Journal of the Korean
Statistical Society
Vol. 24, No. 2, 1995

Asymptotically Distribution-Free Procedure in a Two-Way Layout

Young Hun Choi[1]

ABSTRACT

Main purpose of this article is to consider the asymptotic distribution of the rank transformed F statistic for interaction in a two-way layout. Some theorems and sufficient conditions are derived to have the rank transformed F statistic converged in distribution to a chi-squared random variable with $(I-1)(J-1)$ degrees of freedom divided by $(I-1)(J-1)$. These results will be useful for the other theoretical studies of the rank transform procedure in experimental designs.

KEYWORDS: Two-way layout, F statistic, Limiting distribution, Heteroscedasticity.

1. INTRODUCTION

Conover and Iman(1981) defined the rank transformation(RT) as a procedure in which we simply replace the data with their ranks. The entire set of observations is ranked from smallest to largest, with the smallest observation having rank 1, the second smallest rank 2, and so on. We then apply the usual

[1]Department of Statistics, Hanshin University, Osan, 447–791, Korea.

parametric tests to the ranks. This rank transform approach provides useful methods in analysis of experimental designs, multiple regression, discriminant analysis, cluster analysis, multiple comparisons and so forth, as well as results in a class of nonparametric methods.

Recently many papers have discussed and extended the applicability of the RT's appropriateness in various analysis of variance, basically testing for main effects. In other words, so far the rigorous treatment of the asymptotic null distribution of the rank test statistic for interaction was not performed well even though Thompson(1991) and Akritas(1994) referred to the method detecting interaction in limited situations briefly. Further Fabian(1991) discusses the controversial and sometimes conflicting concepts of interaction, which shows that the concept of interaction is indeed complex, and still open to various interpretations. Thus one particular goal of this paper is thoroughly to introduce the asymptotic theory for the F statistic for testing an interaction effect in a two-way layout without regard to error terms having homoscedasticity or heteroscedasticity for I blocks, J treatments, and N replications, based on the ranks of data, in which nonparametric tests have not generally met with much success.

To begin with, let X_{ijn}, $i = 1,\ldots,I$, $j = 1,\ldots,J$ and $n = 1,\ldots,N$, be independent random variables with the cumulative distribution function F_{ij}. And define $\bar{F}_{i\cdot} = (\sum_{j=1}^{J} F_{ij})/J$, $\bar{F}_{\cdot j} = (\sum_{i=1}^{I} F_{ij})/I$ and $\bar{F}_{\cdot\cdot}$ or $H = (\sum_{i=1}^{I}\sum_{j=1}^{J} F_{ij})/IJ$. Next to define the proposed test statistics, denote $R_{ijn} = R(X_{ijn}) = \sum_{l=1}^{I}\sum_{m=1}^{J}\sum_{k=1}^{N} \mu(X_{ijn} - X_{lmk})$, where $\mu(t) = 1$ for $t \geq 0$ and $\mu(t) = 0$ for $t < 0$. Accordingly denote $R_{ij\cdot} = \sum_{n=1}^{N} R_{ijn}$, $R_{i\cdot\cdot} = \sum_{j=1}^{J}\sum_{n=1}^{N} R_{ijn}$, $R_{\cdot j\cdot} = \sum_{i=1}^{I}\sum_{n=1}^{N} R_{ijn}$ and $R_{\cdots} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{n=1}^{N} R_{ijn}$. Finally denote $\rho_{ijn} = R_{ijn}/(IJN+1)$ and $\rho_{ij\cdot}$, $\rho_{i\cdot\cdot}$, $\rho_{\cdot j\cdot}$ and $\rho_{\cdots}$ will denote $R_{ij\cdot}/(IJN+1)$, $R_{i\cdot\cdot}/(IJN+1)$, $R_{\cdot j\cdot}/(IJN+1)$ and $R_{\cdots}/(IJN+1)$ respectively.

