A BIVARIATE EXTENSION OF THE U STATISTIC[1]

By D. R. Whitney
Ohio State University

1. Summary. Let x, y, and z be three random variables with continuous cumulative distribution functions f, g, and h. In order to test the hypothesis $f = g = h$ under certain alternatives two statistics U, V based on ranks are proposed.

Recurrence relations are given for determining the probability of a given (U, V) in a sample of l x's, m y's, n z's and the different moments of the joint distribution of U and V. The means, second, and fourth moments of the joint distribution are given explicitly and the limit distribution is shown to be normal.

As an illustration the joint distribution of U, V is given for the case $(l, m, n) = (6, 3, 3)$ together with the values obtained by using the bivariate normal approximation. Tables of the joint cumulative distribution of U, V have been prepared for all cases where $l + m + n \leqq 15$.

2. Introduction. Let x, y, and z be three random variables with continuous cumulative distributions f, g, h. We wish to test the hypothesis that $f = g = h$ with the alternative that $f > g, f > h$, or say, $f > g > h$.

To test such a hypothesis with a sample of l x's, m y's, and n z's, we arrange the sample values in ascending order and let U count the number of times a y precedes an x, and V count the number of times a z precedes an x. As a critical region for the hypothesis with the alternative $f > g, f > h$ we propose to use $U \leqq K_1$, $V \leqq K_2$; or with the alternative $g > f > h$, $U \geqq K_3$, $V \leqq K_4$, where the constants K_i are chosen to give the correct significance level. Even if the significance level is fixed the constants K_i are not uniquely determined. A reasonable principle to follow in this case would be to choose

$$P(U \leqq K_2) \doteq P(V \leqq K_2) \quad \text{or} \quad P(U \geqq K_3) \doteq P(V \leqq K_4)$$

according to which alternative is chosen. In particular, if $m = n$ this leads to $K_1 = K_2$ and $K_3 + K_4 = m \cdot n$.

3. Moments of the joint distribution of U and V. We consider sequences of l x's, m y's, n z's and let $T_{lmn}(U, V)$ be the number of such sequences in which a y precedes an x U times and a z precedes an x V times. Omitting the last term in such a sequence leads to the relation

$$(1) \quad T_{lmn}(U, V) = T_{l-1,mn}(U - m, V - n) + T_{l,m-1,n}(U, V) + T_{lm,n-1}(U, V),$$

[1] The U statistic was introduced by H. B. Mann and the author in "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 50–60. The present extension was carried out at the suggestion of J. W. Tukey, Princeton University.

where $T_{lmn}(U, V) = 0$ if $U < 0$; $V < 0$; $U > 0, m = 0$; or $V > 0, n = 0$; and $T_{0mn} = \binom{m+n}{m}$.

Under the hypothesis any of the $(l + m + n)!/l!m!n!$ sequences has equal probability. Hence

$$
(2) \qquad p_{lmn}(U, V) = \frac{l}{l+m+n} p_{l-1,mn}(U - m, V - n) + \frac{m}{l+m+n} p_{l,m-1,n}(U, V) + \frac{n}{l+m+n} p_{lm,n-1}(U, V),
$$

where $p_{lmn}(U, V)$ denotes the probability of a sequence of l x's, m y's, n z's having y precede an x U times and z precede an x V times.

To obtain the mean of U we multiply (2) by U and sum over all U, V. This gives

$$
(3) \qquad E_{lmn}(U) = \frac{l}{l+m+n} E_{l-1,mn}(U) + \frac{m}{l+m+n} E_{l,m-1,n}(U) + \frac{n}{l+m+n} E_{lm,n-1}(U) + \frac{lm}{l+m+n}.
$$

