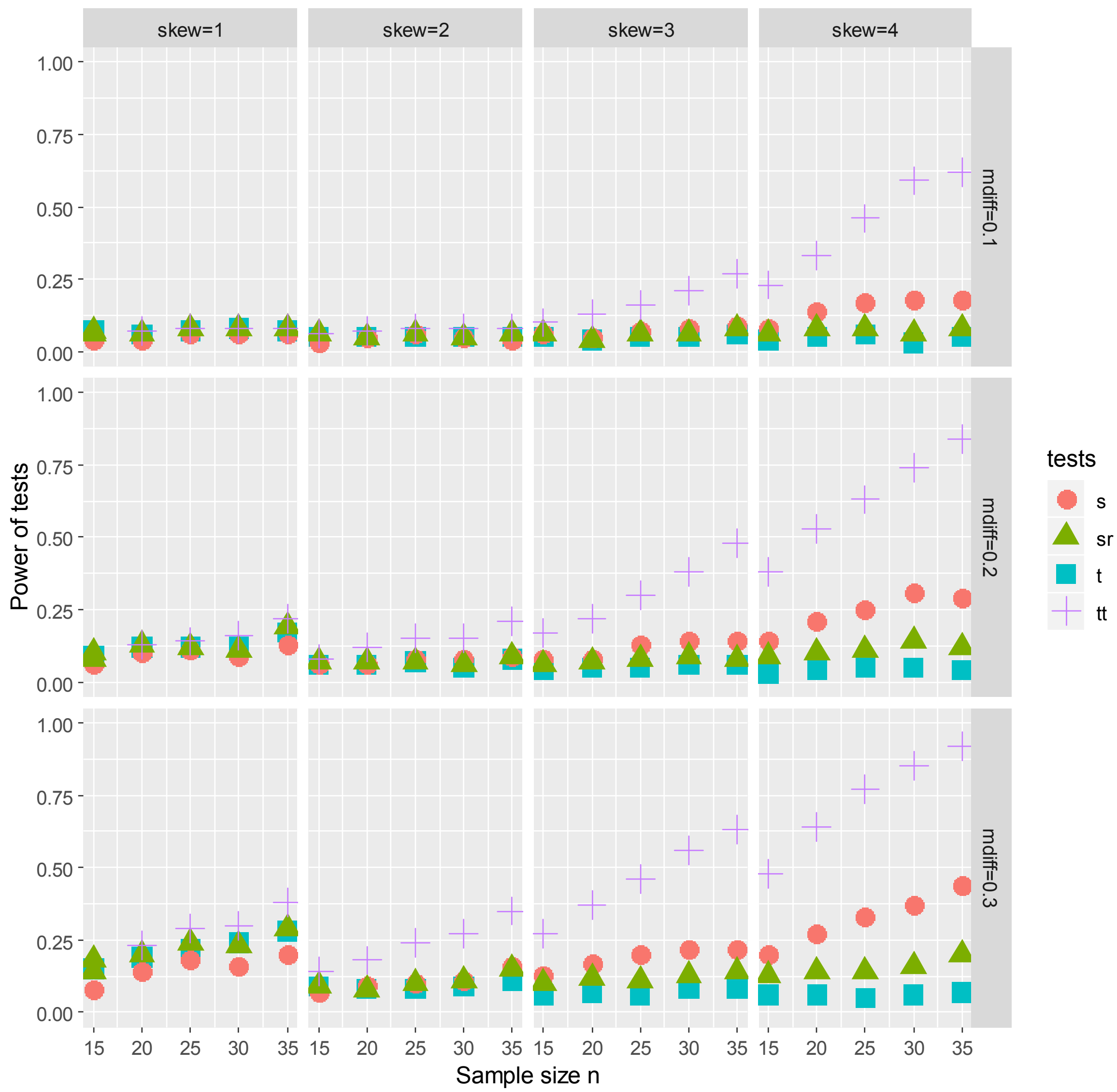


Fig 4.2: Power of gamma distribution with varying mean difference (mdiff) and skewness (skew) when correlation ρ= 0.4.

