Journal of the Korean
Statistical Society
Vol. 23, No. 1, 1994

Rank Transform F Statistic In a 2×2 Factorial Design

Young Hun Choi[1]

ABSTRACT

For a 2×2 factorial design without the restriction of a linear model or without regard to error terms having homoscedasticity, under the null hypothesis of no interaction we can have the rank transformed F statistic for interaction converge in distribution to a chi-squared random variable with one degree of freedom if and only if there is only one main effect.

KEYWORDS: Rank transform, Interaction, Asymptotic distribution, 2×2 factorial design.

1. INTRODUCTION

The rank transform approach has been applied in a variety of circumstances. Much research has been done by statisticians in several fields, especially in the field of experimental designs, to make the rank transform procedure invaluable. However, many papers such as Hora and Conover (1984) primarily dealt with the rank transform procedure for testing for main effects

[1] Department of Statistics, Hanshin University, 411 Yangsan-dong, Osan, 447-791, Korea.

for several experimental designs without interaction. In essence, until recently the rank transform approach did not provide useful solutions for testing for interaction, since the theory for testing for interaction based on rank transform had not yet been developed.

Further when we examine the articles that employed computer generated simulation methods, the contradictory Monte Carlo simulation results conducted by Conover and Iman (1976), who pioneered the rank transform procedure, and by Blair, Sawilowsky, and Higgins (1987) strongly suggest additional research on the rank transform statistic for interaction in a factorial design. The discrepancies of these simulation studies raise serious questions about the usefulness of the rank transform analysis of variance test for analyzing a factorial design due to the growing popularity of the rank transform procedure in many fields. Meanwhile, Thompson (1991) studied the asymptotic properties of the rank transform statistic for testing for interaction in a two-way layout linear model with interaction. However the author did not consider any theoretical studies of the small sample properties of the rank transform test nor did she consider non-additive models. These current major concerns motivated this study of the asymptotic theory for the rank transformed statistic for testing for interaction in a 2×2 factorial design based on the ranks of data without the restriction of a linear model.

Thus the model discussed is a 2×2 factorial design with interaction for 2 blocks, 2 treatments, and N replications without the restriction of an additive linear model or without regard to error terms having homoscedasticity. Throughout this paper, X_{ijn}, $i = 1, \cdots, I$, $j = 1, \cdots, J$, and $n = 1, \cdots, N$, where $I = 2$ and $J = 2$ implicitly, are independent random variables such that X_{ijn} has the continuous distribution function F_{ij} . And let $\bar{F}_{i.}, \bar{F}_{.j}$ and $\bar{F}_{..}$ or H denote the average distribution function for block i, treatment j, and the combination set of all blocks and treatments respectively. In addition, R_{ijn} will denote the rank of random variable X_{ijn} in the combined ranking from 1 to IJN, and $R_{ij.}$, $R_{i..}$, $R.j.$ and $R_{...}$ will denote the sum of ranks for all X_{ijn} having the index (i, j), i, j and the sum of all ranks respectively. Likewise ρ_{ijn} will denote $R_{ijn}/(IJN + 1)$, and $\rho_{ij.}$, $\rho_{i..}$, $\rho_{.j.}$ and $\rho_{...}$ will denote the sum of ranks divided by $(IJN + 1)$ for all X_{ijn} having the index (i, j), i, j and the sum of all ranks respectively.

In Section 2 main results with the hypotheses, two sufficient conditions and a theorem are established. In Section 3 lemmas and proofs are analytically derived for ranks. Many exact, small sample results which will be useful in experimenters and survey investigators are also obtained here for the first time. In Section 4 conclusions are drawn and summarized for a 2×2 factorial design without the restriction of a linear model.

