International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Paired Sample Adaptive Tests and Other Competitors in Location Problem

Chikhla Jun Gogoi[1], Dr. Bipin Gogoi[2]

[1]Research Scholar, Department of Statistics, Dibrugarh University

[2]Professor, Department of Statistics, Dibrugarh University

Abstract: ***One of the more common statistical tests of significance is the test for paired data. Paired data can be obtained from experimental units when measurements are obtained before and after the administration of a treatment. In this paper, comparison is made between adaptive tests, non parametric wilcoxon test and traditional paired sample t test. For comparison of power, Monte – Carlo simulation method is used here. The new adaptive procedure is shown to preserve the size of the test at its nominal level for all continuous distributions. This research work finally comes to the conclusion that in case of normal distribution traditional t test is more powerful than the others. Whereas in case of long tailed or skewed distribution the adaptive tests are more powerful.***

Keywords: adaptive tests, wilcoxon tests, score function, monte-carlo simulation.

1. Introduction

Persons involved in analysing data often choose a statistical procedure after having examined the data. For example, it is not uncommon for a practitioner to transform or smooth data before applying a normal theory test of hypothesis. Such two-staged analyses are termed adaptive since the data determine the transformation used and then the same data are used in the testing procedure.

Many adaptive tests have been developed in an effort to improve the performance of tests of significance. We will consider a test of significance to be "adaptive" if the test procedure is modified after the data have been collected and examined. Adaptive tests of significance have several advantages over traditional tests. They are usually more powerful than traditional tests when used with linear models having long-tailed or skewed distributions of errors. In addition, they are carefully constructed so that they maintain their level of significance. That is, a properly constructed adaptive test that is designed to maintain a significance level of α will have a probability of rejection of the null hypothesis at or near α when the null hypothesis is true. Hence, adaptive tests are recommended because their statistical properties are often superior to those of traditional tests. The adaptive tests have the following properties:

- The actual level of significance is maintained at or near the nominal significance level of α
- If the error distribution is long-tailed or skewed, the adaptive test is usually more powerful than the traditional test, sometimes much more powerful.
- If the error distribution is normal, there is little power loss compared to the traditional tests.
- Adaptive tests are practical.

The adaptive tests automatically reduce the influence of outliers. They are sometimes said to be robust; but to be clear about robustness, we should describe the two kinds of robustness. A test is said to be robust for size if its actual significance level is quite close to the nominal significance level, even when the usual assumptions are not met. For example, a test that is derived by assuming normality of the error distribution would be robust for size if it maintains its level of significance with non-normal errors. A test is said to be robust for power if it has high power relative to other tests when the usual distributional assumptions are not met. Many traditional tests are robust for size with non-normal errors but are not robust for power. Our objective is to study some adaptive tests that are robust for size and robust for power.

2. Objectives

This research work proposes adaptive procedure for analyzing paired data. The procedure uses a function of ordered absolute values of the differences to measure tail heaviness of the underlying distribution. The value of the measure is then used to choose an appropriate signed rank test. The new adaptive procedure is shown to preserve the size of the test at its nominal level for all continuous distributions and typically has nearly the same power as the best signed rank test for a wide range of distributions.

3. Test Procedures

Mathematically, let $d_1,\ldots\ldots,d_n$ i.i.d. *paired* differences from a continuous distribution *F*. The *di* are naturally symmetric about a center μ.We are interested in testing whether this center is equal to a *known* value .Without loss of generality, assume that $\mu = 0$, i.e. we are interested in testing: $H_0 : \mu = 0$ $H_1 : \mu \neq 0$.It's well known that when the *di*'s follow a normal distribution, the optimal test is the t-test. However, if we *know* that *d* comes from a heavy-tailed distribution, e.g. *t*2 or Cauchy, then the signed rank tests with *t*2 (Gastwirth, 1970) or Cauchy scores (Capon, 1961) have higher power than the t-test. We propose an adaptive procedure (denoted by MG) that first uses functions of order statistics of |d*i*| to obtain the information about the tailheaviness of the underlying distribution, and then use it to choose an appropriate signed rank test to analyze the pairs. The procedure strictly keeps the size of the test at the nominal level for all sample sizes, and has about the same power as

the best signed rank test in a wide range of distributions, including the t family.

