NEEDED: An efficient algorithm for computing exact tail probabilities of randomization tests for treatment effects in a randomized block experiment with binary or trinary data.

by

BU-727-M December, 1980

D. S. Robson

ABSTRACT

Binary data in a randomized complete blocks design (RCBD) present an hypothesis testing problem which is amenable to exact randomization test procedures. The astronomical numbers of permutations usually required to compute exact tail probabilities for randomization tests become reduced to manageable numbers of combinations rather than permutations when the response variable can take only the values 0 or 1. Treatment totals in this case reduce to a count of the number of blocks in which the treatment in question scored a "success", and a test statistic such as the treatment sum of squares (or Cochran's Q), which is a symmetric function of the treatment totals, becomes a function of the frequency vector $\underset{\sim}{f}_r = \left(f_r(0),\ f_r(1),\ \cdots,\ f_r(r)\right)$ counting the number of treatments scoring, respectively, 0, 1, $\cdots$ and r "successes" in the r blocks of the design. An algorithm is presented for calculating the conditional permutation distribution of the frequency vector $\underset{\sim}{f}_k$ for the first k blocks, given the vector $\underset{\sim}{f}_{k-1}$, where $\underset{\sim}{f}_1 = \left(f_1(0),\ f_1(1)\right)$ is the given number of failures and successes, respectively, in the first block. Repeated application of

$$P_{H_o}(\underset{\sim}{f}_k) = \sum_{\underset{\sim}{f}_{k-1}} P_{H_o}(\underset{\sim}{f}_{k-1}) P_{H_o}(\underset{\sim}{f}_k | \underset{\sim}{f}_{k-1})$$

produces the H_o-distribution of $\underset{\sim}{f}_r$, and hence the induced H_o-distribution of any specific test statistic which is a function of $\underset{\sim}{f}_r$. This approach, with illustrative

Paper No. BU-727-M in the Biometrics Unit.

examples, is extended to the case of trinary data and the H_o-distribution of a bivariate, triangular array of frequencies $h_r(x,y,z)$ with $x + y + z = r$. The purpose of this report is to provide the basis for development of an efficient computer program to implement exact randomization tests for binary and trinary data in a RCBD.

INTRODUCTION

A randomized complete blocks design (RCBD) in which $\underline{t}$ treatments are independently randomized among the $\underline{t}$ experimental units in each of $\underline{r}$ blocks (replicates) produces an $r \times t$ response matrix $[Y_{ij}]$. Under the null hypothesis (H_o) of no treatment effect there is no statistical association between treatment labels (j) and response (Y_{ij}) within a block (i), and the realized outcome (Y_{i1}, Y_{i2}, $\cdots$, Y_{it}) is regarded as a random (by design) choice of one of the t! (factorial t) H_o-equally likely permutations of this vector. The $\underline{r}$ such choices — one in each block — are (also by design) statistically independent. The H_o-distribution of $[Y_{ij}]$ which assigns equal probability $(t!)^{-r}$ to each $r \times t$ array obtainable from within-row permutations of the realized array is called the <u>permutation-</u> or <u>randomization-distribution</u> generated by the RCBD.

Any statistic T which is a function of the matrix $[Y_{ij}]$ has an induced probability distribution, also called its randomization distribution, which may in principle be completely enumerated by listing the $(t!)^r$ permutations of $[Y_{ij}]$ and calculating the function T for each. If the realized value of T is extreme in the sense that it falls out in the tail of this randomization distribution then the null hypothesis might be rejected.

The exact tail probability would be available from such a complete listing of the $(t!)^r$ outcomes; in practice, however, this number $(t!)^r$ is ordinarily astronomical in magnitude so this approach to exact testing is not feasible.

