**Science Journal of Mathematics & Statistics Published By**

**ISSN:2276-6324 Science Journal Publication**

<http://www.sjpub.org/sjms.html> International Open Access Journal

© Author(s) 2013. CC Attribution 3.0 License.

**Research Article**  **Volume 2013 (2013)**

**Nonparametric Hypothesis Testing Report**

Loc Nguyen

**Vietnam Institute of Mathematics**

**Accepted 17th July, 2013. Not Published Yet**

**Abstract**

This report is the brief survey of nonparametric hypothesis testing. It includes four main sections about hypothesis testing, one additional section discussing goodness-of-fit and conclusion section.

***Sign test section*** gives an overview of nonparametric testing, which begins with the test on sample median without assumption of normal distribution.

***Signed-rank test section*** and ***rank-sum test section*** concern improvements of sign test. The prominence of signed-rank test is to be able to test sample mean based on the assumption about symmetric distribution. Rank-sum test discards the task of assigning and counting plus signs and so it is the most effective method among ranking test methods.

***Nonparametric ANOVA* section** discusses application of analysis of variance (ANOVA) in nonparametric model. ANOVA is useful to compare and evaluate various data samples at the same time.

***Nonparametric goodness-fit-test* section**, an additional section, focuses on different hypothesis, which measure the distribution similarity between two samples. It determines whether two samples have the same distribution without concerning how the form of distribution is.

The *last section* is the conclusion*.*

Note that in this report terms *sample* and *data sample* have the same meaning. A sample contains many data points. Each *data point* is also called an *observation*.

**Keywords:** Overview of nonparametric testing, Nonparametric ANOVA section

**1.1 Sign test**

Nonparametric testing is used in case of without knowledge about sample distribution; concretely, there is no assumption of normality. The nonparametric testing begins with the test on sample *median*. If distribution is symmetric, median is identical to mean. Given the median is the data point at which the left side data and the right side data are of equal accumulate probability.

*P*(*D <* ) *= P*(*D >* ) *=* 0.5

If data is not large and there is no assumption about normality, the median is approximate to population mean. Given null hypothesis *H*0: = and alternative hypothesis *H*1: , the so-called sign test [1, pp. 656-660] is performed as below steps:

1. Assigning plus signs to sample data points whose values are greater than and minus signs to ones whose values are less than . Note that values which equal are not considered. Plus signs and minus signs represent the right side and left side of , respectively.
2. If the number of plus signs is nearly equal to the number of minus signs, then null hypothesis *H*0is true; otherwise *H*0 is false. In other words, that the proportion of plus signs is significantly different from 0.5 cause to rejecting *H*0 in flavor of *H*1.

The reason of *H*0 acceptance is that the probability that data points (or observations) fall in both left side and right side of are of equal value 0.5 and of course, it is asserted that is a real median. Note that terms data point, sample point, sample value and observation are identical.

In the case that alternative hypothesis *H*1: < , if the proportion of plus signs is less than 0.5 then rejecting *H*0 in flavor of *H*1. In the case that alternative hypothesis *H*1: > , if the proportion of plus signs is greater than 0.5 then rejecting *H*0 in flavor of *H*1. Now let *X* be the discrete random variable representing the number of plus signs and suppose that *X* conforms binomial distribution *B*(*X*; *n*; *p*) where *n* and *p* are the total number of sample data points and the probability that plus sign is assigned to a data point, respectively. Because the proportion of plus signs gets 0.5 when *H*0: = is true, the parameter *p* is set to be 0.5. Given the distribution of plus signs is *B*(*X*; *n*; 0.5) and significant level α and let *x* be the instance of *X* where *x* = , there are three following tests [1, pp. 657-660]:

* *H*0: = and *H*1: : In case of *x < n*/2, if 2*P*(*X x*) *< α*  then rejecting *H*0. In case of *x > n*/2, if 2*P*(*X x*) *< α* then rejecting *H*0. This test belongs to two-sided test family.
* *H*0: = and *H*1: < : if *P*(*X x*) *< α* then rejecting *H*0. This test belongs to one-sided test family.
* *H*0: = and *H*1: > : if *P*(*X x*) *< α*  then rejecting *H*0. This test belongs to one-sided test family.

