

LOCALLY MOST POWERFUL RANK TESTS FOR RANDOM EFFECTS IN TWO-WAY EXPERIMENTS

BY ANN CLEMMENS AND Z. GOVINDARAJULU

University of Kentucky

A locally most powerful (LMP) rank test against logistic alternatives is derived for the two-way experiment with random treatment effects and fixed block effects. We tabulate the critical values of the test criterion for $b = 2, 3, 4$, $c = 2, 3, 4$ and $n_{ij} \equiv n = 2(1)10$. We study its asymptotic behavior when the null hypothesis is true.

1. Introduction and Summary. The use of blocks in a design represents an attempt to remove a source of variability in the observations and so makes it possible to obtain a more accurate evaluation of the factor of interest. Observational material is segregated into groups which should be as homogeneous as possible and the effect of the factor of interest is observed on each of the groups individually.

The model one would use for this situation is

$$X_{ijk} = \mu + \beta_i + Y_j + \varepsilon_{ijk},$$
$$k = 1, \dots, n_{ij}, \quad j = 1, \dots, c; \quad \text{and} \quad i = 1, \dots, b,$$

where Y_j and ε_{ijk} are mutually independent random variables. The hypothesis we wish to test is: H_0 : the treatment Y_j produces uniform results, hence its variance is 0.

Traditionally one would assume that the ε_{ijk} are normally distributed with mean zero and variance σ_e^2 ; that the Y_j are also normal, with mean zero and variance σ_T^2 ; and that the block effects, the β_i , are additive. One would then usually use the familiar test criterion $F = \text{MST}/\text{MSE}$.

In this paper we describe a locally most powerful (LMP) rank test of H_0 against logistic alternatives. We also study its asymptotic behavior and find its computational form, and provide a table of critical values of the test criterion for $b = 2, 3, 4$, $c = 2, 3, 4$, and $n_{ij} \equiv n = 2(1)10$.

AMS 1980 Subject Classifications: 62G10, 62K10.

Key words: LMP rank test, null distribution for small and large sample sizes, random effects, two-way experiments.

2. Model and Assumptions. In this section we will present the model of interest, define our notation, and give the regularity assumptions made.

Consider now the model:

$$X_{ijk} = \mu + \beta_i + Y_j + \varepsilon_{ijk},$$

$$k = 1, \dots, n_{ij}; \quad i = 1, \dots, b; \quad j = 1, \dots, c;$$

where, the Y_j and the ε_{ijk} are mutually independent random variables, and β_i is the fixed block effect, with b the number of blocks, c the number of treatments and n_{ij} the number of observations recorded for the (i, j) th cell. Let $n_{i.} = \sum_{j=1}^c n_{ij}$, $n_{.j} = \sum_{i=1}^b n_{ij}$, and $N = \sum_{i=1}^b n_{i.} = \sum_{j=1}^c n_{.j}$.

We will also define $G(y)$ and F to be the distribution functions of the treatment effects (Y_j 's) and the random errors (ε_{ijk} 's) respectively. Appealing to the invariance property of rank tests under translation, we will let $\mu = 0$; furthermore we can assume that $\int_{-\infty}^{\infty} y dG(y) = 0$. We also assume that the block effects β_j are additive, i.e. there is no interaction between block and treatment effects. A further assumption needed is that $\int_{-\infty}^{\infty} y^2 dF(y) = \sigma^2 < \infty$.

With the above set up, we are interested in testing the null and alternative hypotheses given by

$$H_0 : G(y) = \begin{cases} 0 & \text{for } y < 0, \\ 1 & \text{for } y \geq 0, \end{cases}$$

versus,

$H_1 : G(y)$ is a member of the class of nontrivial distribution functions. Let $\bar{G}(y) = G(y/\Delta)$ be a class of nontrivial distribution functions for some small and positive Δ . Then we note that for this class, the statement that the distribution is degenerate at zero is equivalent to the statement that $\Delta = 0$.