Approximations to the RCBD randomization distribution of certain specific test statistics T are available, however, for approximate testing. An example is the chi-square approximation to the randomization distribution of the normalized treatment sum of squares

$$Q = \frac{r \sum_{j=1}^{t} (\bar{Y}_{.j} - \bar{Y}_{..})^2}{\sum_{i=1}^{r} \sum_{j=1}^{t} (Y_{ij} - \bar{Y}_{i.})^2 / r(t-1)} = \frac{t-1}{t} \frac{r\Sigma(\bar{Y}_{.j} - \bar{Y}_{..})^2}{\bar{\sigma}^2}$$

where

$$\sigma_i^2 = \frac{1}{t} \sum_{j=1}^{t} (Y_{ij} - \bar{Y}_{i.})^2$$

is the variance of the j'th entry in block i under the permutation distribution, $-\sigma_i^2/(t-1)$ is the covariance of the j'th and k'th entries in this block and $t\bar{\sigma}^2$ is the expectation of the treatment sum of squares, where

$$\bar{\sigma}^2 = \frac{1}{r} \sum_{i=1}^{r} \sigma_i^2 .$$

The H_o expectation of Q is thus the treatment degrees of freedom, t - 1, and the approximating H_o-distribution is chi-square on t - 1 D.F. [The H_o-variance of Q, however, may be shown to be

$$V_{H_o}(Q) = 2(t-1)\left[1 - \frac{\Sigma\sigma_i^4}{(\Sigma\sigma_i^2)^2}\right]$$

which is somewhat smaller than the chi-square variance 2(t - 1).]

BINARY CASE

Though exact RCBD-randomization distributions are generally intractable, there are important special cases that are of manageable proportions, the simplest of which is the case of a binary response variable where Y_{ij} is either 0 or 1. A treatment total $Y_{\cdot j}$ in this case is simply a count of the number of blocks in which this treatment produced a "success" response. An H_o-sufficient description of the vector $(Y_{\cdot 1}, \cdots, Y_{\cdot t})$ of treatment totals is the frequency distribution

$$f_r(x) \underset{\text{def.}}{\equiv} \#\{j \mid Y_{\cdot j} = x\} \qquad x = 0, 1, \cdots, r$$

or the cumulative frequencies

$$F_r(x) = f_r(0) + f_r(1) + \cdots + f_r(x) .$$

Equivalently, this description could be given by the relative frequency distribution of treatment totals

$$p_r(x) = f_r(x)/t$$

$$P_r(x) = p_r(0) + p_r(1) + \cdots + p_r(x) .$$

A test statistic which is a symmetric function of the treatment totals is then a function of the vector $\left(f_r(0), f_r(1), \cdots, f_r(r)\right)$, so the randomization distribution of any such function is induced by the randomization distribution of this frequency vector. An algorithm for computing the randomization distribution of $\underset{\sim}{f}_r$ would thus permit calculation of the exact significance level of any such test statistic.

An example, again, is the (non-normalized)* treatment sum of squares

---

* Normalization is superfluous if the exact distribution is to be computed.

$$S_2 = \frac{1}{r}\left[\sum_j Y^2_{\cdot j} - rt\bar{Y}^2_{\cdot\cdot}\right]$$

$$= r\left[(0 - \bar{Y}_{\cdot\cdot})^2 f_r(0) + (\tfrac{1}{r} - \bar{Y}_{\cdot\cdot})^2 f_r(1) + \cdots + (\tfrac{r}{r} - \bar{Y}_{\cdot\cdot})^2 f_r(r)\right]$$

$$= rt \sum_{x=0}^{r} (\tfrac{x}{r} - \bar{Y}_{\cdot\cdot})^2 p_r(x)$$

$$= rt\left[2 \sum_{x=0}^{r} \tfrac{x}{r}\Big(1 - P_r(x)\Big) + \bar{Y}_{\cdot\cdot}(1 - \bar{Y}_{\cdot\cdot})\right]$$

which is seen to be a linear function of the frequencies. [S_2 is quadratic in the grand mean $\bar{Y}_{\cdot\cdot}$, but the mean is permutation-invariant and, like the block means $\bar{Y}_{i\cdot}$ and variances $\sigma_i^2 = \bar{Y}_{i\cdot}(1 - \bar{Y}_{i\cdot})$, is a known parameter of the RCBD randomization distribution.] Another example is the test statistic

$$S_1 = \sum_{x=0}^{r} \left|\frac{x}{r} - \bar{Y}_{\cdot\cdot}\right| p_r(x)$$

which has been examined in a simulation study by T. Berggren and, in general, we might consider other linear functions

$$T = \sum_{x=0}^{r} c_x f_r(x) ,$$

where $\underset{\sim}{c}$ is a predetermined vector of constants.

