Ann. Inst. Statist. Math.
Vol. 47, No. 3, 505–523 (1995)

TESTING HOMOGENEITY WITH AN ORDERED ALTERNATIVE IN A TWO-FACTOR LAYOUT BY COMBINING p-VALUES*

RAMAL MOONESINGHE[1] AND F. T. WRIGHT[2]

[1]*National Research Council, 2101 Constitution Avenue, Washington, DC 20418, U.S.A.*
[2]*Department of Statistics, University of Missouri, Columbia, MO 65211, U.S.A.*

(Received March 29, 1993; revised September 9, 1994)

Abstract. Two-factor experiments in which both factors are ordinal are considered. If it is believed *apriori* that the mean response is nondecreasing in each factor with the other held fixed, then one may test for a treatment effect by testing homogeneity with the appropriate ordered alternative. The likelihood ratio test has been developed in the literature, but the level probabilities needed to implement the test have only been determined in a few special cases by Monte Carlo techniques. A test obtained by combining the p-values from a test concerning the rows and a test concerning the columns is studied. Fisher's method of combining p-values is recommended. It is shown that the likelihood ratio test is more powerful, but if one does not want to obtain Monte Carlo estimates of the level probabilities, then the procedure proposed here should be considered.

Key words and phrases: Bivariate trends, combining p-values, Fisher's method, likelihood ratio tests, matrix ordering, order restricted tests, two-moment approximations.

1. Introduction

We consider two-factor experiments in which both factors are ordinal level. In some of these situations, the mean of the dependent variable is believed *apriori* to be nondecreasing in each factor when the other is held fixed. For instance, over suitable ranges it may be reasonable to assume that average yield is a nondecreasing function of moisture and amount of a certain chemical in a fertilizer. Suppose there are R levels of the first factor, C levels of the second factor and μ_{ij} is the mean response with the first factor at level i and the second factor at level j. One could test to determine if there is any treatment effect by testing homogeneity,

$$H_0 : \mu_{ij} \text{ is constant,} \tag{1.1}$$

* This research was supported by the National Institutes of Health under Grant 1R01GM42584-01. Part of this work is taken from the first author's dissertation submitted in partial fulfillment for the Ph.D. degree at the University of Missouri-Columbia.

with the alternative

$$H_1 : \mu_{ij} \leq \mu_{i'j'} \text{ for } 1 \leq i \leq i' \leq R \text{ and } 1 \leq j \leq j' \leq C \quad \text{with at least one inequality.} \tag{1.2}$$

The relation, $\lesssim$, defined on $\Gamma = \{(i,j) : 1 \leq i \leq R \text{ and } 1 \leq j \leq C\}$ by $(i,j) \lesssim (i',j')$ if and only if $1 \leq i \leq i' \leq R$ and $1 \leq j \leq j' \leq C$ is a partial order, which we refer to as the matrix order. Thus, assuming the responses are normally distributed, the work of Bartholomew (1961) gives the likelihood ratio tests of H_0 versus H_1. However, the null distributions of the test statistics involve mixing coefficients which are called level probabilities. In particular, the null distributions are mixtures of chi-squared or beta distributions depending on whether the variances are known or not, see Robertson *et al.* ((1988), Section 2.3). For this partial order, the level probabilities seem to be intractable. For $R = C = 2$, formulas exist, cf. Robertson *et al.* ((1988), p. 84), and Lemke (1983) obtained Monte Carlo estimates of the level probabilities for a few pairs (R, C) for the special case in which the sample sizes are the same for all (i, j).

Following the work of Mudholkar and McDermott (1989) and McDermott and Mudholkar (1993), we develop tests of H_0 versus H_1 by combining p-values. The efficiencies of the procedures developed by McDermott and Mudholkar compared to the efficiencies of the likelihood ratio tests tend to decrease with the number of p-values that are combined. For that reason, we consider combining the p-value from a test of homogeneity in each row with a nondecreasing alternative in each row with the p-value of a test for equality of row effects with the alternative that the row effects are nondecreasing. One could employ the tests developed by McDermott and Mudholkar on the rows and on the columns, but this would involve combining more p-values and we suspect this would cause a decrease in efficiency. Thus, we use the likelihood ratio tests on the rows and on the columns. While there are several ways to combine p-values, we considered the three methods studied by Mudholkar and McDermott (1989), Fisher's, Liptak's and Tippett's. As in their study, we found that Fisher's method gave better power characteristics and so we only report the results for Fisher's method of combining p-values.

If interaction is not present, then order restricted tests for main effects are relatively simple to conduct, see Robertson *et al.* ((1988), pp. 66–70) and Kulatunga and Sasabuchi (1984*a*). However, if interaction cannot be ruled out, then the procedure presented here could be employed while Bartholomew's procedure cannot be implemented unless one wants to obtain Monte Carlo estimates of the level probabilities.

The case of known variance is considered in Section 2. Those results provide large sample tests for one-parameter exponential families and in nonparametric settings. They also lay a foundation for the results in Section 3 which deal with an unknown variance. In Section 4, approximations are given for the p-values needed to implement these tests as well as certain powers for the special case of balanced designs. The results of a Monte Carlo study of the power of this test and the likelihood ratio test are discussed in Section 5. Dykstra and Robertson

(1982) present some data on first year university grade point averages (GPA) as a function of high school percentile ranks and scores on an entrance examination. In Section 6, the procedure developed here is applied to similar data which has been collected more recently.

For balanced designs, known variance and the test obtained by combining p-values, it is conjectured in Section 4 that the minimum of the power function over points in H_1 at a fixed distance from H_0 occurs at one of two types of alternatives and numerical evidence is given to support the conjecture. Also, approximations to the power function at these alternative are provided. In our Monte Carlo study, it was found that the approximations are adequate for most purposes. These approximations provide some assistance in designing experiments with a specified power at a fixed alternative.

In each case considered in Section 5, the likelihood ratio test has the larger power, but to implement it, one must obtain Monte Carlo estimates of the level probabilities. The test based on p-values does not have this drawback. For the variance known case with distance from H_0 chosen so that the minimum power of the test proposed here is about 0.8, the minimum power of the test based on p-values is about 95% of the minimum power of the likelihood ratio test for $(R, C) = (3, 3)$. This percentage seems to decrease with R and C, but is about 88% for $(R, C) = (9, 9)$. Similar results hold for the variance unknown case with moderate sample sizes. However, for a common sample sizes as small as two, the corresponding percentages are 86% for smaller values of R and C and 82% for $(R, C) = (9, 9)$.

2. A test based on p-values: variance known

We suppose that $X_{ijk} \sim N(\mu_{ij}, \sigma^2)$ for $k = 1, \ldots, n_{ij}$, $i = 1, \ldots, R$ and $j = 1, \ldots, C$ are independent random variables and with σ^2 known and H_0 and H_1 defined as in (1.1) and (1.2), we consider testing H_0 versus H_1. Let $N = n_{..} = n_{11} + \cdots + n_{RC}$ be the total sample size and for $1 \le i \le R$ and $1 \le j \le C$, let $n_{i\cdot} = n_{i1} + \cdots + n_{iC}$, $n_{\cdot j} = n_{1j} + \cdots + n_{Rj}$, $\bar{X}_{ij\cdot}$ be the mean of X_{ijk} for $k = 1, 2, \ldots, n_{ij}$,

$$(2.1) \qquad \bar{X}_{i\cdot\cdot} = \sum_{j=1}^{C} n_{ij}\bar{X}_{ij\cdot}/n_{i\cdot}, \qquad \bar{X}_{\cdot j\cdot} = \sum_{i=1}^{R} n_{ij}\bar{X}_{ij\cdot}/n_{\cdot j} \qquad \text{and}$$

$$\bar{X}_{\cdots} = \sum_{i=1}^{R}\sum_{j=1}^{C} n_{ij}\bar{X}_{ij\cdot}/N.$$

One could test H_0 versus H_1 in two stages. First, test H_{01} versus H_{11} with

$$(2.2) \quad H_{01} : \mu_{i1} = \cdots = \mu_{iC} \quad \text{for} \quad 1 \le i \le R$$

and

$$(2.3) \quad H_{11} : \mu_{i1} \le \cdots \le \mu_{iC} \quad \text{for} \quad 1 \le i \le R \quad \text{with at least one inequality.}$$

If H_{01} is rejected, then H_0 is rejected. If H_{01} is not rejected, then with $\mu_{i\cdot} = (n_{i1}\mu_{i1} + \cdots + n_{iC}\mu_{iC})/n_{i\cdot}$ for $1 \le i \le R$, test H_{02} versus H_{12} where

$$(2.4) \qquad H_{02} : \mu_{1\cdot} = \cdots = \mu_{R\cdot} \quad \text{and} \quad H_{12} : \mu_{1\cdot} \le \cdots \le \mu_{R\cdot} \quad \text{with} \quad \mu_{1\cdot} < \mu_{R\cdot}.$$

If H_{02} is rejected, then H_0 is rejected. However, we follow the closely related approach used by Mudholkar and McDermott (1989) and base a test on the p-values from the two stages above. At each stage, we use the likelihood ratio test developed by Bartholomew (1961), but slight modifications are needed in the first stage since H_{01} is not homogeneity.