Note that *P*(…) is accumulated probability of binomial distribution *B*(*X; n; 0.5*), for example, P(*X x*) = . In case that *n* is large enough, for instance n > 10, *B*(*X; n; 0.5*) is approximate to standard normal distribution *N*(*Z*; 0; 1) where *Z =* . Let *z* be the instance of *Z* where *z* = , there are three following tests:

* *H0*: = and *H*1: : if *|z| > zα*/2 then rejecting *H*0 where *zα*/2 is 100*α*/2 percentage point of standard normal distribution.
* *H0*: = and *H*1: < : if *z < -zα*/2 then rejecting *H*0.
* *H0*: = and *H*1: > : if *z > zα*/2 then rejecting *H*0.

In case of pair-test *H*0: – = *d*0 which we need to know how much median shifts from other one , sign test is applied in similar way with a little bit of change. If *d*0 = 0, *H*0 indicates whether equals . We compute all deviations between two samples *X* and *Y* where is sample median of *X* and is sample median of *Y*. Let *di = xi – yi* be the deviation between *x* *Y* and *y* *Y*. Plus signs (minus signs) are assigned to *di* (s) which are greater (less) than *d0*. Now signed test is applied into such plus signs and minus signs by discussed method.

**2.0 Signed-rank test**

Sign test focuses on whether or not the observations are different from null hypothesis but not considers the magnitude of such difference. Wilcoxon signed-rank test [1, pp. 660-663] based on assumption of symmetric and continuous distribution considers both difference and how much difference is. The median is identical to the mean *μ* according to symmetric assumption. It includes four following steps [1, pp. 660-663]:

1. Calculating all deviations between data points and *μ*0, we have *D =* {*d*1, *d*2,…, *dn*} where *di = xi – μ*0and *di 0*.Note that data point *xi* is instance of random variable *X*.
2. Assigning a rank *ri* to each deviation *di* without regard to sign, for instance, rank value 1 and rank value *n* to be assigned to smallest and largest **absolute** deviation (without sign), respectively. If two or more absolute deviations have the same value, these deviations are assigned by average rank. For example, if 3rd, 4th and 5th deviations get the same value, they receive the same rank (3+4+5) / 3 = 4. We have a set of ranks *R =* {*r*1, *r*2,…, *rn*} where *ri* is the rank of *di*.
3. Let *w+* and *w–* be the sum of ranks whose corresponding deviations are positive and negative, respectively. We have *w+ =*  and *w*– *=*  and *w = min*(*w+*, *w*–). Note that *w* is the minimum value between *w+* and *w–*.
4. In flavor of *H*1: *μ*< *μ*0, *H*0 is rejected if *w+* is sufficiently small. In flavor of *H*1: *μ* > *μ*0, *H*0 is rejected if *w–* is sufficiently small. In case of two-sided test *H*1: *μ μ*0, *H*0 is rejected if *w* is sufficiently small. The concept “sufficiently small” is defined via thresholds or pre-computed critical values, see [Walpole, Myers, Myers, Ye 2012, pp. 759] for critical values. The value *w+*, *w–* or *w* is sufficiently small if it is smaller than a certain critical value with respect to significant level *α*.

In case of pair test *H*0: *μ*1– *μ*2 *= d*0, the deviation *di* in step 1 is calculated based *d*0 and two samples *X* and *Y*, so *di = xi – yi* *– d*0where *x* *Y* and *y* *Y*. Note that *μ*1 and *μ*2 are taken from *X* and *Y*, respectively. Steps 2, 3, 4 are performed in similar way.