3.1 T test

Let *di* be the difference between the measurements for the ith pair and let *n* equal the number of pairs. The usual *t* test statistic is $\frac{\bar{d}}{s/\sqrt{n}}$

where $\bar{d}$ is the average of the differences and s^2 is the usual unbiased estimator of the variance of the differences. If the differences are normally distributed then, under the null hypothesis, the test statistic *t* will be distributed as a *t* distribution with *n* — 1 degrees of freedom. This test is popular because it is the most powerful test if the differences are normally distributed.

3.2 Adaptive tests based on ranked scores:

For rank-based tests we let R_i^+ be the rank of |di| among $|d_1|, \ldots, |d_n|$ and we let $\varphi_i = 1$ if di > 0 and $\varphi_i = 0$ if di < 0. With this notation the SR test is based on $\sum \varphi_i R_i^+$, where the sum is over the *n* differences.

In the SR test we sum the ranks, but it is easy to create other rank-based tests that use the sum of a function of the ranks. If we define a general function of R_i^+ as $a(R_i^+)$ the sum over the positive differences is

$\sum \varphi_i a(R_i^+)$ where $a(i)=J(\frac{i}{n+1})$

We perform the test by first calculating the test statistic

$$Z=\frac{\sum_{i=1}^{n} \varphi_i a(R_i^+)-(\frac{1}{2})\sum_{i=1}^{n} a(i)}{\sqrt{(\frac{1}{4})\sum a^2(i)}}$$

Miao and Gastwirth (2009) published an adaptive rank-based test that used a measure of tail-heaviness to determine the set of rank scores to be used to perform the test. Under the null hypothesis the differences are symmetric about zero, so one measure of variability is $s=\sqrt{\sum_{i=1}^{n} \frac{d_i^2}{n}}$. Note that *s* will be sensitive to outliers. Another measure of variability, which is not sensitive to the presence of a few outliers, is $\bar{M} = \frac{median(|d_1|,\ldots,|d_n|)}{0.6745}$

Miao and Gastwirth (2009) proposed that the tail heaviness measure $SM = \frac{\bar{S}}{\bar{M}}$ be used as a measure of tail- heaviness. If the differences are normally distributed about zero with a standard deviation of σ, then $\bar{S}$ and $\bar{M}$ should approximate σ , so *SM* should be close to one. If the distribution has heavy tails, then *s* will be large compared to M, which will produce a ratio that will greatly exceed one. This test, which will be called the MG test, uses SM to select a set of rank scores.

When the data comes from a light-tailed normal distributions or short-tailed uniform distributions, the normal scores test is known to have high power; when the tail-heaviness of the underlying distributions is somewhat medium, like logistic, double exponential or contaminated normal distributions, the Wilcoxon test is highly correlated with the maximum efficiency robust test (Gastwirth, 1966) and the Wilcoxon test should be used. Furthermore, if the data is heavy-tailed, e.g. from a *t*2 then the appropriate signed rank test is the *t*2 scores . We choose the following 3 score functions which include the extreme members of the *t* family of distributions:

$J_1(u)=u$(wilcoxon scores)

$J_2(u)=\frac{3\sqrt{2}}{2}(u\sqrt{(1-u^2)})$ (t_2 scores)

$J_3(u)=\frac{2\tan(\frac{1}{2}\pi u)}{1+\tan^2(\frac{1}{2}\pi u)}$

Conditions for the Adaptive Test:
$sM \geq 2.7$, use the Cauchy scores test;
$2.7 > sM \geq 1.2$, use the *t*2 scores test;
$1.2 > sM \geq 1.02$, use the Wilcoxon test;

4. The Monte Carlo Study

For the simulation study of the t- test , Wicoxon test, t2 score test and Cauchy score test, four families of distributions are selected. These are – the Normal, the Cauchy, the Logistic and the Lognormal distribution.