Hence, in order to derive a LMP rank test for the multiple-block design we consider the hypotheses

$$H_0 : \Delta = 0 \quad \text{versus} \quad H_{\Delta} : \Delta > 0.$$

Let $W_1^{(i)} < \dots < W_{n_i}^{(i)}$ denote the combined ordered sample of the variables X_{ijk} , $k = 1, \dots, n_{ij}$, $j = 1, \dots, c$ within the i th block. Note that since β_i is a fixed effect for i th block, the ranks of X_{ijk} and $X_{ijk} - \beta_i$ are the same. Also let $(Z_1^{(i)}, \dots, Z_{n_i}^{(i)})' = \underline{Z}^{(i)}$ denote the c -sample rank order for the i th block, that is, $Z_k^{(i)} = j$ if $W_k^{(i)} = X_{ijl}$ for some $l = 1, \dots, n_{ij}$ with $\underline{Z} = (\underline{Z}^{(1)'}, \dots, \underline{Z}^{(b)'})'$. The vector \underline{z} will denote any possible realization of the $\Pi_{i=1}^b (n_i! / \Pi_{j=1}^c n_{ij}!)$ possible rank orders.

3. The Locally Most Powerful (LMP) Rank Test Statistic. The LMP rank test statistic for the two-way mixed model described in Section 2 is given in this section.

THEOREM 3.1. *The LMP rank test for H_0 against H_Δ is given by: Reject H_0 when*

$$\begin{aligned} \Psi = & \sum_{i=1}^b \sum_{j=1}^c \sum_{m \neq t}^{n_i} E_0 \left[\frac{f'(W_m^{(i)})f'(W_t^{(i)})}{f(W_m^{(i)})f(W_t^{(i)})} \right] \delta_{jZ_m^{(i)}} \delta_{jZ_t^{(i)}} \\ & + \sum_{i=1}^b \sum_{j=1}^c \sum_{m=1}^{n_i} E_0 \left[\frac{f''(W_m^{(i)})}{f(W_m^{(i)})} \right] \delta_{jZ_m^{(i)}} \\ & + \sum_{j=1}^c \sum_{i \neq k}^b \sum_{m=1}^{n_i} \left\{ E_0 \left[\frac{f'(W_m^{(i)})}{f(W_m^{(i)})} \right] \delta_{jZ_m^{(i)}} \right\} \\ & \cdot \left\{ \sum_{t=1}^{n_k} E_0 \left[\frac{f'(W_t^{(k)})}{f(W_t^{(k)})} \right] \delta_{jZ_t^{(k)}} \right\} \geq K_\alpha \end{aligned} \tag{3.1}$$

where K_α is determined by the level of significance α and $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, provided the conditions below are satisfied:

- i. the density f has a derivative that is absolutely continuous over finite intervals,
- ii. $f''(x)$ is continuous almost everywhere,
- iii. $\int_{-\infty}^{\infty} y^2 dG(y) < \infty$, and
- iv. $E \left| \frac{\partial^2}{\partial X^2} \ln f(X) \right| < \infty$.

PROOF. The proof is too technical and hence is omitted. For details, the reader is referred to Clemmens (1986, pp.27–35).

In comparing expression (3.1) and the test statistic of Govindarajulu (1975), we see that the test statistic for the two-way blocked layout consists of a sum of one-way test statistics (one for each block) plus a second (nontrivial) term which might be labeled as a ‘between-blocks’ contribution.

4. Test Statistic for Logistic Scores. In this section, we will derive the form of the test statistic when the density function f has the logistic form. In addition we will put it into a form which makes calculation of its value easier, whether in the equal or unequal sample size cases.

LEMMA 4.1. *When $f(x) = e^x(1 + e^x)^{-2}$, the two-way LMP rank test for*

H_0 against H_Δ is given by: Reject H_0 when

$$\begin{aligned} \Psi_L = & N + \sum_{j=1}^c n_{.j}^2 - \sum_{i=1}^b \frac{4}{n_{i.}+1} \sum_{j=1}^c n_{ij} S_j^{(i)} \\ & + \sum_{i=1}^b \frac{8}{(n_{i.}+1)(n_{i.}+2)} \sum_{j=1}^c M_j^{(i)} \\ & + \sum_{j=1}^c \frac{4}{(n_{i.}+1)(n_{k.}+1)} \sum_{i \neq k}^b \sum_{j=1}^b S_j^{(i)} S_j^{(k)} \geq K_\alpha \end{aligned} \tag{4.1}$$

where

$$S_j^{(i)} = \sum_{k=1}^{n_i} k \delta_{jZ_k^{(i)}} = \text{sum of the ranks associated with the observations}$$

getting the j^{th} treatment in i^{th} block,

$$M_j^{(i)} = \sum_{k < m}^{n_i} \sum_{m}^{n_i} k(m+1) \delta_{jZ_k^{(i)}} \delta_{jZ_m^{(i)}},$$

and K_α is determined by α .