2. MAIN THEORETICAL RESULTS

Now the hypotheses of our interest are

$$H_0 : \int F_{lm} d(F_{ij} - \bar{F}_{i.} - \bar{F}_{.j} + \bar{F}_{..}) = 0 \text{ for all } i,\ j,\ l \text{ and } m, \quad \text{where } l = 1, \cdots, I \text{ and } m = 1, \cdots, J. \tag{2.1}$$

$$H_a : \int F_{lm} d(F_{ij} - \bar{F}_{i.} - \bar{F}_{.j} + \bar{F}_{..}) \neq 0 \text{ for at least one } i,\ j,\ l \text{ and } m.$$

The following two assumptions are used in the derivation of the asymptotic distribution of the test statistic. The first assumption is

$$\sum_{i=1}^{J} \sum_{j=1}^{J} \int H F_{ij} d(F_{ij} - \bar{F}_{i.} - \bar{F}_{.j} + \bar{F}_{..}) = 0 \text{ where } H = \frac{1}{IJ} \sum_{i'=1}^{I} \sum_{j'=1}^{J} F_{i'j'}. \tag{2.2}$$

The second assumption is

$$\int (F_{ij} - \bar{F}_{i.} - \bar{F}_{.j} + \bar{F}_{..})^2 dH = 0 \text{ for all } i \text{ and } j. \tag{2.3}$$

Note that the assumption (2.3) can be satisfied if and only if $F_{ij} = \bar{F}_{i.}$ or $F_{ij} = \bar{F}_{.j}$ for all i and j. Since the distribution H is a positive measurable function and $(F_{ij} - \bar{F}_{i.} - \bar{F}_{.j} + \bar{F}_{..})^2$ is nonnegative, we can say $\int (F_{ij} - \bar{F}_{i.} - \bar{F}_{.j} + \bar{F}_{..})^2 dH = 0$ for all i and j implies $F_{ij} - \bar{F}_{i.} - \bar{F}_{.j} + \bar{F}_{..} = 0$ for all i and j, which in turn implies $F_{ij} = \bar{F}_{i.}$ or $F_{ij} = \bar{F}_{.j}$ for all i and j. Accordingly overall we can say that the above null hypothesis and two assumptions can be met if and only if $F_{ij} = \bar{F}_{i.}$ or $F_{ij} = \bar{F}_{.j}$ for all i and j.

Comment 1. The first assumption (2.2) is necessary to establish that the limit of the mean of numerator of F statistic under the null hypothesis (2.1) is equal to the expected value of the denominator of F statistic, converging in probability to a positive constant. Meanwhile the second assumption (2.3) is sufficient to have $N^{-1/2}(\rho_{ij.} - \rho_{i..}/J - \rho_{.j.}/I + \rho_{...}/IJ)$ asymptotically identically distributed for each block and treatment.

Comment 2. Except for the second assumption (2.3), which is simplified for a 2×2 factorial design, the results of the remaining parts such as the hypothesis (2.1) and the first assumption (2.2) are applicable for all two-way layout. Furthermore ranks are the special case of general scores for which we can also derive the same results.

The usual F statistic with rank transformed data is

$$F_N = \frac{\dfrac{\sum_{i=1}^{I}\sum_{j=1}^{J}(\rho_{ij.} - \frac{\rho_{i..}}{J} - \frac{\rho_{.j.}}{I} + \frac{\rho_{...}}{IJ})^2}{(I-1)(J-1)N}}{\dfrac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{n=1}^{N}(\rho_{ijn} - \frac{\rho_{ij.}}{N})^2}{IJ(N-1)}}.$$

Let $F_N = (\sigma^2 \cdot Q_N)/[V_N \cdot (I-1)(J-1)]$, where $Q_N = [(I-1)(J-1)\sum_{i=1}^{I}\sum_{j=1}^{J}(\rho_{ij.} - \rho_{i..}/J - \rho_{.j.}/I + \rho_{...}/IJ)^2]/IJN\sigma^2$, $\sigma^2 = \lim_{N\to\infty}[\sum_{i=1}^{I}\sum_{j=1}^{J} Var(\rho_{ij.} - \rho_{i..}/J - \rho_{.j.}/I)]/IJN$ and $V_N = [(I-1)(J-1)\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{n=1}^{N}(\rho_{ijn} - \rho_{ij.}/N)^2]/I^2J^2 (N-1)$. Then the limiting distribution of F_N is given by Theorem 1.

Theorem 1. Let X_{ijn}, $i = 1, \cdots, I$, $j = 1, \cdots, J$ and $n = 1, \cdots, N$, be independent random variables, and let R_{ijn} denote the corresponding ranks. Then under the null hypothesis (2.1) and the assumptions (2.2) and (2.3), the statistic F_N for a 2×2 factorial design converges in distribution to a chi-squared random variable with one degree of freedom.