The study was conducted on computer at the Department of Statistics, Dibrugarh University. To generate the standard normal deviate, the method described in Monte Carlo Method by Hammersly and Handscomb(1964) were used and deviate from the other distributions were generated by using the inverse distribution function on uniform deviates.

In studying the significant levels, we first considered distributions with location parameter equal to zero and with equal scale parameters. Specifically, we considered the distribution functions F(x -μ) , where μ were the location parameters. For each set of sample , the experiment was repeated 10,000 times and proportion of rejection of the true null hypothesis was recorded.

For the power study of the tests, random deviates were generated as above for each group and added to μ . Proportion of rejections based on 10,000 replications at the levels .10, .05 for different combinations of μ were recorded.

Table 1.1: Empirical level and power of different tests under normal distribution

Sample sizes	Sample mean (μ1,μ2)	T test 10% 5%	Wilcoxon test 10% 5%	T2 score test 10% 5%	Cauchy score test 10% 5%
(10,10)	(0,0)	.0971 .0484	.1075 .0524	.1279 .0692	.1017 .0459
	(0,.2)	.1297 .0669 .2126	.1418 .0683 .2271 .1271	.1600 .0887 .2416	.1183 .0568 .1698
	(0,.4)	.1232 .3450 .2250	.3568 .2251 .5060 .3544	.1486 .3637 .2513	.0892 .2588 .1495
	(0,.6)	.4990 .3568	.1026 .0509	.5093 3832	.3694 .2418
	(0,.8)	.1078 .0549	.1903 .1091	.1279 0692	.1014 .0483
		.1913 .0772. 4350	.4333 .3075	.1901 .1204 .4130	.1143 .0776
(30,30)	(0,0)	.1390 .7106 .2421	.7095 .5889	.2949	.2753 .1798
	(0,.2)	.9323 .3839	.9010 .8291	.6702 .5540	.4874 .3560
	(0,.4)			.8669 .7874	.7037 .5788
	(0,.6)				
	(0,.8)				

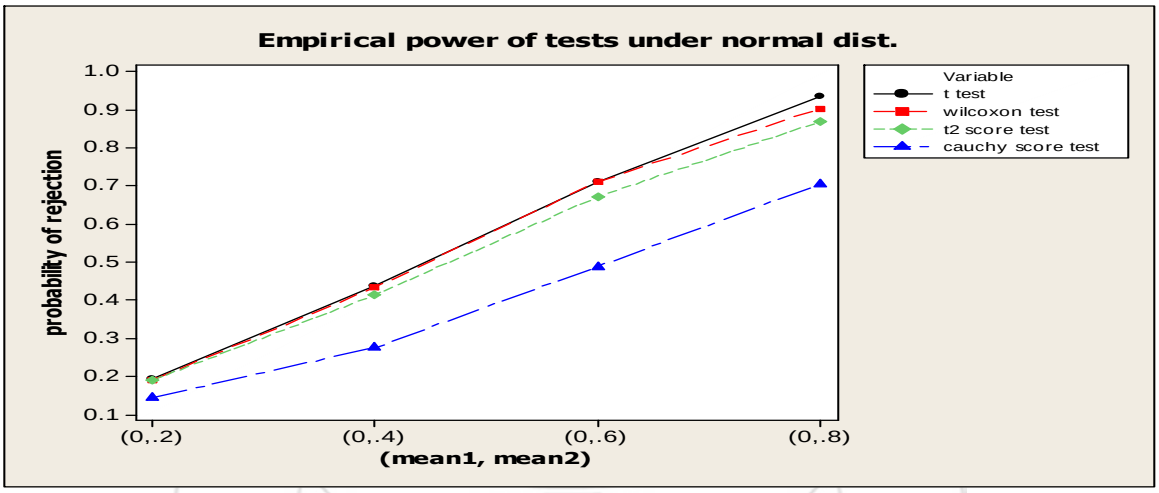

Figure 1.1: Empirical power of tests under normal distribution at 10% level with sample size (30,30)

