International Journal of Reliability and Applications
Vol. 6, No. 1, pp. 53-64, 2005

Multivariate Nonparametric Tests for Grouped and Right Censored Data

Hyo-Il Park*
Department of Statistics, Chong-ju University, Choong-book, Cheong-ju, 360-764, Korea

Jong-Hwa Na
Department of Information and Statistics, Chungbuk National University, Choong-book, Cheong-ju, 361-763, Korea

Seungman Hong
Department of Informational Statistics, Korea University, 208 Seochangri, Jochiwon, Choong-nam, 339-700, Korea

Abstract. In this paper, we propose a nonparametric test procedure for the multivariate, grouped and right censored data for two sample problem. For the construction of the test statistic, we use the linear rank statistics for each component and apply the permutation principle for obtaining the null distribution. For the large sample case, the asymptotic distribution is derived under the null hypothesis with the additional assumption that two censoring distributions are also equal. Finally, we illustrate our procedure with an example and discuss some concluding remarks. In appendices, we derive the expression of the covariance matrix and prove the asymptotic distribution.

Key Words : *Grouped and right censored data, Linear rank statistic, Multivariate data, Nonparametric test, Permutation principle*

1. INTRODUCTION

In survival analysis, very often, statisticians have to deal with the multivariate data, which record more than one kind of event on each subject for the duration of observation. Or sometimes observations may be records of successive event times for the same outcome type. Various types of survival multivariate data have been introduced and discussed in Hougaard (2000). For these data, a lot of nonparametric test procedures have been proposed for comparing equality of distributions between two populations (cf. Puri

*Corresponding Author.
E-mail address: hipark@cju.ac.kr

and Sen, (1971). However sometimes, all or any component(s) of the observation may be right censored for various reasons. In this case, Wei and Lachin (1984) proposed a class of nonparametric tests based on the univariate generalized log-rank type of statistics for each component. Also Park and Desu (1998) and Park (2002) considered some extensions of the two types of median tests for this situation. In this research, however, we consider the case that the observation for the object may be conducted periodically with some fixed time interval or irregularly but with fixed time schedule during the experiment. Therefore the form of data may be categorized along with the time intervals. We call those as the grouped data. In other words, even though the underlying distributions are continuous, the data sets have been changed as a discrete type. For these data, one may apply any one among the usual nonparametric test procedures which have been developed for the untied-value case by adjusting rank or score for the tied value as an *ad hoc* approach. However Puri and Sen (1985) proposed a class of nonparametric tests in the linear model for the grouped data. Also Park (1993) considered nonparametric tests for the right censored data. In this paper, we consider proposing a nonparametric test procedure explicitly for the multivariate and right censored data.

2. MULTIVARIATE TESTS FOR GROUPED AND RIGHT CENSORED DATA

Let $\{X_i = (X_{i1},\ldots,X_{id})^T, i = 1,\ldots,m\}$ and $\{Y_j = (Y_{j1},\ldots,Y_{jd})^T, j = 1,\ldots,n\}$ be two independent d-variate random samples from populations with distribution functions F and G, respectively. $(\cdot)^T$ means the transpose of a vector or matrix. We assume that all the components of each observation vector have non-negative values since we consider the life time random vector. Then it is of interest to test $H_0 : F = G$ assuming that F and G are unknown but continuous.

Since the right censoring schemes must be involved for all components or some component, for each $i = 1,\ldots,m$, $j = 1,\ldots,n$ and $k = 1,\ldots,d$ one may observe $U_{ik} = \min(X_{ik}, C_{ik})$, $\tau_{ik} = I(X_{ik} \le C_{ik})$, $V_{jk} = \min(Y_{jk}, D_{jk})$ and $\gamma_{jk} = I(Y_{jk} \le D_{jk})$, where $\{C_i = (C_{i1},\ldots,C_{id})^T, i = 1,\ldots,m\}$ and $\{D_j = (D_{j1},\ldots,D_{jd})^T, j = 1,\ldots,n\}$ are the independent censoring random samples which are also independent of X's and Y's with unknown arbitrary distribution functions Q_F and Q_G, respectively. Furthermore since we are concerned with the grouped data with finite number of intervals over the positive half real line for each component, each observation for each component may be represented as (U^*_{ik}, τ_{ik}) and (V^*_{jk}, γ_{jk}), where for each $k = 1,\ldots,d$