Proof. From Lemma 5 under the hypothesis (2.1) and the assumptions (2.2) and (2.3), $0 < \sigma^2 < (I-1)(J-1)/3IJ$ exists. Hence it follows from Hájek's (1968) Theorem 2.1 that each $N^{-1/2}(\rho_{ij.} - \rho_{i..}/J - \rho_{.j.}/I + \rho_{...}/IJ)$, for $i = 1, \cdots, I$ and $j = 1, \cdots, J$, converges in distribution to a normal random variable.

Now consider the vector

$$R^{o\prime} = \left[\rho_{11.} - \frac{\rho_{1..}}{J} - \frac{\rho_{.1.}}{I} + \frac{\rho_{...}}{IJ}, \cdots, \rho_{IJ.} - \frac{\rho_{I..}}{J} - \frac{\rho_{.J.}}{I} + \frac{\rho_{...}}{IJ}\right].$$

Every linear combination of $N^{-1/2}R^o$ converges to normality with the covariance terms provided in Lemma 4. Then Hájek and Sidák's (1967) Theorem

V.2.1 and Hora and Iman's (1988) Lemma 2 may be applied to show that each $N^{-1/2}R^o$ for $i = 1, \cdots, I$ and $j = 1, \cdots, J$ converges to a multivariate normal random vector.

Finally using Graybill's(1976) Theorem 4.4.3, Q_N converges in distribution to a chi-squared random variable with one degree of freedom for a 2×2 factorial design. Further from the results (3.6) of Lemma 2 and Lemma 5 and since V_N converges in probability to a constant, it follows that V_N converges in probability to σ^2. Thus this completes the proof.

3. LEMMAS WITH PROOFS

This section consists of five lemmas. The first three lemmas are mainly related to derive the null hypothesis (2.1) and first assumption (2.2) to have $E(F_N) = 1 + O(1/N)$. The next fourth and fifth lemmas are used to construct the additional second assumption (2.3) in order to have $N^{-1/2}(\rho_{ij.} - \rho_{i..}/J - \rho_{.j.}/I + \rho_{...}/IJ)$ asymptotically identically distributed for each block and treatment, and to have Q_N chi-squared.

Lemma 1. Let X_{ijn}, $i = 1, \cdots, I$, $j = 1, \cdots, J$ and $n = 1, \cdots, N$, be independent random variables such that X_{ijn} has the distribution function F_{ij}. Then

$$(1)\ E(R_{ij1}) = IJN\int HdF_{ij} + \frac{1}{2}. \tag{3.1}$$

$$\begin{aligned}(2)\ E(R_{ij1}\cdot R_{i'j'2}) = {} & I^2J^2N^2\int HdF_{ij}\int HdF_{i'j'} - N\sum_{l=1}^{I}\sum_{m=1}^{J}\int F_{lm}dF_{ij}\int F_{lm}dF_{i'j'} \\ & +\frac{5}{2}IJN\int HdF_{ij} - 2IJN\int HF_{i'j'}dF_{ij} + \int F^2_{i'j'}dF_{ij} \\ & +\frac{5}{2}IJN\int HdF_{i'j'} - 2IJN\int HF_{ij}dF_{i'j'} + \int F^2_{ij}dF_{i'j'} \\ & -IJN\int F_{i'j'}dF_{ij}\int HdF_{i'j'} - IJN\int HdF_{ij}\int F_{ij}dF_{i'j'} \\ & +\int F_{i'j'}dF_{ij}\int F_{ij}dF_{i'j'} - \frac{3}{4} \\ & \text{for all } i,\ j,\ i' \text{ and } j',\text{ where } i' = 1,\cdots,I,\ j' = 1,\cdots,J.\end{aligned} \tag{3.2}$$

Proof. (1) Introducing the function, $\mu(t) = 1$ if $t \geq 0$, $\mu(t) = 0$ if $t < 0$, we may write

$$R_{ij1} = \sum_{l=1}^{I} \sum_{m=1}^{J} \sum_{k=1}^{N} \mu(X_{ij1} - X_{lmk})$$

and

$$\begin{aligned} E[\mu(X_{ij1} - X_{lmk})] &= \int F_{lm} dF_{ij} \quad \text{for } (l,\ m,\ k) \neq (i,\ j,\ 1) \\ &= 1 \qquad\qquad\quad \text{for } (l,\ m,\ k) = (i,\ j,\ 1). \end{aligned}$$