Table1.2: Empirical level and power of tests under Cauchy distribution

Sample sizes	(μ1,μ2)	T test 10% 5%	Wilcoxon test 10% 5%	T2 score tests 10% 5%	Cauchy score test 10% 5%
10,10	(0,0)	.0485 .0168	.0992 .0443	.1181 .0611	.0966 .0409
	(0,.2)	.0525 .0187	.1051 .0496	.1268 .0696	.1040 .0467
	(0,.4)	.0626 .0237	.1230 .0625	.1458 .0847	.1234 .0589
	(0,.6)	.0776 .0316	.1517 .0776	.1718 .1094	.1580 .0811
	(0,.8)	.0964 .0434	.1835 .1001	.2200 .1419	.2023 .1073
(30,30)	(0,0)	.0695 .0219	.1013 .0511	.1236 .0686	.1024 .0535
	(0,.2)	.0712 .0234	.1178 .0607	.1448 .0813	.1236 .0652
	(0,.4)	.0788 .0286	.1607 .0934	.2130 .1271	.1907 .1094
	(0,.6)	.0895 .0377	.2301 .1441	.3094 .2101	.2905 .1873
	(0,.8)	.1095 .0486	.3136 .2094	.4254 .3103	.4172 .2912

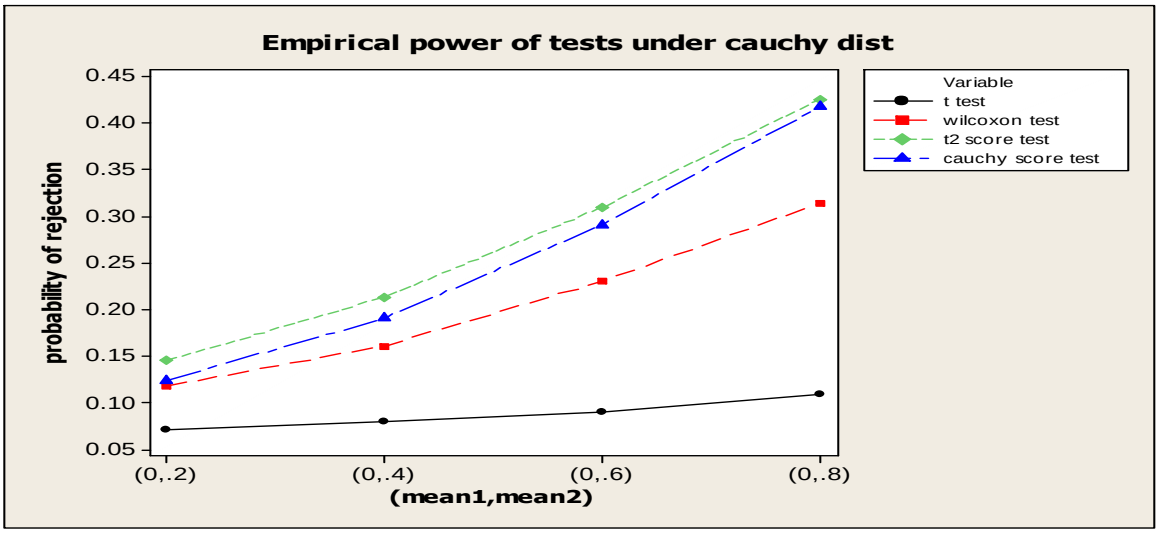

Figure 1.2: Empirical power of tests under Cauchy distribution for 10% level with sample sizes (30,30)