2. LIMITING DISTRIBUTION OF F_N

The rank-transform F test is based on the statistic

$$F_N = \frac{I_N}{W_N},$$

where $I_N = [\sum_{i=1}^{I}\sum_{j=1}^{J}(\rho_{ij\cdot} - \rho_{i\cdot\cdot}/J - \rho_{\cdot j\cdot}/I + \rho_{\cdot\cdot\cdot}/IJ)^2]/[(I-1)(J-1)N]$ and $W_N = [\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{n=1}^{N}(\rho_{ijn} - \rho_{ij\cdot}/N)^2]/[IJ(N-1)]$.

First of all to obtain the limiting distribution, $E(W_N)$ followed by $E(I_N)$ of the statistic F_N will be provided. Namely when we expand the square terms of W_N after subtracting and adding necessary terms, it will be shown that from Choi Y.H.'s(1994) Lemma 2

$$E(W_N) = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(V_{ij} - C_{ij}), \tag{2.1}$$

$$\text{and } \lim_{N\to\infty} E(W_N) = \frac{1}{3} - \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\Big(\int H dF_{ij}\Big)^2,$$

where $V_{ij} = \mathrm{Var}(\rho_{ijn})$ and $C_{ij} = \mathrm{Cov}(\rho_{ijn}, \rho_{ijn'})$ for $n \neq n'$. Note that the expected value of the denominator of F_N converges to a constant as $N \to \infty$.

Next we'll find the expected value of the numerator of I_N, yielding $E(I_N)$ in turn.

$$E\sum_{i=1}^{I}\sum_{j=1}^{J}\Big(\rho_{ij\cdot} - \frac{\rho_{i\cdot\cdot}}{J} - \frac{\rho_{\cdot j\cdot}}{I} + \frac{\rho_{\cdot\cdot\cdot}}{IJ}\Big)^2$$

$$= E\Big(\sum_{i=1}^{I}\sum_{j=1}^{J}\rho_{ij\cdot}^2 - \frac{1}{J}\sum_{i=1}^{I}\rho_{i\cdot\cdot}^2 - \frac{1}{I}\sum_{j=1}^{J}\rho_{\cdot j\cdot}^2 + \frac{1}{IJ}\rho_{\cdot\cdot\cdot}^2\Big)$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\mathrm{Var}(\rho_{ij\cdot}) - \frac{1}{J}\sum_{i=1}^{I}\mathrm{Var}(\rho_{i\cdot\cdot}) - \frac{1}{I}\sum_{j=1}^{J}\mathrm{Var}(\rho_{\cdot j\cdot})$$

$$+ \Big[\sum_{i=1}^{I}\sum_{j=1}^{J}(E\rho_{ij\cdot})^2 - \frac{1}{J}\sum_{i=1}^{I}(E\rho_{i\cdot\cdot})^2 - \frac{1}{I}\sum_{j=1}^{J}(E\rho_{\cdot j\cdot})^2 + \frac{1}{IJ}(E\rho_{\cdot\cdot\cdot})^2\Big] \tag{2.2}$$

Expanding the variances of the first part of (2.2) and adding $[I^2J^2N^4/(IJN+1)^2]\cdot\sum_{i=1}^{I}\sum_{j=1}^{J}[\int Hd(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})]^2$, which is equivalent to the second part of (2.2), give

$$
\begin{aligned}
&= N\frac{(I-1)(J-1)}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(V_{ij}-C_{ij}) \\
&+ N^2\sum_{i=1}^{I}\sum_{j=1}^{J}\Big[C_{ij}-\frac{\sum_{j'=1}^{J}C_{ij,ij'}}{J}-\frac{\sum_{i'=1}^{I}C_{ij,i'j}}{I}+\frac{\sum_{i'=1}^{I}\sum_{j'=1}^{J}C_{ij,i'j'}}{IJ}\Big] \\
&+ \frac{I^2J^2N^4}{(IJN+1)^2}\sum_{i=1}^{I}\sum_{j=1}^{J}\Big[\int Hd(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})\Big]^2, \qquad (2.3)
\end{aligned}
$$

where $C_{ij,ij'}=\mathrm{Cov}(\rho_{ijn},\rho_{ij'n'})$, $C_{ij,i'j}=\mathrm{Cov}(\rho_{ijn},\rho_{i'jn'})$, $C_{ij,i'j'}=\mathrm{Cov}(\rho_{ijn},\rho_{i'j'n'})$, for $i\neq i'$ and $j\neq j'$.