2.1 *Maximum likelihood estimates*

Under H_{01}, the maximum likelihood estimate of μ_{ij} is $\bar{X}_{i\cdot\cdot}$ and we let μ^*_{ij} denote the maximum likelihood estimate of μ_{ij} subject to the restrictions in H_{11}. Because the partial order which determines H_{11} is decomposable, the μ^*_{ij} can be computed separately on the rows, see Robertson *et al.* ((1988), p. 85). Furthermore, on a given row the partial order is a simple order and thus μ^*_{ij} can be computed by the pool-adjacent-violators algorithm on each row separately, see Robertson *et al.* ((1988), p. 8). That is, for each i, the pool-adjacent-violators algorithm is applied to $\bar{X}_{ij\cdot}$ with weights n_{ij}.

Reparameterizing in terms of $\mu_{i\cdot}$ and $\alpha_{ij} = \mu_{ij} - \mu_{i\cdot}$ for $1 \le j \le C-1$ and $1 \le i \le R$, we see that the maximum likelihood estimate of $\mu_{i\cdot}$ under H_{02} is $\bar{X}_{\cdots}$ and we denote the maximum likelihood estimate of $\mu_{i\cdot}$ subject to the restrictions in H_{12} by $\mu^*_{i\cdot}$. The $\mu^*_{i\cdot}$ can be computed by applying the pool-adjacent-violators algorithm to the $\bar{X}_{i\cdot\cdot}$ with weights $n_{i\cdot}$.

2.2 *Test statistics*

For σ^2 known the likelihood ratio test of H_{01} versus H_{11}, which was studied by Kulatunga and Sasabuchi (1984*b*), rejects H_{01} for large values of

$$(2.5) \qquad T_1 = \sum_{i=1}^{R} Y_i \quad \text{where} \quad Y_i = \sum_{j=1}^{C} n_{ij}(\mu^*_{ij} - \bar{X}_{i\cdot\cdot})^2 \quad \text{for} \quad 1 \le i \le R$$

and Y_i for $1 \le i \le R$ are independent. Let $\boldsymbol{n} = (n_{11}, \ldots, n_{RC})'$ and let $\boldsymbol{n}_i = (n_{i1}, \ldots, n_{iC})'$ for $1 \le i \le R$. Kulatunga and Sasabuchi (1984*b*) note that under H_{01},

$$(2.6) \qquad \mathrm{pr}(Y_i/\sigma^2 \ge y) = \sum_{l=1}^{C} P(l, C; \boldsymbol{n}_i)\,\mathrm{pr}(\chi^2_{l-1} \ge y)$$

where χ^2_ν denotes a chi-squared random variable with ν degrees of freedom, $\chi^2_0 \equiv 0$ and the $P(l, C; \boldsymbol{n}_i)$ are the level probabilities for a simple order which are discussed in detail in Robertson *et al.* ((1988), pp. 77–82). If for $C > 5$, one uses the FORTRAN program given in Bohrer and Chow (1978) to compute the level probabilities, then the program due to Sun (1988) for computing orthant probabilities works well. Using (2.6) and the independence of the Y_i, Kulatunga and Sasabuchi (1984*b*) note that

$$\mathrm{pr}(T_1/\sigma^2 \geq t) = \sum_{l=R}^{RC} P(l, RC; \boldsymbol{n}) \,\mathrm{pr}(\chi^2_{l-R} \geq t) \tag{2.7}$$

with $P(l, RC; \boldsymbol{n})$ the convolution $\{P(l, C; \boldsymbol{n}_1)\} * \cdots * \{P(l, C; \boldsymbol{n}_R)\}$.

The likelihood ratio test of H_{02} versus H_{12} rejects H_{02} for large values of

$$T_2 = \sum_{i=1}^{R} n_{i\cdot}(\mu^*_{i\cdot} - \bar{X}_{\cdots})^2$$

and applying Theorem 2.3.1 of Robertson *et al.* (1988) again, under H_{02},

$$\mathrm{pr}(T_2/\sigma^2 \geq t) = \sum_{l=1}^{R} P(l, R; n_{1\cdot}, \ldots, n_{R\cdot}) \,\mathrm{pr}(\chi^2_{l-1} \geq t), \tag{2.8}$$

where $P(l, R; n_{1\cdot}, \ldots, n_{R\cdot})$ are the level probabilities for a simple order which are discussed above.

2.3 *P-values*

T_1 and T_2 are independent because T_1 is a function of $\{\bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} : 1 \leq i \leq R, 1 \leq j \leq C\}$ and T_2 is a function of $\{\bar{X}_{i\cdot\cdot} : 1 \leq i \leq R\}$. Let $\bar{F}_1(t)$ and $\bar{F}_2(t)$ denote (2.7) and (2.8) respectively and note that for $i = 1, 2$, the p-value for the i-th stage is $P_i \equiv \bar{F}_i(T_i/\sigma^2)$. Also, $\bar{F}_i(0) = 1$, $\bar{F}_i$ is nonincreasing and $\bar{F}_i$ is continuous except at $t = 0$. Because $\bar{F}_i$ is not continuous, the null distribution of the p-values are complicated slightly. Let $p_i = \bar{F}_i(0+)$ and for $0 \leq y \leq p_i$, let x_{iy} be such that $\bar{F}_i(x_{iy}) = y$. Thus,

$$\mathrm{pr}(P_i \geq y \mid T_i > 0) = 1 - y/p_i \quad \text{if} \quad 0 \leq y \leq p_i$$

and the conditional distribution of $P_i \mid T_i > 0$ is uniform on $(0, p_i)$. With $P'_i = P_i/p_i$, the conditional distribution of $P'_i \mid T_i > 0$ is uniform on $(0, 1)$.

Without loss of generality, we may base our test statistic on the P'_i. In particular, the proposed test rejects H_0 for large values of

$$\Psi_T = -2 \sum_{i=1}^{2} \log P'_i. \tag{2.9}$$

Conditional on $T_i > 0$, $-2 \log P'_i$ has a chi-squared distribution with two degrees of freedom. Thus, conditioning on whether T_i is positive or not, we see that under H_0,

$$\begin{aligned} \mathrm{pr}(\Psi_T \geq c) = {} & p_1 p_2 \,\mathrm{pr}(\chi^2_4 \geq c) + p_1(1 - p_2) \,\mathrm{pr}(\chi^2_2 \geq c - 2 \log p_2) \\ & + (1 - p_1) p_2 \,\mathrm{pr}(\chi^2_2 \geq c - 2 \log p_1) \\ & + (1 - p_1)(1 - p_2) I(2 \log p_1 + 2 \log p_2 \geq c), \end{aligned} \tag{2.10}$$

where $I(A)$ is the indicator of A. Of course, the last term in (2.10) is zero for $c \geq 0$. Table 1 gives the critical values for Ψ_T for $2 \leq R, C \leq 9$, selected α and balanced designs.

