VOLUME 5, NUMBER 2
DECEMBER 2009

ISSN 1693-5098

JOURNAL OF QUANTITATIVE METHODS

JOURNAL DEVOTED TO THE MATHEMATICAL AND STATISTICAL APPLICATIONS IN VARIOUS FIELDS

QUANTITATIVE METHODS

QUANTITATIVE METHODS provide a forum for communication between statisticians and mathematicians with the users of applied of statistical and mathematical techniques among a wide range of areas such as operations research, computing, actuary, engineering, physics, biology, medicine, agriculture, environment, management, business, economics, politics, sociology, and education. The journal will emphasize on both the relevance of numerical techniques and quantitative ideas in applied areas and theoretical development, which clearly demonstrate significant, applied potential.

Vol. 5, No. 2 December 2009 QUANTITATIVE METHODS ISSN 1693-5098

TABLE OF CONTENTS

JOURNAL OF QUANTITATIVE METHODS
Vol. 5, No. 2, December 2009 ISSN 1693-5098

USE OF RANKS FOR TESTING FIXED TREATMENT EFFECTS IN BASIC LATIN SQUARE DESIGN

Sigit Nugroho
University of Bengkulu, Indonesia
snugroho@unib.ac.id
(Received 10 Junel 2009, accepted 10 October 2009, published December 2009)

Abstracts. Friedman, Kruskall-Wallis, Durbin, and Anderson have developed nonparametric test in the design of experiments for fixed treatment effects. In this paper, such statistics using ranks to test fixed treatment effects in the basic latin square design is proposed.

Keywords : Latin square, rank test, fixed treatment effect.

1. Introduction

Friedman (1937) developed rank test for fixed treatment effects in the basic randomized complete block design. Friedman statistic is defined as

$$F = \frac{12}{bk(k+1)} \sum_{j=1}^{k} \left[R_j - \frac{b(k+1)}{2} \right]^2 \quad (1)$$

or equivalently

$$F = \frac{12}{bk(k+1)} \sum_{j=1}^{k} R_j^2 - 3b(k+1) \quad (2)$$

where b is the number of blocks while k is the level of treatments used. R_j is the total ranks of the observations having j-th treatment.

Durbin (1951) also developed Friedman's idea in the Balanced Incomplete Block Design. Durbin statistic is as follows:

$$D = \frac{12(t-1)}{rt(k-1)(k+1)} \sum_{j=1}^{t} \left[R_j - \frac{r(k+1)}{2} \right]^2 \quad (3)$$

which is equivalent to

$$D = \frac{12(t-1)}{rt(k-1)(k+1)} \sum_{j=1}^{t} R_j^2 - 3\frac{r(t-1)(k+1)}{k-1} \quad (4)$$

where t is the number of treatments used, k is the size of the blocks, b is the number of blocks used in the experiment, r is the frequency each treatment applied, and λ the number of blocks where i-th and j-th treatment appeared together in a block, with $\lambda(t-1)=r(k-1)$ and $kb = rt$.

For the completely randomized design, Kruskal-Wallis introduced the test in 1952. Kruskal-Wallis test statistic is as follows:

$$W = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{\left[R_i - (1/2)n_i(N+1) \right]^2}{n_i} \quad (5)$$

for $m = 2, 3, ..., 2t$. and is equal to 0 for other m.

It could be verified the above probability density function for $t = 2$

Table 1. Possible cell total ranks when $t = 2$.

Ranking by		Column	
		1	2
Row	1	2	3
	2	3	4

The probability density function is

$$P(R(X_{ij}^{(k)}) = m) = \begin{cases} 1/4 & , \quad m = 2 \\ 2/4 & , \quad m = 3 \\ 1/4 & , \quad m = 4 \\ 0 & , \quad others \end{cases}$$

for $t = 3$ results in total ranks as in the table given below

Table 2. Possible cell total ranks when $t = 3$.

Ranking by		Column		
		1	2	3
Row	1	2	3	4
	2	3	4	5
	3	4	5	6

and the probability density function is

$$P(R(X_{ij}^{(k)}) = m) = \begin{cases} 1/9 & , \quad m = 2 \\ 2/9 & , \quad m = 3 \\ 3/9 & , \quad m = 4 \\ 2/9 & , \quad m = 5 \\ 1/9 & , \quad m = 6 \\ 0 & , \quad others \end{cases}$$

Similarly, for $t = 4$ will result in total ranks as in the table given below

Table 3. Possible cell total ranks when $t = 4$.

