ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:2 - Sayı/No: 1 : 73-76 (2001)

ARAŞTIRMA MAKALESİ/**RESEARCH ARTICLE**

SUPERIORITY TEST BY THE WAY OF BERNOULLI TRIALS

Adil KORKMAZ[1], Süleyman GÜNAY[2]

ABSTRACT

In this study it has been proposed a new nonparametric method for superiority test by pointing out preference-ranking problem facing in Friedman test.

Key Words: Friedman Test, Superiority Test and Bernoulli Trials.

BERNOULLİ DENEMELERİYLE ÜSTÜNLÜK TESTİ

ÖZ

Bu çalışmada Friedman testinde karşılaşılan tercihleri rütbelendirme sorununa değinilerek üstünlük testi için yeni bir yöntem önerilmektedir.

Anahtar Kelimeler: Friedman Testi, Üstünlük Testi ve Bernoulli Denemeleri.

1. INTRODUCTION

Kruskal-Wallis test (Kruskal, 1952) and Friedman test are the nonparametric types of analysis of variance. Kruskal-Wallis test is a generalised type of Mann-Whitney U test (Mann-Whitney, 1947). Friedman test has been used to reply for the question whether some alternatives are more preferable than the others are. If some alternatives are more preferable than the others it can be argued those are superior. However if none of the alternatives are more preferable than the others are, it can be argued those are equivalent.

Suppose that there are n alternatives that would be preferred by m individuals. Someone appoints the first or last rank for the most or least preferable alternative respectively and the other ranks between the first and the last for the other alternatives. Obviously the rank, which shows the order of any individual's preference under the condition that all the alternatives are equivalent, is a random variable that has discrete uniform distribution. Now a question comes immediately: What is the reality behind the alternatives whether some of them are more preferable then the others or all of them are equivalent. Using all the individual's preference ranks, Friedman test has been replied to this question.

However there may be a preference-ranking problem in that test. This problem, which has been raised in the time of data collecting for the test, may be two kinds. **The one** could be raised from a situation that the alternatives are equivalent. In those states any individual cannot appoint any rank to any alternative because he/she is on an indifference situation that the alternatives are not more or less preferable than the others. **The two** is that there may be a lot of alternatives. In mentioned situation it may be easy to appoint the first, the second or the third ranks to the corresponding alternatives but may not be easy to appoint all the ranks to all the alternatives. **Those** may be characterised as a preference-ranking problem during the time of the data collecting.

1 Akdeniz Üniversitesi İktisadi ve İdari Bilimler Fakültesi İktisat Bölümü ANTALYA.
2 Hacettepe Üniversitesi Fen Bilimleri Enstitüsü İstatistik Bölümü Beytepe-ANKARA.

Received: 10 February 2000 ; **Accepted:** 15 February 2001.

However there is no preference-ranking problem in the method proposed here. Because nobody does not appoint any rank to any alternative in the present method. As it would be understood, he/she only says that any alternative is or is not preferable (one can put "important" into the place of "preferable").

2. PROPOSED METHOD

If somebody says that any alternative is important then a Bernoulli trial is successful; reversely if he/she says that any alternative is not important then the Bernoulli trial is unsuccessful. Nobody compares any two alternatives and only considers single alternative while he/she decides. Proposed method has been used those Bernoulli random trials for the superiority test against Friedman test.

Suppose that there are m homogeneous individuals as $I_1, I_2, \ldots, I_m$ and n alternatives as $A_1, A_2, \ldots, A_n$. If g^{th} individual says that i^{th} alternative is important then the Bernoulli trial $x_{gi}=1$ and otherwise $x_{gi}=0$. Those random variables may be defined for $g=1,2,\ldots,m$ and $i=1,2,\ldots,n$ which has been shown as a contingency table below:

Where

$$X_{g\bullet} = \sum_{i=1}^{n} X_{gi}, \quad X_{\bullet i} = \sum_{g=1}^{m} X_{gi}, \quad X_{\bullet\bullet} = \sum_{g=1}^{m} \sum_{i=1}^{n} X_{gi}$$

3.PRELIMINARY CONDITIONS

The first preliminary condition given is that the individuals are homogeneous. This means that $P(x_{gi}=1)=P(x_{hi}=1)$ for $g \neq h=1,2,\ldots,m$ and $i=1,2,\ldots,n$. Obviously if the individuals are not homogeneous, the superiority test cannot be possible because in that case individual effect may not be eliminated. Under the first preliminary condition, the main question is whether the alternatives are homogeneous or not. For the reason of that the test hypothesis can be expressed as follows:

H_0: The alternatives are homogeneous.