This and a similar equation for $E_{lmn}(V)$, together with the obvious initial conditions, give

$$
(4) \qquad E_{lmn}(U) = \frac{lm}{2}, \qquad E_{lmn}(V) = \frac{ln}{2}.
$$

The recurrence relations for the higher moments about the mean are obtained by multiplying (2) by $(U - \frac{1}{2}lm)^i(V - \frac{1}{2}ln)^j$ and summing over all U, V. Using $u = U - \frac{1}{2}lm$, $v = V - \frac{1}{2}ln$,

$$
(5) \qquad \begin{aligned} E_{lmn}(u^i v^j) &= \frac{l}{l+m+n} \sum_{\alpha=0}^{i} \sum_{\beta=0}^{j} \binom{i}{\alpha}\binom{j}{\beta}\left(\frac{m}{2}\right)^{i-\alpha}\left(\frac{n}{2}\right)^{j-\beta} E_{l-1,mn}(u^\alpha v^\beta) \\ &\quad + \frac{m}{l+m+n} \sum_{\alpha=0}^{i} \binom{i}{\alpha}(-1)^{i-\alpha}\left(\frac{l}{2}\right)^{i-\alpha} E_{l,m-1,n}(u^\alpha v^j) \\ &\quad + \frac{n}{l+m+n} \sum_{\beta=0}^{j} \binom{j}{\beta}(-1)^{j-\beta}\left(\frac{l}{2}\right)^{j-\beta} E_{lm,n-1}(u^i v^\beta). \end{aligned}
$$

For $i + j \leqq 4$ the solutions of (5) satisfying the initial conditions $E_{0mn}(u^i v^j) = E_{l00}(u^i v^j) = 0$ are found to be

$$
\begin{aligned} E_{lmn}(u^2) &= \frac{1}{12} lm(l + m + 1), \\ E_{lmn}(uv) &= \frac{1}{12} lmn, \\ E_{lmn}(u^2 v) &= E_{lmn}(uv^2) = 0, \end{aligned}
$$

$$
\begin{aligned}
E_{lmn}(u^4) &= \frac{1}{240}\, lm \\
&\quad \cdot (l + m + 1)(5l^2m + 5lm^2 - 2l^2 - 2m^2 + 3lm - 2l - 2m), \\
E_{lmn}(u^3v) &= \frac{1}{240}\, lmn(5l^2m + 5lm^2 - 2l^2 - 2m^2 + 3lm - 2l - 2m), \\
E_{lmn}(u^2v^2) &= \frac{1}{720}\, lmn \\
&\quad \cdot (5l^3 + 5l^2m + 5l^2n + 15lmn + 14l^2 \\
&\qquad + 3lm + 3ln - 6mn + 7l - 2m - 2n - 2).
\end{aligned} \tag{6}
$$

From symmetry considerations it follows that $E_{lmn}(u^iv^j) = 0$ if $i + j$ is odd.

4. Limit distribution of u and v. Let $F(l, m, n)$ be a function of integers l, m, n and define an operator Ψ by

$$
\begin{aligned}
\Psi F(l, m, n) \equiv{}& l[F(l, m, n) - F(l - 1, m, n)] \\
&+ m[F(l, m, n) - F(l, m - 1, n)] + n[F(l, m, n) - F(l, m, n - 1)].
\end{aligned} \tag{7}
$$

This permits (5) to be rewritten as

$$
\begin{aligned}
\Psi E_{lmn}(u^iv^j) ={}& l \sum_{\substack{\alpha=0 \\ \alpha+\beta<i+j}}^{i} \sum_{\beta=0}^{j} \binom{i}{\alpha}\binom{j}{\beta}\left(\frac{m}{2}\right)^{i-\alpha}\left(\frac{n}{2}\right)^{j-\beta} E_{l-1,mn}(u^\alpha v^\beta) \\
&+ m \sum_{\alpha=0}^{i-1} \binom{i}{\alpha}(-1)^{i-\alpha}\left(\frac{l}{2}\right)^{i-\alpha} E_{l,m-1,n}(u^\alpha v^j) \\
&+ n \sum_{\beta=0}^{j-1} \binom{j}{\beta}(-1)^{j-\beta}\left(\frac{l}{2}\right)^{j-\beta} E_{lm,n-1}(u^iv^\beta).
\end{aligned} \tag{8}
$$

In order to work with equation (8) we need these properties:

(a) If $\Psi F(l, m, n)$ is a polynomial of degree t in all the variables, α in l, β in m, γ in n, then $F(l, m, n)$ is a polynomial of degree t in all the variables, α in l, β in m, γ in n.