BINARY ALGORITHM

The permutation distribution of $\underset{\sim}{f}_r$ may be computed recursively by exploiting the readily computable multihypergeometric form of the conditional distribution of entries in the 2 X k table:

number in k'th	number of successes in first k - 1 blocks: 0	1	...	k - 1	Total
0	$F_k(0)$	$F_k(1) - F_{k-1}(0)$	...	$F_k(k-1) - F_{k-1}(k-2)$	$t - Y_{k.}$
1	$F_{k-1}(0) - F_k(0)$	$F_{k-1}(1) - F_k(1)$		$F_{k-1}(k-1) - F_k(k-1)$	$Y_{k.}$
Total	$F_{k-1}(0)$	$F_{k-1}(1) - F_{k-1}(0)$	...	$F_{k-1}(k-1) - F_{k-1}(k-2)$	t

Conditional on both the row totals (number of failures and successes in the k'th block) and the column totals $\left(f_{k-1}(x)\right)$ the distribution generated by randomly selecting the $Y_{k.}$ treatments to be assigned the score "1" in the k'th block is the multihypergeometric:

$$P_{H_o}(\underset{\sim}{F}_k \mid \underset{\sim}{F}_{k-1}) = \frac{\prod_{x=0}^{k-1} \binom{F_{k-1}(x) - F_{k-1}(x-1)}{F_k(x) - F_{k-1}(x-1)}}{\binom{t}{Y_{k.}}} .$$

Multiplying by $P_{H_o}(\underset{\sim}{F}_{k-1})$ and summing over $\underset{\sim}{F}_{k-1}$ gives $P_{H_o}(\underset{\sim}{F}_k)$.

BINARY EXAMPLE

Numerical example with $r = 3$ and $t = 10$:

		1	2	3	4	5	6	7	8	9	10	Total = $Y_{i.}$
	1	0	0	0	1	0	0	1	0	0	0	2
$[Y_{ij}]$ =	2	0	1	0	0	0	0	1	0	1	0	3
	3	1	0	1	0	1	1	0	0	1	0	5
$Y_{.j}$ =	Total	1	1	1	1	1	1	2	0	2	0	10

Derivation of $P_{H_o}(\underset{\sim}{f}_2)$:

2^{nd} \ 1^{st}	0	1	
0	$F_2(0)$	$F_2(1) - F_1(0)$	7
1	$F_1(0) - F_2(0)$	$F_1(1) - F_2(1)$	3
	$f_1(0) = 8$	$f_1(1) = 2$	10

$$P = \frac{\binom{8}{F_2(0)}\binom{2}{F_2(1) - F_1(0)}}{\binom{10}{7}}$$

NOTE:

$f_2(0) = F_2(0)$, $f_2(1) = [F_1(0) - F_2(0)] + [F_2(1) - F_1(0)]$, $f_2(2) = t - f_2(0) - f_2(1)$.

$f_2(0)$	$f_2(1)$	$f_2(2)$	$P = P_{H_o}(\underset{\sim}{f}_2 \mid \underset{\sim}{f}_1) = P_{H_o}(\underset{\sim}{f}_2)$
5	5	0	$56/120 = 7/15$
6	3	1	$56/120 = 7/15$
7	1	2	$8/120 = 1/15$

Derivation of $P_{H_o}(\underset{\sim}{f}_3 \mid \underset{\sim}{f}_2)$ for each $\underset{\sim}{f}_2$:

3^{rd}	First 2: 0	1	2	
0	$F_3(0)$	$F_3(1) - F_2(0)$	$F_3(2) - F_2(1)$	5
1	$F_2(0) - F_3(0)$	$F_2(1) - F_3(1)$	$f_3(3)$	5
	$f_2(0)$	$f_2(1)$	$f_2(2)$	

$$P = \frac{\binom{f_2(0)}{f_3(0)}\binom{f_2(1)}{F_3(1) - F_2(0)}\binom{f_2(2)}{f_3(3)}}{\binom{10}{5}}$$

