Journal of the Kerala Statistical Association
Vol.24, December 2013, 1-11 *ISSN 2249-4553*

Identities Emerging from Two Sample U-Statistics

Sharada V. Bhat* and I.D. Shetty
Department of Statistics
Karnatak University, Dharwad

ABSTRACT

In this paper, we derive some mathematical identities which involve combinatorial coefficients. The well known theory of two-sample U Statistics is used to derive the identities.

Key words and Phrases: *U-Statistics, two-sample location problem, combinatorial coefficients, ordered ranks.*

1 Introduction

Identities are useful in simplifying many algebraic expressions. They provide simple alternate expressions to solve complex algebraic expressions. Riordan (1968) contains many such fundamental identities, Joshi and Balakrishnan (1981) provide statistical derivations of some such identities and Baiju and Thomas (2007) describe some identities using well established theories of order statistics and U-statistics based on certain linear functions of order statistics.

In this paper, we derive some identities using two-sample U-statistics and ordered ranks. The two-sample U-statistics is described in section 2 and two-sample

*email: bhat_sharada@yahoo.com

U-Statistics from which these identities emerge are described in section 3. The identities derived from mean and variance of these U-statistics is given in section 4 and identities emerging from ordered ranks is given is section 5.

2 Definition of Two-sample U-Statistics

U-Statistics is a class of unbiased estimators of parameters of a population. They are often used as test statistics although they emphasize estimation. Randles and Wolfe (1979) describes the two-sample U-Statistics as follows:

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be the observations of two independent random samples drawn from cumulative distribution functions (cdf) $F(x)$ and $G(y)$ respectively. A parameter θ is said to be estimable of degree (b, d) for distributions (F, G) in a family $\boldsymbol{F}$ if b and d are the smallest sample sizes for which there exists an estimator of θ that is unbiased for every $(F, G) \in \boldsymbol{F}$. That is, there is a function $h(.\ ;\ .)$ such that $E_{F,G}\left[h(X_1, ..., X_b; Y_1, ..., Y_d)\right] = \theta$ for every $(F, G) \in \boldsymbol{F}$, where $h(.\ ;\ .)$ is called as two-sample kernel and is symmetric in it's X_i components and separately symmetric in it's Y_j components. Under these conditions a two sample U-statistic, for $m \geq b$ and $n \geq d$ has the form

$$U(X_1, ..., X_m; Y_1, ..., Y_n) = \left(\begin{pmatrix} m \\ b \end{pmatrix} \begin{pmatrix} n \\ d \end{pmatrix} \right)^{-1} \sum_{\alpha} h(X_{i_1}, ..., X_{i_b}, Y_{j_1}, ..., Y_{j_d}) \ , \tag{2.1}$$

where $\sum_\alpha$ is the collection of all subsets of $b(d)$ integers chosen without replacement from the integers {1,. . . ,m} {1,. . . ,n}.

3 Some two sample U-Statistics for location problem

Suppose $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are independent random samples from populations with absolutely continuous distribution functions F(x) and G(y) having probability density functions (pdf) $f(x)$ and $g(y)$ respectively. Then the two sample location problem is to test $H_0 : F\left(x\right) \equiv G\left(x\right)$ against the alternative$H_1 : G\left(x\right) =$

$F(x-\theta)$, $\theta>0$ or $\theta<0$ or $\theta\neq 0$, -∞<x<∞, that is, two distributions differ only in their location parameter.

Suppose b and d are some fixed positive integers such that $1 \leq b \leq m$ and $1 \leq d \leq n$. For testing $H_0 : \theta = 0$ against$H_1 : \theta > 0$, Shetty and Bhat (1993) proposed

$$SB_1 = [\binom{m}{b}\binom{n}{d}]^{-1} \sum_{\alpha} \phi_1 (X_{i_1}, ..., X_{i_b}; Y_{j_1}, ..., Y_{j_d}), \tag{3.1}$$

where $\sum_{\alpha}$is sum over all $\binom{m}{b}\binom{n}{d}$possible sub samples,

$$\phi_1(X_1, ..., X_b; Y_1, ..., Y_d) = \begin{cases} 1, & if \quad M_1 \leq M_2 \\ 0, & otherwise \end{cases}, \tag{3.2}$$

M_1= median of $(X_1, ..., X_b)$, M_2= median of $(Y_1, ..., Y_d)$, and b and d are odd positive integers.