PROOF. The first four additive terms of the two-way statistics are directly analogous to those in the logistic form of the one-way statistic derived in Clemmens and Govindarajulu (1990) requiring only the addition of the block index i to the variables and appropriate summation.

The final term of (4.1) is derived from the fact that in the logistic case

$$\begin{aligned} \sum_{k=1}^{n_i} E_0 \left[\frac{f'(W_k^{(i)})}{f(W_k^{(i)})} \right] \delta_{jZ_k^{(i)}} &= \sum_{k=1}^{n_i} E_0(1 - 2U_k^{(i)}) \delta_{jZ_k^{(i)}} \\ &= n_{ij} - \frac{2}{n_{i.}+1} S_j^{(i)} \end{aligned} \tag{4.2}$$

where $U_k^{(i)}$ are standard uniform order statistics in a sample of size $n_{i.}$.

A test statistic written in terms of ranks rather than the δ functions is more convenient for computational purposes. Such a form for the test statistic Ψ_L requires that the ranks within each cell must be ordered, smallest to largest. Then we define $R_{jk}^{(i)}$ to be the k^{th} smallest rank in cell (i, j) .

COROLLARY 4.1.1. *The test statistic Ψ_L defined in (4.1) can, for the*

case of logistic scores, be written as

$$\begin{aligned} \Psi_L = & \sum_{j=1}^c n_{.j}^2 - \sum_{i=1}^b \frac{2n_{i.}(2n_{i.} + 1)}{3(n_{i.} + 2)} + 4 \sum_{j=1}^c \left[\sum_{i=1}^b \frac{S_j^{(i)}}{n_{i.} + 1} \right]^2 \\ & - 4 \sum_{i=1}^b \sum_{j=1}^c \frac{n_{.j} S_j^{(i)}}{n_{i.} + 1} - 4 \sum_{i=1}^b \sum_{j=1}^c \frac{S_j^{(i)^2}}{(n_{i.} + 1)^2(n_{i.} + 2)} \\ & + 8 \sum_{i=1}^b \sum_{j=1}^c \sum_{k=1}^{n_{ij}} \frac{(n_{ij} - k) R_{jk}^{(i)}}{(n_{i.} + 1)(n_{i.} + 2)}. \end{aligned} \tag{4.3}$$

PROOF. It can be shown that

$$\sum_{j=1}^c M_j^{(i)} = \frac{1}{2} \sum_{j=1}^c S_j^{(i)^2} - \frac{n_{i.}(n_{i.} + 1)(2n_{i.} + 1)}{12} + \sum_{j=1}^c \sum_{k=1}^{n_{ij}} (n_{ij} - k) R_{jk}^{(i)}$$

and it can easily be shown that

$$\sum_{i \neq k}^b \sum_{i \neq k}^b \frac{S_j^{(i)} S_j^{(k)}}{(n_{i.} + 1)(n_{k.} + 1)} = \left[\sum_{i=1}^b \frac{S_j^{(i)}}{n_{i.} + 1} \right]^2 - \sum_{i=1}^b \left[\frac{S_j^{(i)}}{n_{i.} + 1} \right]^2.$$

Substituting these expressions into (4.1) and combining terms yields expression (4.3). Note that all but one of these terms are either constant or dependent solely on the sums of the ranks within each cell; this is the most attractive feature of this form for computation.

Further simplification results when we consider the case when the sample size is the same in each cell, i.e. $n_{ij} = n$; in this case,

$$\begin{aligned} \Psi_L = & bnc - b^2 cn^2 - \frac{2bcn(2cn + 1)}{3(nc + 2)} + \frac{4bcn^2}{cn + 2} + \\ & \frac{4}{(nc + 1)} \sum_{j=1}^c \left[\sum_{i=1}^b S_j^{(i)} \right]^2 \\ & - \frac{4}{(nc + 1)^2(nc + 2)} \sum_{j=1}^c \sum_{i=1}^b S_j^{(i)^2} \\ & - \frac{8}{(nc + 1)(nc + 2)} \sum_{j=1}^c \sum_{i=1}^b \sum_{k=1}^n k R_{jk}^{(i)}. \end{aligned} \tag{4.4}$$

It should be noted that despite the seeming complexity of the constants and leading multipliers, the three stochastic terms are relatively easy to accumulate.