Table 1. Critical values for Ψ_T and Ψ_E with balanced designs.

R	α	$C=2$	$C=3$	$C=4$	$C=5$	$C=6$	$C=7$	$C=8$	$C=9$
2	0.100	5.694	6.051	6.160	6.209	6.234	6.249	6.259	6.266
2	0.050	7.426	7.779	7.888	7.936	7.961	7.977	7.986	7.993
2	0.025	9.097	9.448	9.557	9.604	9.630	9.645	9.655	9.662
2	0.010	11.244	11.594	11.702	11.750	11.775	11.790	11.800	11.806
2	0.001	16.456	16.803	16.911	16.959	16.984	17.000	17.010	17.016
3	0.100	6.660	6.855	6.900	6.915	6.922	6.925	6.927	6.928
3	0.050	8.376	8.570	8.615	8.630	8.637	8.640	8.642	8.643
3	0.025	10.036	10.231	10.275	10.290	10.297	10.301	10.302	10.304
3	0.010	12.174	12.368	12.412	12.428	12.435	12.438	12.440	12.441
3	0.001	17.374	17.567	17.612	17.626	17.633	17.636	17.638	17.640
4	0.100	7.054	7.160	7.177	7.181	7.183	7.184	7.184	7.184
4	0.050	8.767	8.872	8.889	8.893	8.895	8.896	8.896	8.896
4	0.025	10.425	10.530	10.547	10.551	10.553	10.554	10.554	10.554
4	0.010	12.560	12.665	12.682	12.687	12.689	12.689	12.690	12.690
4	0.001	17.756	17.860	17.878	17.882	17.884	17.885	17.885	17.885
5	0.100	7.257	7.313	7.319	7.321	7.321	7.321	7.321	7.321
5	0.050	8.968	9.023	9.030	9.031	9.031	9.032	9.032	9.032
5	0.025	10.625	10.681	10.687	10.688	10.689	10.689	10.689	10.689
5	0.010	12.760	12.815	12.822	12.823	12.823	12.823	12.823	12.823
5	0.001	17.954	18.009	18.016	18.018	18.018	18.017	18.018	18.018
6	0.100	7.375	7.404	7.406	7.407	7.407	7.407	7.407	7.407
6	0.050	9.085	9.114	9.116	9.116	9.117	9.117	9.117	9.117
6	0.025	10.742	10.770	10.773	10.773	10.773	10.773	10.773	10.773
6	0.010	12.876	12.905	12.907	12.907	12.907	12.907	12.907	12.907
6	0.001	18.070	18.099	18.101	18.101	18.101	18.101	18.102	18.102
7	0.100	7.450	7.464	7.465	7.465	7.465	7.465	7.465	7.465
7	0.050	9.159	9.174	9.175	9.175	9.175	9.175	9.175	9.175
7	0.025	10.815	10.830	10.831	10.831	10.831	10.831	10.831	10.831
7	0.010	12.949	12.964	12.965	12.965	12.965	12.965	12.965	12.965
7	0.001	18.143	18.158	18.159	18.159	18.159	18.159	18.159	18.159
8	0.100	7.500	7.508	7.508	7.508	7.508	7.508	7.508	7.508
8	0.050	9.209	9.217	9.217	9.217	9.217	9.217	9.217	9.217
8	0.025	10.865	10.873	10.873	10.873	10.873	10.873	10.873	10.873
8	0.010	13.000	13.007	13.008	13.008	13.008	13.008	13.008	13.008
8	0.001	18.193	18.201	18.201	18.201	18.201	18.201	18.201	18.201
9	0.100	7.537	7.540	7.540	7.540	7.540	7.540	7.540	7.540
9	0.050	9.246	9.249	9.249	9.249	9.249	9.249	9.249	9.249
9	0.025	10.902	10.905	10.905	10.905	10.905	10.905	10.905	10.905
9	0.010	13.036	13.040	13.040	13.040	13.040	13.040	13.040	13.040
9	0.001	18.229	18.232	18.232	18.232	18.232	18.232	18.232	18.232

3. A test based on p-values: variance unknown

Let X_{ijk}, H_0, H_{01}, H_{02}, H_1, H_{11}, H_{12}, μ_{ij}^* and $\mu_{i.}^*$ be defined as in Section 2 except that σ^2 is unknown. A test of H_0 versus H_1 is developed by combining p-values from the tests of H_{01} versus H_{11} and H_{02} versus H_{12}. The likelihood ratio test of H_{01} versus H_{11} rejects H_{01} for large values of

$$E_1 = \sum_{i=1}^{R}\sum_{j=1}^{C} n_{ij}(\mu_{ij}^* - \bar{X}_{i..})^2 \Bigg/ \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{n_{ij}} (X_{ijk} - \bar{X}_{i..})^2, \tag{3.1}$$

and the likelihood ratio test of H_{02} versus H_{12} rejects H_{02} for large values of

$$E_2 = \sum_{i=1}^{R} n_{i\cdot}(\mu_{i.}^* - \bar{X}_{...})^2 \Bigg/ \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{n_{ij}} (X_{ijk} - \bar{X}_{...})^2. \tag{3.2}$$

We show that under H_{01}, E_1 and E_2 are independent. Under H_{01},

$$\sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{n_{ij}} (X_{ijk} - \bar{X}_{i..})^2, \qquad \bar{X}_{1..}, \ldots, \bar{X}_{R..} \tag{3.3}$$

are complete and sufficient and thus E_1, which is ancillary, is independent of (3.3). Writing the denominator in E_2 as

$$\sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{n_{ij}} (X_{ijk} - \bar{X}_{i..})^2 + \sum_{i=1}^{R} n_{i\cdot}(\bar{X}_{i..} - \bar{X}_{...})^2,$$

it is clear that E_2 is a function of the quantities in (3.3) and thus independent of E_1.

From Robertson *et al.* ((1988), p. 70), it follows that if H_{01} holds, then with $B(a, b)$ denoting a beta random variable and $B(0, b) \equiv 0$,

$$\text{pr}(E_2 \geq t) = \sum_{l=1}^{R} P(l, R; n_{1\cdot}, \ldots, n_{R\cdot})\, \text{pr}\left(B\left(\frac{1}{2}l - \frac{1}{2}, \frac{1}{2}N - \frac{1}{2}l\right) \geq t\right). \tag{3.4}$$

Also, Kulatunga and Sasabuchi (1984*b*) note that under H_{01},

$$\text{pr}(E_1 \geq t) = \sum_{l=R}^{RC} P(l, RC; \boldsymbol{n})\, \text{pr}\left(B\left(\frac{1}{2}l - \frac{1}{2}R, \frac{1}{2}N - \frac{1}{2}l\right) \geq t\right). \tag{3.5}$$

One may combine the p-values associated with E_1 and E_2 and again we found that Fisher's method is preferred in this setting. For $i = 1, 2$, let P_i be the p-value associated with E_i; let $p_i = \text{pr}(E_i > 0) = \text{pr}(T_i > 0)$; let $P_i' = P_i/p_i$ and reject H_0 for large values of

$$\Psi_E = -2\sum_{i=1}^{2} \log P_i'. \tag{3.6}$$

Under H_0, the tail probabilities are given by (2.10). It is interesting that Ψ_T and Ψ_E have the same critical values for a given level of significance.

4. Approximating p-values and powers: balanced designs

Computing p-values via (2.7) and (2.8) in the variance known case or via (3.4) and (3.5) in the variance unknown case can be tedious. For balanced designs and simple orders, Bartholomew (1961), for the variance known case, and Sasabuchi and Kulatunga (1985), for the unknown variance case, showed that moment approximations can be used in most practical applications. For known variance, balanced designs and simple orders, Bartholomew (1961) also obtained a two-moment approximation to the minimum of the power function at points that satisfy the ordering and are a fixed distance from homogeneity. In this section, we assume the sample sizes are equal and show how these results can be modified for the tests developed here. Singh and Wright (1987) obtained approximations for the minimum powers in the case of unknown variance, but they are more cumbersome to apply. Thus, approximations to minimum powers are not discussed for unknown variance. If the degrees of freedom on the variance estimator is not too small, the power in the variance known case should not be too much larger than in the unknown variance case.