$x =$	0	1	2	
				5
				5
$f_2(x) =$	5	5	0	

$$P = \frac{\binom{5}{f_3(0)}\binom{5}{F_3(1) - 5}\binom{0}{0}}{252} \cdot \frac{7}{15}$$

$f_3(0)$	$f_3(1)$	$f_3(2)$	$f_3(0)$	$P = P_{H_o}(\underset{\sim}{f}_3 \mid \underset{\sim}{f}_2) \cdot P_{H_o}(\underset{\sim}{f}_2)$
0	10	0	0	(1/252)(7/15)
1	8	1	0	(25/252)(")
2	6	2	0	(100/252)(")
3	4	3	0	(100/252)(")
4	2	4	0	(25/252)(")
5	0	5	0	(1/252)(")

$x =$	0	1	2	
				5
				5
$f_2(x) =$	6	3	1	

$$P = \frac{\binom{6}{f_3(0)}\binom{3}{F_3(1) - 6}\binom{1}{f_3(3)}}{252} \cdot \frac{7}{15}$$

$f_3(0)$	$f_3(1)$	$f_3(2)$	$f_3(3)$	P
1	8	1	0	(6/252)(7/15)
2	6	2	0	(45/252)(")
2	7	0	1	(15/252)(")
3	5	1	1	(60/252)(")
3	4	3	0	(60/252)(")
4	2	4	0	(15/252)(")
4	3	2	1	(45/252)(")
5	1	3	1	(6/252)(")

$x =$	0	1	2	
				5
				5
$f_2(x) =$	7	1	2	

$$P = \frac{\binom{7}{f_3(0)}\binom{1}{F_3(1) - 7}\binom{2}{f_3(3)}}{252} \cdot \frac{1}{15}$$

$f_3(0)$	$f_3(1)$	$f_3(2)$	$f_3(3)$	P
2	6	2	0	(21/252)(1/15)
3	4	3	0	(35/252)(")
3	5	1	1	(70/252)(")
4	3	2	1	(70/252)(")

					T	
Summing to get $P_{H_o}(\underset{\sim}{f}_3)$:				$3780\ P_{H_o}(\underset{\sim}{f}_3)$	$c_x = \lvert x - r\bar{Y}_{..}\rvert$	$c_x = (x - r\bar{Y}_{..})^2$
0	10	0	0	0	0	0
1	8	1	0	217	2	2
2	6	2	0	1036	4	4
3	4	3	0	1155	6	6
4	2	4	0	280	8	8
5	0	5	0	7	10	10
2	7	0	1	105	4	6
3	5	1	1	490	6	8
4	3	2	1	385	8	10
5	1	3	1	42	10	12
4	4	0	2	35	8	12
5	2	1	2	21	10	14
				3780		

Induced distributions:

$T_1 = \sum_x \lvert x - r\bar{Y}_{..} \rvert f_r(x)$	$P_{H_o}(T_1)$	$T_2 = \sum_x (x - r\bar{Y}_{..})^2 f_r(x)$	$P_{H_o}(T_2)$
0	7	0	7
2	217	2	217
$\boxed{4}$	$\boxed{1141}$	$\boxed{4}$	$\boxed{1036}$
6	1645	6	1260
8	700	8	770
10	70	10	392
	3780	12	77
		14	21
			3780

The boxed case is the realized outcome in the illustrative data set. The reader may check that when T_2 is normalized to Q, the mean and variance of this induced distribution agree with the formulas given earlier for $E_{H_o}(Q)$ and $V_{H_o}(Q)$.

TRINARY CASE

Considerably more extensive computations are required to obtain the RCBD permutation distribution if treatment response can fall into any one of three different categories. If r and t are not too large, however, these computations may still be feasible; the calculations are trivial, for example, in the case illustrated below with $r = 3$ and $t = 5$. A greater variety of test statistics might be employed in this situation since the summary statistic for each treatment is now 2-dimensional, consisting of a count of the number of responses in each of the 3 categories (but with the total count being fixed at r). An H_o-sufficient description of these treatment responses is now a bivariate frequency distribution supported on $(r+1)(r+2)/2$ lattice points in the plane $x + y + z = r$ where x, y and z are the number of blocks in which the treatment response fell in the first, second and third category, respectively. Thus, if