5. Asymptotic form for Ψ_L under H_0 . One can show (see Clemmens (1986 pp.39-40)) that $E_0\Psi_L = 0$. Further, we will show that the statistic Ψ_L has asymptotically a chi-square distribution. We will also find an asymptotic form of the statistic with which to compute values that can be compared with tabled chi-squared values.

We noted earlier that the statistic Ψ_L can be considered as a sum of statistics one for each of the b blocks plus a term which can be described as a 'between-blocks' contribution. Notationally we can write

$$\Psi_L = \sum_{i=1}^b T^{(i)} + \sum_{i \neq k}^b \sum_{i \neq k}^b T^{(i,k)},$$

where $\sum_{i=1}^b T^{(i)}$ includes the first two terms of (3.1) and $\sum_{i \neq k}^b \sum_{i \neq k}^b T^{(i,k)}$ represents the third term of (3.1). Referring to Govindarajulu (1975, Section 2) and substituting from (4.2) allow us to state that

$$\begin{aligned} \Psi_L &= \sum_{i=1}^b \left[\sum_{j=1}^c \sum_{m=1}^{n_i} \sum_{t=1}^{n_i} E_0[(1 - 2U_m^{(i)})(1 - 2U_t^{(i)})] \delta_{jZ_m^{(i)}} \delta_{jZ_t^{(i)}} + \frac{n_i}{3} \right] \\ &+ \sum_{i \neq k}^b \sum_{j=1}^c \sum_{m=1}^{n_i} \left[\sum_{m=1}^{n_i} E_0(1 - 2U_m^{(i)}) \delta_{jZ_m^{(i)}} \right] \cdot \left[\sum_{t=1}^{n_i} E_0(1 - 2U_t^{(i)}) \delta_{jZ_t^{(i)}} \right] \\ &= \sum_{j=1}^c \sum_{i=1}^b \sum_{k=1}^b \sum_{m=1}^{n_i} \sum_{t=1}^{n_i} E_0 \left[(1 - 2U_m^{(i)})(1 - 2U_t^{(k)}) \right] \delta_{jZ_m^{(i)}} \delta_{jZ_t^{(k)}} - \frac{N}{3}. \end{aligned}$$

Thus one can write

$$\Psi_L = R^* + \tilde{R} - \frac{N}{3}$$

where

$$R^* = \sum_{j=1}^c \left[\sum_{i=1}^b \sum_{t=1}^{n_i} E_0(1 - 2U_t^{(i)}) \delta_{j, Z_t^{(i)}} \right]^2$$

and the rest is \tilde{R} . Or, since asymptotically only R^* is stochastic,

$$R^* = \Psi_L - \tilde{R} + \frac{N}{3}. \tag{5.1}$$

We appeal to Lemma 3.6.1 and Theorem 3.6.3 in Clemmens (1986) to note that asymptotically $\frac{\tilde{R}}{N} - \sum_{i=1}^b \sum_{j=1}^c \frac{n_{ij}(n_{ij}-1)}{3N(n_i-1)} \rightarrow 0$ in probability as $N \rightarrow \infty$.

In the equal sample size case, when $n_{ij} = n$, we will need a term for comparison with chi-squared tables. Thus we can say that the distribution of $3c\Psi_L^*/b = 3cR^*/Nb$ is asymptotically equivalent to a central chi-squared

variable with $c-1$ degrees of freedom. The form required for table comparisons will be

$$\frac{3c\Psi_L^*}{b} = \frac{3\Psi_L}{bn} - \frac{c(n-1)}{(cn-1)} + c. \quad (5.2)$$

Remark. Since $E_0\Psi_L = 0$, and since $c(n-1)/(cn-1) \rightarrow 1$, we can deduce that $E_0(3c\Psi_L^*/b) \rightarrow c-1$ in the equal sample size case, identical to the expected value of a central chi-squared variable with $c-1$ degrees of freedom.