Table: Contingency Table of Bernoulli Trials

	Row↓	A_1	A_2	...	A_n	The Sum of the Row
Column→		1	2	...	n	
I_1	1	x_{11}	x_{12}	...	x_{1n}	$x_{1\bullet}$
I_2	2	x_{21}	x_{22}	...	x_{2n}	$x_{2\bullet}$
...	...	...	...	...	...	...
I_m	m	x_{m1}	x_{m2}	...	x_{mn}	$x_{m\bullet}$
The Sum of the Column		$x_{\bullet 1}$	$x_{\bullet 2}$	...	$x_{\bullet n}$	$x_{\bullet\bullet}$

This means that $P(x_{gi}=1)=P(x_{gj}=1)$ for $g=1,2,\ldots,m$ and $i \neq j=1,2,\ldots,n$. Let the common probability be p for all the intersections of all the individuals and all the alternatives. Then the following equation immediately comes:

$$P(x_{gi}=1)=p, \qquad g=1,2,\ldots,m;\ i=1,2,\ldots,n$$

The first preliminary condition is not one and only one condition for the superiority test. The **second** preliminary condition, which is exactly necessary for the superiority test, may be called as the independency property of the alternatives. Under this condition, the following relation may be written:

$$P(x_{gi}=1, x_{gj}=1)=P(x_{gi}=1)\cdot P(x_{gj}=1),$$

$$g=1,2,\ldots,m;\ i \neq j=1,2,\ldots,n$$

This means that if some individual prefers some alternative with a probability, obviously and independently he/she also prefers another one with the same probability under the homogeneity condition.

4. CENTRAL LIMIT THEOREM FOR BERNOULLI TRIALS

Suppose that the random variables $z_1, z_2, \ldots, z_k$ are independent Bernoulli trials with the probability p. Let $q=1-p$. The characteristic function of the standardised sum of $z_1, z_2, \ldots, z_k$ may be written like that:

$$\left(p \cdot e^{i \cdot \frac{q \cdot t}{\sqrt{k \cdot p \cdot q}}} + q \cdot e^{-i \cdot \frac{p \cdot t}{\sqrt{k \cdot p \cdot q}}}\right)^k$$

Simplifying the following equation is obtained:

$$\left(1 - \frac{t^2}{2 \cdot k} + Ok^{-3/2}\right)^k$$

The result converges to $e^{-\frac{t^2}{2}}$ for $k \to \infty$. This means that standardised sum of $z_1, z_2, \ldots, z_k$ is asymptotically standard normal variable. This is also a lemma of De Moivre-Laplace theorem (Gnedenko, 1976) because the sum of of $z_1, z_2, \ldots, z_k$ is a binomial random variable.

5.THE STATISTIC FOR H_0

From the above theorem, each of $x_{\bullet 1}, x_{\bullet 2}, \ldots, x_{\bullet n}$ have independently and asymptotically standard normal distribution. The expected value and the variance of mentioned binomial random variables might be written as follows:

$$E(x_{\bullet i})=m \cdot p, Var(x_{\bullet i})=m \cdot p \cdot (1-p), \qquad i=1,2,\ldots,n$$

Under the homogeneity condition the random variables $\frac{x_{\bullet i} - m \cdot p}{\sqrt{m \cdot p \cdot (1-p)}}$ which is standardised type of the random variable $x_{\bullet i}$-m·p for i=1,2,...,n and also their quadratic sum would be small. That sum may be written like that:

$$S = \sum_{i=1}^{n} \frac{(x_{\bullet i} - m \cdot p)^2}{m \cdot p \cdot (1-p)} \qquad (1)$$

It is clear that S has chi-square distribution with n degrees of freedom for m→∞ for the reason of well-known definition of chi-square random variable. S cannot use as a statistic because p is unknown (Mood-Graybill, 1963). The uniformly minimum varianced unbiased estimator of p is as follows (Casella-Berger, 1990; İnal-Günay, 1993):

$$\hat{P} = \frac{x_{\bullet\bullet}}{m \cdot n} \qquad (2)$$

This result may be derived from the following expressions. Suppose that $z_1, z_2, \ldots, z_k$ are Bernoulli trials with p. Let $u1 = z_1 + z_2 + \ldots + z_k$.