(b) If P_t, Q_t are polynomials of degree t and $l, m, n \to \infty$ so that $F(l, m, n) \to F_0$ and $\dfrac{\Psi P_t}{Q_t} \to c$, then $\dfrac{P_t}{Q_t} \to \dfrac{c}{t}$.

Leaving the proof of these statements to a later section (Section 5), we shall apply them now to equation (8). Since $E_{lmn}(u^iv^j) = 0$ for $i + j$ odd we consider only the case $i + j = 2r$. For $r = 1, 2$, $E_{lmn}(u^iv^j)$ is a polynomial of degree $3r$, of degree at most $2r$ in l, i in m, and j in n. If we assume this to hold for $i + j < 2r$, then from (8) $\Psi E_{lmn}(u^iv^j)$, $i + j = 2r$, has these properties and hence $E_{lmn}(u^iv^j)$ does also.

In what follows there are two cases according as i and j are both even or both odd. We give only the first case explicitly. Replacing i and j in (8) by $2i$ and $2j$, we obtain

$$\begin{aligned}\Psi E_{lmn}(u^{2i}v^{2j}) &= l\Bigg[\binom{2i}{2i-2}\left(\frac{m}{2}\right)^2 E_{l-1,mn}(u^{2i-2}v^{2j}) \\ &\quad+\binom{2i}{2i-1}\binom{2j}{2j-1}\left(\frac{m}{2}\right)\left(\frac{n}{2}\right) E_{l-1,mn}(u^{2i-1}v^{2j-1}) \\ &\quad+\binom{2j}{2j-2}\left(\frac{n}{2}\right)^2 E_{l-1,mn}(u^{2i}v^{2j-2})\Bigg] \\ &\quad+m\left[\binom{2i}{2i-2}\left(\frac{l}{2}\right)^2 E_{l,m-1,n}(u^{2i-2}v^{2j})\right] \\ &\quad+n\left[\binom{2j}{2j-2}\left(\frac{l}{2}\right)^2 E_{lm,n-1}(u^{2i}v^{2j-2})\right]+P_{3(i+j)-1}(l,m,n),\end{aligned} \tag{9}$$

where $P_{3(i+j)-1}(l, m, n)$ indicates a polynomial of degree $3(i + j) - 1$ in l, m, n, which is also of degree at most $2(i + j)$ in l, $2i$ in m, and $2j$ in n. This may be reduced to

$$\begin{aligned}\Psi E_{lmn}(u^{2i}v^{2j}) &= \tfrac{1}{4}lm(l+m)i(2i-1)E_{lmn}(u^{2i-2}v^{2j}) \\ &\quad+lmn\cdot i\cdot jE_{lmn}(u^{2i-1}v^{2j-1})+\tfrac{1}{4}ln(l+n)j(2j-1)E_{lmn}(u^{2i}v^{2j-2}) \\ &\quad+P_{3(i+j)-1}(l,m,n).\end{aligned} \tag{10}$$