Now after expressing each covariances of the second part of (2.3) as integral types in terms of distribution function, we can simplify the results by combining the same terms in order to generate the following desirable equation mainly by using integration by parts in several places. Then we can get

$$
\begin{aligned}
E(I_N) &= \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(V_{ij}-C_{ij})+\frac{N}{(I-1)(J-1)}\cdot\frac{-N}{(IJN+1)^2}\cdot\Big[\\
&\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{l=1}^{I}\sum_{m=1}^{J}\int F_{lm}dF_{ij}\int F_{lm}d(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot}) \\
&+4IJ\sum_{i=1}^{I}\sum_{j=1}^{J}\int HF_{ij}d(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot}) \\
&+2IJ\sum_{i=1}^{I}\sum_{j=1}^{J}\int HdF_{ij}\int F_{ij}d(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot}) \\
&-\frac{2}{N}\sum_{i=1}^{I}\sum_{j=1}^{J}\int F_{ij}^2d(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})
\end{aligned}
$$

$$+\frac{1}{IJN}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{i'=1}^{I}\sum_{j'=1}^{J}\Big[\Big(\int F_{ij}dF_{ij}\Big)^2-\Big(\int F_{ij'}dF_{ij}\Big)^2$$

$$-\Big(\int F_{i'j}dF_{ij}\Big)^2-\Big(\int F_{i'j'}dF_{ij}\Big)^2\Big]\Big]$$

$$+\frac{I^2J^2N^3}{(I-1)(J-1)(IJN+1)^2}\sum_{i=1}^{I}\sum_{j=1}^{J}\Big[\int Hd(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})\Big]^2. \quad (2.4)$$

Thus, substituting (2.1) into the first part of (2.4) gives

$$\lim_{N\to\infty}E(I_N)=\Big[\frac{1}{3}-\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\Big(\int HdF_{ij}\Big)^2\Big]-\frac{1}{(I-1)(J-1)I^2J^2}\cdot\Big[$$

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{l=1}^{I}\sum_{m=1}^{J}\int F_{lm}dF_{ij}\int F_{lm}d(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})$$

$$+4IJ\sum_{i=1}^{I}\sum_{j=1}^{J}\int HF_{ij}d(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})$$

$$+2IJ\sum_{i=1}^{I}\sum_{j=1}^{J}\int HdF_{ij}\int F_{ij}d(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})\Big]$$

$$+\lim_{N\to\infty}\frac{N}{(I-1)(J-1)}\sum_{i=1}^{I}\sum_{j=1}^{J}\Big[\int Hd(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})\Big]^2. \quad (2.5)$$

Note that the first part of (2.5) is the limit of the mean of the denominator of F_N, and the additional terms represent a noncentrality factor. The last term in (2.5) will increase without bound as $N\to\infty$ unless $\int Hd(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})=0$ for all i and j is true. Further the remaining terms in (2.5) must necessarily equal zero under the null hypothesis if the limiting distribution of F_N is to be the usual F distribution. Hence first of all from (2.5) we can derive the null hypothesis $\int F_{lm}d(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})=0$ for all i, j, l and m which implies $\int Hd(F_{ij}-\bar{F}_{i\cdot}-\bar{F}_{\cdot j}+\bar{F}_{\cdot\cdot})=0$ for all i and j.

Lemma 1. $\int F_{lm} d(F_{ij} - \bar{F}_{i\cdot} - \bar{F}_{\cdot j} + \bar{F}_{\cdot\cdot}) = 0$ for all i, j, l and m is equivalent to $\int F_{lm} d(F_{ij} - F_{ij'} - F_{i'j} + F_{i'j'}) = 0$ for all i, j, i', j', l and m, where i', $l = 1, \ldots, I$ and j', $m = 1, \ldots, J$.