4.1 *Approximating p-values when the variance is known*

In this subsection, we consider two-moment gamma approximations for the p-values of T_1 and T_2 which are relatively simple to use with balanced designs. For balanced designs, formulas and numerical values for the first two cumulants of the null distribution of T_2/σ^2, which we denote by $\kappa_i^*(R)$ for $i = 1, 2$, are given in (3.2.2) and Table A.15 of Robertson *et al.* (1988) and $p_2 = \text{pr}(T_2 > 0) = 1 - 1/R$. They show that the cumulants of the conditional null distribution of T_2/σ^2 given that $T_2 > 0$ are

$$\kappa_1 = \kappa_1^*(R)/p_2 \quad \text{and} \quad \kappa_2 = \kappa_2^*(R)/p_2 - (1 - p_2)(\kappa_1^*(R)/p_2)^2. \tag{4.1}$$

Let $\rho_2 = \kappa_2/\kappa_1$, $b_2 = \kappa_1/\rho_2$ and G_b, $\bar{G}_b$ and g_b denote the distribution function, the tail probabilities and the density of a gamma distribution with parameters b and 1. The p-value corresponding to t_2', an observed value of T_2/σ^2, is

$$\bar{F}_2(t_2') \approx p_2 \bar{G}_{b_2}(t_2'/\rho_2). \tag{4.2}$$

For $5 \le R \le 40$, Table A.14 of Robertson *et al.* (1988) gives the values of ρ_2 and b_2.

One sees from (2.5), (2.6) and the independence of the Y_i, that the cumulants of the null distribution of T_1/σ^2 are $\tau_i^*(R, C) = R\kappa_i^*(C)$ for $i = 1, 2$ and $p_1 = \text{pr}(T_1 > 0) = 1 - (1/C)^R$. Replacing $\kappa_i^*(R)$ by $\tau_i^*(R, C)$ and p_2 by p_1 in (4.1), gives τ_1 and τ_2. With $\rho_1 = \tau_2/\tau_1$ and $b_1 = \tau_1/\rho_1$, the p-value corresponding to t_1', an observed value of T_1/σ^2, is

$$\bar{F}_1(t_1') \approx p_1 \bar{G}_{b_1}(t_1'/\rho_1). \tag{4.3}$$

4.2 *Approximating powers when the variance is known*

Expressions for the power functions of the likelihood ratio tests for the simple ordering, $\mu_1 \leq \cdots \leq \mu_k$, only have been derived for $k = 3$ and 4, see Robertson *et al.* ((1988), Section 2.5). Approximations for the minimum of the powers at a fixed distance from the null hypothesis have been provided for the simple order with balanced designs. Bartholomew (1961) conjectured that with the distance to homogeneity fixed, the minimum power of the likelihood ratio test at points satisfying the ordering occurs at $\mu_1 < \mu_2 = \cdots = \mu_k$ and at $\mu_1 = \cdots = \mu_{k-1} < \mu_k$, see Singh and Wright (1989) for additional numerical evidence. Approximations to these minimum powers are of interest because (1) the gain in power due to the order restriction is smallest at these points and (2) if an experiment is designed to have a specified power at these points, then the power is at least as large at all points in H_1 with the same Δ.

The analogous approximations are given here for the test obtained in Section 2. Let n denote the common value of n_{ij}, $w = n/\sigma^2$, $\bar{\mu}$ denote the average of the μ_{ij} and

$$\Delta = \left\{ w \sum_{i=1}^{R} \sum_{j=1}^{C} (\mu_{ij} - \bar{\mu})^2 \right\}^{1/2} . \tag{4.4}$$

We seek the alternatives in H_1 with fixed $\Delta > 0$ which give the minimum power for Ψ_T.

The symmetry properties of the power functions of order restricted tests for normal means play an important role in this search. For instance, consider the likelihood ratio test of homogeneity of normal means with the nondecreasing alternative, $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$ and $\mu_1 < \mu_k$, equal sample sizes, and variances which are known and equal. By negating all the observations it is easily seen that the power at an alternative, $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$, is the same as that at the alternative $-\mu = (-\mu_1, -\mu_2, \ldots, -\mu_k)$ with the alternative hypothesis changed to $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_k$ and $\mu_1 > \mu_k$. By relabeling, the alternative can be changed to the original nondecreasing alternative. Thus, the powers at μ and $-\mu^{(r)} = (-\mu_k, -\mu_{k-1}, \ldots, -\mu_1)$ are the same for testing homogeneity with a nondecreasing alternative. This is also true for the case of unknown but equal variances. It is also helpful to note that, whether the variances are known or unknown, the power of the likelihood ratio test of homogeneity against an ordered alternative is not changed if the same constant is added to each of the underlying means. Hence, for equal variances whether they are known or unknown, the power of the likelihood ratio test of homogeneity with a nondecreasing alternative is the same at the alternatives (ae_j, be_{k-j}) and (ae_{k-j}, be_j) where a and b are real numbers and e_i is a vector of dimension i with a 1 in each component.

Because of Bartholomew's conjecture concerning minimum powers for the simple order, in our search for mean vectors with a minimum power for fixed $\Delta > 0$, we consider alternatives of the form $\mu_{11} = a$ and $\mu_{ij} = b$ for $(i,j) \neq (1,1)$, or alternatively with $\mu_{RC} = b$ and $\mu_{ij} = a$ for $(i,j) \neq (R,C)$, where $a < b$. We call such mean configurations type I alternatives and applying the symmetry and invariance properties given in the last paragraph, one can show that the power of

Ψ_T is the same at these two types of alternatives. This is also true for the test Ψ_E. In fact an argument like that given above shows that the power of Bartholomew's likelihood ratio test for the case of known (unknown) variances, which we denote by T (E), is the same at these two types of alternatives. Suppose that $\mu_{11} = a$ and $\mu_{ij} = b$ otherwise. For $i = 2, \ldots, R$, Y_i/σ^2 has the same distribution as under H_{01}, the means in the first row are the form that Bartholomew conjectured would yield the minimum power for Y_1/σ^2 and the $\mu_{i\cdot}$ are also of that form.

We also considered alternatives for which T_2/σ^2 has its null distribution and T_1/σ^2 has minimum power for such alternatives. In this case, the $\mu_{i\cdot}$ are the same. It is easy to show that if the alternative is in H_1 and one of the rows is constant then they all are and the mean configuration is in H_0 which means $\Delta = 0$. Thus, none of the rows are constant. Furthermore, since

$$\sum_{j=1}^{C} \mu_{\alpha j} = \sum_{j=1}^{C} \mu_{\beta j} \quad \text{and} \quad \mu_{\alpha j} \le \mu_{\beta j} \quad \text{for} \quad 1 \le \alpha < \beta \le R \quad \text{and} \quad 1 \le j \le C,$$

all of the rows must be the same and not constant. Because of Bartholomew's conjecture, we consider alternatives with all of the rows of the form $(a, a, \ldots, a, b)$ or all of the rows of the form $(a, b, \ldots, b)$ and call these type II alternatives. Again, one can show that the power of Ψ_T is the same at these two kinds of type II alternatives and this is also true for the test Ψ_E. In fact an argument like that given above shows that the power of Bartholomew's likelihood ratio test for the case of known (unknown) variances, i.e. T (E), is the same at these two types of alternatives.

In addition, we considered type III alternatives, which are those that yield a null distribution for T_1/σ^2 and minimum power for T_2/σ^2. It is straightforward to show that these alternatives are of the form $\mu_{ij} = a$ for $1 \le i \le R-1$ and $1 \le j \le C$ and $\mu_{Rj} = b$ for $1 \le j \le C$ or of the form $\mu_{1j} = a$ for $1 \le j \le C$ and $\mu_{ij} = b$ for $2 \le i \le R$ and $1 \le j \le C$. Again, Ψ_T has the same powers at these two kinds of type III alternatives and this is also true of Ψ_E, T and E.