Now we write
$$\lambda_{lmn}^{\alpha\beta} \equiv \frac{E_{lmn}(u^\alpha v^\beta)}{\sigma_u^\alpha \sigma_v^\beta};$$

then dividing (10) by $\sigma_u^{2i}\sigma_v^{2j} = [\frac{1}{12}lm(l+m+1)]^i[\frac{1}{12}ln(l+n+1)]^j$ gives

$$\begin{aligned}\frac{\Psi E_{lmn}(u^{2i}v^{2j})}{\sigma_u^{2i}\sigma_v^{2j}} &= \frac{\frac{1}{4}lm(l+m)i(2i-1)}{\frac{1}{12}lm(l+m+1)}\lambda_{lmn}^{2i-2,2j} \\ &\quad+\frac{lmn\cdot i\cdot j}{\frac{1}{12}\sqrt{lm(l+m+1)ln(l+n+1)}}\lambda_{lmn}^{2i-1,2j-1} \\ &\quad+\frac{\frac{1}{4}ln(l+n)j(2j-1)}{\frac{1}{12}ln(l+n+1)}\lambda_{lmn}^{2i,2j-2}+\frac{P_{3(i+j)-1}(l,m,n)}{\sigma_u^{2i}\sigma_v^{2j}}.\end{aligned} \tag{11}$$

Let

$$\rho(l,m,n) \equiv \frac{E_{lmn}(u,v)}{\sigma_u\sigma_v} = \sqrt{\frac{mn}{(l+m+1)(l+n+1)}}$$

and use $\rho(l, m, n) \to \rho_0$ to mean $l, m, n \to \infty$ in such a way that

$$\sqrt{\frac{mn}{(l+m+1)(l+n+1)}} \to \rho_0.$$

We then have for $\rho(l, m, n) \to \rho_0$

$$\lambda_{lmn}^{11} \to \rho_0,\ \lambda_{l,m,n}^{40} \to 3,\ \lambda_{lmn}^{31} \to 3\rho_0,\ \lambda_{lmn}^{22} \to 1+2\rho_0^2. \tag{12}$$

Dropping the l, m, n to denote the limiting values, (12) are just special cases i.e., $i + j = 2$ or 4, of

$$\lambda^{2i,2j} = \frac{(2i)!(2j)!}{2^{i+j}} \sum_{\alpha=0}^{\min(i,j)} \frac{(2\rho_0)^{2\alpha}}{(i-\alpha)!(j-\alpha)!(2\alpha)!}, \tag{13}$$

$$\lambda^{2i+1,2j+1} = \frac{(2i+1)!(2j+1)!}{2^{i+j}} \rho_0 \sum_{\alpha=0}^{\min(i,j)} \frac{(2\rho_0)^{2\alpha}}{(i-\alpha)!(j-\alpha)!(2\alpha+1)!}.$$

Inductively then, we assume for $\alpha + \beta < 2(i + j)$ that $\lambda_{lmn}^{\alpha\beta}$ satisfies (13) for $\rho(l, m, n) \to \rho_0$. Since $P_{3(i+j)-1}(l, m, n)$ in (11) has degree at most $2(i + j)$ in l, $2i$ in m, $2j$ in n, we obtain

$$\lim_{\rho(l,m,n)\to\rho_0} \frac{\Psi E_{lmn}(u^{2i}v^{2j})}{\sigma_u^{2i}\sigma_v^{2j}}$$

$$= 3i(2i-1)\frac{(2i-2)!(2j)!}{2^{i+j-1}} \sum_{\alpha=0}^{\min(i-1,j)} \frac{(2\rho_0)^{2\alpha}}{(i-1-\alpha)!(j-\alpha)!(2\alpha)!} \tag{14}$$

$$+ 12i\cdot j\rho_0 \frac{(2i-1)!(2j-1)!}{2^{i+j-2}} \rho_0 \sum_{\alpha=0}^{\min(i-1,j-1)} \frac{(2\rho_0)^{2\alpha}}{(i-1-\alpha)!(j-1-\alpha)!(2\alpha+1)!}$$

$$+ 3j(2j-1)\frac{(2i)!(2j-2)!}{2^{i+j-1}} \sum_{\alpha=0}^{\min(i,j-1)} \frac{(2\rho_0)^{2\alpha}}{(i-\alpha)!(j-1-\alpha)!(2\alpha)!},$$

and this reduces to

$$\lim_{\rho(l,m,n)\to\rho_0} \frac{\Psi E_{lmn}(u^{2i}v^{2j})}{\sigma_u^{2i}\sigma_v^{2i}} = 3(i+j)\frac{(2i)!(2j)!}{2^{i+j}} \sum_{\alpha=0}^{\min(i,j)} \frac{(2\rho_0)^{2\alpha}}{(i-\alpha)!(j-\alpha)!(2\alpha)!}. \tag{15}$$