Proof. Since $\int F_{lm} d(F_{ij} - \bar{F}_{i\cdot} - \bar{F}_{\cdot j} + \bar{F}_{\cdot\cdot}) = 0$ implies $\int F_{lm} dF_{ij} = \int F_{lm} d(\bar{F}_{i\cdot} + \bar{F}_{\cdot j} - \bar{F}_{\cdot\cdot})$, it follows that

$$\begin{aligned}
&\int F_{lm} d(F_{ij} - F_{ij'} - F_{i'j} + F_{i'j'}) \\
&= \int F_{lm} d(\bar{F}_{i\cdot} + \bar{F}_{\cdot j} - \bar{F}_{\cdot\cdot}) - \int F_{lm} d(\bar{F}_{i\cdot} + \bar{F}_{\cdot j'} - \bar{F}_{\cdot\cdot}) \\
&\quad - \int F_{lm} d(\bar{F}_{i'\cdot} + \bar{F}_{\cdot j} - \bar{F}_{\cdot\cdot}) + \int F_{lm} d(\bar{F}_{i'\cdot} + \bar{F}_{\cdot j'} - \bar{F}_{\cdot\cdot}) = 0.
\end{aligned}$$

Conversely, since $\int F_{lm} d(F_{ij} - F_{ij'} - F_{i'j} + F_{i'j'}) = 0$ implies $\int F_{lm} dF_{ij} = \int F_{lm} d(F_{ij'} + F_{i'j} - F_{i'j'})$, it follows that

$$\begin{aligned}
&\int F_{lm} d(F_{ij} - \bar{F}_{i\cdot} - \bar{F}_{\cdot j} + \bar{F}_{\cdot\cdot}) \\
&= \int F_{lm} d(F_{ij'} + F_{i'j} - F_{i'j'}) - \int F_{lm} d(F_{ij'} + \bar{F}_{i'\cdot} - F_{i'j'}) \\
&\quad - \int F_{lm} d(\bar{F}_{\cdot j'} + F_{i'j} - F_{i'j'}) + \int F_{lm} d(\bar{F}_{\cdot j'} + \bar{F}_{i'\cdot} - F_{i'j'}) = 0.
\end{aligned}$$

This completes the proof.

Traditionally two factors are said to be interact if the difference in mean responses for two levels of one factor is not constant across levels of the second factor. Thus $\int F_{lm} d(F_{ij} - F_{ij'} - F_{i'j} + F_{i'j'}) = 0$ for all i, j, i', j', l and m satisfies the rank analogue of the definition of no interaction since $\int F_{lm} d(F_{ij} - F_{ij'} - F_{i'j} + F_{i'j'}) = EF_{lm}(X_{ij}) - EF_{lm}(X_{ij'}) - EF_{lm}(X_{i'j}) + EF_{lm}(X_{i'j'}) = 0$. In other words, $P(X_{lm} \leq X_{ij}) - P(X_{lm} \leq X_{ij'}) - P(X_{lm} \leq X_{i'j}) + P(X_{lm} \leq X_{i'j'}) = 0$ for all i, j, i', j', l and m.

In essence, here note that from Lemma 1, the hypotheses which are more interpretable can be derived as follows.

$$H_0 : \int F_{lm} d(F_{ij} - F_{ij'} - F_{i'j} + F_{i'j'}) = 0 \text{ for all } i, j, i', j', l \text{ and } m. \tag{2.6}$$

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

Further from (2.5) we can establish the first necessary assumption which is used in the derivation of the asymptotic distribution of the test statistic. Namely the first assumption to be used is

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \int H F_{ij} d(F_{ij} - \bar{F}_{i\cdot} - \bar{F}_{\cdot j} + \bar{F}_{\cdot\cdot}) = 0. \tag{2.7}$$

Now under the null hypothesis (2.6) and the condition (2.7), we can say that $E(I_N)$ converges to $1/3 - (1/IJ) \sum_{i=1}^{I} \sum_{j=1}^{J} (\int H dF_{ij})^2$ from (2.5). When summarized, $E(I_N) - E(W_N) = O(1/N)$ under the null hypothesis (2.6) and the assumption (2.7). By combining (2.1) and (2.4) and since W_N converges in probability to a positive constant which will be shown in Theorem 2, we can say that $E(F_N) = 1 + O(1/N)$.