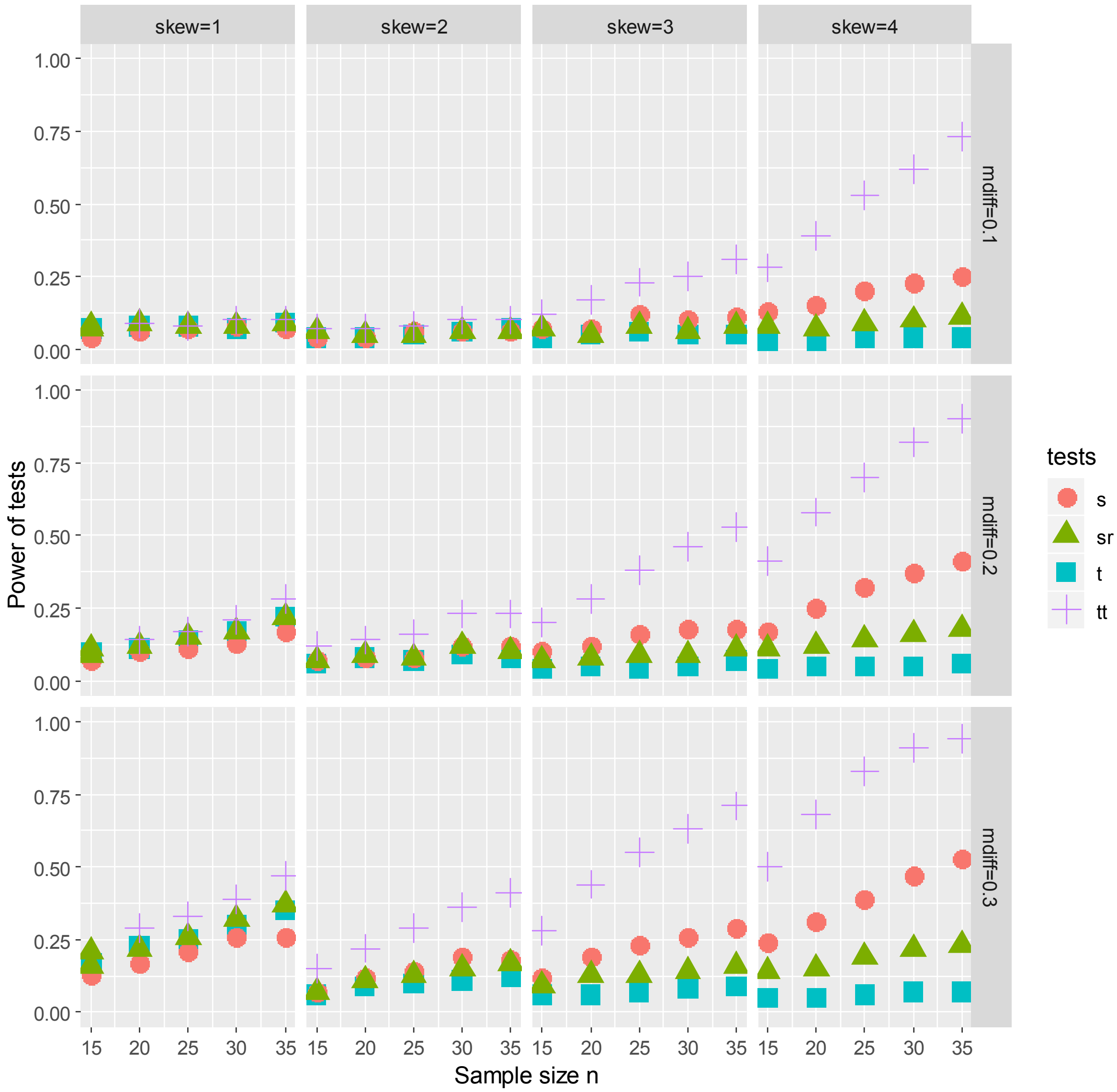


Fig 4.3: Power of gamma distribution with varying mean difference (mdiff) and skewness (skew) when correlation ρ= 0.6.