Shetty and Bhat (1994) proposed

$$SB_2 = [\binom{m}{b}\binom{n}{d}]^{-1} \sum_{\alpha} \phi_2 (X_{i_1}, ..., X_{i_b}; Y_{j_1}, ..., Y_{j_d}) \tag{3.3}$$

and

$$SB_3 = [\binom{m}{b}\binom{n}{d}]^{-1} \sum_{\beta} \phi_3 (X_{i_1}, ..., X_{i_d}; Y_{j_1}, ..., Y_{j_b}), \tag{3.4}$$

where $\sum_{\beta}$is sum over all $\binom{m}{d}\binom{n}{b}$possible sub samples,

$$\phi_2(X_1, ..., X_b; Y_1, ..., Y_d) = \begin{cases} 1, & if \quad X_{(b)} \leq M_2 \\ 0, & otherwise \end{cases} \tag{3.5}$$

$$\phi_3(X_1, ..., X_b; Y_1, ..., Y_d) = \begin{cases} 1, & if \quad M_3 \leq Y_{(1)} \\ 0, & otherwise \end{cases} \tag{3.6}$$

M_3= median of $(X_1, ..., X_d)$, $X_{(b)}$= maximum of $(X_1, ..., X_b)$, $Y_{(1)}$= minimum of $(Y_1, ..., Y_b)$ and d is an odd positive integer.

Shetty et al.(1997) proposed a two-sample U-Statistic with kernel being the function of sample quantiles which is given by

$$SB_4 = [\binom{m}{b}\binom{n}{d}]^{-1} \sum_{\alpha} \phi_4 (X_{i_1}, ..., X_{i_b}; Y_{j_1}, ..., Y_{j_d}) \tag{3.7}$$

and

$$\phi_4(X_1,...,X_b;Y_1,...,Y_d)=\begin{cases}1 & if \quad X_{(k_1)b}\le Y_{(k_2)d}\\ 0 & otherwise\end{cases} \tag{3.8}$$

where,

$$\mathrm{k}_1=\begin{cases}b\beta, & if \quad b\beta \quad \text{is an integer}\\ [b\beta]+1, & if \quad b\beta \quad \text{is not an integer},\end{cases}$$

$$\mathrm{k}_2=\begin{cases}d\gamma, & if \quad d\gamma \quad \text{is an integer}\\ [d\gamma]+1, & if \quad d\gamma \quad \text{is not an integer},\end{cases}$$

β^{th} quantile of a sample size n is defined as the r^{th}order statistic, where

$$r=\begin{cases}n\beta, & if \quad n\beta \quad \text{is an integer}\\ [n\beta]+1, & if \quad n\beta \quad \text{is not an integer},\end{cases}$$

$X_{(k_1)b}$= k_1^{th}order statistic of $(X_1,...,X_b)$ and $Y_{(k_2)d}$=k_2^{th} order statistic of $(Y_1,....,Y_d)$. All these two sample U-Statistics are expressed in alternative forms using ordered ranks and their properties are also studied. An extensive study of these statistics is carried out in Bhat (1996).

4 Identities from mean and variance of two sample U-Statistics

In this section, we present some identities and their proofs based on mean and variance of two sample U statistics defined in section 3.

Identity 4.1.

$$\left(d!/(q!)^2\right)\ \sum_{i=p+1}^{b}\binom{b}{i}B\left(i+q+1,\,b-i+q+1\right)\ =1/2 \tag{4.1}$$

for b and d being odd positive integers.

Proof.

The mean of SB_1 under H_0 is obviously $1/2$.