Meanwhile, the second assumption to be used to have $\sigma^2 = \lim_{N \to \infty} [\sum_{i=1}^{I} \sum_{j=1}^{J} \mathrm{Var}(\rho_{ij\cdot} - \rho_{i\cdot\cdot}/J - \rho_{\cdot j\cdot}/I)]/IJN = \lim_{N \to \infty} [\mathrm{Var}(\rho_{ij\cdot} - \rho_{i\cdot\cdot}/J - \rho_{\cdot j\cdot}/I)]/N = [(I-1)(J-1)/IJ] \cdot [1/3 - (1/IJ) \sum_{i=1}^{I} \sum_{j=1}^{J} (\int H dF_{ij})^2]$ is

$$\begin{aligned}
2 &\int (F_{ij} - \bar{F}_{i\cdot} - \bar{F}_{\cdot j} + \bar{F}_{\cdot\cdot})(F_{i'j'} - \bar{F}_{i'\cdot} - \bar{F}_{\cdot j'} + \bar{F}_{\cdot\cdot}) dH \\
&- \Big[\mathrm{Var}\big(H(X_{ij})\big) - \frac{1}{J} \sum_{j=1}^{J} \mathrm{Var}\big(H(X_{ij})\big) \\
&\quad - \frac{1}{I} \sum_{i=1}^{I} \mathrm{Var}\big(H(X_{ij})\big) + \mathrm{Var}\big(H(X_{i'j})\big) \\
&\quad - \frac{1}{J} \sum_{j=1}^{J} \mathrm{Var}\big(H(X_{i'j})\big) + \mathrm{Var}\big(H(X_{ij'})\big) \\
&\quad - \frac{1}{I} \sum_{i=1}^{I} \mathrm{Var}\big(H(X_{ij'})\big) + \mathrm{Var}\big(H(X_{i'j'})\big)\Big] = 0,
\end{aligned}$$

for all i, j, i' and j'. (2.8)

Thus $0 < \sigma^2 < (I-1)(J-1)/3IJ$ exists, and we can show that a condition of Hájek's(1968) Theorem 2.1 is satisfied. Namely with the assumption (2.8), $\mathrm{Var}(\rho_{ij\cdot} - \rho_{i\cdot\cdot}/J - \rho_{\cdot j\cdot}/I + \rho_{\cdots}/IJ) = N\sigma^2 \to \infty$ as $N \to \infty$, and $\max(c_{ijn} - \bar{c})^2$ exists because c_{ijn} does not depend on N or IJN; that is $c_{ijn} = c_{ij}$ for all i and j. Therefore $\mathrm{Var}(\rho_{ij\cdot} - \rho_{i\cdot\cdot}/J - \rho_{\cdot j\cdot}/I + \rho_{\cdots}/IJ) = N\sigma^2 > K_\varepsilon \cdot \max(c_{ijn} - \bar{c})^2$, which entails $\max_x |P(R - ER < x\sqrt{\mathrm{Var}R}) - \Phi(x)| < \varepsilon$, where $R = \rho_{ij\cdot} - \rho_{i\cdot\cdot}/J - \rho_{\cdot j\cdot}/I + \rho_{\cdots}/IJ$ and Φ is the c.d.f. of the standard normal distribution. Hence it follows that each $N^{-1/2}(\rho_{ij\cdot} - \rho_{i\cdot\cdot}/J - \rho_{\cdot j\cdot}/I + \rho_{\cdots}/IJ)$, for $i = 1, \ldots, I$ and $j = 1, \ldots, J$, converges in distribution to a normal random variable.

Comment 1. The assumption (2.8) which is applicable to a two-way layout is simplified to $\int(F_{ij} - \bar{F}_{i\cdot} - \bar{F}_{\cdot j} + \bar{F}_{\cdot\cdot})^2 dH = 0$ for all i and j, the assumption (2.3) of Choi Y.H.(1994), in case of a 2×2 factorial design which is just a special case of a general two-way layout from Lemma 2. Further note that when there are no main effects($F_{ij} = F_{i'j'}$ for all i, j, i' and j') or if and only if there is one main effect in a two-way layout in which sampling is from populations with equal variances and having shift of means($X_{ijn} \sim (\mu_{ij}, \sigma^2)$ and $F_{ij} = \bar{F}_{i\cdot}$ or $F_{ij} = \bar{F}_{\cdot j}$ for all i and j), the null hypotheses (2.6) and two assumptions (2.7) and (2.8) are all satisfied. In addition, when both main effects are present($F_{ij} \neq \bar{F}_{i\cdot}$ and $F_{ij} \neq \bar{F}_{\cdot j}$ for at least one i and j) in a general two-way layout, the assumption (2.8) is necessary in the derivation of the asymptotic chi-squared distribution of the test statistic.