So, $E(R_{ij1}) = \sum_{l=1}^{I} \sum_{m=1}^{J} \sum_{k=1}^{N} \int F_{lm} dF_{ij} - \int F_{ij} dF_{ij} + 1 = IJN \int H dF_{ij} + \frac{1}{2}.$

(2) In a similar manner,

$$R_{ij1} \cdot R_{i'j'2} = \sum_{l=1}^{I} \sum_{m=1}^{J} \sum_{k=1}^{N} \sum_{l'=1}^{I} \sum_{m'=1}^{J} \sum_{k'=1}^{N} \mu(X_{ij1} - X_{lmk}) \cdot \mu(X_{i'j'2} - X_{l'm'k'}). \quad (3.3)$$

Note that we have

$$E[\mu(X_{ij1} - X_{lmk}) \cdot \mu(X_{i'j'2} - X_{l'm'k'})] = \int F_{lm} dF_{ij} \cdot \int F_{l'm'} dF_{i'j'}$$

when $(l,\ m,\ k) \neq (i,\ j,\ 1)$ or $(i',\ j',\ 2)$ and $(l',\ m',\ k') \neq (i,\ j,\ 1),\ (i',\ j',\ 2)$ or $(l,\ m,\ k)$.

Therefore accounting for the above terms by removing and replacing with the correct terms after adding the intersection terms subtracted more than once, it follows that

$$\begin{aligned} &E(R_{ij1} \cdot R_{i'j'2}) = \\ &N^2 \sum_{l=1}^{I} \sum_{m=1}^{J} \sum_{l'=1}^{I} \sum_{m'=1}^{J} \int F_{lm} dF_{ij} \int F_{l'm'} dF_{i'j'} - N \sum_{l=1}^{I} \sum_{m=1}^{J} \int F_{ij} dF_{ij} \int F_{lm} dF_{i'j'} \\ &- N \sum_{l=1}^{I} \sum_{m=1}^{J} \int F_{i'j'} dF_{ij} \int F_{lm} dF_{i'j'} - N \sum_{l=1}^{I} \sum_{m=1}^{J} \int F_{lm} dF_{ij} \int F_{ij} dF_{i'j'} \\ &- N \sum_{l=1}^{I} \sum_{m=1}^{J} \int F_{lm} dF_{ij} \int F_{i'j'} dF_{i'j'} - N \sum_{l=1}^{I} \sum_{m=1}^{J} \int F_{lm} dF_{ij} \int F_{lm} dF_{i'j'} \\ &+ 2 \int F_{ij} dF_{ij} \int F_{ij} dF_{i'j'} + \int F_{i'j'} dF_{ij} \int F_{ij} dF_{i'j'} \\ &+ \int F_{ij} dF_{ij} \int F_{i'j'} dF_{i'j'} + 2 \int F_{i'j'} dF_{ij} \int F_{i'j'} dF_{i'j'} \end{aligned}$$

$$+[N\sum_{l=1}^{I}\sum_{m=1}^{J}\int F_{lm}dF_{i'j'} - \int F_{i'j'}dF_{i'j'} - \int F_{ij}dF_{i'j'}]$$
$$+[N\sum_{l=1}^{I}\sum_{m=1}^{J}(\int F_{lm}dF_{i'j'} - \int F_{lm}F_{ij}dF_{i'j'}) - (\int F_{i'j'}dF_{i'j'} - \int F_{i'j'}F_{ij}dF_{i'j'})$$
$$-(\int F_{ij}dF_{i'j'} - \int F_{ij}^2 dF_{i'j'})]$$
$$+[N\sum_{l=1}^{I}\sum_{m=1}^{J}(\int F_{lm}dF_{ij} - \int F_{lm}F_{i'j'}dF_{ij}) - (\int F_{ij}dF_{ij} - \int F_{ij}F_{i'j'}dF_{ij})$$
$$-(\int F_{i'j'}dF_{ij} - \int F_{i'j'}^2 dF_{ij})]$$
$$+[N\sum_{l=1}^{I}\sum_{m=1}^{J}\int F_{lm}dF_{ij} - \int F_{ij}dF_{ij} - \int F_{i'j'}dF_{ij}]$$
$$+[N\sum_{l=1}^{I}\sum_{m=1}^{J}\{1 - \int(F_{ij} + F_{i'j'})dF_{lm} + \int F_{ij}F_{i'j'}dF_{lm}\}$$
$$-\{1 - \int(F_{ij} + F_{i'j'})dF_{ij} + \int F_{ij}F_{i'j'}dF_{ij}\}$$
$$-\{1 - \int(F_{ij} + F_{i'j'})dF_{i'j'} + \int F_{ij}F_{i'j'}dF_{i'j'}\}]$$
$$+[1 + \int F_{ij}dF_{i'j'} + \int F_{i'j'}dF_{ij} + 0].$$