$$U^*_{ik} = \sum_{l=0}^{s_k} I_l^k Z_{ikl} \text{ and } V^*_{jk} = \sum_{l=0}^{s_k} I_l^k Z_{jkl}$$

where $I_l^k = \left\{ z ; a_l^k \le z < a_{l+1}^k \right\}$, $l = 0,\ldots,s_k$ with $0 = a_0^k < a_1^k < \ldots < a_{s_k}^k < a_{s_k+1}^k = \infty$ and

$$Z_{ikl} = \begin{cases} 1 & if\ U_{ik} \in I_l^k \\ 0 & otherwise \end{cases} \text{ and } Z_{jkl} = \begin{cases} 1 & if\ V_{jk} \in I_l^k \\ 0 & otherwise \end{cases}.$$

Based on these data, we consider to propose a nonparametric test procedure for testing $H_0 : F = G$. For this purpose, let $N = m + n$ and F_{kN}, any consistent estimate of F_k which is the k th marginal distribution function of F based on the combined sample. Also let c_{1kl} and c_{0kl} be the scores of the uncensored and censored observations for the k th component in the l th subinterval such as

$$c_{1kl} = \frac{1}{F_{kN}(a_{l+1}) - F_{kN}(a_l)} \int_{F_{kN}(a_l)}^{F_{kN}(a_{l+1})} \phi(u)du \text{ and } c_{0kl} = \frac{1}{1 - F_{kN}(a_l)} \int_{F_{kN}(a_l)}^{1} \phi(u)du,$$

where $\phi(\bullet)$ is a square integrable function and can be chosen with the consideration of power. One may choose $\phi(u) = -f_k'(F_k^{-1}(u)) / f_k(F_k^{-1}(u))$ for $0 < u < 1$ if he or she has any knowledge about underlying distribution F, where f_k is the density of F_k and f_k' is the first derivative of f_k for each $k = 1, \ldots, d$. Also F_k^{-1} is the inverse image of F_k. Therefore we note that the score functions may be completely determined by ϕ and the choice of ϕ may depend on the underlying distribution function. As an example, for a logistic distribution, $\phi(u) = 2u - 1$. In later chapter, we will discuss more about scores. For the univariate case, Park(1993) proposed a class of locally most powerful tests based on the following linear rank statistics

$$S_{kN} = \sum_{i=1}^{m} \sum_{l=0}^{s_k} \{\tau_{ik} c_{1kl} Z_{ikl} + (1 - \tau_{ik}) c_{0kl} Z_{ikl}\} \qquad (2.1)$$

for testing $H_0^k : F_k = G_k$, where G_k is the k th marginal distribution of G. Park used the likelihood ratio principle to derive the linear rank statistics (2.1). In this study, we will use the permutation principle (cf. Good, 2000) for the development of the test procedure. Then we note that the resulting tests become conditional and we have to assume that $Q_F = Q_G$ under $H_0 : F = G$. Let P_N be the permutation distribution function under H_0 based on the two samples, (U_{ik}^*, τ_{ik}) and (V_{jk}^*, γ_{jk}), $i = 1, \ldots, m$, $j = 1, \ldots, n$ and $k = 1, \ldots, d$. Let $E(S_{kN} \mid P_N)$ be the null expectation of S_{kN} based on P_N. Then for large values of $| S_{kN} - E(S_{kN} \mid P_N) |$, one may reject $H_0^k : F_k = G_k$ in favor of $H_0^k : F_k \neq G_k$. In order to decide the critical value for any given significance level α, we need the null distribution of S_{kN}. It may be possible to obtain the null distribution of

S_{kN} for small or reasonable sample sizes under P_N. However for the large sample case, we have to consider the asymptotic normality using the large sample approximation. For this matter, you may refer to Park (1993) for the univariate case.