6. Calculation Methods. To find the exact distribution of Ψ_L it would be necessary to determine all possible partitions of ranks of the c sub-samples of sizes n_{i1}, \dots, n_{ic} for each of the blocks and compute the value of the test statistic for each partition. When c, b , and the n_{ij} are small, this can be done but in practice we found that this direct method became unwieldy rather quickly, i.e. when N is relatively small. Therefore we turned to computer simulation in order to generate the approximate distribution and critical values of Ψ_L .

To simulate the distribution of Ψ_L values, we need to draw samples repeatedly from the population of X 's which is valid when the null hypothesis holds. Recall that our model is

$$X_{ijk} = \mu + \beta_i + Y_j + \varepsilon_{ijk},$$

where we assumed $\mu = 0$ and $\int_{-\infty}^{\infty} y_j dG(y_j) = 0$. Under the null hypothesis the Y_j 's are constant and equal to 0, so $X_{ijk} = \beta_i + \varepsilon_{ijk}$; since we rank all the observations within each block separately, without loss of generality, we can set the block effect $\beta_i \equiv 0$.

Now, since the distribution of Ψ_L when H_0 is true is free of the underlying distribution of the ε_{ijk} again without loss of generality, we can assume that ε_{ijk} are standard uniform random variables.

7. Tables of Critical Values of Ψ_L . In this section we present tables of simulated critical values of the Ψ_L statistic for the $b = 2, 3, 4$, $c = 2, 3, 4$ and $n = 2(1)10$ and for the most commonly used values of α , namely .01 (where available), 0.05, and 0.10. Since this statistic has a discrete distribution, we will not be able to give critical regions with exactly the nominal value of α when N is small; in these cases we will provide the value or values which come closest and give the actual probability.

The number of distributions we could investigate is enormous; so we must limit our labors to those most likely to be of use. Our expectation is that experimenters will find the partitions with equal sample sizes to be of most interest and so in the tables for simulated distributions we include critical values only for distributions of this type. We also need these equal sample size distributions in order to investigate asymptotic behavior.

With the exception of the case of $b = c = n = 2$ all of the critical values for Ψ_L given in this section were derived from simulated distributions (The simulated values agree with the exact values for the case of $b = c = n = 2$).

For our simulation study, we used the IMSL library function GGUW to generate b random samples of size nc from the uniform (0,1) distribution to produce one possible value of the statistic Ψ_L . For each combination of b, c and n , we repeated the above procedure a minimum of 15,000 times; for large values of n , we used as many as 140,000 iterations.

8. Asymptotic Behavior of Ψ_L . According to the results of Section 5, when $n_j = n$ for all j , $3c\Psi_L^*/b$ is asymptotically distributed as a central chi-square with $c - 1$ degrees of freedom. Equation (5.2) gives us a method for converting Ψ_L to a variable asymptotically equal to $3c\Psi_L^*/b$; by comparing these to the appropriate values of chi-square distribution, we can determine the value of n beyond which the asymptotic behavior holds.

Let $W_L = 3c\Psi_L^*/b = \frac{3\Psi_L}{bn} - \frac{c(n-1)}{(nc-1)} + c$. When comparing the values of W_L to the χ^2 values we noticed that the value of n is the most important factor in determining how rapidly the distribution approaches the asymptotic distribution.

Critical values of W_L approach the corresponding χ^2 values more rapidly for $\alpha = 0.10$ than for $\alpha = 0.05$ and more rapidly for $\alpha = 0.05$ than for $\alpha = 0.01$.

From the comparisons we have made, we conclude that for W_L the chi-square distribution would serve as a good approximation for the critical values for any combination of b and c provided $n \geq 10$. For example, when $b = c = 4$, $W_{L, .05} = 7.7$ and $\chi_{3,0.05}^2 = 7.8$.

REFERENCES

- CLEMMENS, A. E. (1986). Some nonparametric contributions to linear models: Random Effects Models. Doctoral dissertation submitted to the University of Kentucky.
- CLEMMENS, A. E. and GOVINDARAJULU, Z. (1990). A certain locally most powerful test for one-way random effects model: null distribution and power considerations. *Comm. Statist.* A19, 4139–4151.
- GOVINDARAJULU, Z. (1975). Locally most powerful rank order tests for one-way random effects model. *Studia Scientiarum Mathematicarum Hungarica* 10, 47–60.