When simplifying by combining the same terms, we can obtain the result.

The result of Lemma 2 is used to show that the expected value of the denominator of F_N, converges in probability to a positive constant, $1/3 - (IJ)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J}(\int HdF_{ij})^2$, without the need for any assumptions.

Lemma 2. Assume the condition of Lemma 1 and let $V_{ij} = Var(\rho_{ijn})$ and $C_{ij} = Cov(\rho_{ijn}, \rho_{ijn'})$ for $n \neq n'$. Further denote V_N by $[(I-1)(J-1)/I^2J^2(N-1)]\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{n=1}^{N}(\rho_{ijn} - \bar{\rho}_{ij.})^2$ where $\bar{\rho}_{ij.} = \rho_{ij.}/N$. Then

(1) $$E(R_{ij1}^2) = I^2J^2N^2\int H^2 dF_{ij} - IJN\int HdF_{ij}^2 + 3IJN\int HdF_{ij} - N\sum_{l=1}^{I}\sum_{m=1}^{J}\int F_{lm}^2 dF_{ij} + \frac{1}{6}. \quad (3.4)$$

(2) $$\lim_{N\to\infty}\sum_{i=1}^{I}\sum_{j=1}^{J}(V_{ij} - C_{ij}) = \frac{1}{3}IJ - \sum_{i=1}^{I}\sum_{j=1}^{J}(\int HdF_{ij})^2. \quad (3.5)$$

(3) $$\lim_{N\to\infty} E(V_N) = \frac{(I-1)(J-1)}{IJ}\left[\frac{1}{3} - \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(\int HdF_{ij})^2\right]. \quad (3.6)$$

Proof. (1) When modifying (3.3) by simply changing the index, we can write

$$R_{ij1}^2 = \sum_{l=1}^{I}\sum_{m=1}^{J}\sum_{k=1}^{N}\sum_{l'=1}^{I}\sum_{m'=1}^{J}\sum_{k'=1}^{N} \mu(X_{ij1} - X_{lmk}) \cdot \mu(X_{ij1} - X_{l'm'k'}).$$

Note that we have

$$E[\mu(X_{ij1} - X_{lmk}) \cdot \mu(X_{ij1} - X_{l'm'k'})] = \int F_{lm} F_{l'm'} dF_{ij}$$

when $(l,\ m,\ k) \neq (i,\ j,\ 1)$ and $(l',\ m',\ k') \neq (i,\ j,\ 1)$ or $(l,\ m,\ k)$.

Therefore by using the same procedure as we have employed in proving (3.2) of Lemma 1

$$\begin{aligned} E(R_{ij1}^2) = {} & \\ & N^2 \sum_{l=1}^{I}\sum_{m=1}^{J}\sum_{l'=1}^{I}\sum_{m'=1}^{J} \int F_{lm} F_{l'm'} dF_{ij} - N \sum_{l=1}^{I}\sum_{m=1}^{J} \int F_{ij} F_{lm} dF_{ij} \\ & - N \sum_{l=1}^{I}\sum_{m=1}^{J} \int F_{lm} F_{ij} dF_{ij} - N \sum_{l=1}^{I}\sum_{m=1}^{J} \int F_{lm}^2 dF_{ij} + 2 \int F_{ij}^2 dF_{ij} \\ & + N \sum_{l=1}^{I}\sum_{m=1}^{J} \int F_{lm} dF_{ij} - \int F_{ij} dF_{ij} + N \sum_{l=1}^{I}\sum_{m=1}^{J} \int F_{lm} dF_{ij} - \int F_{ij} dF_{ij} \\ & + N \sum_{l=1}^{I}\sum_{m=1}^{J} \int F_{lm} dF_{ij} - \int F_{ij} dF_{ij} + 1. \end{aligned}$$