For the multivariate case, we consider using a quadratic form for test statistic by combining the linear rank statistics (2.1) for testing $H_0 : F = G$. Therefore we need the expectations, variances and covariances under $H_0 : F = G$. For this purpose, we introduce some more notation. Let n_{1l}^k and n_{0l}^k be the numbers of observations whose values are uncensored and censored contained in the subinterval $[a_l^k, a_{l+1}^k)$, respectively. Under P_N, Park (1993) obtained both as follows: for each k, we have that under H_0

$$E(S_{kN} \mid P_N) = m\left\{\sum_{l=0}^{s_k} c_{1kl}\frac{n_{1l}^k}{N} + \sum_{l=0}^{s_k} c_{0kl}\frac{n_{0l}^k}{N}\right\} = m(\mu_{1k} + \mu_{0k})$$

and

$$V(S_{kN} \mid P_N) = \frac{mn}{(N-1)}\left\{\sum_{l=0}^{s_k} c_{1kl}^2\frac{n_{1l}^k}{N} + \sum_{l=0}^{s_k} c_{0kl}^2\frac{n_{0l}^k}{N} - (\mu_{1k} + \mu_{0k})^2\right\}.$$

Also for $Cov(S_{kN}, S_{k'N} \mid P_N)$, we denote $n_{11ll'}^{kk'}$, $n_{10ll'}^{kk'}$, $n_{01ll'}^{kk'}$ and $n_{00ll'}^{kk'}$ as the numbers of observations whose values are uncensored for both components, uncensored for the kth component but censored for the k'th component, censored for the kth component but uncensored for the k'th component and censored for both components, respectively. Then the covariance between the kth and k'th components under H_0, is of the form

$$Cov\left(S_{kN}, S_{k'N} \middle| P_N\right) = \frac{mn}{N-1}\left\{\sum_{l=0}^{s_k}\sum_{l'=0}^{s_{k'}}\left[c_{1kl}c_{1k'l'}\frac{n_{11ll'}^{kk'}}{N} + c_{1kl}c_{0k'l'}\frac{n_{10ll'}^{ll'}}{N}\right.\right.$$
$$\left.\left. + c_{0kl}c_{1k'l'}\frac{n_{01ll'}^{kk'}}{N} + c_{0kl}c_{0k'l'}\frac{n_{00ll'}^{kk'}}{N}\right] - (\mu_{1k} + \mu_{0k})(\mu_{1k'} + \mu_{0k'})\right\}.$$

The derivation of $Cov(S_{kN}, S_{k'N} \mid P_N)$ will appear in Appendix. Now we define vectors and matrix in order to propose a class of test statistics for testing $H_0 : F = G$. Let

$$S_N = (S_{1N}, S_{2N}, \text{K}, S_{dN})^T$$

and

$$V_N = \begin{pmatrix} V(S_{1N} \mid P_N)\, Cov(S_{1N}, S_{2N} \mid P_N) \cdots Cov(S_{1N}, S_{dN} \mid P_N) \\ Cov(S_{1N}, S_{2N} \mid P_N)\, V(S_{2N} \mid P_N) \cdots Cov(S_{2N}, S_{dN} \mid P_N) \\ \vdots \\ Cov(S_{1N}, S_{dN} \mid P_N)\, Cov(S_{2N}, S_{dN} \mid P_N) \cdots V(S_{dN} \mid P_N) \end{pmatrix}.$$

We assume that V_N is positive definite. This guarantees the existence of the inverse V_N^{-1} of V_N. Then we may propose a class of test statistics for testing $H_0 : F = G$

$$T_N = (S_N^T - E(S_N^T \mid P_N))V_N^{-1}(S_N - E(S_N \mid P_N)).$$

Then we may reject $H_0 : F = G$ for large values of T_N in favor of $H_1 : F \neq G$. Then for any given significance level α, in order to decide the critical value $C_N(\alpha)$, we need the distribution of T_N under H_0. Then we may obtain the null distribution function of T_N based on the permutation principle for small or reasonable sample sizes. For large sample sizes, it is natural to consider the large sample approximation. Now we need the following assumption:

Assumption. As $N \to \infty$, n/N converges to a constant $\lambda \in (0, 1)$.