Theorem 1: u_1 is sufficient for p.

Proof

While u_1 is given the conditional probability of $z_1, z_2, \ldots, z_k$ is below:

$$\frac{P(z_1, z_2, \ldots, z_k)}{P(u_1)} = \frac{p^{z_1} \cdot (1-p)^{1-z_1} \cdot p^{z_2} \cdot (1-p)^{1-z_2} \cdots p^{z_k} \cdot (1-p)^{1-z_k}}{\binom{k}{u_1} \cdot p^{u_1} \cdot (1-p)^{k-u_1}} = \frac{1}{\binom{k}{u_1}}$$

While $z_1+z_2+\ldots+z_k$ is given the conditional probability of $z_1, z_2, \ldots, z_k$ is independent of p. Obviously this means that u1 is sufficient for p.

Theorem 2: $\frac{u_1}{k}$ is an unbiased estimator for p.

Proof

It is unbiased estimator because its expected value is equal to p.

Theorem 3: $\frac{u_1}{k}$ is the minimum varianced estimator for p.

Proof

Because both Cramer-Rao boundary for any p estimator and the variance of $\frac{u_1}{k}$ are the same. Using likelihood function Cramer-Rao boundary may be derived by for any p estimator. Likelihood function of Bernoulli trials is as follows:

$$L(Z_1, Z_2, \ldots, Z_k) = p^{\sum_{\alpha=1}^{k} z_\alpha} \cdot (1-p)^{k - \sum_{\alpha=1}^{k} z_\alpha}$$

Suppose that τ(p) is a function of the parameter p. Let $T(z_1, z_2, \ldots, z_k)$ be an estimator of τ(p). Cramer-Rao boundary of an estimator of τ(p) is as follows (İnal-Günay, 1993):

$$V[T(Z_1, Z_2, \ldots, Z_k)] \geq \frac{-[\tau'(p)]^2}{E\left[\frac{\partial^2 \log L}{\partial p^2}\right]}$$

The nominator and the denominator of Cramer-Rao inequality may be found as follows:

$$\log L = \sum_{\alpha=1}^{k} z_\alpha \cdot \log p + \left(k - \sum_{\alpha=1}^{k} z_\alpha\right) \cdot (1-p),$$

$$E\left\{\frac{\partial^2 \log L}{\partial p^2}\right\} = -\frac{k}{p \cdot (1-p)}$$

This is the denominator of Cramer-Rao inequality. if τ(p)=p then the nominator is minus one. So,

$$V[T(Z_1, Z_2, \ldots, Z_k)] \geq \frac{p \cdot (1-p)}{k}$$

On the other hand, using the rule about the sum of independent random variables, the variance of $\frac{u_1}{k}$ may also be derived directly as $\frac{p \cdot (1-p)}{k}$. This means that the variance of $\frac{u_1}{k}$ is the minimum variance of an estimator for p. So, $\frac{u_1}{k}$ is the minimum varianced estimator for p.

The completeness property of u1 may also be proved.

Lemma 1

Two equations may be written for the superiority test by the way of Bernoulli trials:

k=m·n,

$u_1 = x_{\bullet\bullet}$

So, uniformly minimum varianced unbiased estimator of p is $\frac{x_{\bullet\bullet}}{m \cdot n}$.

Theorem 4: $\frac{u_1}{k}$ is consistent estimator for p.

Proof

Putting $E(\hat{p}) = p$ and $(\hat{p}) = \frac{p \cdot (1-p)}{m \cdot n}$ into Chebyshev inequality it could be found that the absolute value of $\hat{p} \cdot p$ is zero for m→∞.