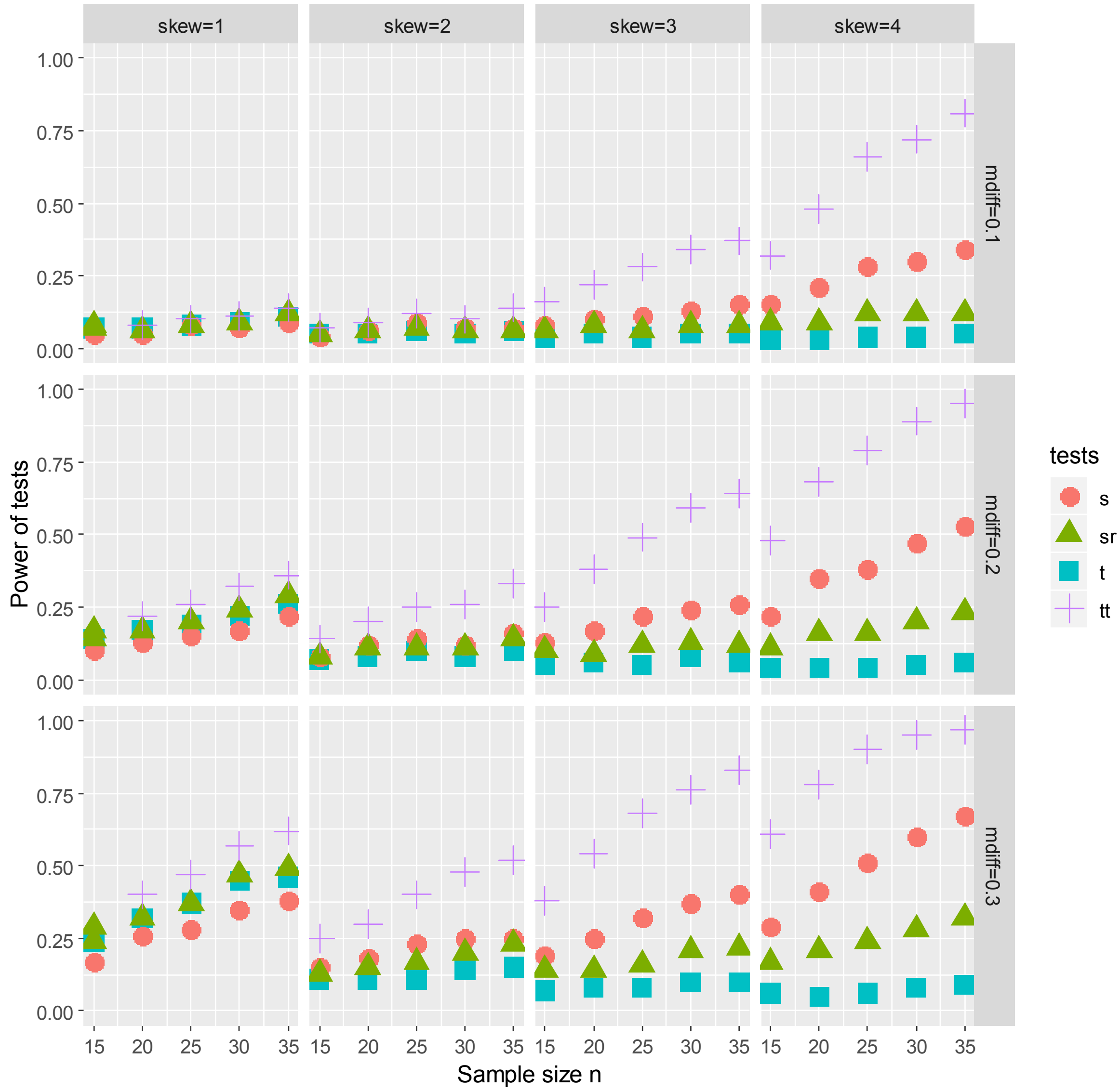