Ranking by		Column			
		1	2	3	4
Row	1	2	3	4	5
	2	3	4	5	6
	3	4	5	6	7
	4	5	6	7	8

and the probability density function is

or can be written as

$$W = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \tag{6}$$

where n_i is the number of observations having i-th treatment, $N = \sum_{i=1}^{k} n_i$. R_i is the total ranks of the observations having i-th treatment.

Friedman alternative test is given by Anderson (1959). Anderson suggested the statistic

$$A = \frac{k}{b} \sum_{l=1}^{k} \sum_{j=1}^{k} \left(D_{lj} - \frac{b}{k} \right)^2 \tag{7}$$

where D_{lj} is the number of blocks where j-th treatment having ranks l, b is the number of blocks and k is the treatment level used.

In general, the nonparametric procedures for analyzing complex experimental designs are awkward and tedious. Until better methods are made available in those areas, the experimenter is almost forced to use parametric procedures with their often unrealistic assumptions. (Conover, 1971).

Notation
Latin square model can be written as

$$X_{ij}^{(k)} = \mu + \beta_i + \gamma_j + \tau_{(k)} + \varepsilon_{ij}^{(k)} \tag{8}$$

$i = 1,2,...,t \quad j = 1,2,...,t \ dan \ k = 1,2,...,t$

where $X_{ij}^{(k)}$ is the observation in i-th row blocking and j-th column blocking and receiving k-th treatment; μ is the general average, β_i is the i-th row blocking effects, γ_j is the j-th column blocking effects, $\tau_{(k)}$ is the k-th treatment effects which is equal to $\tau_{(k)}$ if k-th treatment is in i-th row blocking and in j-th column blocking and is equal to zero when it is absent in that cell, i.e. in i-th row blocking and in j-th column blocking and $\varepsilon_{ij}^{(k)}$ is the experimental error in i-th row blocking and in j-th column blocking having k-th treatment.

$R(X_{ij}^{(k)})$ is the (i,j)-th cell rank. according to i-th row blocking and j-th column blocking added. This method of ranking follows from the way we usually do randomization in the latin square design. In the $t \times t$ latin square, where each treatment is applied once in each row blocking and each column blocking, therefore, ranking in each row and column blocking can take a possible value of $1, 2, ..., t$; so that $R(X_{ij}^{(k)})$ may vary from $2, 3, ..., 2t$. As we know that the distribution of this ranking either in row or column blocking is discrete uniform. Notice that row blockings are independent of column blockings.

Probability Density Function of the (i,j)-th Cell Total Ranks

Based on the way we rank and calculating total rank of each (i,j)-th cell, the probability of the (i,j)-th cell having k-th treatment is equal to m is :

$$P(R(X_{ij}^{(k)}) = m) = \frac{\min(m, t+1) - \max(m, t+1) + t}{t^2} \tag{9}$$

$$P(R(X_{ij}^{(k)})=m)=\begin{cases}1/16 & , \quad m=2\\ 2/16 & , \quad m=3\\ 3/16 & , \quad m=4\\ 4/16 & , \quad m=5\\ 3/16 & , \quad m=6\\ 2/16 & , \quad m=7\\ 1/16 & , \quad m=8\\ 0 & , \quad others\end{cases}$$

For $t \geq 5$, it can be verified similarly and finally will get the general formula for the probability density function as in (9).

Mean and Variance of the (i,j)-th Cell Total Ranks, $R(X_{ij}^{(k)})$

After getting the probability density function for the *(i,j)*-th cell, which depends on the size of the square, then the mean (expected value) and variance ot the total rank for the *(i,j)*-th cell can be calculated. Based on the definition of the expected value for the discrete random variables, its mean can be written as

$$E(R(X_{ij}^{(k)}))=\sum_{m=2}^{2t} m\cdot P(R(X_{ij}^{(k)})=m) \tag{10}$$

It can be verified using the definition as stated in (10), the expected value for the series ot *t* (where *t* is the size of the square) are tabled as follows:

Table 4. Expected value of *(i,j)-th* cell total ranks

t	$E(R(X_{ij}^{(k)}))$
2	3
3	4
4	5
5	6
...	...
p	p+1

Therefore, in general, the expected value of the total rank for the (i,j)-th cell in (10) can be written as

$$E(R(X_{ij}^{(k)}))=t+1 \tag{11}$$

Thus, the expected value only depends on *t* which is also called the level of the treatment used or the size of row or column blocking.