To provide information about where the minimum power occurs, Monte Carlo estimates, based on 100,000 iterations, of the power of Ψ_T with $\alpha = 0.05$, $(R, C) = (2,2)$, $(2,3)$, $(3,2)$, $(2,4)$, $(4,2)$ and $(3,3)$ and $\Delta = 1$, 2, 3 and 4 were obtained for all mean configurations with just two distinct values. Because the power of Ψ_T is not changed if the same value is added to all μ_{ij}, we may take the smaller of the two values to be zero and the other value is determined by Δ and the form of the alternative. Type II alternatives have the smallest estimates of powers except in the two cases of $(R, C) = (4, 2)$ with $\Delta = 1$ and 2, and in these two cases, the type III alternatives have the smallest estimates of power. For $(R, C) = (4, 2)$ and $\Delta = 1$, the estimates of power for the type I, II and III alternatives are 0.156, 0.158 and 0.152, respectively, and for $(R, C) = (4, 2)$ and $\Delta = 2$, the estimates of power for the type I, II and III alternatives are 0.410, 0.410 and 0.408, respectively. This provides some indication that our search can be limited to type I, II and III alternatives. Incidentally, for $(R, C) = (9, 2)$ and $\Delta = 1$, 2, 3 and 4, the type I alternatives have smaller estimates of power than the other two types of alternatives.

In addition, random searches over points in H_1 with $\Delta = 3$ were conducted. This value of Δ was chosen because it gives powers about 0.8. Such powers are large enough to be of practical interest, but not so large as to obscure the differences in power. In particular, with $(R, C) = (2,2)$, $(2,3)$ and $(3,2)$, $\mu_{11} = 0$ and $\mu_{RC} = 1$, $RC - 2$ pseudo random uniform variables on the interval $(0,1)$ were generated. Next, for each permutation of these $RC-2$ values, $\mu_{12}, \ldots, \mu_{1C}, \mu_{21}, \ldots$ and $\mu_{R(C-1)}$ were set equal to the permuted values provided the resulting mean configuration was in H_1. These values of μ_{ij} were multiplied by the appropriate constant to make $\Delta = 3$, and based on 10,000 iterations, Monte Carlo estimates of the power of Ψ_T at these alternatives were obtained. This was repeated 1,000 times. With R, C and $\Delta = 3$ fixed, let $\hat{p}$ denote the smallest estimated power for the alternatives with two distinct values. Each of the estimates of power in these 1,000 iterations which were less than

$$p_u = \hat{p} + 3[\hat{p}(1 - \hat{p})(1/10000 + 1/100000)]^{1/2}$$

were recorded along with the corresponding mean configuration. (If $\tilde{p}$ is one of the 1,000 power estimates and the corresponding true power is close to the true power associated with $\hat{p}$, then with high probability $\tilde{p}$ will be less than $\hat{p} + 3\sigma_{\hat{p}-\tilde{p}}$ which is approximately equal to p_u.) For the case $(R, C) = (2,2)$ with $\Delta = 3$, the smallest estimated power for alternatives with two distinct values occurred at a type II alternative and the estimated power was 0.809. All of the points identified in the random search were "close" to being of type II. For instance, the smallest power estimate obtained in the search was 0.810 and occurred at $(\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}) = (0, 2.9674, 0.00886, 3.0406)$. For the case $(R, C) = (2,3)$ with $\Delta = 3$, the smallest estimated power for alternatives with two distinct values occurred at a type II alternative and the estimated power was 0.753. All of the points identified in this random search were also "close" to being of type II. For instance, the smallest power estimate obtained in this search was 0.755 and occurred at $(\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23}) = (0, 2.5906, 2.6286, 0.00815, 2.6501, 2.6816)$. For the case $(R, C) = (3,2)$ with $\Delta = 3$, the smallest estimated power for alternatives with two distinct values occurred at a type II alternative and the estimated power was 0.763. The points identified in the random search were again "close" to being of type II. The smallest power estimate obtained in this search was 0.773 and occurred at $(\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, \mu_{31}, \mu_{32}) = (0, 2.4178, 0.0753, 2.5552, 0.1200, 2.5623)$.

Based on the information given above, we limited our search for the points with minimum power to alternatives of type I, II and III. With $R, C = 3$, 5 and 9, $\alpha = 0.05$ and $\Delta = 3$, Monte Carlo estimates of the power of Ψ_T were obtained for these three types of alternatives using 10,000 iterations. In none of these cases did the type III alternative have the smallest estimated power. In fact, the only cases we found for which the type III alternatives had the smallest estimates of power were $(R, C) = (4,2)$ with $\Delta = 1$ and 2. In the next section it is observed that the minimum powers seem to be smaller in the case $(R, C) = (A, B)$ than in $(R, C) = (B, A)$ with $A > B$ and thus we recommend labelling so that $R \geq C$. For $R \geq C$, we conjecture that the minimum powers in H_1 at a fixed distance, Δ, from H_0 occur either at type I or type II alternatives. Hence, we develop approximations to the power of Ψ_T at these types of alternatives.

We consider type I alternatives first, that is $a = \mu_{11} < \mu_{12} = \cdots = \mu_{1C} = b$ and $b = \mu_{i1} = \cdots = \mu_{iC}$ for $2 \le i \le R$. Then

$$(4.5)\quad \Delta^2 = w(RC-1)(b-a)^2/(RC) \quad \text{or} \quad b = a + \Delta[RC/((RC-1)w)]^{1/2}.$$

Using the argument in Barlow *et al.* ((1972), p. 162),

$$(4.6)\qquad Y_1/\sigma^2 \approx Y_1'/\sigma^2 + \bar{\chi}_2^2(\Delta_1)$$

where Y_1'/σ^2 has distribution given by (2.6) with C and $\boldsymbol{n}_1$ replaced by $C-1$ and the $C-1$ dimensional constant vector with each $n_{1j} = n$, $\Delta_1^2 = w(C-1)(b-a)^2/C$, $\bar{\chi}_2^2(\Delta_1)$ has distribution and cumulants given by (3.60) and (3.61) in Barlow *et al.* (1972), and Y_1'/σ^2, $Y_2, \ldots, Y_R$, $\bar{\chi}_2^2(\Delta_1)$ are independent. Denoting the cumulants of $\bar{\chi}_2^2(\Delta_1)$ by $\delta_i(\Delta_1)$, the approximate cumulants of T_1/σ^2 for this alternative are

$$(4.7)\qquad \gamma_i \equiv (R-1)\kappa_i^*(C) + \kappa_i^*(C-1) + \delta_i(\Delta_1) \quad \text{for} \quad i = 1, 2.$$

Thus, with $\rho_3 = \gamma_2/\gamma_1$, $b_3 = \gamma_1/\rho_3$ and $t > 0$,

$$(4.8)\qquad \mathrm{pr}(T_1/\sigma^2 > t \mid T_1 > 0) \approx \bar{G}_{b_3}(t/\rho_3).$$

Robertson *et al.* ((1988), p. 153) noted that for powers typically of interest in designing experiments, $\mathrm{pr}(T_1 > 0)$ is close to one and so the unconditional and conditional probabilities are nearly the same.