$X_{\cdot j}$ = number of blocks in which treatment j responded "A"

$Y_{\cdot j}$ = ______ " ______ "B"

$Z_{\cdot j}$ = ______ " ______ "C"

then the frequency distribution is

$$h_r(x,y,z) = \#\{j \mid X_{\cdot j} = x,\ Y_{\cdot j} = y,\ Z_{\cdot j} = z = r - x - y\} .$$

In order to provide the statistician with flexibility in the choice of a test statistic, it would be desirable to generate the permutation distribution of the triangular array $\underset{\sim}{h}_r$, from which the induced distribution of a test statistic may be obtained.

An important special case of a trinary response arises when A = 0, B = 1 and C = 2; i.e., when a treatment response within a block is either 0, 1 or 2 "successes". The treatment sum of squares is again a reasonable test statistic in this case, and its induced distribution will be illustrated in the numerical example.

TRINARY ALGORITHM

A recursive construction of $P(\underset{\sim}{h}_r)$ may again be used, now exploiting the simple form of the conditional distribution in a 3 X n contingency table with fixed row totals ($X_{k\cdot}$, $Y_{k\cdot}$, $Z_{k\cdot} = t - X_{k\cdot} - Y_{k\cdot}$) and with column totals being the $n \le t$ non-zero components of $\underset{\sim}{h}_{k-1}$. Since $\underset{\sim}{h}_k$ is a partition of the integer t there can be at most t non-zero components of $\underset{\sim}{h}_k$ for any k, $1 \le k \le r$, and since $\underset{\sim}{h}_{k-1}$ sums to $t = X_{k\cdot} + Y_{k\cdot} + Z_{k\cdot}$ the conditional probability of the 3 X n table entries $X_k(x,y,z)$, $Y_k(x,y,z)$ and $Z_k(x,y,z)$ becomes

$$P(\underset{\sim}{X}_k, \underset{\sim}{Y}_k, \underset{\sim}{Z}_k \mid \underset{\sim}{h}_{k-1}) = \frac{X_{k\cdot}!\,Y_{k\cdot}!\,Z_{k\cdot}!}{t!} \prod_{x,y,z} \frac{h_{k-1}(x,y,z)!}{X_k(x,y,z)!\,Y_k(x,y,z)!\,Z_k(x,y,z)!} .$$

The frequency distribution $\underset{\sim}{h}_k$ may be compiled from this 3 X n table using the relation

$$h_k(x,y,z) = X_k(x-1,y,z)h_{k-1}(x-1,y,z) + Y_k(x,y-1,z)h_{k-1}(x,y-1,z)$$
$$+ Z_k(x,y,z-1)h_{k-1}(x,y,z-1)$$

where $X_k(x-1,y,z)$, for example, is the number of treatments having the outcome history $(x-1,y,z)$ in the first $k-1$ blocks $(x-1+y+z = k-1)$ and having the outcome "A" in the k'th block. This table thus uniquely determines $\underset{\sim}{h}_k$, so the probability calculated above uniquely determines $P(\underset{\sim}{h}_k | \underset{\sim}{h}_{k-1})$.

<u>NUMERICAL ILLUSTRATION</u>

	Treatment					Total		
Block	1	2	3	4	5	#A	#B	#C
1	B	A	C	B	C	1	2	2
2	A	A	A	C	A	4	0	1
3	A	B	B	C	B	1	3	1
Tr. Total								
#A	2	2	1	0	1	6		
#B	1	1	1	1	1		5	
#C	0	0	1	2	1			4

<u>Recursive step at k = 2</u>: $\underset{\sim}{h}_2 = [h_2(200),\ h_2(020),\ h_2(002),\ h_2(110),\ h_2(101),\ h_2(011)]$