But it is also given by

$$\mathrm{E}\left(\mathrm{SB}_1\right)=\int_{-\infty}^{\infty}\mathrm{F}_{M_1}\left(x\right)\ \mathrm{d}\ \mathrm{F}_{M_2}\left(x\right),$$

where

$$F_{M_1}(x)=\sum_{i=p+1}^{b}\left(\begin{array}{c} b \\ i \end{array}\right)[F(x)]^i\ \ \left[\overline{F}\left(x\right)\right]^{b-i},$$

$$F_{M_2}(x)=\sum_{i=q+1}^{d}\left(\begin{array}{c} d \\ i \end{array}\right)[F(x)]^i\ \ \left[\overline{F}\left(x\right)\right]^{d-i},$$

$$p=(b-1)/2,\qquad q=(d-1)/2\qquad\text{and}\quad\ \overline{F}\left(\mathrm{x}\right)\ =1-F\left(x\right)\ .$$

Therefore,

$$E\left(SB_1\right)=\int_{-\infty}^{\infty}\sum_{i=p+1}^{b}\left(\begin{array}{c} b \\ i \end{array}\right)[F(x)]^i\ \ \left[\overline{F}\left(x\right)\right]^{b-i}\left(x\right)\ \left(d!/(q!)^2\right)\ [F(x)]^q\ \ \left[\overline{F}\left(x\right)\right]^q dF(x)$$

$$=\sum_{\mathrm{i=p+1}}^{\mathrm{b}}\left(\begin{array}{c} b \\ i \end{array}\right)\left(d!/(q!)^2\right)\int_{-\infty}^{\infty}[F(x)]^{i+q}\ \ \left[\overline{F}\left(x\right)\right]^{b-i+q}dF\left(x\right)$$

$$=\left(d!/(q!)^2\right)\sum_{i=p+1}^{b}\left(\begin{array}{c} b \\ i \end{array}\right)B\left(i+q+1,b-i+q+1\right),$$

where $\mathrm{B}(x,y)=\Gamma\left(x\right)\ \Gamma\left(y\right)\ /\,\Gamma(x+y)$. Therefore, we get the identity (4.1).

Identity 4.2.

$$b^2\zeta_{10}(SB_1)/\left(d^2\zeta_{01}(SB_1)\right)=1, \tag{4.2}$$

where

$$\zeta_{10}(SB_1)=cov\left[\phi_1(X_1,...,X_b;Y_1,...,Y_d)\ ,\ \phi_1(X_1,X_{b+1},...,X_{2b-1};Y_1,....,Y_{2d})\right]$$

and

$$\zeta_{01}(SB_1)=cov\left[\phi_1(X_1,...,X_b;Y_1,...,Y_d)\ ,\ \phi_1(X_{b+1},...,X_{2b};Y_1,Y_{d+1},...,Y_{2d-1})\right].$$

Proof. While deriving the asymptotic variance of SB_1 under the null hypothesis, certain expressions are evaluated. We have

$$\zeta_{10}(SB_1) = E\left[\phi_1(X_1, ..., X_b; Y_1, ..., Y_d)\ ,\ \phi_1(X_1, X_{b+1}, ..., X_{2b-1}; Y_{d+1}, ..., Y_{2d})\right] - E\left[SB_1\right]^2$$

$$= \int_{-\infty}^{\infty} P^2\left(med\ (x, X_2, ..., X_b) \leq M_2\right) dF(x) - (1/4)\ . \tag{4.3}$$

Also

$$P(med\ (x, X_2, ..., X_b) \leq M_2) = d!(b-i)![K_1 + \cdots + K_b], \tag{4.4}$$

where

$$K_1 = P[x \leq Y_1,\ \ X_2 \leq X_3 \leq ... \leq x \leq ... \leq X_b,\ \ Y_2 \leq Y_3 \leq ... \leq Y_1 \leq ... \leq Y_d],$$

x in the middle position and Y_1 in the middle position.