A similar approximation is needed for T_2. For the alternative being considered, $\mu_{1\cdot} = [(C-1)b + a]/C$ and $\mu_{i\cdot} = b$ for $i = 2, \ldots, R$ and the associated weights, which are the reciprocals of the variances of the $\bar{X}_{i\cdot\cdot}$, are $nC/\sigma^2 = Cw$. With $\Delta_2^2 = w(R-1)(b-a)^2/(RC)$ and applying the same argument from Barlow *et al.* (1972) used above, the nonnull cumulants of T_2/σ^2 are approximated by

$$(4.9)\qquad \lambda_i \equiv \kappa_i^*(R-1) + \delta_i(\Delta_2) \quad \text{for} \quad i = 1, 2.$$

With $\rho_4 = \lambda_2/\lambda_1$, $b_4 = \lambda_1/\rho_4$ and $t > 0$,

$$(4.10)\qquad \mathrm{pr}(T_2/\sigma^2 > t \mid T_2 > 0) \approx \bar{G}_{b_4}(t/\rho_4).$$

Next, the results above are used to approximate the distribution of the P_i'. Conditional on $T_1 > 0$, the nonnull distribution of T_1/σ^2 is approximated by that of $\rho_3 U$ where U is a gamma random variable with parameters b_3 and 1, cf. (4.8). Applying (4.3), the distribution function of $-2\log P_1'$ conditional on $T_1 > 0$ is

$$(4.11)\quad \begin{aligned} \mathrm{pr}(-2\log P_1' \le t \mid T_1 > 0) &= \mathrm{pr}(P_1' \ge e^{-t/2} \mid T_1 > 0) \\ &\approx \mathrm{pr}(\bar{G}_{b_1}(\rho_3 U/\rho_1) \ge e^{-t/2}) \\ &= \mathrm{pr}(U \le \rho_1 G_{b_1}^{-1}(1 - e^{-t/2})/\rho_3) \\ &= G_{b_3}(\rho_1 G_{b_1}^{-1}(1 - e^{-t/2})/\rho_3) \equiv H_1(t). \end{aligned}$$

Similarly, the approximate distribution function of $-2\log P_2'$ conditional on $T_2 > 0$ is

$$G_{b_4}(\rho_2 G_{b_2}^{-1}(1-e^{-t/2})/\rho_4) \equiv H_2(t) \tag{4.12}$$

and its approximate conditional density is

$$h_2(t) = \frac{g_{b_4}[\rho_2 G_{b_2}^{-1}(1-e^{-t/2})/\rho_4]\rho_2 e^{-t/2}}{2\rho_4 g_{b_2}[G_{b_2}^{-1}(1-e^{-t/2})]}. \tag{4.13}$$

Convoluting $-2\log P_1'$ and $-2\log P_2'$, one can approximate the power function of Ψ_T at this alternative by

$$\mathrm{pr}(\Psi_T \geq c \mid T_1 > 0, T_2 > 0) \approx 1 - \int_0^c H_1(c-t)h_2(t)dt, \tag{4.14}$$

where c is the critical value for Ψ_T given by (2.10). In Section 5, the accuracy of this approximation is studied by Monte Carlo techniques.

Next, a similar approximation is obtained for the power of Ψ_T with type II alternatives. In particular, we consider $\mu_{ij} = a$ for $1 \leq i \leq R$ and $1 \leq j \leq C-1$ and $\mu_{iC} = b$ for $1 \leq i \leq R$. Then,

$$\Delta^2 = wR(C-1)(b-a)^2/C \quad \text{or} \quad b = a + \Delta[C/(R(C-1)w)]^{1/2}. \tag{4.15}$$

Using the argument in Barlow *et al.* ((1972), p. 162), each Y_i/σ^2 is approximated by the variable in (4.6) except that $b-a$ is determined by (4.15) rather than (4.5). Thus, the approximate cumulants for T_1/σ^2 in this case are

$$\begin{aligned} &\eta_i \equiv R(\kappa_i^*(C-1) + \delta_i(\Delta_1)) \quad \text{for } i = 1,2 \quad \text{and} \\ &\mathrm{pr}(T_1/\sigma^2 > t \mid T_1 > 0) \approx \bar{G}_{b_5}(t/\rho_5), \end{aligned} \tag{4.16}$$

where Δ_1 is as in (4.6), $\rho_5 = \eta_2/\eta_1$, $b_5 = \eta_1/\rho_5$ and $t > 0$. Applying the argument that led to (4.11) gives

$$\mathrm{pr}(-2\log P_1' \leq t \mid T_1 > 0) \approx G_{b_5}(\rho_1 G_{b_1}^{-1}(1-e^{-t/2})/\rho_5) \equiv H_3(t), \tag{4.17}$$

and $-2\log P_2'$ conditional on $T_2 > 0$ has a chi-square distribution with two degrees of freedom. Convoluting $-2\log P_1'$ and $-2\log P_2'$ and using $\mathrm{pr}(T_1 > 0) \approx 1$ and $\mathrm{pr}(T_2 > 0) = 1 - 1/R$, we obtain

$$\begin{aligned} \mathrm{pr}(\Psi_T \geq c) \approx 1 &- \left(\frac{1}{2} - \frac{1}{2R}\right)\int_0^c H_3(c-t)\exp(-t/2)dt \\ &- \frac{1}{R}H_3\left(c - 2\log\left(1 - \frac{1}{R}\right)\right). \end{aligned} \tag{4.18}$$

The accuracy of this approximation is studied in Section 5.

4.3 *Approximating p-values when the variance is unknown*

Sasabuchi and Kulatunga (1985) studied moment approximations for the null distribution of the likelihood ratio test of homogeneity with a simply ordered alternative with unknown variance. Singh and Wright (1988) found that for most practical purposes the two-moment approximation is adequate. We consider two-moment beta approximations for the p-values of E_1 and E_2. The notation follows that of Robertson *et al.* ((1988), p. 124). With

$$p_2 = \text{pr}(E_2 > 0) = \text{pr}(T_2 > 0) = 1 - 1/R, \tag{4.19}$$
$$a_2 = \frac{\kappa_1^*(R)}{(N-1)p_2}, \qquad b_2 = \frac{\kappa_2^*(R) + [\kappa_1^*(R)]^2}{(N-1)(N+1)p_2}$$

$\bar{H}_{cd}$ the tail function of a beta distribution with parameters c and d,

$$c_2 = a_2(a_2 - b_2)/(b_2 - a_2^2) \quad \text{and} \quad d_2 = (1 - a_2)(a_2 - b_2)/(b_2 - a_2^2), \tag{4.20}$$
$$\text{pr}(E_2 \geq y) \approx p_2 \bar{H}_{c_2 d_2}(y) \quad \text{for} \quad y > 0. \tag{4.21}$$

To use this technique on E_1, one needs to find $p_1 = \text{pr}(E_1 > 0)$, $a_1 = E(E_1 \mid E_1 > 0)$ and $b_1 = E(E_1^2 \mid E_1 > 0)$. However, from the argument given on p. 124 of Robertson *et al.* (1988), one sees that

$$p_1 = 1 - 1/C^R, \qquad a_1 = \frac{R\kappa_1^*(C)}{(N-R)p_1} \quad \text{and} \tag{4.22}$$
$$b_1 = \frac{R\kappa_2^*(C) + [R\kappa_1^*(C)]^2}{(N-R)(N-R+2)p_1}.$$

With c_1 and d_1 defined as c_2 and d_2 in (4.20) with a_2 and b_2 replaced with a_1 and b_1,

$$\text{pr}(E_1 \geq y) \approx p_1 \bar{H}_{c_1 d_1}(y) \quad \text{for} \quad y > 0. \tag{4.23}$$

5. Results of a power study

A study of the powers of the tests based on Ψ_T and Ψ_E as well as Bartholomew's likelihood ratio tests, which we denote by T and E for the cases of variance known or unknown respectively, was conducted for balanced designs and selected R and C. For the chosen pairs (R, C), Monte Carlo estimates of the level probabilities needed to implement Bartholomew's tests were obtained based on 100,000 iterations and using them, the approximate $\alpha = 0.05$ critical values were computed. The corresponding critical values for Ψ_T, which are the same as for Ψ_E, were taken from Table 1.

Because we want to be able to compare the minimum powers of the tests, we first need some indication of which alternatives in H_1 with fixed $\Delta > 0$ yield minimum powers for T with a balanced design. Considering the simple loop ordering, Singh and Schell (1992) based on numerical evidence, conjecture that for

$R = C = 2$, the minimum powers occur at alternatives of the form $\mu_{11} = \mu_{12} = a$ and $\mu_{21} = \mu_{22} = b$, which is type III, or alternatively at $\mu_{11} = \mu_{21} = a$ and $\mu_{12} = \mu_{22} = b$, which is type II. (Of course, for $R = C$, T, as well as E, has the same powers for type II and III alternatives.) For $\Delta = 1, 2, 3, 4$, $(R, C) = (2, 2), (2, 3), \ldots, (2, 6), (3, 3), (3, 4), \ldots, (3, 6), (4, 4), (4, 5)$, $\alpha = 0.05$ and alternatives in H_1 with exactly two distinct values, we obtained Monte Carlo estimates of power based on 100,000 iterations. In each case with $R = 2$, the type III alternatives had the smallest power. However, in the case $(R, C) = (2, 9)$, which was also considered, type I and III alternatives had the smallest powers and the estimates, to three decimal places, for the type I alternatives were the same or slightly smaller than those for the type III alternatives. In all the other cases which were considered, i.e. the ones with $2 < R \leq C$, the type I alternatives have the smaller estimates of power.