Now combining the same terms yields the desired result.

$$\begin{aligned} (2)\ V_{ij} - C_{ij} &= [E\rho_{ij1}^2 - (E\rho_{ij1})^2] - [E(\rho_{ij1} \cdot \rho_{ij2}) - E\rho_{ij1} \cdot E\rho_{ij2}] \\ &= E\rho_{ij1}^2 - E(\rho_{ij1} \cdot \rho_{ij2}) \end{aligned}$$

From (3.4) and Iman, Hora and Conover's (1984) Lemma 3

$$\begin{aligned} = \frac{I^2 J^2 N^2}{(IJN+1)^2} \Big[& \int H^2 dF_{ij} - \Big(\int H dF_{ij}\Big)^2 + \frac{1}{IJN} \int H dF_{ij}^2 \\ & + \frac{1}{I^2 J^2 N} \sum_{l=1}^{I}\sum_{m=1}^{J} \Big(\int F_{lm} dF_{ij}\Big)^2 - \frac{1}{IJN} \int H dF_{ij} - \frac{1}{I^2 J^2 N} \sum_{l=1}^{I}\sum_{m=1}^{J} \int F_{lm}^2 dF_{ij} \Big]. \end{aligned}$$

Thus taking the summations into the above result gives

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(V_{ij}-C_{ij}) = \frac{I^2J^2N^2}{(IJN+1)^2}\left[\frac{1}{3}IJ - \sum_{i=1}^{I}\sum_{j=1}^{J}(\int HdF_{ij})^2 + \frac{2}{IJN}\sum_{i=1}^{I}\sum_{j=1}^{J}\int HdF_{ij}^2 + \frac{1}{I^2J^2N}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{l=1}^{I}\sum_{m=1}^{J}(\int F_{lm}dF_{ij})^2 - \frac{3}{2N}\right].$$

The lemma follows immediately as $N \to \infty$.

$$\begin{aligned}(3)\ E[\rho_{ijn} - \bar{\rho}_{ij.}]^2 &= E[(\rho_{ijn} - E\rho_{ijn}) - (\bar{\rho}_{ij.} - E\bar{\rho}_{ij.}) + (E\rho_{ijn} - E\bar{\rho}_{ij.})]^2\\ &= E[(\rho_{ijn} - E\rho_{ijn}) - (\bar{\rho}_{ij.} - E\bar{\rho}_{ij.})]^2\\ &= Var(\rho_{ijn}) + Var(\bar{\rho}_{ij.}) - 2Cov(\rho_{ijn}, \bar{\rho}_{ij.})\\ &= V_{ij} + (1/N^2)[NV_{ij} + N(N-1)C_{ij}] - (2/N)[V_{ij} + (N-1)C_{ij}]\\ &= [(N-1)/N](V_{ij} - C_{ij}).\end{aligned}$$

Substituting the above result into $E(V_N)$, followed by combining (3.5), yields the result (3.6) easily.

The result of Lemma 3 is central to show that the mean of numerator of F_N, which may be cumbersome to get but can be simplified mainly by using integration by parts in several places in order to yield convenient forms, converges to a positive constant, $1/3 - (IJ)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J}(\int HdF_{ij})^2$, under the null hypothesis (2.1) and the assumption (2.2).

Lemma 3. Assume the condition of Lemma 1. Then

$$ER_{ij.} - E\frac{R_{i..}}{J} - E\frac{R_{.j.}}{I} + E\frac{R_{...}}{IJ} = IJN^2\int Hd(F_{ij} - \bar{F}_{i.} - \bar{F}_{.j} + \bar{F}_{..}). \quad (3.7)$$

Proof. From (3.1) note that

$$ER_{ij.} = IJN^2\textstyle\int HdF_{ij} + \frac{1}{2}N, \quad ER_{i..} = IJ^2N^2\int Hd\bar{F}_{i.} + \frac{1}{2}JN,$$
$$ER_{.j.} = I^2JN^2\textstyle\int Hd\bar{F}_{.j} + \frac{1}{2}IN, \quad ER_{...} = I^2J^2N^2\int Hd\bar{F}_{..} + \frac{1}{2}IJN.$$

Therefore when we combine the above results, the lemma follows easily.