From this it follows that

$$\lim_{\rho(l,m,n)\to\rho_0} \frac{E_{lmn}(u^{2i}v^{2j})}{\sigma_u^{2i}\sigma_v^{2j}} = \frac{(2i)!(2j)!}{2^{i+j}} \sum_{\alpha=0}^{\min(i,j)} \frac{(2\rho_0)^{2\alpha}}{(i-\alpha)!(j-\alpha)!(2\alpha)!}, \tag{16}$$

and in like manner for $E_{lmn}(u^{2i+1}v^{2j+1})$. Therefore the moments of the limit distribution are those of a bivariate normal distribution. Hence the variables

$$\frac{U - \frac{lm}{2}}{\sqrt{\frac{1}{12}lm(l+m+1)}}, \qquad \frac{V - \frac{ln}{2}}{\sqrt{\frac{1}{12}ln(l+n+1)}}$$

have in the limit a joint normal distribution with means 0, variances 1, and correlation coefficient ρ_0, where

$$\rho_0 = \lim_{l,m,n\to\infty} \sqrt{\frac{mn}{(l+m+1)(l+n+1)}}.$$

5. Properties of Ψ.

Lemma 1. *If*

$$F(x, y, z) = \sum_{i=0}^{\lambda}\sum_{j=0}^{\mu}\sum_{k=0}^{\nu} a_{ijk}x^i y^j z^k,$$

then

$$\Psi F(x, y, z) = \sum_{i=0}^{\lambda} \sum_{j=0}^{\mu} \sum_{k=0}^{\nu} A_{ijk} x^i y^j z^k,$$

where

$$A_{ijk} = \sum_{\alpha=i}^{\lambda} (-1)^{\alpha-i} a_{\alpha jk} \binom{\alpha}{i-1} + \sum_{\beta=j}^{\mu} (-1)^{\beta-j} a_{i\beta k} \binom{\beta}{j-1} + \sum_{\gamma=k}^{\nu} (-1)^{\gamma-k} a_{ij\gamma} \binom{\gamma}{k-1}.$$

This follows from a straightforward application of the definition of Ψ.

LEMMA 2. *If $F(x, y, z)$ is a polynomial in x, y, z of degree σ, of degree λ in x, μ in y, ν in z, then so is ΨF.*

This follows from the representation of ΨF in Lemma 1.

LEMMA 3. *For any polynomial $F(x, y, z)$ there exists a polynomial $G(x, y, z)$ such that $\Psi G = F$.*

Let the coefficients of F be denoted by A_{ijk} and the unknown coefficients of G by a_{ijk}. The lemma will follow if we can solve the equations in Lemma 1 for a_{ijk}, $i = 0, 1, \cdots, \lambda; j = 0, 1, \cdots, \mu; k = 0, 1, \cdots, \nu$. For $i + j + k$ a maximum for all the i, j, k of A_{ijk}, we have

$$A_{ijk} = a_{ijk} \binom{i}{i-1} + a_{ijk} \binom{j}{j-1} + a_{ijk} \binom{k}{k-1} = (i + j + k)a_{ijk}.$$

By induction we assume that the equation can be solved for the a_{ijk} for all i, j, k such that $i + j + k > t$. Then for $i + j + k = t$ we have $A_{ijk} = (i + j + k)a_{ijk}$ plus a's whose subscripts add to more than t. Hence the a_{ijk} can be determined.