2nd block \ 1st block				
	$X_2(100)$	$X_2(010)$	$X_2(001)$	$X_{2.} = 4$
	$Y_2(100)$	$Y_2(010)$	$Y_2(001)$	$Y_{2.} = 0$
	$Z_2(100)$	$Z_2(010)$	$Z_2(001)$	$Z_{2.} = 1$
	$h_1(100) = 1$	$h_1(010) = 2$	$h_1(001) = 2$	$t = 5$

$$P = \frac{4!\,0!\,1!\,1!\,2!\,2!}{5! \prod_{\substack{\text{all}\\ \text{entries}}} (\text{entry})!} = \frac{.8}{\prod_{\substack{\text{all}\\ \text{entries}}} (\text{entry})!}$$

$$\begin{array}{ccc|l} 1 & 2 & 1 & \\ 0 & 0 & 0 & P = .4 \\ 0 & 0 & 1 & \\ \hline & & & \end{array} \qquad \begin{array}{ccc|l} 1 & 1 & 2 & \\ 0 & 0 & 0 & P = .4 \\ 0 & 1 & 0 & \\ \hline & & & \end{array} \qquad \begin{array}{ccc|l} 0 & 2 & 2 & \\ 0 & 0 & 0 & P = .2 \\ 1 & 0 & 0 & \\ \hline & & & \end{array}$$

$\underset{\sim}{h}_2 = [1,0,1,2,1,0]$ (i) $\qquad \underset{\sim}{h}_2 = [1,0,0,1,2,1]$ (ii) $\qquad \underset{\sim}{h}_2 = [0,0,0,2,3,0]$ (iii)

Recursive step at k = r = 3:

1st and 2nd blocks / 3rd block		Total
#A's		$X_{3.} = 1$
#B's	$P = \dfrac{1!\,3!\,1!\,\Pi h_2(x,y,z)!}{5!\,\Pi(\text{entry})!} = P(\underset{\sim}{h}_3 \mid \underset{\sim}{h}_2)$	$Y_{3.} = 3$
#C's		$Z_{3.} = 1$
Total	$h_2(200)h_2(020)h_2(002)h_2(110)h_2(101)h_2(011)$	$t = 5$

The (triangular) lattice supporting $h_3(x,y,z)$ is the set of 10 points:

$\{(x,y,z)\} = \{(300),\ (210),\ (201),\ (120),\ (030),\ (111),\ (021),\ (102),\ (012),\ (003)\}$.

Conditioning first on $\underset{\sim}{h}_2$ = (i) above, we must enumerate the 13 possible 3 x 6 contingency table outcomes with marginal totals:

							1
	$P = \dfrac{0.1}{\Pi(\text{entry})!}$						3
							1
$\underset{\sim}{h}_2$ =	1	0	1	2	1	0	5

beginning with

	1	0	0	0	0	0	1	
	0	0	1	2	0	0	3	P = .05
	0	0	0	0	1	0	1	
$\underset{\sim}{h}_2$ =	1	0	1	2	1	0	5	

$$\underset{\sim}{h}_3 = [1,\ 0,\ 0,\ 2,\ 0,\ 0,\ 0,\ 1,\ 1,\ 0]$$

and next

	1		0	0	0		1
	0		1	1	1		3
	0		0	1	0		1
$\underset{\sim}{h}_2$ =	1	0	1	2	1	0	5

$$\underset{\sim}{h}_3 = [1, 0, 0, 1, 0, 2, 0, 0, 1, 0], \qquad P(\underset{\sim}{h}_3 | \underset{\sim}{h}_2) = 0.1 .$$

The remaining possible outcomes for case (i) are (omitting null columns):

$$\begin{array}{cccc|} 1&0&0&0\\ 0&0&2&1\\ 0&1&0&0\\ \hline \multicolumn{4}{c|}{P = .05} \end{array} \quad \begin{array}{cccc|} 0&1&0&0\\ 1&0&2&0\\ 0&0&0&1\\ \hline \multicolumn{4}{c|}{.05} \end{array} \quad \begin{array}{cccc|} 0&1&0&0\\ 1&0&1&1\\ 0&0&1&0\\ \hline \multicolumn{4}{c|}{.10} \end{array} \quad \begin{array}{cccc|} 0&1&0&0\\ 0&0&2&1\\ 1&0&0&0\\ \hline \multicolumn{4}{c|}{.05} \end{array} \quad \begin{array}{cccc|} 0&0&1&0\\ 1&1&1&0\\ 0&0&0&1\\ \hline \multicolumn{4}{c|}{.10} \end{array} \quad \begin{array}{cccc|} 0&0&1&0\\ 1&1&0&1\\ 0&0&1&0\\ \hline \multicolumn{4}{c|}{.10} \end{array}$$