Then we obtain the asymptotic result in the following Theorem.

Theorem 2.1. With the above assumption, under H_0 with $Q_F = Q_G$, T_N converges in distribution to a chi-square distribution with d degrees of freedom.

Proof. Let

$$S_{kN}^* = \sum_{i=1}^{m} \sum_{l=1}^{s_k} \{\tau_{il} c_{ikl} Z_{ikl} + (1 - \tau_{il}) c_{0kl} Z_{ikl}\}.$$

Then it is easy to check that

$$E(S_{kN}^* | P_N) = m\left\{\sum_{l=1}^{s_k} c_{1kl} \frac{n_{1l}^k}{N} + \sum_{l=1}^{s_k} c_{0kl} \frac{n_{0l}^k}{N}\right\} = m(\mu_{1k} + \mu_{0k})$$

and

$$V(S_{kN}^* | P_N) = \frac{mn}{N-1}\left\{\sum_{l=1}^{s_k} c_{1kl}^2 \frac{n_{1l}^k}{N} + \sum_{l=1}^{s_k} c_{0kl}^2 \frac{n_{0l}^k}{N} - (\mu_{1k} + \mu_{0k})^2\right\}.$$

Also we have

$$Cov\left(S^*_{kN}, S^*_{kN}\middle|P_N\right) = \frac{mn}{N}\left[\sum_{l=1}^{s_k}\sum_{l'=1}^{s_{k'}}\left\{c_{1kl}c_{1k'l'}\frac{n^{kk'}_{11ll'}}{N} + c_{1kl}c_{0k'l'}\frac{n^{kk'}_{10ll'}}{N}\right.\right.$$
$$\left.\left. + c_{0kl}c_{1k'l'}\frac{n^{kk'}_{01ll'}}{N} + c_{0kl}c_{0k'l'}\frac{n^{kk'}_{00ll'}}{N}\right\} - (\mu_{1k}+\mu_{0k})(\mu_{1k'}+\mu_{0k'})\right].$$

Finally for each k, we define

$$\overline{S}_{kN} = \frac{S_{kN} - E\left(S_{kN}\middle|P_N\right)}{\sqrt{V\left(S_{kN}\middle|P_N\right)}} \text{ and } \overline{S}^*_{kN} = \frac{S^*_{kN} - E(S^*_{kN} \mid P_N)}{\sqrt{V(S^*_{kN} \mid P_N)}}.$$

Park (1993) have shown that

$$\lim_{N\to\infty} E\left[\left(\overline{S}_{kN} - \overline{S}^*_{kN}\right)^2\middle|P_N\right] = 0.$$

Therefore we see that for each k

$$\overline{S}_{kN} - \overline{S}^*_{kN} \xrightarrow{p} 0 \text{ under } P_N \text{ almost everywhere,}$$

where $\xrightarrow{p}$ implies convergence in probability. Since S^*_{kN} consists of m independent and identically distributed random variables, $\sum_{l=1}^{s_k}\{\tau_{il}c_{ikl}Z_{ikl} + (1-\tau_{il})c_{ikl}Z_{ikl}\}$, from the central limit theorem with Assumption, the asymptotic distribution of $\overline{S}^*_{kN}$ is standard normal. Therefore the asymptotic distribution of $\overline{S}_{kN}$ is also standard normal. For the derivation of the asymptotic distribution of T_N, let $\overline{S}_N = (\overline{S}_{1N}, \mathrm{K}, \overline{S}_{dN})^T$ and $\overline{S}^*_N = (\overline{S}^*_{1N}, \mathrm{K}, \overline{S}^*_{dN})^T$. Also let $a = (a_1, \Lambda, a_d)^T$ a column vector whose components are real scalars. Then we have that

$$E\left[\left(a^T\overline{S}_N - a^T\overline{S}^*_N\right)^2\middle|P_N\right]$$
$$= E\left[\left\{\sum_{k=1}^{d} a_k\left(S_{kN} - S^*_{kN}\right)\right\}^2\middle|P_N\right]$$
$$= \sum_{k=1}^{d} a_k^2 E\left[\left(S_{kN} - \overline{S}^*_{kN}\right)^2\middle|P_N\right] + \sum\sum_{k\neq l} a_k a_l E\left[\left(S_{kN} - \overline{S}^*_{kN}\right)\left(S_{lN} - \overline{S}^*_{lN}\right)\middle|P_N\right].$$