LEMMA 4. *If $\Psi[F(x, y, z) - G(x, y, z)] = 0$, then $F(x, y, z) - F(0, 0, 0) = G(x, y, z) - G(0, 0, 0)$.*

Let $t = x + y + z$. The lemma is true for $t = 0$, and we assume it to be true for all x, y, z such that $x + y + z < t$. Then for $x + y + z = t$,

$$\Psi[F(x, y, z) - G(x, y, z)] = 0$$

gives

$$(x + y + z)[F(x, y, z) - G(x, y, z)] - x[F(x - 1, y, z) - G(x - 1, y, z)] - y[F(x, y - 1, z) - G(x, y - 1, z)] - Z[F(x, y, z - 1) - G(x, y, z - 1)] = 0.$$

Using our induction assumption,

$$(x + y + z)[F(x, y, z) - G(x, y, z)] = (x + y + z)[F(0, 0, 0) - G(0, 0, 0)],$$

and the lemma follows.

6. Distribution of u and v in a particular case with $l = 6, m = n = 3$ Using. the relation (1), the table of $T_{633}(U, V)$ (Table 1) was obtained. In this case $E(U) = E(V) = 9$, $\sigma_u^2 = \sigma_v^2 = 15$, $\sigma_{uv} = 4.5$, $\rho = 0.3$.

TABLE 1

$T_{633}(U, V)$

V \ U	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
9	14	15	32	55	78	103	150	155	178	200	178	155	150	103	78	55	32	15	14
8	17	18	41	65	91	112	158	160	194	178	173	144	139	95	71	46	30	14	14
7	16	20	42	66	85	108	146	158	160	155	144	122	110	74	52	38	23	11	10
6	25	24	48	71	95	114	170	146	158	150	139	110	103	64	49	33	21	10	10
5	20	20	39	58	75	98	114	108	112	103	95	74	64	43	31	22	13	6	5
4	19	18	37	51	74	75	95	85	91	78	71	52	49	31	23	14	9	4	4
3	15	16	32	56	51	58	71	66	65	55	46	38	33	22	14	10	6	3	3
2	14	15	30	32	37	39	48	42	41	32	30	23	21	13	9	6	4	2	2
1	10	12	15	16	18	20	24	20	18	15	14	11	10	6	4	3	2	1	1
0	20	10	14	15	19	20	25	16	17	14	14	10	10	5	4	3	2	1	1

TABLE 2

$$\sum_{V=0}^{k} \sum_{U=0}^{h} p_{633}(U, V)$$

k \ h	0	1	2	3	4	5	6	7	8	9
9	9 (12)	18 (21)	36 (37)	62 (59)	96 (91)	137 (132)	191 (181)	242 (237)	298 (295)	351 (352)
8	8 (10)	17 (19)	33 (33)	56 (52)	86 (80)	120 (114)	166 (156)	210 (201)	256 (249)	298 (295)
7	7 (9)	15 (16)	29 (28)	48 (45)	73 (67)	102 (95)	139 (128)	174 (165)	210 (201)	242 (237)
6	7 (8)	13 (14)	24 (23)	41 (36)	61 (54)	84 (76)	113 (101)	139 (128)	166 (156)	191 (181)
5	5 (6)	10 (11)	19 (18)	32 (28)	46 (41)	63 (57)	84 (76)	102 (95)	120 (114)	137 (132)
4	4 (5)	8 (8)	15 (13)	24 (21)	35 (30)	46 (41)	61 (54)	73 (67)	86 (80)	96 (91)
3	3 (3)	6 (6)	11 (10)	17 (14)	24 (21)	32 (28)	41 (36)	48 (45)	56 (52)	62 (59)
2	2 (2)	4 (4)	8 (6)	11 (10)	15 (13)	19 (18)	24 (23)	29 (28)	33 (33)	36 (37)
1	2 (1)	3 (3)	4 (4)	6 (6)	8 (8)	10 (11)	13 (14)	15 (16)	17 (19)	18 (21)
0	1 (1)	2 (1)	2 (2)	3 (3)	4 (5)	5 (6)	7 (8)	7 (9)	8 (10)	9 (12)