Meanwhile, the variance of the total rank for the *(i,j)*-th cell, by definition, and using the information above is equal to

$$\mathrm{var}(R(X_{ij}^{(k)}))=\sum_{m=2}^{2t}(m-(t+1))^2\cdot P(R(X_{ij}^{(k)})=m) \tag{12}$$

Table 5. Variance of *(i,j)-th* cell total ranks

t	$\mathrm{var}(R(X_{ij}^{(k)}))$
2	2/4
3	12/9
4	40/16
5	100/25
6	210/36
7	392/49
...	...
p	$\dfrac{\frac{p}{2}\left(\sum_{i=1}^{p} i^2 - \sum_{i=1}^{p} i\right)}{p^2}$

Therefore, in general, when the level of the treatment or the size of row or column blocking is equal to t, the *(i,j)*-th cell total rank variance is equal to

$$\mathrm{var}(R(X_{ij}^{(k)})) = \frac{\frac{1}{2}\left(\sum_{i=1}^{t} i^2 - \sum_{i=1}^{t} i\right)}{t} \qquad (13)$$

As in getting its mean or its expected value, its variance only depends on the level of the treatment used in the experiment or the size of row or column blocking, t.

In addition, equation (13) which is also the variance of the *(i,j)*-th cell total rank can be written as a recursive function. The recursive function stated is

$$f(t+1) = \frac{t}{t+1} f(t) + \frac{1}{2} t \; ; \; f(2) = \frac{2}{4} \qquad (15)$$

for $t = 2, 3, 4, ...$

It can be concluded that the (i,j)-th cell total rank has mean and variance which depends on t and forms a series, which we name them LS-t mean and variance series.

Treatment Total Rank, $R(X^{(k)})$

For the k-th treatment ($k=1,2,...,t$), its treatment total rank is the sum of all (i,j)-th cell total rank receiving k-th treatment, and is denoted by $R(X^{(k)})$. Notice that between treatments are independent, and so are between row and column blockings, and between treatments and blockings, too.

The mean of the treatment total rank is

$$\begin{aligned} E(R(X^{(k)})) &= E\left(\sum_{l=1}^{t} R(X_{ij}^{(k_l)})\right) \\ &= \sum_{l=1}^{t} E(R(X_{ij}^{k_l})) \qquad (15) \\ &= \sum_{l=1}^{t} (t+1) = t(t+1) \end{aligned}$$

The variance of the treatment total rank is given by the followings:

$$\begin{aligned}\operatorname{var}(R(X^{(k)})) &= \operatorname{var}(\sum_{l=1}^{t} R(X_{ij}^{(k_l)})) \\ &= \sum_{l=1}^{t} \operatorname{var}(R(X_{ij}^{k_l})) \\ &= \sum_{l=1}^{t} \frac{\frac{1}{2}\left(\sum_{i=1}^{t} i^2 - \sum_{i=1}^{t} i\right)}{t} \\ &= \frac{1}{2}\left(\sum_{i=1}^{t} i^2 - \sum_{i=1}^{t} i\right)\end{aligned} \tag{16}$$

Whenever t is large enough, we can use Central Limit Theorem to show that Standard Normal distribution can be used as an approximate distribution to the following random variable

$$\frac{R(X^{(k)}) - E(R(X^{(k)}))}{\sqrt{\operatorname{var}(R(X^{(k)}))}} \tag{17}$$

Hence, we can use Chi-square distribution with k degrees of freedom as an approximate distribution the following random variable

$$S = \sum_{k=1}^{t} \frac{(R(X^{(k)}) - E(R(X^{(k)})))^2}{\operatorname{var}(R(X^{(k)}))} \tag{18}$$

if $R(X^{(k)})$ are independent to each other. But $R(X^{(k)})$ are dependent, since $\sum_{k=1}^{t} R(X^{(k)}) = t^2(t+1)$. Knowing the first $t-1$ $R(X^{(k)})$, we can find the last $R(X^{(k)})$ automatically since its total is already known. Thus, the random variable stated in (18) is therefore having Chi-square distribution with $t-1$ degrees of freedom. Statistics in (18) can also be written as

$$S = \frac{2}{\left(\sum_{i=1}^{t} i^2 - \sum_{i=1}^{t} i\right)} \sum_{k=1}^{t} (R(X^{(k)}) - t(t+1))^2 \tag{19}$$

S-statistics

In general, for the $t \times t$ latin square, the number of possible composition of ranks are $(t!)^{2t}$. For $t > 3$ building the table that gives possible value of s, frequency for each s occur, probability and cumulative probability, and p-value of s is tedious. Therefore, as long as there is no available exact table, chi-square distribution with $t-1$ degrees of freedom can be used as an approximate distribution.

The exact table for $t = 2$ can be easily constructed, since it is only used $(2!)^2(2!)^2 = 16$ possible rank composition. The number, the probability, the cumulative probability, and the p-value of s are given in Table 6.