From the Cauchy-Schwarz inequality, under P_N, since

$$E\left[\left(\overline{S}_{kN}-\overline{S}^{*}_{kN}\right)\left(\overline{S}_{k'N}-\overline{S}^{*}_{k'N}\right)\Big|P_N\right]\le E\left[\left(\overline{S}_{kN}-\overline{S}^{*}_{kN}\right)^2\Big|P_N\right]^{1/2} E\left[\left(\overline{S}_{k'N}-\overline{S}^{*}_{k'N}\right)^2\Big|P_N\right]^{1/2},$$

we see that under P_N

$$\lim_{N\to\infty} E\left[\left(a^T\overline{S}_N-a^T\overline{S}^{*}_N\right)^2\Big|P_N\right]=0$$

for every column vector a whose components are all real. Therefore we have that for every column vector a

$$a^T\overline{S}_N-a^T\overline{S}^{*}_N\xrightarrow{p}0 \text{ under } P_N \text{ almost everywhere.}$$

For $k\neq k'$, let

$$\rho_{kk'N}=Cov\left(S_{kN},S_{k'N}|P_N\right)\Big/\sqrt{V\left(S_{kN}|P_N\right)V\left(S_{k'N}|P_N\right)}.$$

Then it is easy to see that with Assumption and the fact that ϕ is square integrable

$$\rho_{kk'N}\xrightarrow{p}\rho_{kk'},$$

where $\rho_{kk'}$ is the limit correlation coefficient between S_{kN} and $S_{k'N}$ for $k\neq k'$. Let R be the limiting correlation matrix whose elements are 1 for $k=k'$ and $\rho_{kk'}$ for $k\neq k'$. Also let R^* be the limiting correlation matrix $\overline{S}^{*}_N$. Then it is obvious that $R=R^*$.

Since for every column vector a, $a^T\overline{S}^{*}_N$ converges in distribution to a normal distribution with mean 0 and variance a^TRa, $a^T\overline{S}_N$ also converges in distribution to a normal distribution with mean 0 and variance a^TRa. Thus $\overline{S}_N$ converges in distribution to a multivariate normal with 0 mean vector and covariance matrix R with applying the Cramer-Wold's device (cf. Billingsley, 1985). Thus we have the result.

3. AN EXAMPLE AND CONCLUDING REMARKS

For illustration of our test procedure, we show an example using the data of Makuch et al. (1991). The data reveal the number of days before HIV-1 positivity in culture was registered in an assay. The patients were divided into two groups according to taking drug 1 or 2. During a 4-month period, this in vitro procedure was performed repeatedly on samples monthly from patients with AIDS. Thus each observation consists of four consecutive event times and so can be considered as a four variate datum. We note that some components of each observation were censored. Also we note that for each component, there are a lot of tied values since each observation records the number of days during a 30-days span. Therefore the data come into the realm of our study. For more detailed discussion of this data, you may refer to Makuch et al. (1991). Then from the data, it is of interest to see whether there is any difference in days among the four occasions between the two groups. For the Wilcoxon score for all components, $\phi_k(u) = u$ with the Kaplan-Meier estimate, we obtain that

$$S_{47}^T - E(S_{47}^T \mid P_{47}) = (1.69,-0.25,0.51,-1.00)$$

$$V_{47} = \begin{pmatrix} 0.71 & 0.19 & 0.33 & 0.19 \\ 0.19 & 0.80 & 0.22 & 0.20 \\ 0.33 & 0.22 & 0.70 & 0.28 \\ 0.19 & 0.20 & 0.28 & 0.65 \end{pmatrix}$$

$$T_{47} = 7.91.$$

Then the corresponding p-value is 0.095. If we use the median score for each k such as