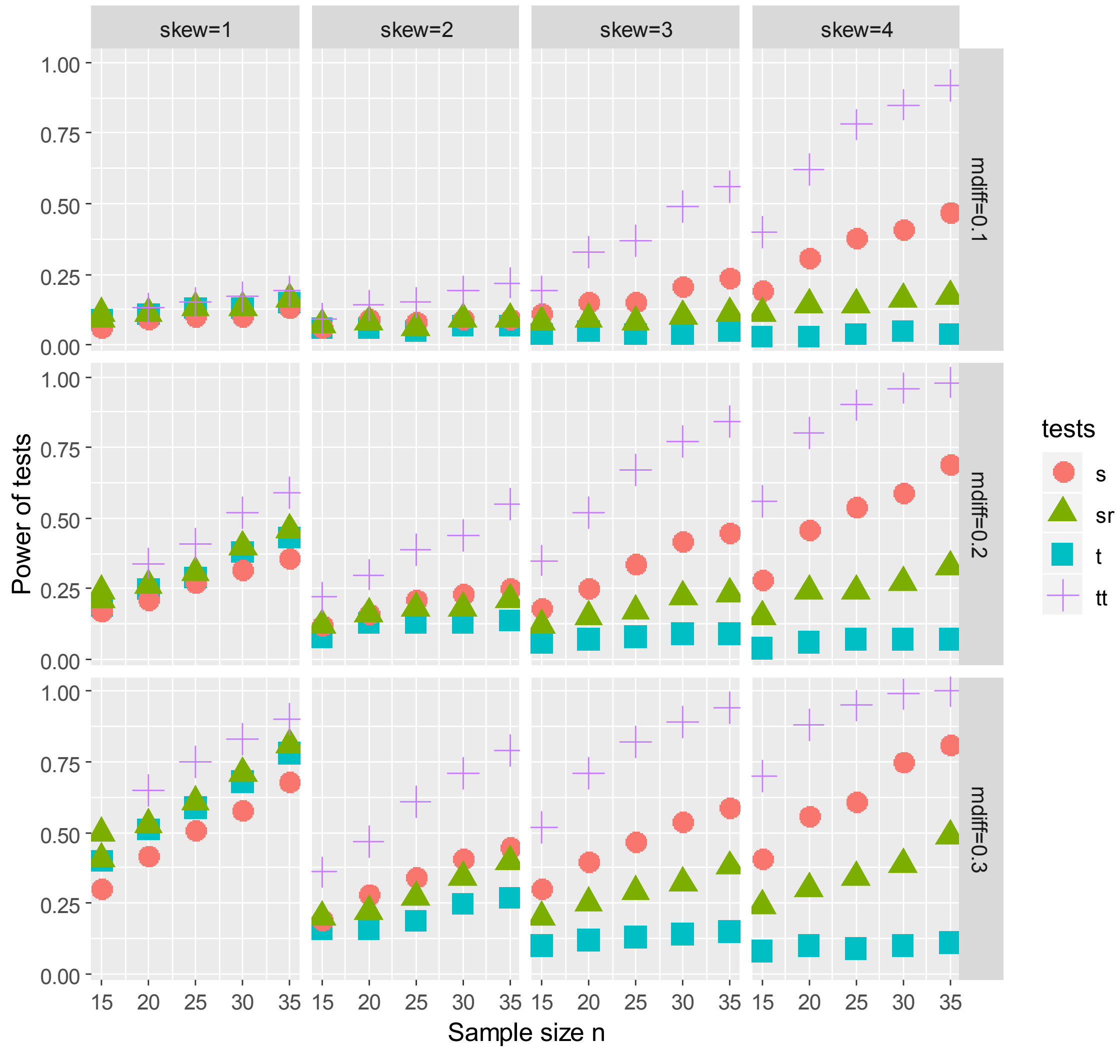