For the simply ordered case, $\mu_1 \leq \cdots \leq \mu_k$, with $\Delta > 0$ fixed, Bartholomew conjectured that the maximum power occurs when $\mu_2 - \mu_1 = \cdots = \mu_k - \mu_{k-1} > 0$. While the conjecture does not seem to be correct, cf. Singh and Wright (1989), it is adequate for practical purposes. Because of this conjecture, we included alternatives with μ_{ij} proportional to $i + j$ and $\Delta > 0$ fixed. The mean vectors labeled type IV alternatives are of this form.

Using the IMSL routine QUADS, the approximations to the powers of Ψ_T given in (4.14) and (4.18), which we denote by $\tilde{\pi}(\mu, \Psi_T)$, were computed for $\alpha = 0.05$ and values of Δ incremented by 0.1 and the values of Δ which gave powers near 0.8 were used in the rest of the study. Using 10,000 iterations and a significance level of 0.05, Monte Carlo estimates of the powers of T, E, Ψ_T and Ψ_E, which we denote by $\hat{\pi}(\mu, T)$, $\hat{\pi}(\mu, E)$, $\hat{\pi}(\mu, \Psi_T)$ and $\hat{\pi}(\mu, \Psi_E)$, were obtained. Let n denote the common sample size. For the case of known variance, the power depends on n only through the factor $w = n/\sigma^2$ in Δ. For convenience, we took $n = \sigma^2 = 1$. For the case of unknown variance, the degrees of freedom for the variance estimator is $RC(n-1)$ and so to contrast with the variance known case, in which one could think of the degrees of freedom as infinite, we chose $n = 2$.

Table 2 contains these power estimates for $(R, C) = (3, 3), (3, 5), (3, 9), (5, 5), (5, 9)$ and $(9, 9)$. In these cases, we conjecture that T has minimum powers at type I alternatives and Ψ_T has minimum powers at type I or type II alternatives. Thus, the power estimates are given for type I, II and IV alternatives.

First, we note that for type I alternatives, the largest discrepancy between $\tilde{\pi}(\mu, \Psi_T)$ and $\hat{\pi}(\mu, \Psi_T)$ is 0.011 and that these differences are within sampling error. Thus, it seems that the approximation to powers given in (4.14) performs quite well. For type II alternatives, the largest discrepancy between $\tilde{\pi}(\mu, \Psi_T)$ and $\hat{\pi}(\mu, \Psi_T)$ is 0.030. While these differences are larger than for type I alternatives, the approximation for type II alternatives given in (4.18) is adequate for most practical purposes.

The powers of T and E are symmetric in R and C, but this is not true for Ψ_T and Ψ_E. We found that if $R < C$, the minimum powers are larger than if R and C were interchanged. However, the powers for type IV alternatives are larger if $R > C$. For instance, if $(R, C) = (9, 3)$ and $\Delta = 4.0$, for the variance known case, the Monte Carlo estimate of power for type I and IV alternatives are 0.809 and

Table 2. Approximations to the powers of Ψ_T, $\tilde{\pi}(\mu, \Psi_T)$, at type I and II alternatives and Monte Carlo estimates of the powers of Ψ_T, Ψ_E, T and E, $\hat{\pi}(\mu, \Psi_T)$, $\hat{\pi}(\mu, \Psi_E)$, $\hat{\pi}(\mu, T)$ and $\hat{\pi}(\mu, E)$, at type I, II and IV alternatives with $\alpha = 0.05$. Type I alternatives are proportional to $\mu_{11} = 0$ and $\mu_{ij} = 1$ otherwise; type II alternatives are proportional to $\mu_{ij} = 0$ for $1 \le i \le R$, $1 \le j \le C-1$ and $\mu_{iC} = 1$ for $1 \le i \le R$; and type IV alternatives are proportional to $\mu_{ij} = i + j$.

	alternative		Variance Known				Variance Unknown, $n_{ij} \equiv 2$		
R, C	type	Δ	$\tilde{\pi}(\mu, \Psi_T)$	$\hat{\pi}(\mu, \Psi_T)$	$\hat{\pi}(\mu, T)$	$\frac{\hat{\pi}(\mu, \Psi_T)}{\hat{\pi}(\mu, T)}$	$\hat{\pi}(\mu, \Psi_E)$	$\hat{\pi}(\mu, E)$	$\frac{\hat{\pi}(\mu, \Psi_E)}{\hat{\pi}(\mu, E)}$
3, 3	I	3.3	0.805	0.805	0.831	0.969	0.633	0.702	0.902
3, 3	II	3.3	0.759	0.787	0.833	0.945	0.596	0.711	0.838
3, 3	IV	3.3	—	0.904	0.914	0.989	0.840	0.884	0.950
3, 5	I	3.6	0.801	0.812	0.851	0.954	0.686	0.772	0.889
3, 5	II	3.6	0.773	0.795	0.860	0.924	0.669	0.787	0.850
3, 5	IV	3.6	—	0.933	0.946	0.986	0.901	0.940	0.959
3, 9	I	4.0	0.830	0.840	0.883	0.951	0.745	0.843	0.874
3, 9	II	4.0	0.811	0.841	0.894	0.941	0.745	0.853	0.873
3, 9	IV	4.0	—	0.960	0.975	0.985	0.949	0.968	0.980
5, 5	I	3.9	0.794	0.799	0.874	0.914	0.691	0.817	0.842
5, 5	II	3.9	0.811	0.829	0.891	0.930	0.729	0.835	0.873
5, 5	IV	3.9	—	0.960	0.978	0.982	0.950	0.971	0.978
5, 9	I	4.4	0.831	0.836	0.918	0.911	0.759	0.881	0.862
5, 9	II	4.4	0.856	0.858	0.926	0.927	0.804	0.901	0.892
5, 9	IV	4.4	—	0.961	0.991	0.970	0.980	0.992	0.988
9, 9	I	4.8	0.810	0.816	0.930	0.877	0.745	0.913	0.816
9, 9	II	4.8	0.888	0.882	0.947	0.931	0.842	0.931	0.904
9, 9	IV	4.8	—	0.996	0.998	0.998	0.995	0.996	0.999

0.966, respectively. The corresponding values taken from Table 2 are 0.840 and 0.960. For $R = 9$, $C = 3$ and $\Delta = 4.0$ in the variance unknown case, the Monte Carlo estimates are 0.682 and 0.964, which should be compared with 0.737 and 0.949. If one only knows that μ satisfies H_1, then we recommend labelling so that $R < C$, but if in addition, one believes μ is in the "middle" of H_1, then label so that $R > C$.

In every case considered, the likelihood ratio test has larger power. The maximum likelihood estimates subject to the restriction in H_1 can be obtained by the iterative algorithm in Dykstra and Robertson (1982) and thus the only impediment to implementing T or E is the level probabilities, which could be estimated by Monte Carlo techniques. If one doesn't wish to obtain such estimates of the level probabilities, Ψ_T or Ψ_E can be used. In the variance known case, the loss in minimum power due to using Ψ_T rather than T ranges from around 5% for the $(3, 3)$, $(3, 5)$ and $(3, 9)$ cases to around 12% for the $(9, 9)$ case. The powers for the two tests with type IV alternatives are very close. For the variance unknown case with small sample sizes of 2, the decreases in minimum power are greater. In particular, the losses are about 14% for the $(3, 3)$, $(3, 5)$ and $(3, 9)$ cases and about 18% for the $(9, 9)$ case. However, part of the differences between the variance known and unknown cases is due to the fact that the powers are smaller. For instance in the case $(3, 9)$, if Δ is increased to 4.3, the estimated powers for Ψ_E and E are 0.803 and 0.893 and $\hat{\pi}(\mu, \Psi_E)/\hat{\pi}(\mu, E) = 0.899$. In fact, in each case considered in Table 2, if Δ is increased to make $\hat{\pi}(\mu, \Psi_E) \approx 0.8$, then the ratios $\hat{\pi}(\mu, \Psi_E)/\hat{\pi}(\mu, E)$ are increased about two or three per cent.