$$\phi_k(u) = \begin{cases} 1 & if \quad u \leq 0.5 \\ 0 & otherwise \end{cases}$$

with the Kaplan-Meier estimate, then we have

$$S_{47}^T - E(S_{47}^T \mid P_{47}) = (0.45,-0.61,0.00,2.83)$$

$$V_{47} = \begin{pmatrix} 0.96 & 0.14 & 0.34 & 0.62 \\ 0.14 & 0.62 & 0.07 & 0.14 \\ 0.34 & 0.07 & 0.75 & 0.21 \\ 0.62 & 0.14 & 0.21 & 0.89 \end{pmatrix}$$

$$T_{47} = 14.71$$

and the corresponding p-value is 0.003, which shows that there may exist any significant difference between the two groups. All the numerical calculations were carried out using S-PLUS (cf. S-PLUS 4 Programmers Guide, 1997). Also we note that Makuch et al. (1991) obtained 0.055 as its p-value based on the Gehan score for all components.

For the univariate data, Prentice (1978) obtained a similar form of the linear rank statistics (2.1) in no tied value case among uncensored observations with applying the marginal likelihood and proposed a nonparametric test procedure. Also there is another type of linear rank statistics which is so-called the log-rank type of statistics (cf. Gehan, 1965). Mehrotra et al. (1982) showed that the two types of linear rank statistics are algebraically equivalent. For the grouped data, also Neuhaus (1993) proposed a nonparametric test procedure based on the modified log-rank statistics and applied the permutation principle to obtain the critical value. Park (1997) showed that the two types of linear rank statistics for grouped data are also equivalent. Therefore one may consider another test procedure for multivariate data with possibly censored components using the modified log-rank statistics considered by Neuhaus (1993) for the grouped data. Then one has to estimate the expectation and covariance matrix in this case.

Already, it was pointed out that the choice of the scores depends on the underlying distribution F for the considerations of optimality property such as locally most powerfulness of tests for the univariate case. Therefore for the considerations of the power of test, we may allow to vary the scores from component to component along with the marginal distribution functions. This point makes us to be able to apply our procedure to the data which contain the categorical component or components. Therefore we may apply our tests to the various type of data. Also Prentice (1978) obtained the linear rank statistics (2.1) in no tied value case for uncensored observations by applying the marginal likelihood. Prentice used the expected value scores. Park (2000) also showed that the two types of scores are asymptotically equivalent.

ACKNOWLEGEMENT

This study was supported by a grant of Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (02-PJ1-PG3-10411-0002).

APPENDIX

In this appendix, we derive the covariance $Cov(S_{kN}, S_{k'N} \mid P_N)$. For this it is enough to consider $Cov(S_{1N}, S_{2N} \mid P_N)$ since the rest of them are actually the same except notations. From now on, we omit the superscripts since we consider the first two components only. Then we note that

$$E\{S_{1N}, S_{2N} | P_N\}$$

$$= E\sum_{i=1}^{m}\left\{\sum_{l_1=0}^{s_1}\{\tau_{i1}c_{11l_1}Z_{i1l_1} + (1-\tau_{i1})c_{01l_1}Z_{i1l_1}\}\sum_{l_2=0}^{s_2}\{\tau_{i2}c_{12l_2}Z_{i2l_2} + (1-\tau_{i2})c_{02l_2}Z_{i2l_2}\}\middle| P_N\right\}$$

$$+ 2E\sum\sum_{1\le i\ne j\le m}\left\{\sum_{l_1=0}^{s_1}\{\tau_{i1}c_{11l_1}Z_{i1l_1} + (1-\tau_{i1})c_{01l_1}Z_{i1l_1}\}\right.$$

$$\left.\sum_{l_2=0}^{s_2}\left\{\tau_{j2}c_{12l_2}Z_{j2l_2}+(1-\tau_{j2})c_{02l_2}Z_{j2l_2}\right\}\right|P_N\Bigg\}$$