Lemma 2. If we have a linear model in which sampling is from populations with having shift of means and equal variances, $X_{ijn} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijn}$, then the assumption (2.8) will be satisfied if and only if there is one main effet such as treatment or block effect.

Proof. If variances are identical for all i and j, the assumption (2.8) is simplified to $\int(F_{ij} - \bar{F}_{i\cdot} - \bar{F}_{\cdot j} + \bar{F}_{\cdot\cdot})(F_{i'j'} - \bar{F}_{i'\cdot} - \bar{F}_{\cdot j'} + \bar{F}_{\cdot\cdot})dH = 0$ for all i, j, i' and j' which implies $F_{ij} - \bar{F}_{i\cdot} - \bar{F}_{\cdot j} + \bar{F}_{\cdot\cdot} = 0$ or $F_{i'j'} - \bar{F}_{i'\cdot} - \bar{F}_{\cdot j'} + \bar{F}_{\cdot\cdot} = 0$ for all i, j, i' and j' which in turn implies that $F_{ij} = \bar{F}_{i\cdot}$ or $F_{ij} = \bar{F}_{\cdot j}$ for all i and

j.

Therefore we can conclude that the assumption (2.8) is true if and only if $F_{ij} = \bar{F}_{i\cdot}$ or $F_{ij} = \bar{F}_{\cdot j}$ for all i and j. In other words, (2.8) can be satisfied if and only if there is one main effet such as treatment or block effect. This result agrees with that reported by Thompson(1991).

In conclusion, let $F_N = \sigma^2 \cdot Q_N/[V_N \cdot (I-1)(J-1)]$. Then the main result of the first theorem reveals that under the null hypothesis the statistic Q_N has a limiting chi-squared distribution under the additional assumptions (2.7) and (2.8). The second theorem shows that under the null hypothesis and the condition (2.7), V_N converges in probability to a positive constant. The third theorem shows that the statistic F_N converges in distribution to an $(I-1)(J-1)$ degree of freedom chi-squared random variable divided by $(I-1)(J-1)$.

2.1 Limit Theorem 1 for Q_N

Theorem 1. Let X_{ijn}, $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $n = 1, \ldots, N$ be independent random variables such that R_{ijn} represent the corresponding ranks. Further let $\rho_{ijn} = R_{ijn}/(IJN+1)$. Then under the null hypothesis (2.6) and the assumptions (2.7) and (2.8), the statistic

$$Q_N = \frac{(I-1)(J-1)\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\rho_{ij\cdot} - \frac{\rho_{i\cdot\cdot}}{J} - \frac{\rho_{\cdot j\cdot}}{I} + \frac{\rho_{\cdot\cdot\cdot}}{IJ}\right)^2}{IJN\sigma^2}$$

has a limiting distribution as $N \to \infty$ that is chi-squared with $(I-1)(J-1)$ degrees of freedom, where

$$\sigma^2 = \lim_{N\to\infty} \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} \mathrm{Var}\left(\rho_{ij\cdot} - \frac{\rho_{i\cdot\cdot}}{J} - \frac{\rho_{\cdot j\cdot}}{I}\right)}{IJN} \text{ exists.}$$

Proof. First note that σ^2 can be expressed as $\sigma^2 = [(I-1)(J-1)/IJ] \cdot [1/3 - (1/IJ)\sum_{i=1}^{I}\sum_{j=1}^{J}(\int H dF_{ij})^2]$. Now suppose the following vector

$$\boldsymbol{S}^{*'} = \left[\rho_{11\cdot} - \frac{\rho_{1\cdot\cdot}}{J} - \frac{\rho_{\cdot 1\cdot}}{I} + \frac{\rho_{\cdot\cdot\cdot}}{IJ}, \ldots, \rho_{IJ\cdot} - \frac{\rho_{I\cdot\cdot}}{J} - \frac{\rho_{\cdot J\cdot}}{I} + \frac{\rho_{\cdot\cdot\cdot}}{IJ}\right].$$