Let *W+* be random variables of *w+*. If *n* 15 then *W+* approaches normal distribution with mean and variance . We can normalize *W+* so as to define critical region via percentage point *zα* of normal standard distribution,

**3.0 Rank-sum test**

Rank-sum test [1, pp. 665-667] is a variant of signed-rank test. Suppose there are two samples *X* = {*x*1, *x*2,…, } and *Y =* {*y*1, *y*2,…, } and the null hypothesis is specified as *H*0: *μ*1= *μ*2 where *μ*1 and *μ*2 are taken from *X* and *Y*, respectively. We assign ranks to such *n*1 *+ n*2 data points according to their values, for instance, rank value 1 and rank value *n*1+ *n*2 to be assigned to smallest and largest sample value. If two or more data points have the same value, these points are assigned by average rank. For example, if 3rd, 4th and 5th data points get the same value, they receive the same rank (3+4+5) / 3 = 4.Let *R =* {*r*1, *r*2,…, } be the set of these ranks. Let *w*1 and *w*2 be the sum of ranks corresponding to *n*1 data points in *X* and *n*2 data points in *Y*, respectively.

and

Where *ri* is a rank of a data point in the set *X Y* and *ri =*

We have . There are three following tests [1, pp. 665-667]:

* Rejecting *H*0 in flavor of alternative *H*1: *μ*1< *μ*2 if *w*1 is sufficiently small.
* Rejecting *H*0 in flavor of alternative *H*1: *μ*1> *μ*2 if *w*2 is sufficiently small.
* In case of two-sided test with *H*1: *μ*1 *μ*2 if the minimum of *w*1 and *w*2 is sufficiently small then rejecting *H*1.

Rank-sum test has two advantages in comparison of signed-rank test:

* There is no need to calculate deviations among samples and to count the number of plus signs and minus signs.
* Samples can has different number of data points, for instance, |*X*| = *n*1 *n*2 = |*Y*|

Setting *u*1 *= w*1 – and *u*2 *= w*1 – and suppose that *u*1 and *u*2 are instances of random variables *U*1 and *U*2, respectively. If both *n*1and *n*2are greater than8, variable *U*1 (or *U*2) is approximate to normal distribution with mean and variance . We can normalize *U*1 (*U*2) so as to define critical region via percentage point *zα* of normal standard distribution, .

**4.0 Nonparametric ANOVA**

In many applications, we process various samples (*X*, *Y*, *Z*, etc.) where each sample is a set of observations (data points) which relate to a concrete method, a way or an approach that creates or produces these observations. Such concrete method is called treatment. In other words, we consider a matrix of observations and each row represents a mono-sample attached to a treatment, for instance, *X* or *Y* or *Z*, etc. For convenience, matrix of observations is call multi-sample or sample, in short. Treatments are grouped into categories which are called factors. If sample has only one factor, it is single-factor sample; otherwise, it is called several-factor sample. Following table is an example of single-factor sample.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Treatment 1 | *y*11 | *y*12 | *y*13 | *=*( *y*11 *+ y*12 *+ y*13) / 3 |
| Treatment 2 | *y*21 | *y*22 | *y*23 | *=*( *y*21 *+ y*22 *+ y*23) / 3 |
| Treatment 3 | *y*31 | *y*32 | *y*33 | *=*( *y*31 *+ y*32 *+ y*33) / 3 |
|  |  |  |  | *=*( *+ +* ) / 3 |

Let *Yij* be the random variable representing *jth* data point of *ith* treatment [2, p. 472].

Where *μ* so-call overall mean is the mean over whole sample, called treatment effect denotes the parameter of *ith* treatment and denotes the random error. There is the assumption that random error is independently distributed and conforms normal distribution; moreover, it has mean 0 and variance *σ*2. Let *μi = μ + τi* be the treatment mean of *ith*treatment. The objective of analysis of variance (ANOVA) [2, pp. 468-490] is to analyze statistics about treatment mean, treatment effect, random error so as to take out conclusions about such statistics. Basically, ANOVA focuses on characteristics relating to deviation, variability, sum of squares, mean square, etc. An typical application of ANOVA is to test whether *k* treatment means *μ*1, *μ*2,…, *μk* are equal; it means that we test the following hypotheses:

*H*0: *μ*1 = *μ*2 = …= *μk*

*H*1: *μ*1 *μ*2 … *μk*

Due to *μi* = *μ* + *τi*, this test is re-written:

*H*0: *τ*1 = *τ* 2 = …= *τk* = 0

*H*1: *τ i* 0 for at least one treatment

If *H*0 is true, treatments have no effect on whole sample. Let *yij* be the instance of random variable *Yij*. Let *yi*, , *y* and be the sum of observations of treatment *i*, the average of observations of treatment *i*, the sum of whole observations and the average of whole observations.

, , ,

Where *k* is the number of treatments, *ni* is the number of observations under treatment and *N* = *n*1 + *n*2 +… + *nk* is the total number of observations. Let *SST*, *SSTreatment* and *SSE* [2, pp. 474-475] be the total sum of squares, treatment sum of squares and error sum of squares. Please pay attention to *SST*, *SSTreatment* and *SSE* because they are main research objects in ANOVA. We have [2, p. 475]:

Following is the sum of squares identity [2, p. 475]:

*SST = SSTreatment* + *SSE*

Treatment sum of squares *SSTreatment* is very important because it reflects treatment effects *τi* (s) and treatment means *μi* (s). The expected values of treatment sum of squares and error sum of squares are computed as below [2, p. 474]:

*SST* and *SSTreatment* and *SSE* have *N –* 1 and *k –* 1 degrees of freedom, respective because there are *N* observations over whole sample and *k* treatments. So *SSE* has *N – k =* (*N –* 1) *–* (*k –* 1) due to *SSE = SST – SSTreatment*. Based on degrees of freedom, treatment mean square *MSTreatment* and error mean square *MSE* is determined as below [2, pp. 474-475]:

If null hypothesis *H*0: *τ*1 = *τ* 2 = … = *τ k* = 0 is true, *MSTreatment* is an unbiased estimate of variance σ2 due to

Moreover *MSE* is always an unbiased estimate of variance *σ*2 due to *E*(*MSE*) = = *σ*2. So *MSTreatment* and *MSE* conform chi-square distribution and the ratio of *MSTreatment* to *MSE* conforms F-distribution with *k* – 1 and *n*(*k* – 1) degrees of freedom [2, p. 475]:

Hypothesis *H*0: *τ*1 = *τ*2 = …= *τk* = 0 is rejected if the ratio *F*0 > *fα*, *k-*1, *n*(*k-*1) where *fα*, *k-*1, *n*(*k-*1) is the 100*α* percentage point of *F-*distribution with *k–*1 and *N–k* degrees of freedom.

We have already discussed about parametric ANOVA with normality assumption, now nonparametric ANOVA is the next topic. Nonparametric ANOVA has no normality assumption of random error but the independence of random error is required. The Kruskal-Wallis [2, pp. 589-591] [1, pp. 668-669] test is a popular nonparametric test. Suppose treatment *i* has *ni* observations and there are *k* treatment, let *N = n*1 + *n*2 +…+ *nk* be the total of observations. Kruskal-Wallis test assigns ranks to such *N* observations according to their values, for instance, rank value 1 and rank value *N* to be assigned to smallest and largest sample value. If two or more observations have the same value, these observations are assigned by average rank. For example, if 3rd, 4th and 5th observations get the same value, they receive the same rank (3+4+5) / 3 = 4*.* Let *Rij* be the rank of observation *Yij*. If null hypothesis *H*0: is true, which means that all treatments have the same mean, then ranks spread over all treatments equally. In other words, the expected value of *Rij* (s) is nearly equal to the mid-point of *N* ranks, so we have [2, p. 589]:

*E*(*Rij*) = (*N* + 1) / 2

Let = be average rank of treatment *i*, the expected value of is determined as below [2, p. 589]:

If the null hypothesis *H*0: is true, the average rank does not shift from its expected value (*N*+1) / 2 much. The difference between and its expected value (*N* + 1) / 2 is determined by following statistic [2, p. 589]:

This formula is transformed into more practical format as below [2, p. 590]:

Where is the sum of ranks under treatment *i*. It is proved that statistic *K* approaches chi-square distribution with *k –* 1 degrees of freedom where *k* is the number of treatments. Null hypothesis *H*0: is rejected in flavor of alternative hypothesis *H*1: if *K >* .

