RESEARCH ARTICLE OPEN ACCESS

Comparison of Adaptive Test with the Traditional t test in Paired Sample for Location Problem.

[1]Chikhla Jun Gogoi*, [2]Bipin Gogoi
[1]. Research Scholar, Department of Statistics, Dibrugarh University, Assam.
[2]. Professor, Department of Statistics, Dibrugarh University, Assam.
Corresponding Author: Chikhla Jun Gogoi

ABSTRACT
In paired sample case for location problem usually paired sample t test is more preferred provided the parent distribution is normal. In this paper it is tried to discuss an exceptional case, that the parent distribution in non normal. Here Adaptive test is suggested for the particular case. An example is given here to clarify the discussion.

Keywords: Paired t test, Adaptive test, power.

Date of Submission: 30-08-2017 Date of acceptance: 20-09-2017

I. 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 which will have a probability of rejection of the null hypothesis at or near 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:

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

II. OBJECTIVE:

In this paper it is tried to make a comparison between Paired t test and Adaptive test for paired sample data in case of Location problem using one example.

III. TEST PROCEDURES:

Mathematically, let $d_1,......,d_n$ .be the paired differences from a continuous distribution F. The d_i 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, we assume that: $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$.It's well known that when the *di'*s follow a normal distribution, the optimal test is the t-test.

3.1 T test:

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

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 test:

The doubling of Data:

Let d_i be the difference of ith pair and let μ be the mean of the differences. We used here an Adaptive one sided test of H_0: $\mu = 0$ versus H_a:$\mu > 0.$ Now to construct a symmetric empirical distribution, we will double the sample differences by defining the double differences

$$d_{D,i} = d_i \;\; For\; i = 1,2,\dots,n$$

$$= -d_{i-n} \;\; For\; i = n+1, n+2,\dots,2n$$

Now to determine the appropriate weights for the observations, we standardize them by using a robust estimate of the standard deviation. Let iqr_D be the interquartile range based on the doubled differences $\{d_{D,1}, d_{D,2} \dots, d_{i,2n}\}$. *Now* $s_D = \frac{iqr_D}{1.349}$. The standardised values for the doubled differences are given by $z_{D,i} = d_i/S_D$ for i= 1, 2, ..., 2n

Smoothing and Weighting:

Let $Z_D = \{Z_{D,1},\dots,Z_{D,2n}\}$. In order to get a smoothed cdf at a point z we have to calculate $\widehat{F_h}(z; Z_D) = \frac{1}{2n}\sum_{i=1}^{2n} \emptyset\,(\frac{z - z_{D,i}}{h})$

Where $\emptyset(.)$ is the cdf of standard normal distribution and h is a smoothing constant and $h = 1.26\, n^{-1/3}$ is the smoothing constant.

For the ith standardised observation $z_{D,i}$ the value $\emptyset^{-1}[\widehat{F_h}(z_{D,i}; z_D)]$ is the corresponding normal score. The weights of Adaptive test can be calculated as

$$w_i = \frac{\emptyset^{-1}[\widehat{F_h}(z_{D,i}; z_D)]}{z_{D,i}} \quad \text{for i= 1,2,...,n}$$

The appropriate model for paired data is $d_i = \beta + \varepsilon_i$, and the usual one sided t test is based on the test statistic for testing $H_0: \beta = 0$ versus $H_a: \beta > 0$. With the Adaptive approach we weight each observation to obtain the WLS model $w_i d_i = w_i\beta + w_i\varepsilon_i$

Substituting $y_i^* = w_i d_i$, $x_i^* = w_i$ and $\varepsilon_i^* = w_i\varepsilon_i$. The transformed model can be written as

$y_i^* = x_i^*\beta + \varepsilon_i^*$. Using OLS methods on the transformed data, we will get the WLS estimate of β as

$$b = \frac{\sum_{i=1}^{n} x_i^* Y_i^*}{\sum_{i=1}^{n}(x_i^*)^2} = \frac{\sum_{i=1}^{n} w_i^2 d_i}{\sum_{i=1}^{n} w_i^2}$$

The Adaptive test statistics is,

$$t = \frac{b\sqrt{\sum_{i=1}^{n} w_i^2}}{s^*}$$

where $s^* = \sqrt{SSE^*/(n-1)}$

$$SSE^* = \sum_{i=1}^{n} w_i^2 d_i^2 - \left(\sum_{i=1}^{n} w_i^2 d_i\right)^2 / \sum_{i=1}^{n} w_i^2$$

Example 1:

A study is run to evaluate the effectiveness of an exercise program in reducing systolic blood pressure in patients with pre-hypertension (defined as a systolic blood pressure between 120-139 mmHg or a diastolic blood pressure between 80-89 mmHg). A total of 15 patients with pre-hypertension enrolled in the study, and their systolic blood pressures are measured. Each patient then participates in an exercise training program where they learn proper techniques and execution of a series of exercises. Patients are instructed to do the exercise program 3 times per week for 6 weeks. After 6 weeks, systolic blood pressures are again measured. The data are shown below.

Table 1:

Patient	Systolic Blood Pressure Before Exercise Program	Systolic Blood Pressure After Exercise Program	Difference (d_i) (Before-After)
1	125	118	7
2	132	134	-2
3	138	130	8
4	120	124	-4
5	125	105	20
6	127	130	-3
7	136	130	6
8	139	132	7
9	131	123	8
10	132	128	4
11	135	126	9
12	136	140	-4
13	128	135	-7
14	127	126	1
15	130	132	-2

Set up hypotheses and determine level of significance.

H_0: The median difference is zero versus

H_1: The median difference is not zero at $\alpha=0.05$

$\sum_{i=1}^{15} d_i=48$, $\bar{d} = 3.2$

$S^2=1/(n-1) \sum_{i=1}^{15}(d_i - \bar{d})^2=(1/14) \sum_{i=1}^{15}(d_i - 3.2)^2$

$= 50.3142$

S=7.093

$t=\frac{3.2}{7.093/\sqrt{15}}$

$=1.748$

Power of t test is 0.9678

The calculated $IQR_D=12.50$

$S_D=9.26612$

$h=1.26*(15)^{-1/3}$

$=0.5109$

The WLS estimate of β is b=0.830712/33.0572=.02513

$SSE=\frac{1805.93-(0.830712)^2}{33.0572}$

$=54.61$

$S^*=\sqrt{3.9}=1.975$

Now $t=\frac{0.02513 \sqrt{33.0572}}{1.975}= 0.07316$

Power of Adaptive test is .95021

IV. CONCLUSION

Since the power of t test is more than the Adaptive test in this particular example, so the parent distribution of the data assumed to be symmetric and hence the traditional paired t test is suitable for the example otherwise if the power of Adaptive test becomes more than the t test, we may assume that the parent distribution of that particular data set is not symmetric and accordingly the Adaptive test would be suitable for the data.

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.

International Journal of Engineering Research and Applications (IJERA) is **UGC approved** Journal with Sl. No. 4525, Journal no. 47088. Indexed in Cross Ref, Index Copernicus (ICV 80.82), NASA, Ads, Researcher Id Thomson Reuters, DOAJ.

Chikhla Jun Gogoi. "Comparison of Adaptive Test with the Traditional t test in Paired Sample for Location Problem." International Journal of Engineering Research and Applications (IJERA), vol. 7, no. 9, 2017, pp. 62–63.