Putting $\hat{p}$ into the place p and using (1) the test statistic may now be defined as follows:

$$Q = \frac{\sum_{j=1}^{n} \left(x_{\bullet j} - m \cdot \hat{p}\right)^2}{m \cdot \hat{p} \cdot \left(1 \cdot \hat{p}\right)} \qquad (3)$$

Putting (2) into (3), the following result is obtained:

$$Q = \frac{\sum_{j=1}^{n} \left(x_{\bullet j} - \frac{x_{\bullet\bullet}}{n}\right)^2}{\frac{x_{\bullet\bullet}}{n} \cdot \left(1 - \frac{x_{\bullet\bullet}}{m \cdot n}\right)}$$

It is clear that for the reason of consistency property of $\hat{p}$ while $m \to \infty$, Q converges to S while $m \to \infty$. As it would be known that S converges to χ^2_n while $m \to \infty$. In the case the following relation may be written:

$$Q \to S \to \chi^2_n \qquad \text{while } m \to \infty$$

Suppose that c is a critical value of chi-square distribution with n degrees of freedom for significance level α. Using Neyman-Pearson lemma for the most powerful test, while m is large the decision function $\Psi(Q)$ for H_0 might be defined as follows:

$$\Psi(Q) = \begin{cases} 1 & Q > c \\ 0 & Q \leq c \end{cases}$$

As it would be understood, H_0 must be rejected while $\Psi(Q)=1$.

6. RESULT

The superiority test by the way Bernoulli trials presented here may be used for the situation that the homogeneous individuals have preference-ranking problem while they prefer the alternatives. If there were no preference-ranking problem, obviously Friedman test would be appropriate for superiority test. However sometimes the alternatives may be equivalent and somebody cannot prefer an alternative among the others. In that time it is possible to say that somebody is in indifference situation, which he/she cannot appoint any rank to the alternatives but he/she can say that an alternative is or is not preferable. In those situations the superiority test by the way of Bernoulli trials is the most applicable technique for the homogeneity test while Friedman test is inapplicable. For the reason of that the superiority test by the way of Bernoulli trials is not a competitor for Friedman test, but it is a subsidiary for that in the case there is individual's preference-ranking problem during data collecting.

REFERENCES

Casella, G.-Berger, R. L., (1990), *Statistical Inference,* Duxbury Press, Section 7, Belmont.

Gnedenko, B. V., (1976), *The Theory of Probability,* Mir Publishers, 76, Moscow.

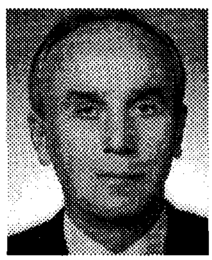
İnal, C.-Günay, S., (1993), *Olasılık ve Matematiksel İstatistik,* H.Ü. Fen Fakültesi, 385-423, Beytepe.

Kruskal, W. H., (1952), "A Nonparametric Test for the Several Sample Problem", *The Annals of Mathematical Statistics,* 23, (4), 525-540.

Mann, H. B.-Whitney, D. R., (1947), "On a Test Whether One of Two Random Variables is Stochastically Larger Than the Other", *The Annals of Mathematical Statistics,* 18, (1), 50-60.

Mood, A. M.-Graybill, F. A., (1963), *Introduction to the Theory of Statistics,* McGraw-Hill, 144, New York.

Adil Korkmaz, was born in Yomra (Trabzon) in 1956. He was graduated from Electrics Faculty of Istanbul Technical University and Economics Faculty of Istanbul University. He received a M.Sc. degree from the Branch of Economics at Graduate Institute Social Sciences, Istanbul University and Ph.D. degree from the Branch of Statistics at Graduate Institute Science, Hacettepe University. Since 2000, he has been working in Akdeniz University.



Süleyman Günay, was born in Düzce in 1949. He was graduated from the Mathematics Department of Ankara University in 1968. He was appointed as a research assistant at the Statistics Department of Hacettepe University and received Ph.D. degree in 1972. He was appointed as an associate professor in 1978 being head of Statistics Department in 1980. He had worked in Warwick University for one year via postdoctorate scholarship of TUBITAK. In 1988, he was appointed as professor and vice dean of Science Faculty of Hacettepe University. He was an editor of Journal of Hacettepe Natural Sciences and Engineering for a period of three years. He has been working as a lecturer at the Branch of Applied Statistics, Statistics Department of Hacettepe University.