Table 6. Possible value of s, frequency, probability, cumulative probability, and p-value of s when $t = 2$.

s	# s	P(S=s)	P(S≤s)	P(S≥s)
0	6	0.375	0.375	1.000
2	8	0.500	0.875	0.625
8	2	0.125	1.000	0.125

When $t = 3$ there will be $(3!)^6 = 46656$ possible compositions in arranging the ranks. The table is given below.

Table 7. Possible value of s, frequency, probability, cumulative probability, and p-value of s when $t = 3$.

S	# s	P(S=s)	P(S≤s)	P(S≥s)
0.0	2040	0.0437	0.0437	1.0000
0.5	10080	0.2160	0.2598	0.9563
1.5	7920	0.1698	0.4295	0.7402
2.0	6570	0.1408	0.5703	0.5705
3.5	8280	0.1775	0.7478	0.4297
4.5	3180	0.0682	0.8160	0.2522
6.0	1980	0.0424	0.8584	0.1840
6.5	3240	0.0694	0.9279	0.1416
8.0	936	0.0201	0.9479	0.0721
9.5	1080	0.0231	0.9711	0.0521
10.5	792	0.0170	0.9880	0.0289
12.5	180	0.0039	0.9919	0.0120
13.5	120	0.0026	0.9945	0.0081
14.0	180	0.0039	0.9983	0.0055
15.5	72	0.0015	0.9999	0.0017
18.0	6	0.0001	1.0000	0.0001

Hypothesis of no fixed treatment effect versus alternative hypothesis that there are treatment effects can be tested using S statistic. We do reject the null hypothesis whenever S is large enough, or when *p-value* is less than the level significance of the test.

Examples

First, use the data given below

C 18	A 22	B 17
A 16	B 14	C 19
B 24	C 16	A 18

Ranking according to row blocking, as it does in Friedman test gives

C 2	A 3	B 1
A 2	B 1	C 3
B 3	C 1	A 2

And then if we do ranking according to column blocking results

C 2	A 3	B 1
A 1	B 1	C 3
B 3	C 2	A 2

Thus, the cell total ranks are

C 4	A 6	B 2
A 3	B 2	C 6
B 6	C 3	A 4

Finally, we will get, treatment total ranks easily, A = 6+3+4 = 13; B = 2+2+6 = 10; and C = 4+6+3 = 13.
And the value of S can be calculated using (19)

$$S_{hitung} = \frac{(13-12)^2 + (10-12)^2 + (13-12)^2}{\frac{1}{2}(14-6)} = 1.5$$

From table 7 we get *p-value* = *0.7402*. If we use 5% level of significance for testing the hypothesis, then null hypothesis is not rejected, since *p-value* = *0.7402* > *0.05*. It can be interpreted as A, B, and C having the same treatment effect. If chi-square distribution with 2 degrees of freedom is used to approximate the *p-value* = *0.4724* which is also resulting not to reject the null hypothesis.

Second example is using the following data

C 28	A 7	B 17
A 6	B 16	C 34
B 18	C 26	A 8

The calculated S-statistic = 18, gives the *p-value* according to table 7 is equal to *0.0001* resulting the rejection of the null hypothesis; thus, treatment A, B, and C having different fixed treatment effects. At least two different treatments have different effects. Note that $P(\chi^2_2 = 0.0001)$ which is so close to the exact table.

References

[1] Anderson, R.L. (1959). Use of contingency tables in the analysis of consumer preference studies. *Biometrics.* 15, 582-590.
[2] Conover, W.J. (1971). *Practical Nonparametric Statistics*. New York:John Wiley & Sons. 462p.
[3] Durbin, J. (1951). Incomplete blocks in ranking experiments. *British Journal of Psychology. Statistical Section,* 4, 85-90.
[4] Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association,* 32, 675-701.
[5] Gibbons, J.D. (1985). *Nonparametric Statistical Inference*. New York:Mercel Dekker, Inc.
[6] Govindarajulu, Z. (2007). *Nonparametric Inference*. World Scientific Publishing Co. Pte. Ltd. Singapore.
[7] Kruskal, W.H. and W.A. Wallis. (1952). Use of ranks on one-criterion variance analysis. *Journal of the American Statistical Association,* 47, 583-621.
[8] Randles, R.H, and D.A. Wolfe. (1979). *Introduction to the Theory of Nonparametric Statistics*. New Jersey:John Wiley & Sons. Hoboken.
[9] Sprent, P. and N.C. Smeeton. (2001). *Applied Nonparametric Statistical Methods*. New York:Chapmann & Hall.