$$\begin{array}{cccc|} 0&0&1&0\\ 1&0&1&1\\ 0&1&0&0\\ \hline \multicolumn{4}{c|}{P = .10} \end{array} \quad \begin{array}{cccc|} 0&0&1&0\\ 0&1&1&1\\ 1&0&0&0\\ \hline \multicolumn{4}{c|}{.10} \end{array} \quad \begin{array}{cccc|} 0&0&0&1\\ 1&1&1&0\\ 0&0&1&0\\ \hline \multicolumn{4}{c|}{.10} \end{array} \quad \begin{array}{cccc|} 0&0&0&1\\ 1&0&2&0\\ 0&1&0&0\\ \hline \multicolumn{4}{c|}{.05} \end{array} \quad \begin{array}{cccc|} 0&0&0&1\\ 0&1&2&0\\ 1&0&0&0\\ \hline \multicolumn{4}{c|}{.05} \end{array}$$

This same array of tables and P-values apply in case (ii) where $\underset{\sim}{h}_2 = [1, 0, 0, 1, 2, 1]$, but the column labels differ in case (ii) so the resulting $\underset{\sim}{h}_3$ values also differ. The first of the 13 tables, for example, now becomes

$$\begin{array}{cccccc|} 1&0&0&0&0&0\\ 0&0&0&1&2&0\\ 0&0&0&0&0&1\\ \hline \end{array}$$

$$\underset{\sim}{h}_3 = [1, 0, 0, 1, 0, 2, 0, 0, 1, 0] \qquad P(\underset{\sim}{h}_3 | \underset{\sim}{h}_2) = .05 .$$

In case (iii) only 4 configurations are possible:

(x,y,z) =	(110)	(101)	
#A =	1	0	1
#B =	1	2	3
#C =	0	1	1
$h_2(x,y,z)$ =	2	3	5

$$\begin{pmatrix} x\\ y\\ z \end{pmatrix} = \begin{pmatrix} 3&2&2&1&0&1&0&1&0&0\\ 0&1&0&2&3&1&2&0&1&0\\ 0&0&1&0&0&1&1&2&2&3 \end{pmatrix}$$

$$\underset{\sim}{h}_3 = (0\ 1\ 0\ 1\ 0\ 2\ 0\ 1\ 0\ 0)$$

$$P(\underset{\sim}{h}_3 | \underset{\sim}{h}_2) = 0.3$$

1	0
0	3
1	0

$$\underset{\sim}{h}_3 = (0\ 1\ 0\ 0\ 0\ 4\ 0\ 0\ 0\ 0)$$
$$P(\underset{\sim}{h}_3 | \underset{\sim}{h}_2) = 0.1$$

0	1
2	1
0	1

$$\underset{\sim}{h}_3 = (0\ 0\ 1\ 2\ 0\ 1\ 0\ 1\ 0\ 0)$$
$$P(\underset{\sim}{h}_3 | \underset{\sim}{h}_2) = 0.3$$

0	1
1	2
1	0

$$\underset{\sim}{h}_3 = (0\ 0\ 1\ 1\ 0\ 3\ 0\ 0\ 0\ 0)$$
$$P(\underset{\sim}{h}_3 | \underset{\sim}{h}_2) = 0.3$$

The following table of $P(\underset{\sim}{h}_3)$ is obtained by calculating $P(\underset{\sim}{h}_2) \cdot P(\underset{\sim}{h}_3 | \underset{\sim}{h}_2)$ and summing over the three cases (i), (ii) and (iii) of $\underset{\sim}{h}_2$. The boxed row of this table shows the original, realized outcome and its probability under H_o.