Table1.3: Empirical level and power of tests under Logistic Distribution

Sample sizes	(μ1,μ2)	T test 10% 5%	Wilcoxon test 10% 5%	T2 score tests 10% 5%	Cauchy score test 10% 5%
(10,10)	(0,0)	.0930 .0428	.1046 .0480	.1233 .0669	.1031 .0464
	(0,.2)	.1044 .0516	.1157 .0539	.1342 .0732	.1074 .0492
	(0,.4)	.1331 .0706	.1465 .0720	.1635 .0938	.1278 .0598
	(0,.6)	.1778 .1003	.1924 .1071	.2124 .1307	.1573 .0799
	(0,.8)	.2395 .1431	.2520 .1501	.2740 .1777	.1981 .1105
(30,30)	(0,0)	.1131 .0562	.0999 .0493	.1267 .0648	.1008 .0483
	(0,.2)	.1353 .0631	.1315 .0660	.1482 .0842	.1143 .0577
	(0,.4)	.1798 .0831	.2180 .1317	.2324 .1448	.1687 .0901
	(0,.6)	.2980 .1183	.3557 .2429	.3538 .2505	.2508 .1562
	(0,.8)	.4266 .2676	.5199 .3888	.5144 .3854	.3674 .2448

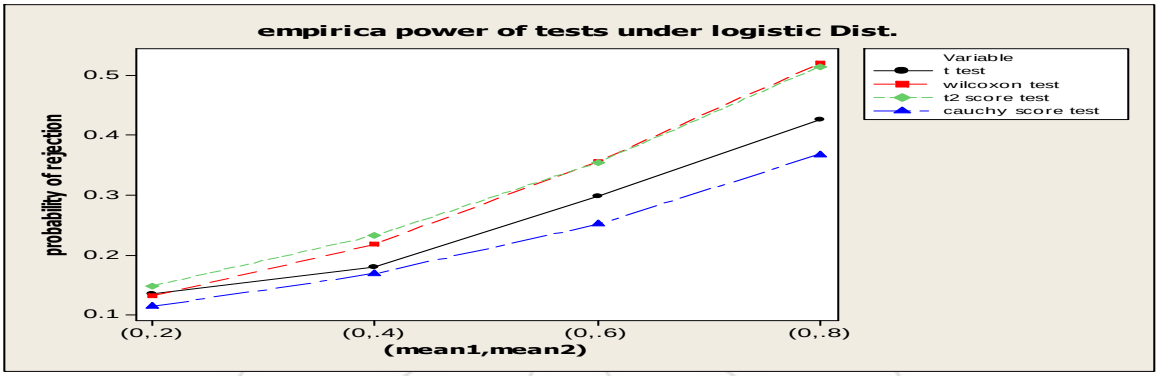

Figure 1.3: Empirical power of tests under Logistic distribution at 10% level with sample size (30,30)

Table1.4: Empirical level and power of tests under lognormal distribution:

Sample sizes	(μ1,μ2)	T test 10% 5%	Wilcoxon test 10% 5%	T2 score test 10% 5%	Cauchy score test **10% 5%**
(10,10)	(0,0)	.0782 .0338	.1044 .0518	.1285 .0684	.1027 .0441
	(0,.2)	.0985 .0437	.1357 .0695	.1659 .0943	.1370 .0664
	(0,.4)	.1490 .0766	.2075 .1190	.2504 .1629	.2259 .1255
	(0,.6)	.2246 .1317	.3076 .1834	.3607 .2540	.3386 . 2079
	(0,.8)	.3105 .2052	.4117 .2633	.4805 .3589	.4534 .2972
(30,30)	(0,0)	.0945 .0365	.1046 .0507	.1261 .0632	.0976 .0448
	(0,.2)	.1124 .0498	.1746 .1002	.2219 .1389	.1977 .1178
	(0,.4)	.1700 .0860	.3645 .2464	.4641 .3460	.4395 .3152
	(0,.6)	.2460 .1471	.5825 .4536	.706 .5870	.6896 .5556
	(0,.8)	.3390 .2218	.7568 .6481	.8647 .7829	.8517 .7539