Next Lemma 4 shows that the second assumption (2.3) is sufficient to have $N^{-1/2}(\rho_{ij.} - \rho_{i..}/J - \rho_{.j.}/I + \rho_{...}/IJ)$ asymptotically identically distributed for each block and treatment.

Lemma 4. Assume the condition of Lemma 1. Then under the null hypothesis (2.1) and the assumption (2.3),

$$\lim_{N\to\infty} Cov\left(N^{-1/2}(\rho_{ij.} - \frac{\rho_{i..}}{J} - \frac{\rho_{.j.}}{I} + \frac{\rho_{...}}{IJ}), N^{-1/2}(\rho_{i'j'.} - \frac{\rho_{i'..}}{J} - \frac{\rho_{.j'.}}{I} + \frac{\rho_{...}}{IJ})\right)$$
$$= \frac{1}{(I-1)(J-1)} \cdot \sigma^2 \quad \text{for } i \neq i' \text{ and } j \neq j'.$$

Proof. First of all by definition of covariance

$$Cov\left[N^{-1/2}(\rho_{ij.} - \frac{\rho_{i..}}{J} - \frac{\rho_{.j.}}{I} + \frac{\rho_{...}}{IJ}), N^{-1/2}(\rho_{i'j'.} - \frac{\rho_{i'..}}{J} - \frac{\rho_{.j'.}}{I} + \frac{\rho_{...}}{IJ})\right] \tag{3.8}$$
$$= \frac{1}{N(IJN+1)^2}\Big[E\{(R_{ij.} - \frac{R_{i..}}{J} - \frac{R_{.j.}}{I} + \frac{R_{...}}{IJ})(R_{i'j'.} - \frac{R_{i'..}}{J} - \frac{R_{.j'.}}{I} + \frac{R_{...}}{IJ})\}$$
$$-E(R_{ij.} - \frac{R_{i..}}{J} - \frac{R_{.j.}}{I} + \frac{R_{...}}{IJ}) \cdot E(R_{i'j'.} - \frac{R_{i'..}}{J} - \frac{R_{.j'.}}{I} + \frac{R_{...}}{IJ})\Big]$$

Substituting the result (3.7) and applying the null hypothesis (2.1) give

$$= \frac{1}{N(IJN+1)^2}\left[E\{(R_{ij.} - \frac{R_{i..}}{J} - \frac{R_{.j.}}{I} + \frac{R_{...}}{IJ})(R_{i'j'.} - \frac{R_{i'..}}{J} - \frac{R_{.j'.}}{I} + \frac{R_{...}}{IJ})\}\right].$$

Evaluating the above equation is tedious but straightforward. Namely when we combine the same terms after expanding and substituting the result (3.2) of Lemma 1, we can find that under the null hypothesis (2.1) and the assumption (2.3) and by the expression of σ^2 which will be clearly provided in Lemma 5,

$$\lim_{N\to\infty}(3.8) = \frac{1}{IJ}\left[\frac{1}{3} - \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(\int H dF_{ij})^2\right] = \frac{1}{(I-1)(J-1)} \cdot \sigma^2.$$

Finally Lemma 5 states that the hypothesis (2.1) and assumptions (2.2) and (2.3) are sufficient to have $N^{-1/2}(\rho_{ij.} - \rho_{i..}/J - \rho_{.j.}/I + \rho_{...}/IJ)$ converged to normality.