Permutation distribution $P_{H_o}(\underset{\sim}{h}_3)$ of $h_3(x,y,z)$ and the induced distribution of $T_2 = \Sigma(y + 4z)h_3(x,y,z)$

$h_3(x,y,z)$ \ (x,y,z)	(300)	(210)	(201)	(120)	(030)	(111)	(021)	(102)	(012)	(003)	$100\ P_{H_o}(\underset{\sim}{h}_3)$	T_2
	1	0	0	2	0	0	0	1	1	0	2	49
	1	0	0	1	0	2	0	0	1	0	6	47
	1	0	0	2	0	1	0	0	0	1	2	53
	0	1	0	2	0	0	0	2	0	0	2	41
	0	1	0	1	0	2	0	1	0	0	8	39
	0	0	1	2	0	1	0	1	0	0	8	37
	0	2	0	1	0	0	0	1	1	0	4	47
	0	**2**	**0**	**0**	**0**	**2**	**0**	**0**	**1**	**0**	**6**	**45**
	0	2	0	1	0	1	0	0	0	1	4	51
	0	1	1	0	0	2	1	0	0	0	6	39
	0	1	1	1	0	1	0	0	1	0	12	43
	0	1	1	2	0	0	0	0	0	1	2	49
	0	0	2	2	0	0	0	0	1	0	2	41
	1	0	0	1	0	1	1	1	0	0	4	45
	1	0	0	0	0	3	1	0	0	0	2	43
	0	2	0	0	0	1	1	1	0	0	4	43
	0	1	1	1	0	0	1	1	0	0	4	41
	0	0	2	1	0	1	1	0	0	0	4	37
	0	1	0	0	0	4	0	0	0	0	4	37
	0	0	1	1	0	3	0	0	0	0	8	35
	0	1	0	1	0	2	0	1	0	0	6	39
$Eh_3(x,y,z) =$	.16	.80	.52	.96	0	1.48	.24	.44	.32	.08		

In this special case A = 0, B = 1, C = 2 "successes", the uncorrected sum of squares of treatment total successes is

$$T_2 = \sum_{x,y,z} [0^2 \cdot x + 1^2 \cdot y + 2^2 \cdot z] h_3(x,y,z)$$

and its induced permutation distribution is

T_2 =	35	37	39	41	43	$\boxed{45}$	47	49	51	53
$P_{H_o}(T_2)$ =	.08	.16	.20	.08	.18	$\boxed{.10}$	.10	.04	.04	.02

so the exact tail probability of the realized $T_2 = 45$ is P = .30. This may be compared to the nominal value

$$P = e^{-\frac{Q}{2}}(1 + \frac{Q}{2}) = .231$$

for

$$Q = \frac{T - 33.8}{2} = 5.6 \ .$$

DISCUSSION

Computational efficiency using minimal storage is the main goal in developing a program to implement these algorithms. The ultimate objective is a program which could be incorporated into a statistical package to provide permutation significance levels for built-in or user-specified test statistics for the RCBD. Approximations by asymptotic distributions become accurate for surprisingly small values of r and/or t, and the program should have the capability of computing such approximate tail probabilities and making a decision of whether to compute the corresponding exact probability. In the binary case an algorithm is available (T. Berggren's M.S. thesis) for computing the H_o-mean and covariance matrix of $\underset{\sim}{f}_r$, and these results could be readily extended to the n-ary case to provide the basic data needed in most asymptotic approximations.

Quaternary or higher n-ary cases are probably intractable with respect to exact randomization test computations except for very small r and t. About once a year in my consulting practice, however, I do encounter an RCBD with a quaternary response variable — usually as the number of successes in 3 trials (as when an experimental unit is a jar containing 3 treated, binary-responding insects) which cannot be viewed as 3 i.i.d. Bernoulli trials. The experimenter and the statistician are presently at a loss to determine the most appropriate statistical method for testing treatment effects in such small-n-ary cases.

Berggren's thesis illustrates another context for application of RCBD permutation tests with n-ary data. His development is couched in the mark-recapture setting, using $n - 1$ capture devices on a population of t individuals over r capture periods, where individuals caught and removed during a period are returned to the population at the end of that period. Assuming independence between individuals and periods, he thereby reduces this sampling model to that of an RCBD, and considers only the binary case $n = 2$. Any proposed development of the trinary case would thus have utility also in this mark-recapture setting using two different, competing capture methods.

The doubtful assumptions of independence in that setting point up the need for <u>exact</u> tests of association or interaction in a two-way table with binary data, as a possible direction for further research and development of statistical computing algorithms.