$$K_i = P[X_i \leq Y_1,\ \ X_2 \leq X_3 \leq .... \leq X_i \leq .... \leq x,\ \ Y_2 \leq Y_3 \leq ... \leq Y_1 \leq .... \leq Y_d,]$$

for $i = 2, ..., b$ and X_i in the middle position and Y_1 in the middle position.
After evaluating the expressions for$K_1, \ldots, K_b$ and substituting in (4.4) and (4.3) we get

$$\zeta_{10}(SB_1) = \left(d!/(b-1)!\right)^2 K\left(b, d\right) / \left(p!^2 q!^2 (b+d-1) \begin{pmatrix} b+d-2 \\ p+q \end{pmatrix}\right)^2, \tag{4.5}$$

where

$$K(b,d) = \sum_{i=0}^{p+q} \begin{pmatrix} b+d-1 \\ i \end{pmatrix}^2 B(2i+1,\ 2b+2d-2i-1)$$
$$+ \sum_{i \neq i'=1}^{p+q} \sum \begin{pmatrix} b+d-1 \\ i \end{pmatrix} \begin{pmatrix} b+d-1 \\ i' \end{pmatrix} B(i+i'+1,\ \ 2b+2d-i-i'-1) - (1/4)\,.$$

Similarly,

$$\zeta_{01}(SB_1) = E[\phi_1(X_1, ..., X_b; Y_1, ..., Y_d)\ ,\ \phi_1(X_{b+1}, ..., X_{2b}\ ;\ Y_1, Y_{d+1}, ..., Y_{2d-1}] - (1/4)$$

$$= (b!/(d-1)!)^2 K(b,d) / \left(p!^2 q!^2 (b+d-1) \begin{pmatrix} b+d-2 \\ p+q \end{pmatrix} \right)^2 . \quad (4.6)$$

Therefore

$$\zeta_{10}(SB_1)/\zeta_{01}(SB_1) = (d!/(b-1)!)^2 / \ (b!/(d-1)!)^2 = (d/b)^2 . \quad (4.7)$$

Hence, we get the identity (4.2).

Identity 4.3.

$$\zeta_{10}(\mathrm{SB}_2) - \zeta_{01}(\mathrm{SB}_3) = 0 , \quad (4.8)$$

where

$$\zeta_{10}(SB_2) = cov\left[\phi_2(X_1,...,X_b;Y_1,...,Y_d) \ , \ \phi_2(X_1,X_{b+1},...,X_{2b-1};Y_{d+1},....,Y_{2d})\right]$$

and

$$\zeta_{01}(SB_3) = cov\left[\phi_3(X_1,...,X_b;Y_1,...,Y_d) \ , \ \phi_3(X_{b+1},....,X_{2b};Y_1,Y_{d+1}....,Y_{2d-1})\right].$$

Proof. Under the null hypothesis

$$E(SB_2) = \int_{-\infty}^{\infty} [F(x)]^b \, dF_{M_2}(x)$$

$$= \left(d!/(q!)^2\right) B(b+q+1, q+1) \quad (4.9)$$

and

$$E(SB_3) = \int_{-\infty}^{\infty} \left[\overline{F}(x)\right]^b dF_{M_3}(x),$$

where

$$F_{M_3}(x) = \sum_{i=q+1}^{d} \begin{pmatrix} d \\ i \end{pmatrix} [F(x)]^i \left[\overline{F}(x)\right]^{d-i} .$$

Therefore ,

$$E(SB_3) = \left(d!/(q!)^2\right) \int_{-\infty}^{\infty} [F(x)]^q \left[\overline{F}(x)\right]^{b+q} dF(x)$$

$$= \left(d!/(q!)^2\right) B(q+1, b+q+1) \ . \quad (4.10)$$

Since the kernel $\phi_2(.;.)$can be obtained from $\phi_3(.;.)$ by replacing(X_i)'s by $(-X_i)$'s, (Y_j)'s by $(-Y_j)$'s and interchanging labels we get$E(SB_2) = E(SB_3)$.