Lemma 5. Assume the condition of Lemma 1. Then under the null hypothesis (2.1) and the assumptions (2.2) and (2.3),

$$\sigma^2 = \lim_{N\to\infty}\sum_{i=1}^{I}\sum_{j=1}^{J} Var(\rho_{ij.} - \frac{\rho_{i..}}{J} - \frac{\rho_{.j.}}{I})/IJN = \lim_{N\to\infty} Var(\rho_{ij.} - \frac{\rho_{i..}}{J} - \frac{\rho_{.j.}}{I})/N$$

$$= \frac{(I-1)(J-1)}{IJ}\left[\frac{1}{3} - \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(\int H dF_{ij})^2\right].$$

Proof. When using the results (3.2), (3.7) and integration by parts in several places in order to yield short forms, we can establish that

$$\begin{aligned}\sigma^2 &= \frac{1}{IJN}\lim_{N\to\infty}\sum_{i=1}^{I}\sum_{j=1}^{J} Var(\rho_{ij.} - \frac{\rho_{i..}}{J} - \frac{\rho_{.j.}}{I}) \\ &= \frac{(I-1)(J-1)}{I^2J^2}\lim_{N\to\infty}\sum_{i=1}^{I}\sum_{j=1}^{J}(V_{ij} - C_{ij})\end{aligned}$$

under the hypothesis (2.1) and the assumption (2.2). Thus substituting (3.5) gives $\sigma^2 = [(I-1)(J-1)/IJ]\cdot[1/3-(1/IJ)\sum_{i=1}^{I}\sum_{j=1}^{J}(\int HdF_{ij})^2]$. On the other hand, $\lim_{N\to\infty} N^{-1}Var(\rho_{ij.}-\rho_{i..}/J-\rho_{.j.}/I)$ is equivalent to $\lim_{N\to\infty} N^{-1}E(\rho_{ij.}-\rho_{i..}/J-\rho_{.j.}/I+\rho_{...}/IJ)^2$ under the null hypothesis (2.1), which is in turn equivalent to $[(I-1)(J-1)/IJ]\cdot[1/3-(1/IJ)\sum_{i=1}^{I}\sum_{j=1}^{J}(\int HdF_{ij})^2]$ under the second assumption (2.3) by using the same method as we established in proving Lemma 4. We then obtain the desired result.

4. CONCLUSIONS

One particular interest of this paper is to study the asymptotic theory of the rank transformed statistic, computed on ranks, for testing for interaction in a 2×2 factorial design without the restriction of an additive linear model or without regard to error terms having homoscedasticity. Then the results can be summarized as follows. If and only if there is only one main effect ($F_{ij} = \bar{F}_{i.}$ or $F_{ij} = \bar{F}_{.j}$ for all i and j), under the null hypothesis of no interaction, provided as equation (2.1) in this paper, we can have the rank transformed F statistic for interaction converge in distribution to a chi-squared random variable with one degree of freedom.

REFERENCES

(1) Blair, R. C., Sawilowsky, S.S. and Higgins, J.J.(1987). Limitations of the Rank Transform Statistic in Tests for Interactions. *Communications in Statistics, Part B - Simulation and Computation,* 16, 1133-1145.

(2) Conover, W. J. and Iman, R. L. (1976). On Some Alternative Procedure Using Ranks for the Analysis of Experimental Designs. *Communications in Statistics*, A5, 1349-1368.

(3) Graybill, F. A.(1976). *Theory and Application of the Linear Model*, North Scituate, Mass., Duxbury Press.

(4) Hájek, J.(1968). Asymptotic Normality of Simple Linear Rank Statistics under Alternatives. *The Annals of Mathematical Statistics,* 39, 325-346.

(5) Hájek, J. and Sidák, Z. (1967). *Theory of Rank Tests*, New York, Academic Press.

(6) Hora, S. C. and Conover, W.J. (1984). The F Statistic in the Two-Way Layout with Rank-Score Transformed Data. *Journal of the American Statistical Association*, 79, 668-673.

(7) Hora, S. C. and Iman, R. L. (1988). Asymptotic Relative Efficiencies of the Rank Transformation Procedure in Randomized Complete Block Design. *Journal of the American Statistical Association*, 83, 462-470.

(8) Iman, R. L., Hora, S. C. and Conover, W. J. (1984). Comparison of Asymptotically Distribution Free Procedures for the Analysis of Complete Blocks. *Journal of the American Statistical Association*, 79, 674-685.

(9) Thompson, G. L. (1991). A Note on the Rank Transform for Interactions. *Biometrika*, 78, 3, 697-701.