**5.0 Nonparametric goodness-fit-test**

Goodness-fit-test is the test that determines whether a sample conforms specified distribution or whether two samples have the same distribution. Although Kolmogorov–Smirnov goodness-fit-test being a kind of nonparametric testing does not consider the sample distribution, it is based on the definition of Kolmogorov distribution. Kolmogorov distribution is continuous distribution whose accumulative distribution function is defined as below [3]:

The critical value *Kα* at significant level *α* is 100*α* percentage point satisfying equation:

Nonparametric Kolmogorov–Smirnov test is to determine whether two samples have the same distribution regardless of the underlying distribution. Given *X =* {*x*1, *x2*,…, *xn*} and *Y =* {*y*1, *y*2,…, *yn*} are two testing samples, the null hypothesis *H*0 is that *X* and *Y* have the same distribution. Let *FX* and *FY* be the empirical distribution functions of *X* and *Y*, respectively. Note that empirical distribution function is accumulative function which increases gradually according to the order of values.

Let *D* be the maximum absolute deviation between *FX* and *FY* over whole samples *X* and *Y*.

It is easy to recognize that the process to find out *D* is iterative process browsing all pairs of observation (*xi*, *yi*) *X Y*. It is proved that *D* conforms *K* distribution. Therefore, the null hypothesis *H*0 is rejected at significant level *α* if *D > Kα*.

**6.0 Conclusion**

Now we had a general and detailed point of view about nonparametric testing. We can draw two main comments from research over this domain:

1. Firstly, nonparametric model is less efficient than parametric model because it lacks valuable information under sample when it has no knowledge about the distribution. All properties of distribution such as mean, variance, standard deviation, median, mode, skewness, kurtosis, etc are essential information of which nonparametric model does not take advantages. However, nonparametric testing is very useful and appropriate to cases that knowledge of distribution cannot be extracted or sample does not conform normal distribution. In case that underlying distribution is ignored and nonparametric testing is the best choice. Therefore, we conclude that the most important thing is to choose appropriate model (parametric or nonparametric) which is adaptive to testing situation and testing requirement.
2. Secondly, nonparametric model is often based on ranking. Ranking process aims to transform origin sample into simpler sample so-called ranking sample. Ranking sample is the set of ranks; thus, each rank is assigned to respective observation from origin sample. Because nonparametric model does not know valuable information of origin sample such as mean, variance, standard deviation; it will exploit ranking sample to discover such valuable information. Therefore, nonparametric testing, in turn, applies parametric methods into the ranking sample. Concretely, nonparametric testing assumes that statistic (s) on ranking sample conform some pre-defined distributions. For example, sign test assumes that the number of plus signs in ranking data conforms binominal distribution, signed-rank test and sum-rank test apply Wilcoxon distribution into ranking data and nonparametric goodness-fit-test is based on Kolmogorov distribution. We conclude that parametric testing and nonparametric testing have a strongly mutual relationship and so, we should take advantages of both of them.

**Reference**

1. R. E. Walpole, R. H. Myers, S. L. Myers and K. Ye, Probability & Statistics for Engineers & Scientists, 9th ed., D. Lynch, Ed., Boston, Massachusetts: Pearson Education, Inc., 2012, p. 816.
2. D. C. Montgomery and G. C. Runger, Applied Statistics and Probability for Engineers, 3rd Edition ed., New York, NY: John Wiley & Sons, Inc, 2003, p. 706.
3. Wikipedia, "Kolmogorov–Smirnov test," Wikimedia Foundation, 25 March 2016. [Online]. Available: http://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov\_test. [Accessed 2013].