Also,

$$\zeta_{10}(SB_2) = E\left(\phi_2(X_1,...,X_b;Y_1,...,Y_d)\ ,\ \phi_2(X_1,X_{b+1},...,X_{2b-1};Y_{d+1},....,Y_{2d})\right) - E^2[SB_2]$$

$$= \int_{-\infty}^{\infty} P^2\left(\max\left(x,X_2,...,X_b\right) \leq M_2\right) dF(x) - E^2(SB_2) \tag{4.11}$$

and

$$\zeta_{01}(SB_3) = E\left(\phi_3(X_1,...,X_d;Y_1,...,Y_b)\ ,\ \phi_3(X_{d+1},...,X_{2d};Y_1,Y_{b+1},....,Y_{2b-1})\right) - E^2[SB_3]$$

$$= \int_{-\infty}^{\infty} P^2\left(M_3 \leq \min\left(y,Y_2,...,Y_b\right)\right) dF(x) - E^2(SB_3). \tag{4.12}$$

Since $E[SB_2] = E[SB_3]$ under H_0 and from symmetry, we have

$$P\left(\max\left(x,X_2,...,X_b\right) \leq M_2\right) = P\left(M_3 \leq \min\left(y,Y_2,...,Y_d\right)\right).$$

From (4.9) through (4.12) we get the identity (**??**).

Identity 4.4.

$$k\binom{d}{k}\sum_{i=k}^{b}\binom{b}{i}\ B\left(i+k,\ 2b-i-k+1\right) = 1/2 \tag{4.13}$$

or

$$\left[d!/\left((k-1)!(d-k)!\right)\right]\ \textstyle\sum_{i=k}^{b}\binom{b}{i}\ B\left(i+k,\ 2b-i-k+1\right) = 1/2.$$

Proof. Under the null hypothesis

$$\begin{aligned} E(SB_4) &= P(X_{(k_i)} \leq Y_{(k_2)}) \\ &= \int_{-\infty}^{\infty} P(X_{(k_i)} \leq y) dF_{Y_{(k_2)}}(y) \\ &= k_2\binom{d}{k_2}\sum_{i=k_1}^{b}\binom{b}{i}\ B\left(i+k_2,\ b+d-i-k_2+1\right)), \end{aligned} \tag{4.14}$$

When $b = d$ and $k_1 = k_2 = k$, we have $E(SB_4) = 1/2$. Therefore we get identity (4.13).

Identity 4.5. For $i = 1,3,5,...,b+d-1$,

$$e\left(SB_1\left(b,d\right)\right)/e\left(SB_1\left(i,b+d-i\right)\right) = 1, \tag{4.15}$$

where

$$e\left(SB_1\left(b,d\right)\right) = \left[\left(b!d!\right)/\left(\left(p!q!\right)^2\sigma_{b,d}\right)\right]\int_{-\infty}^{\infty}\left[F\left(x\right)\right]^{p+q}\left[\overline{F}\left(x\right)\right]^{p+q}\left[f\left(x\right)\right]^2 dx,$$

$$\sigma_{b,d}^2 = \left(d!\right)^2 \left(b!\right)^2 K\left(b,d\right) / \left[\left(p!\right)^4 \left(q!\right)^4 \left(b+d-1\right)^2 \left(\begin{array}{c} b+d-2 \\ p+q \end{array} \right)^2 \lambda\left(1-\lambda\right) \right]$$

and

$$0 < \lambda = \lim_{N \to \infty} \left(m/N\right) < 1, \qquad N = m+n.$$

Proof. Under the null hypothesis

$$\sigma_{b,d}^2 = \left(b^2/\lambda\right) \zeta_{10}\left(SB_1\right) + \left(d^2/\left(1-\lambda\right)\right) \zeta_{01}\left(SB_1\right)$$

$= b^2 \zeta_{10}\left(SB_1\right) / \left(\lambda\left(1-\lambda\right)\right)$ or $d^2 \zeta_{01}\left(SB_1\right) / \left(\lambda\left(1-\lambda\right)\right)$ by identity 4.2.
It is worth to note that $e\left(SB_1\left(b,d\right)\right)$ depends on $\left(b+d\right)$ and underlying distribution $F\left(x\right)$. Thus for $1 \leq b \leq m$, $1 \leq d \leq n$, b,d being odd positive integers, given $F\left(x\right)$, we get $e\left(SB_1\left(b,d\right)\right) = e\left(SB_1\left(i,b+d-i\right)\right)$ for$i = 1,3,5,...,b+d-1$. Therefore, we get identity (4.15).