$$=mE\left\{\sum_{l_1=0}^{s_1}\left\{\tau_{i1}c_{11l_1}Z_{i1l_1}+(1-\tau_{i1})c_{01l_1}Z_{i1l_1}\right\}\sum_{l_2=0}^{s_2}\left\{\tau_{12}c_{12l_2}Z_{12l_2}+(1-\tau_{12})c_{02l_2}Z_{12l_2}\right\}\middle|P_N\right\}$$

$$+m(m-1)E\left\{\sum_{l_1=0}^{s_1}\left\{\tau_{11}c_{11l_1}Z_{11l_1}+(1-\tau_{11})c_{01l_1}Z_{11l_1}\right\}\sum_{l_2=0}^{s_2}\left\{\tau_{22}c_{12l_2}Z_{22l_2}+(1-\tau_{22})c_{02l_2}Z_{22l_2}\right\}\middle|P_N\right\}$$

Therefore it is enough to obtain the expressions of the following 8 terms

$$\sum_{l_1=0}^{s_1}\sum_{l_2=0}^{s_1}c_{11l_1}c_{12l_2}E\left\{\tau_{11}Z_{11l_1}\tau_{12}Z_{12l_2}\middle|P_N\right\},\ \sum_{l_1=0}^{s_1}\sum_{l_2=0}^{s_1}c_{11l_1}c_{02l_2}E\left\{\tau_{11}Z_{11l_1}(1-\tau_{12})Z_{12l_2}\middle|P_N\right\}$$

$$\sum_{l_1=0}^{s_1}\sum_{l_2=0}^{s_1}c_{01l_1}c_{12l_2}E\left\{(1-\tau_{11})Z_{11l_1}\tau_{12}Z_{12l_2}\middle|P_N\right\}$$

$$\sum_{l_1=0}^{s_1}\sum_{l_2=0}^{s_1}c_{01l_1}c_{02l_2}E\left\{(1-\tau_{11})Z_{11l_1}(1-\tau_{12})Z_{12l_2}\middle|P_N\right\},$$

and

$$\sum_{l_1=0}^{s_1}\sum_{l_2=0}^{s_1}c_{11l_1}c_{12l_2}E\left\{\tau_{11}Z_{11l_1}\tau_{22}Z_{22l_2}\middle|P_N\right\},\ \sum_{l_1=0}^{s_1}\sum_{l_2=0}^{s_1}c_{11l_1}c_{02l_2}E\left\{\tau_{11}Z_{11l_1}(1-\tau_{22})Z_{22l_2}\middle|P_N\right\}$$

$$\sum_{l_1=0}^{s_1}\sum_{l_2=0}^{s_1}c_{01l_1}c_{12l_2}E\left\{(1-\tau_{11})Z_{11l_1}\tau_{22}Z_{22l_2}\middle|P_N\right\}$$

$$\sum_{l_1=0}^{s_1}\sum_{l_2=0}^{s_1}c_{01l_1}c_{02l_2}E\left\{(1-\tau_{11})Z_{11l_1}(1-\tau_{22})Z_{22l_2}\middle|P_N\right\}.$$

We note that

$$E[\tau_{11}Z_{11l_1}\tau_{12}Z_{12l_2}\mid P_N]=\frac{n_{11l_1l_2}}{N},\qquad E[\tau_{11}Z_{11l_1}(1-\tau_{12})Z_{12l_2}\mid P_N]=\frac{n_{10l_1l_2}}{N},$$

$$E[\ (1-\tau_{11})Z_{11l_1}\tau_{12}Z_{12l_2}\mid P_N]=\frac{n_{01l_1l_2}}{N}$$

and

$$E[\ (1-\tau_{11})Z_{11l_1}(1-\tau_{12})Z_{12l_2}\mid P_N]=\frac{n_{00l_1l_2}}{N}.$$