TABLE 1. Critical Values of Ψ_L for Equal Size n

b	c	n	$\Psi_{L,0.01}$	$\Psi_{L,0.05}$	$\Psi_{L,0.10}$
2	2	2	- - -	4.2	(0.058) 1.8 (0.168)
		3	8.9 (0.013)	4.4 (0.060)	4.1 (0.091)
		4	12.2	5.8 (0.051)	4.0 (0.098)
		5	16.2 (0.011)	8.2	5.1 (0.099)
		6	19.3	10.1	5.9
		7	23.6	11.5	7.2
		8	27.2	13.7	8.3
		9	31.6	15.8	9.3
		10	37.2	17.2	10.2
		2	3	2	6.2
3	10.6			6.8	4.6
4	16.9			10.0	6.8
5	19.9			11.8	8.4
6	25.4			14.0	9.3
7	32.8			17.6	11.9
8	36.0			20.5	13.9
9	40.0			21.2	14.0
10	47.4			25.4	17.5
2	4			2	7.2
		3	12.2	7.6	5.3
		4	18.6	11.5	7.9
		5	24.6	14.9	10.2
		6	29.6	18.0	12.5
		7	35.8	21.1	14.6
		8	41.5	42.1	16.7
		9	46.0	27.4	18.9
		10	52.3	30.7	21.0
		3	2	2	10.1 (0.008)
3	12.4			6.7	4.4
4	18.9			10.7	6.3 (0.101)
5	23.5			11.6	7.5 (0.101)
6	29.3			15.4	9.7
7	33.4			18.5	11.4
8	39.2			20.7	12.6
9	45.9			24.2	14.9
10	53.8			27.5	16.6
3	3			2	9.9
		3	17.5	10.5	6.8
		4	25.9	14.7	9.2
		5	33.5	19.3	12.8
		6	39.6	22.5	15.1
		7	46.4	26.8	17.6
		8	53.8	29.7	20.3

436 *LOCALLY MOST POWERFUL RANK TESTS*

b	c	n	$\Psi_{L,.01}$	$\Psi_{L,.05}$	$\Psi_{L,.10}$
		9	60.1	32.8	22.3
		10	73.0	38.5	25.3
3	4	1	4.2 (.016)	3.2(0.05)	2.9 (.074)
					2.2 (.147)
		2	12.6	7.9	5.6
		3	21.1	12.7	8.8
		4	29.4	17.8	12.1
		5	37.6	22.1	15.5
		6	45.7	27.2	18.6
		7	54.5	31.9	21.9
		8	61.7	36.7	25.0
		9	69.8	40.9	28.1
		10	78.8	46.1	31.5
4	2	2	9.6(.021)	5.9(.072)	3.1(.122)
		3	16.7	9.8	6.8
		4	24.2	12.2	7.4
		5	35.2	17.8	10.9
		6	41.1	20.1	13.4
		7	48.3	25.4	15.3
		8	54.0	28.0	16.7
		9	63.4	32.0	19.4
		10	69.1	36.3	21.0
4	3	1	6.0(.005)	4.5(.043)	4.0(.073)
			4.5(.043)	4.0(.073)	2.5(.127)
		2	14.6	8.5	5.8
		3	24.6	14.4	9.6
		4	33.5	19.0	12.7
		5	43.7	24.3	16.2
		6	54.2	30.0	20.0
		7	63.5	35.3	23.5
		8	71.9	40.4	26.5
		9	81.7	46.4	30.2
		10	93.4	51.3	33.8
4	4	1	6.7	4.8	3.2
		2	17.6	10.7	7.5
		3	29.8	17.5	12.0
		4	38.9	23.5	16.2
		5	50.7	30.0	20.4
4	4	6	61.1	35.9	24.3
		7	73.3	42.7	29.0
		8	83.4	48.4	33.1
		9	93.5	54.9	37.4
		10	104.9	61.6	42.0

DEPARTMENT OF STATISTICS
UNIVERSITY OF KENTUCKY
817 PATTERSON OFFICE TOWER
LEXINGTON, KENTUCKY 40506-0027
USA