Under the null hypothesis (2.6) and the assumption (2.8), we can say that the covariance matrix of $\boldsymbol{S}^*$ is $N\sigma^2\boldsymbol{D}$, where $\boldsymbol{D} = \|d_{ij,i'j'}\|$, $d_{ij,i'j'} = 1$ if $i = i'$ and $j = j'$, $d_{ij,i'j'} = -1/(J-1)$ if $i = i'$ and $j \neq j'$, $d_{ij,i'j'} = -1/(I-1)$ if $i \neq i'$ and $j = j'$, $d_{ij,i'j'} = 1/(I-1)(J-1)$ otherwise. Then we can establish that $\boldsymbol{S} = \boldsymbol{S}^*/\sqrt{N}$ has a IJ-variate normal limiting distribution.

Next let $\boldsymbol{A}^*$, $\boldsymbol{\Sigma}$ and $\boldsymbol{A}$ be $IJ \times IJ$ matrix whose rows and columns are indexed by the ordered pairs (i,j) and (l,m), where the second index, $1 \leq j,m \leq J$, runs faster than the first index, $1 \leq i,l \leq I$. In addition, let $\theta(i,l) = 1$ or 0 as to whether $i = l$ or $i \neq l$ and define $\boldsymbol{A}^*$ with elements $\theta(i,l)\theta(j,m) - [1/I]\theta(j,m) - [1/J]\theta(i,l) + 1/IJ$. Further define $\boldsymbol{\Sigma}$ and $\boldsymbol{A}$ as $[IJ\sigma^2/(I-1)(J-1)] \cdot \boldsymbol{A}^*$ and $[(I-1)(J-1)/IJ\sigma^2] \cdot \boldsymbol{A}^*$ respectively.

Note here from the fact that $\boldsymbol{A}^*$ is idempotent and symmetric, $\boldsymbol{A\Sigma} = \boldsymbol{A}^* \cdot \boldsymbol{A}^* = \boldsymbol{A}^*$. Thus $(\boldsymbol{A\Sigma})(\boldsymbol{A\Sigma}) = \boldsymbol{A}^* \cdot \boldsymbol{A}^* = \boldsymbol{A}^* = \boldsymbol{A\Sigma}$, which satisfies the condition($\boldsymbol{A\Sigma}$ is an idempotent matrix) of Graybill's(1976) Theorem 4.4.3. Further note that $p = \text{Rank}[\boldsymbol{A\Sigma}] = \text{tr}[\boldsymbol{A\Sigma}] = \text{tr}[\boldsymbol{A}^*] = (I-1)(J-1)$. Finally $\boldsymbol{S}'\boldsymbol{A}\boldsymbol{S}$ is identical to Q_N. Therefore Q_N will converge in distribution to a chi-squared random variable with $(I-1)(J-1)$ degrees of freedom.

2.2 Limit Theorem 2 for V_N

Theorem 2. If H_0 in (2.6) is true and if the condition (2.7) and $(1/M)\sum_{i=1}^{M} \rho_i^2 < K_0$, $M = IJN$, hold, the estimator

$$V_N = \frac{(I-1)(J-1)\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{n=1}^{N}\left(\rho_{ijn} - \frac{\rho_{ij\cdot}}{N}\right)^2}{I^2J^2(N-1)}$$

is unbiased for

$$\sigma_N^2 = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\text{Var}\left(\frac{\rho_{ij\cdot} - \frac{\rho_{i\cdot\cdot}}{J} - \frac{\rho_{\cdot j\cdot}}{I}}{\sqrt{N}}\right)$$

and V_N converges in probability to σ^2 as $N \to \infty$.