Table 2 gives the cumulative distribution $\sum_{V=0}^{k} \sum_{U=0}^{h} p_{633}(U, V)$. The numbers have all been multiplied by 1000. The figures in parentheses are the values obtained by using $(U - 9 - \frac{1}{2})/\sqrt{15}$, $(V - 9 - \frac{1}{2})/\sqrt{15}$ as random variables from a bivariate normal distribution of means zero, variances one, and correlation coefficient 0.3.

7. Example. Suppose that y, x, z denote the lengths of life of rats that have been exposed to insecticides of supposedly decreasing toxicity. We would then be interested in the hypothesis $g = f = h$ under the alternative $g > f > h$.

For a sample of 3 y's, 6 x's, and 3 z's, a critical region of size .044 is found from the preceding table to be $U \geqq 12$, $V \leqq 6$. In an experiment the sequence

TABLE 2—*Continued*

$$\sum_{V=0}^{k} \sum_{U=0}^{h} p_{633}(U, V)$$

k \ h	10	11	12	13	14	15	16	17	18
9	400	440	478	502	520	533	541	544	548
	(404)	(447)	(482)	(508)	(526)	(537)	(544)	(548)	(551)
8	338	369	398	418	431	441	447	450	452
	(336)	(370)	(398)	(417)	(430)	(439)	(444)	(446)	(449)
7	272	296	318	332	342	349	353	355	357
	(268)	(293)	(313)	(327)	(337)	(345)	(346)	(348)	(349)
6	213	230	246	256	263	268	271	272	274
	(204)	(222)	(236)	(245)	(252)	(255)	(258)	(259)	(260)
5	151	162	173	179	184	187	189	190	190
	(147)	(159)	(168)	(174)	(178)	(180)	(182)	(183)	(183)
4	106	113	119	124	127	129	130	130	131
	(101)	(108)	(114)	(118)	(120)	(121)	(122)	(123)	(123)
3	68	72	76	79	81	82	83	83	83
	(65)	(70)	(73)	(75)	(76)	(77)	(78)	(78)	(78)
2	39	42	44	45	46	47	47	47	48
	(40)	(43)	(44)	(45)	(46)	(47)	(47)	(47)	(47)
1	20	21	22	23	23	23	24	24	24
	(23)	(24)	(25)	(26)	(26)	(26)	(26)	(27)	(27)
0	10	10	11	11	12	12	12	12	12
	(13)	(13)	(13)	(14)	(14)	(14)	(14)	(14)	(14)

yyxxxyxzxzzx was obtained. For this sample $U = 15$, $V = 4$, and consequently we presume the toxic effects to be as supposed.

For a sample of 7 y's, 6 x's, 8 z's, we first compute

$$E(U) = 21, \qquad E(V) = 28, \qquad \sigma_U^2 = 49, \qquad \sigma_V^2 = 60, \qquad \rho^2 = 4/15.$$

The critical region can be written as

$$\frac{U - E(U)}{\sigma_U} \geqq h, \qquad \frac{V - E(V)}{\sigma_V} \leqq -k,$$

where h, k are to be determined to give a significance level of 5%, say, and subject to

$$P\left(\frac{U - E(U)}{\sigma_U} \geqq h\right) \doteq P\left(\frac{V - E(V)}{\sigma_V} \leqq -k\right).$$

With the normal approximation to the distribution of U or V the last condition implies $h = k$. Then entering Pearson's table for the normal bivariate distribution with $\rho = -.52$ (interpolating between $-.50$ and $-.55$) we find that $h = k \doteq .37$ are the desired values. This gives a 5% critical region of $U \geqq 24$, $V \leqq 25$.