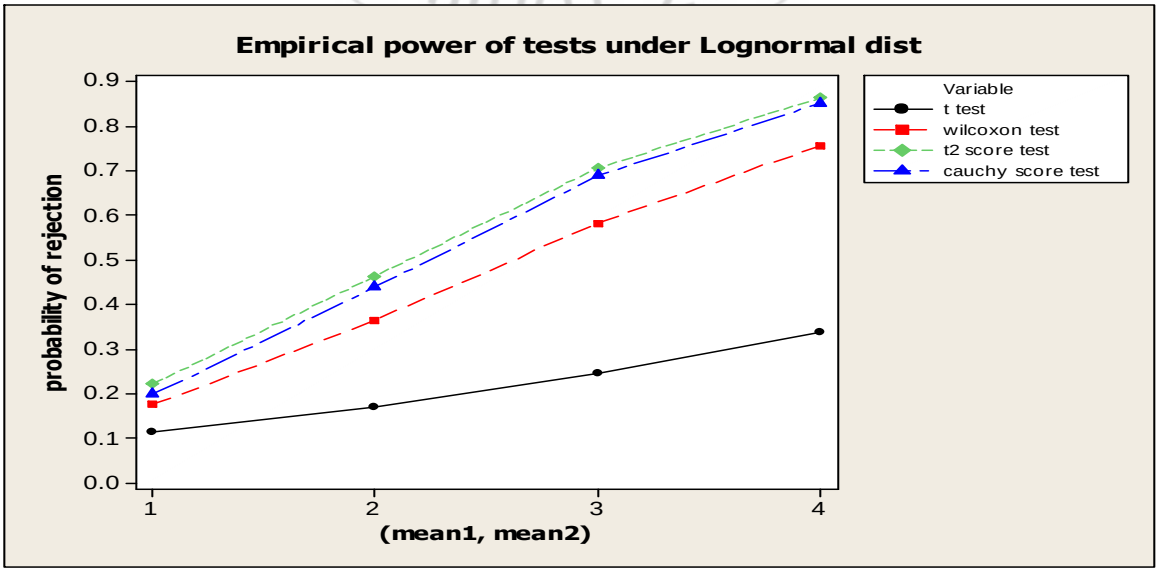

Figure 1.4: Empirical power of tests under lognormal distribution at 10% level with sample size (30,30)

5. Discussion

For comparison purposes we have considered various combinations of sample sizes with equal. We have also considered different ses of μ_i's for the study.

From Tables 1.1 – 1.4 it is observed that the parametric t-test maintain the nominal level except the Cauchy and skew distribution lognormal. In these cases t-test seems to be conservatives. Wilcoxon test and all other score based test are found to be robust against the distributions in terms of the level concerned.

Table 1.1 shows the power of tests under normal distribution. We have seen that power of t- test is higher than the other tests in this distribution in presence of various combinations of location parameters. Power of Wilcoxon test is found to be slightly less than the t-test but more than other score base tests..

Table 1.2 gives the power of tests statistics under Cauchy distribution. Here ,we observe that t2 test is more powerful than other tests at 10% and 5% level of significance.

Table 1.3 displays the power of tests under logistic distribution. We have seen that Wilcoxon , t2 test are more powerful than other two tests with at 10% and 5% level.

Table 1.4 shows the power of tests under lognormal distribution. Here t2 score test and Cauchy score test are more powerful than the all other tests at 10% and 5% level

References

[1] Capon, J. (1961). Asymptotic Efficiency of Certain Locally Most Powerful Rank Tests. *The Annals of Mathematical Statistics* **32** 88–100.
[2] Chernoff, H., Gastwirth, J. L., and Johns Jr., M. V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *The Annals of Mathematical Statistics* **38** 352–372.
[3] D'Agostino, R. B. and Cureton, E. E. (1973). A Class of Simple linear Estimators of the Standard Deviation of the Normal Distribution. *Journal of American Statistical Association* **68** 207–210.
[4] Weston Solutions of Michigan, INC. (2004). Phase I Summary Report for Detroit Lead Assessment Project, Great Lakes Smelting – 1640 East Euclid Street, Detroit, Wayne County,Michigan.
[5] Freidlin, B., Miao, W., and Gastwirth, J. L. (2003). On the Use of the Shapiro-Wilk Test in Two-Stage Adaptive Inference for Paried Data from Moderate to Very Heavy Tailed Distributions. *Biometrical Journal* **45** 887–900.
[6] Gastwirth, J. L. (1966). On Robust Procedures. *Journal of the American Statistical Association* **61** 929–948.