Proof. i) From (2.4), we note that under the null hypothesis (2.6) and the condition (2.7),

$$\sigma_N^2 = \frac{(I-1)(J-1)}{I^2J^2}\sum_{i=1}^{I}\sum_{j=1}^{J}(V_{ij} - C_{ij}) + O\Big(\frac{1}{N}\Big).$$

And we obtain

$$E(V_N) = \frac{(I-1)(J-1)}{I^2J^2}\sum_{i=1}^{I}\sum_{j=1}^{J}(V_{ij} - C_{ij}).$$

Thus $E(V_N) - \sigma_N^2 \to 0$ as $N \to \infty$, which establishes asymptotic unbiasedness.

ii) Now to establish weak consistency, note that

$$V_N = \frac{(I-1)(J-1)}{I^2J^2}\Big[\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{n=1}^{N}\rho_{ijn}^2/(N-1) - \sum_{i=1}^{I}\sum_{j=1}^{J}\Big(\frac{\rho_{ij\cdot}}{\sqrt{N(N-1)}}\Big)^2\Big].$$

Furthermore from Hájek's(1968) Equation 2.22,

$$\lim_{N\to\infty} V_N = \frac{(I-1)(J-1)}{I^2J^2}\Big[IJ\int_0^1 t^2dt - \sum_{i=1}^{I}\sum_{j=1}^{J}\Big(\int HdF_{ij}\Big)^2\Big].$$

Therefore V_N converges in probability to a constant σ^2 provided in Theorem 1.

iii) However from the condition that for every M_0 there exists K_0 such that $(1/M)\sum_{i=1}^{M}\rho_i^2 < K_0$ for all $M > M_0$,

$$\frac{(I-1)(J-1)}{I^2J^2(N-1)}\cdot\sum_{i=1}^{M}\rho_i^2 < \frac{(I-1)(J-1)N}{IJ(N-1)}\cdot K_0.$$

Thus V_N is bounded from below by zero and from above by $[(I-1)(J-1)N/IJ(N-1)]\cdot K_0$.

iv) Overall since $E(V_N)-\sigma_N^2 \to 0$ as $N \to \infty$, V_N converges in probability to a constant and is bounded from below and above, it follows that V_N converges in probability to σ^2.

2.3 Limit Theorem 3 for Statistic F_N

Theorem 3. Under the conditions of Theorem 1, the statistic F_N converges in distribution to an $(I-1)(J-1)$ degree of freedom chi-squared random variable divided by $(I-1)(J-1)$.

Proof. Using Q_N and V_N defined in Theorem 1 and Theorem 2 respectively, we have $F_N = \sigma^2 \cdot Q_N/[V_N \cdot (I-1)(J-1)]$. Now Theorem 1 provides that Q_N converges in probability to $\chi^2_{(I-1)(J-1)}$. Theorem 2 says that V_N converges in probability to the σ^2 defined in Theorem 1. Since the numerator converges in distribution, the ratio converges in distribution to $\chi^2_{(I-1)(J-1)}$ by Slutsky's Theorem.

3. CONCLUSIONS

First of all when there are no main effects without any limitations or if and only if there is one main effect in a two-way layout in which sampling is from populations with equal variances and having shift of means, then under the null hypothesis of no interaction the rank transformed F statistic converges in distribution to a chi-squared random variable with $(I-1)(J-1)$ degrees of freedom divided by $(I-1)(J-1)$.

Secondly when both main effects are present($F_{ij} \neq \bar{F}_{i\cdot}$ and $F_{ij} \neq \bar{F}_{\cdot j}$ for at least one i and j) in a general two-way layout with error terms having heteroscedasticity, then under the null hypothesis of no interaction and the assumptions, which are provided as equations (2.7) and (2.8), the rank transformed F statistic for interaction converges in distribution to an $(I-1)(J-1)$ degree of freedom chi-squared random variable divided by $(I-1)(J-1)$.

REFERENCES

(1) Akritas, M.G. and Arnold, S.F. (1994). Fully Nonparametric Hypotheses for Factorial Designs. *Journal of the American Statistical Association*, **89**, 336–343.

(2) Choi, Y.H. (1994). Rank Transform F Statistic in a 2×2 Factorial Design. *Journal of the Korean Statistical Society*, **23**, 103–114.

(3) Conover, W.J. and Iman, R.L. (1981). Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics. *The American Statistician*, **35**, 124–128.

(4) Fabian, V. (1991). On the Problem of Interactions in the Analysis of Variance. *Journal of the American Statistical Association*, **86**, 362–374.

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

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

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