Fig 4.4: Power of gamma distribution with varying mean difference (mdiff) and skewness (skew) when correlation ρ= 0.8.



4. Noted Consistency and limitation of underlying tests

In all simulations in Tables 1.1—1.3, there is no definite effect of correlation, skewness, sample size and paired distributions (beta, exponential or gamma) on Type I error rate. All tests are comparable with the estimated Type I error rates ranging from 0.03 to 0.06 for desired level of 0.05 (5% level of significance). It appears that all tests are relatively consistent in controlling Type I error rate, with maximum underestimation of 0.02 or maximum over estimation of 0.01 due to the deviation of the estimated rates from 0.05.

While consistency has been noted for all tests in controlling Type I error rates, all conventional tests—untransformed paired t-test (t), Sign test (s) and Signed rank test (sr) are subject to some sort of limitations in power concern. For a given mean difference in paired beta distributions, (Table 2), the power of all tests increase due to increase in sample sizes and values of correlation. It appears that the tt- test demonstrates the highest power among all underlying tests for a given mean difference and correlation with increasing sample size. The t test demonstrates the second highest power or comparable with sr, and the s test demonstrates the lowest power.

For a given mean difference and correlation for paired exponential distributions (Table 3), power of all tests increase with increasing sample sizes. It is also evident that the power of all tests increase with increasing mean difference and values of the correlation, irrespective of the sample sizes. Overall, the tt- test demonstrates the highest power, and the s test the second highest with a few exception when correlation is 0.2 and mean difference is 0.2. In all simulations, the t test demonstrates the lowest power.

In Table 4, for skewness varying between 1 and 4 and correlation in the paired gamma distributions 0.2, t-test exhibits extremely poor power and breaks down with increasing skewness. The power of s test decreases for increasing skewness between 1 and 2, and increases for skewness between 3 and 4 with increasing sample size. It appears that the tt test has the highest power among all underlying tests, and the s test retains the second highest power except for skewness=1, where sr test seems to be better than s test. Similar pattern in power as exhibited in Table 4 continues to follow in all simulations where correlation ranges between 0.4 and 0.8, and skewness between 1 and 4 for selected values of mean difference. Figures 4.1-4.4 demonstrate clearly that tt test outperforms all other tests in power, with s test retaining the second best with a very few exceptions when skewness=1, where sr has higher or comparable power as s test.

5. Conclusion

The t test, in particular, breaks down for skewed and highly correlated paired samples. In the presence of higher skewness, the s test outperforms sr test, while tt test outperforms all the competitive tests. Therefore, this study leads to the overall conclusion that the log-transformed paired t-test (tt) would be the better choice over untransformed paired t-test (t), Sign test (s) and Singed rank test (sr) given paired samples for correlated and skewed distributions. If the concern of the test lies in controlling the Type I error rates, then any of the underlying tests could be undertaken.

References

[1]. Sprent, P. and Smeeton, N.C. 2000. Applied Non-parametric Statistical Methods. Third Edition, Chapman and Hall/CRC.

[2]. Wackerly, D., Mendenhall, W. and Scheaffer, R.L. 2008. Mathematical Statistics with Applications. Seventh Edition, Thomson Brooks/Cole.

[3]. G. M. Clarke and D. Cooke, The Nature of Parametric and Non-Parametric tests. A Basic Course in Statistics, 4th edn. Arnold,1998.

[4]. M.M. Desu and D. Raghavarao, Nonparametric Statistical Methods for Complete and Censored Data. Chapman and Hall/CRC (2004).

[5]. P.H. Kvam and B. Vidakovic, Nonparametric Statistics with Applications to Science and Engineering, Wiley, 2007.

[6]. Hogg, R.V., McKean, J.W. and Craig, A.T. 2013. *Introduction to Mathematical Statistics*. Seventh Edition, Prentice Hall.

[7]. I. Akeyede and S. G. Akinyemi. 2010. Power Comparison of Sign and Wilcoxon Sign Rank Test Under Non Normal, African Journal of Physical Sciences, Devon Science Publication.

[8]. A. Imam, M. Usman and M. A. Chiawa, On Consistency and Limitation of paired t-test, Sign and Wilcoxon Sign Rank Test, IOSR Journal of Mathematics (IOSA-JM), Volume 10, Issue 1, 2014.
[9]. M. Usman and A. I. Maksha, An Efficient Alternative To t-Test For One Sample Nonnormal Data, Journal of Applied Science & technology, Auchi Polytechnic, Nigeria, 2010.
[10]. Pagano, M. and Gauvreau, K. 2000. Principles of Biostatistics. Second Edition, Duxbury.
[11]. Rosner, B. 2016. Fundamentals of Biostatistics. Eight Edition, Cengage Learning.
[12]. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.