In summary, the likelihood ratio test has greater power than the test developed here, but is much more complicated to use. In many situations, the tests based on Ψ_T and Ψ_E provide satisfactory alternatives to those based on T and E.

6. A numerical example

The use of the test developed here for the case of unknown variance is illustrated on some data concerning the first-year GPA of entering freshmen at the University of Iowa. Because the original data in Dykstra and Robertson (1982) is not available, a variance estimator cannot be computed. Thus, data from 1990 is considered. Table 3 gives the summary statistics for the GPA of the 2511 students when cross-classified according to high school percentile rank (HSR) and composite score on the ACT entrance examination (ACTC). The $R = 8$ intervals for HSR and the $C = 6$ intervals on ACTC were obtained by collapsing some of the intervals in Dykstra and Robertson (1982) so that none of the cells are empty. As they noted, it is reasonable to assume that the mean GPA is nondecreasing in each variable with the other fixed. Thus, one could test H_0 versus H_1 to determine if there is a statistically significant effect due to ACTC and HSR. The first entry in each cell is n_{ij}, the second entry is $\bar{X}_{ij\cdot}$ and the third entry is μ^*_{ij}. In the last column, which is labeled Row Statistics, the entries are $n_{i\cdot}$, $\bar{X}_{i\cdot\cdot}$ and $\mu^*_{i\cdot}$, respectively. The overall mean for the 2511 observations is 2.5857 and the sum of the squared observations is 18,488.4671.

From the SAS procedure GLM, the observed significance level of the F-test for interaction is less than 0.0001. Thus, the simpler test for main effects mentioned

Table 3. First-year GPA of freshmen entering the University of Iowa in the fall of 1990. The students are cross-classified according to high school percentile rank (HSR) and composite score on the ACT entrance examination (ACTC). The first entry in a cell is the number of students in that cell, the second is the average GPA for the students in the cell and the third is μ^*_{ij}, the estimate of the cell mean subject to H_{01}. The entries in the last column are defined similarly except the third entry is $\mu^*_{i\cdot}$.

HSR	ACTC						Row
	1–15	16–18	19–21	22–24	25–27	28–36	Statistics
	1	4	8	9	15	1	38
(0.9, 1.0)	3.8300	1.6950	1.6338	1.3511	1.6653	2.3300	1.6618
	1.6291	1.6291	1.6291	1.6291	1.6653	2.3300	1.6618
	1	7	8	14	28	10	68
(0.8, 0.9]	2.5500	1.8700	2.1037	1.7136	1.9057	1.7310	1.8696
	1.8696	1.8696	1.8696	1.8696	1.8696	1.8696	1.8696
	2	17	23	36	49	12	139
(0.7, 0.8]	1.4900	1.8176	1.7657	1.7642	2.0820	2.1717	1.9142
	1.4900	1.7766	1.7766	1.7766	2.0820	2.1717	1.9142
	11	39	64	59	56	13	242
(0.6, 0.7]	1.8500	2.0741	2.1720	1.9559	2.4407	2.1662	2.1508
	1.8500	2.0698	2.0698	2.0698	2.3890	2.3890	2.1508
	13	35	85	120	90	19	362
(0.5, 0.6]	1.9715	2.2180	2.3796	2.3316	2.2238	2.3805	2.2947
	1.9715	2.2180	2.3125	2.3125	2.3125	2.3805	2.2947
	14	39	116	171	128	44	512
(0.4, 0.5]	1.9057	2.4338	2.4379	2.5776	2.6859	2.4386	2.5318
	1.9057	2.4338	2.4379	2.5776	2.6227	2.6227	2.5318
	3	24	89	166	172	90	544
(0.3, 0.4]	1.6800	2.2854	2.5853	2.7480	2.7666	2.8508	2.7180
	1.6800	2.2854	2.5853	2.7480	2.7666	2.8508	2.7180
	3	9	18	101	209	266	606
(0.0, 0.3]	2.1833	2.6300	3.0794	2.9902	3.0281	3.3453	3.1524
	2.1833	2.6300	3.0037	3.0037	3.0281	3.3453	3.1524

in the introduction would not seem to be appropriate. The value of the test statistic for homogeneity within the rows is $E_1 = 0.04142$ and the value of the test statistic for homogeneity of the row averages is $E_2 = 0.24240$. All of the beta tail probabilities in (3, 4) and (3, 5) are 0 to five decimal places. As was expected, the hypothesis of homogeneity is rejected, and in fact, this is the case for any reasonable value of α. Since the beta tail probabilities are so small there is no need to compute the level probabilities, $P(l, R; n_{1\cdot}, \ldots, n_{R\cdot})$ or $P(l, RC; \boldsymbol{n})$.

Acknowledgements

The authors thank two referees for suggestions which markedly improved this work. We also thank Tim Robertson and Jerald Dallam, the Registrar at the University of Iowa, for the data studied in Section 6.

References

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inferences Under Order Restrictions*, Wiley, New York.

Bartholomew, D. J. (1961). A test of homogeneity of means under restricted alternatives (with discussion), *J. Roy. Statist. Soc. Ser. B*, **24**, 239–281.

Bohrer, R. and Chow, W. (1978). Weights for one-sided multivariate inferences, *Appl. Statist.*, **27**, 100–104.

Dykstra, R. L. and Robertson, T. (1982). An algorithm for isotonic regression for two or more independent variables, *Ann. Statist.*, **10**, 708–711.

Kulatunga, D. D. S. and Sasabuchi, S. (1984*a*). A test of simultaneous homogeneity against ordered alternatives in multifactorial designs, *Comm. Statist. A—Theory Method*, **13**, 3173–3183.

Kulatunga, D. D. S. and Sasabuchi, S. (1984*b*). A test of homogeneity of mean vectors against multivariate isotonic alternatives, *Mem. Fac. Sci. Kyushu Univ. Ser. A*, **38**, 151–161.

Lemke, J. H. (1983). Estimation and testing for two-way contingency tables within order restricted inference parameter spaces, Ph.D. Thesis, Pennsylvania State University.

McDermott, M. P. and Mudholkar, G. S. (1993). A simple approach to testing homogeneity of order constrained means, *J. Amer. Statist. Assoc.*, **88**, 1371–1379.

Mudholkar, G. S. and McDermott, M. P. (1989). A class of test for equality of ordered means, *Biometrika*, **76**, 161–168.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, Wiley, New York.

Sasabuchi, S. and Kulatunga, D. D. S. (1985). Some approximations for the null distribution of the $\bar{E}^2$ statistic used in order restricted inference, *Biometrika*, **72**, 476–480.

Singh, B. and Schell, M. J. (1992). On power functions of the likelihood ratio test for the simple loop order in normal means: Unequal sample sizes, *Statist. Probab. Lett.*, **14**, 253–267.

Singh, B. and Wright, F. T. (1987). Approximations to the powers of some order restricted tests with slippage alternatives, *Biometrika*, **74**, 863-870.

Singh, B. and Wright, F. T. (1988). Two-moment approximations to some null distributions on order restricted inference: Unequal sample sizes, *Canad. J. Statist.*, **16**, 269–282.

Singh, B. and Wright, F. T. (1989). The power functions of the likelihood ratio tests for a simply ordered trend in normal means, *Comm. Statist. Theory Methods*, **18**, 2351–2392.

Sun, H.-J. (1988). A FORTRAN subroutine for computing normal orthant probabilities, *Comm. Statist. Simulation Comput.*, **17**, 1097–1111.