Also we note that

$$E[\tau_{11}Z_{11l_1}\tau_{22}Z_{22l_2} \mid P_N] = P\{\tau_{11}=1,\ Z_{11l_1}=1,\ \tau_{22}=1,\ Z_{22l_2}=1 \mid P_N\}$$
$$= P\{\tau_{11}=1,\ Z_{11l_1}=1, \mid \tau_{22}=1,\ Z_{22l_2}=1, H_N\}P\{\tau_{22}=1,\ Z_{22l_2}=1 \mid P_N\}$$
$$= \frac{n_{11\cdot l_2}}{N}\left[\frac{n_{11l_1\cdot}-1}{N-1}\frac{n_{11l_1l_2}}{n_{11\cdot l_2}} + \frac{n_{11l_1\cdot}}{N-1}\frac{n_{11\cdot l_2}-n_{11l_1l_2}}{n_{11\cdot l_2}}\right]$$
$$= \frac{n_{11l_1\cdot}}{N-1}\frac{n_{11\cdot l_2}}{N} - \frac{1}{N-1}\frac{n_{11l_1l_2}}{N}$$

where $n_{11l_1\cdot} = \sum_{l=1}^{s_2} n_{11l_1l}$ and $n_{11\cdot l_2} = \sum_{l=1}^{s_1} n_{11ll_2}$.

In the following the other notations with dot are similarly defined. Then with the same arguments used for $E\,[\tau_{11}Z_{11l_1}\tau_{22}Z_{22l_2} \mid P_N]$, we obtain that

$$E[\tau_{11}Z_{11l_1}(1-\tau_{22})Z_{22l_2} \mid P_N] = \frac{n_{10l_1\cdot}}{N-1}\frac{n_{10\cdot l_2}}{N} - \frac{1}{N-1}\frac{n_{10l_1l_2}}{N} \quad ,$$
$$E[(1-\tau_{11})Z_{11l_1}\tau_{22}Z_{22l_2} \mid P_N] = \frac{n_{01l_1\cdot}}{N-1}\frac{n_{01\cdot l_2}}{N} - \frac{1}{N-1}\frac{n_{01l_1l_2}}{N}$$

and

$$E[(1-\tau_{11})Z_{11l_1}(1-\tau_{22})Z_{22l_2} \mid P_N] = \frac{n_{00l_1\cdot}}{N-1}\frac{n_{00\cdot l_2}}{N} - \frac{1}{N-1}\frac{n_{00l_1l_2}}{N} \quad .$$

Using the expressions for all terms, with straightforward but tedious calculation leads to the conclusion.

REFERENCES

Billingsley, P. (1985). *Probability and measure,* second edition, Wiley and Sons, Inc., New York.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, **52**, 203-223.

Good, P. (2000). *Permutation test- A pratical guide to resampling methods for testing hypotheses.* Springer-Verlag, New York.

Hougaard, P. (200). *Analysis of multivariate survival data.* Springer-Verlag, New York.

Makuch, R. W., Escobar, M. and Merill, S. (1991). A two sample test for incomplete multivariate data. *Applied Statistics*, **40**, 202-212.

Mehrotra, K. G., Michalek, J. E. and Mihalko, D. (1982). A relationship between forms of linear rank procedures for censored data.. *Biometrika*, **69**, 674-676.

Neuhaus, G. (1993). Conditional rank tests for the two-sample problem under random censorship. *Annals of Statistics*, **21**, 1760-1779.

Park, H. I. (1993). Nonparametric rank-order tests for the right censored and grouped data in linear model. *Communications in Statistics-Theory and Methods*, **22**, 3143-3158.

Park, H. I. (2000). A study on the scores for right censored data. *Journal of Statistical Planning and Inference*, **86**, 101-111.

Park, H. I. (2002). Multivariate percentile tests for incomplete data. *Annals of the Institute of Statistical Mathematics*. 934-944.

Park, H. I. and Desu, M. M. (1998). Multivariate control test for right censored data. *Communications in Statistics-Theory and Methods*, **27**, 1923-1935.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, **63**, 291-298.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric methods in multivariate analysis*. Wiley and Sons, Inc., New York.

Puri, M. L. and Sen, P. K. (1985). *Nonparametric methods in general linear models*. Wiley and Sons, Inc., New York.

S-PLUS 4 Programmer's Guide, (1997). Data Analysis Product Division, Mathsoft Inc., Seattle, Washington.

Wei, L. J. and Lachin, J. M. (1984). Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *Journal of American Statistical Association*, **79**, 653-661.