5 Identities based on Ordered Ranks

In this section, we present some identities based on the ordered ranks of two sample U statistics defined in section 3. Suppose that $X_{(1)} \leq X_{(2)} \leq ..... \leq X_{(m)}$ and $Y_{(1)} \leq Y_{(2)} \leq ..... \leq Y_{(n)}$ are the order statistics corresponding to X and Y sample observations respectively. Let $R_{(i)}\left(S_{(j)}\right)$ be the rank of $X_{(i)}\left(Y_{(j)}\right)$ in the joint ranking of X and Y observations. Then we have the following identities.

Identity 5.1. $\sum_{i=1}^{m} \sum_{j=0}^{q} \left(\begin{array}{c} i-1 \\ p \end{array} \right) \left(\begin{array}{c} m-i \\ p \end{array} \right) \left(\begin{array}{c} R_{(i)}-i \\ q-j \end{array} \right) \left(\begin{array}{c} n-R_{(i)}+i \\ q+1+j \end{array} \right)$

$$= \sum_{j=1}^{n} \sum_{i=0}^{p} \left(\begin{array}{c} j-1 \\ q \end{array} \right) \left(\begin{array}{c} n-j \\ q \end{array} \right) \left(\begin{array}{c} S_{(j)}-j \\ p+1+i \end{array} \right) \left(\begin{array}{c} m-S_{(j)}+j \\ p-i \end{array} \right). \qquad (5.1)$$

Proof. Choose a sub sample of size b from the X sample such that $X_{(i)}$ is the median. For a fixed i, this can be done in $\left(\begin{array}{c} i-1 \\ p \end{array} \right) \left(\begin{array}{c} m-i \\ p \end{array} \right)$ ways. Similarly choose a sub sample of size d from Ysample such that $Y_{(j)}$ is the median and is

greater than$X_{(i)}$. Each such sub-sample pair results in$\phi_1(.\,;\,.) = 1$. Using the fundamental rules of counting, we get the identity (5.1).

Identity 5.2.

$$\begin{pmatrix} m \\ b \end{pmatrix}\begin{pmatrix} n \\ d \end{pmatrix} SB_2 = \sum_{j=1}^{n}\begin{pmatrix} j-1 \\ q \end{pmatrix}\begin{pmatrix} n-j \\ q \end{pmatrix}\begin{pmatrix} S_{(j)}-j \\ b \end{pmatrix}. \tag{5.2}$$

Proof. Choose a sub sample of size d from the Y sample such that $Y_{(j)}$ is the median. For a fixed j this can be done in $\begin{pmatrix} j-1 \\ q \end{pmatrix}\begin{pmatrix} n-j \\ q \end{pmatrix}$ ways. The number of X observations less than $Y_{(j)}$ will be$\big(S_{(j)}-j\big)$. A sub sample of size b from the X observations can be chosen in $\begin{pmatrix} S_{(j)}-j \\ b \end{pmatrix}$ ways and we get identity (5.2).

Identity 5.3.

$$\begin{pmatrix} m \\ d \end{pmatrix}\begin{pmatrix} n \\ b \end{pmatrix} SB_3 = \sum_{i=1}^{m}\begin{pmatrix} i-1 \\ q \end{pmatrix}\begin{pmatrix} m-i \\ q \end{pmatrix}\begin{pmatrix} n-R_{(i)}+i \\ b \end{pmatrix}. \tag{5.3}$$

Proof. For a fixed$i$, $X_{(i)}$can be chosen as median of sub sample of size d from Xobservations in $\begin{pmatrix} i-1 \\ q \end{pmatrix}\begin{pmatrix} m-i \\ q \end{pmatrix}$ ways. The number of Yobservations greater than $X_{(i)}$ is $\big(n-R_{(i)}+i\big)$. A sub sample of size b from these Y observations can be chosen in $\begin{pmatrix} n-R_{(i)}+i \\ b \end{pmatrix}$ ways. For each i, $\begin{pmatrix} i-1 \\ q \end{pmatrix}\begin{pmatrix} m-i \\ q \end{pmatrix}\begin{pmatrix} n-R_{(i)}+i \\ b \end{pmatrix}$ sub-sample pairs for which $\phi_3(.\,;\,.) = 1$. Then by the fundamental rule of counting, we get the identity (5.3).

Identity 5.4.

$$\sum_{i=1}^{m}\sum_{j=0}^{k_2-1}\begin{pmatrix} i-1 \\ k_1-1 \end{pmatrix}\begin{pmatrix} m-i \\ b-k_1 \end{pmatrix}\begin{pmatrix} R_{(i)}-i \\ k_2-j-1 \end{pmatrix}\begin{pmatrix} n-R_{(i)}+i \\ d-k_2+j+1 \end{pmatrix}$$

$$= \sum_{j=1}^{n}\sum_{i=0}^{k_1-1}\begin{pmatrix} j-1 \\ k_2-1 \end{pmatrix}\begin{pmatrix} n-j \\ d-k_2 \end{pmatrix}\begin{pmatrix} S_{(j)}-j \\ b-k_1+i+1 \end{pmatrix}\begin{pmatrix} m-S_{(j)}+j \\ k_1-i-1 \end{pmatrix}. \tag{5.4}$$

Proof. Choose a sub sample of size b from the X sample such that $X_{(i)}$ is the k_1^{th} order statistic ($(b\beta)^{th}$quantile). For a fixed i, this can be done in $\begin{pmatrix} i-1 \\ k_1-1 \end{pmatrix}$

$\begin{pmatrix} m-i \\ b-k_1 \end{pmatrix}$ ways. Now choose a sub sample of size d from Ysample such that $Y_{(j)}$ is the k_2^{th} order statistic ($(d\gamma)^{th}$quantile) and is greater than $X_{(i)}$. Each such sub-sample pair results in $\phi_4(.\,;\,.) = 1$. The $Y_{(j)}$ can be selected in $\begin{pmatrix} R_{(i)}-i \\ k_2-j-1 \end{pmatrix}\begin{pmatrix} n-R_{(i)}+i \\ d-k_2+j+1 \end{pmatrix}$ ways. Thus using the fundamental rules of counting, we get the identity (5.4).

Acknowledgement: We thank the referee for his useful comments.

References

Baiju, K.V. and Thomas, P.Y. (2007). On some identities proved using Statistical mathematics, Journal of the Kerala Statistical Association 18, 34-44.

Bhat, S.V. (1996). Studies in Nonparametric Inference. An unpublished Ph.D thesis submitted to Karnatak University, Dharwad.

Joshi, P.C. and Balakrishnan, N. (1981). Applications of order Statistics in combinatorial identities, Journal of Combinatorics, Information and System Sciences 6, 271-278.

Randles, R. H. and Wolfe, D.A. (1979). Introduction to Theory of Nonparametric Statistics, John Wiley and Sons, New York.

Riordan, J. (1968). Combinatorial Identities, John Wiley and Sons, New York.

Shetty, I.D. and Bhat, S.V. (1993). Some Competitors of Mood's median test for location alternatives, Karnatak University, Journal of Science 37, 138-146.

Shetty, I.D. and Bhat, S.V. (1994). A Note on the Generalization of Mathisen's median test, Statistics and Probability Letters 19, 199-204.

Shetty, I.D, Sengupta, D. and Bhat, S. V (1997). A General class of two-sample Nonparametric tests based on sub-sample quantiles, Karnatak University, Journal of Science 41, 104-125.