A Set of Nonparametric Tests for Experiments with Lattice-Ordered Means: Theory, Programs and Applications

March 19, 2004

Matthew Strand
National Jewish Medical and Research Center
Division of Biostatistics
1400 Jackson St., M222a
Denver, CO 80206
e-mail: strandm@njc.org
Phone: (303) 398-1862
Fax: (303) 270-2131

Shalender Bhasin
Charles Drew University
Los Angeles, CA

Hsiao-Pin Liu
University of Colorado Health Sciences Center
Denver, CO

Key Words: ordered-restricted inference, Kendall, myostatin, testosterone

Summary

In many factorial experiments where the factors have levels that are ordinal or quantitative, a researcher may predict that the mean response in certain treatments will be higher or lower than those in other treatments. One type of order that may be anticipated is called *lattice order*, where average response tends to increase (or decrease) as the levels of any one of the factors is increased, holding the others fixed. A Kendall-type statistic, which measures the degree of lattice order in the data, can also be used to carry out a test involving lattice-ordered means. In this article, tests for individual factors are developed to complement the overall test of lattice order, and the methods are then applied to relevant and current data. Programs in R and FORTRAN are included to carry out the tests.

1 Introduction

Researchers that perform factorial experiments or observational studies often have a good idea about how some of the treatment means will be ordered. For example, in a 2x2 clinical trial, some aspect of health may be measured on individuals that are taking two types of vitamin supplements (no/yes for each). Taking either one of the supplements is expected to improve health, and taking both is expected to improve health even more. Less certain is how taking supplement "A" only compares with taking supplement "B" only. In an observational study, lung cancer rates may be determined for groups of people classified by smoking status (no/yes) and amount of occupational exposure to hazardous airborne particles (none/low/high). It may be anticipated that an increase in exposure to particles and/or smoking increases the incidence of cancer. It may not be readily clear how the rate for smokers with a lower level of particle exposure compares with the rate for non-smokers with a higher level of particle exposure.

For such experiments and studies, most researchers would probably use ANOVA to analyze the data. This includes an overall test for some difference in treatment means and tests for main and interactive effects of the factors. One shortcoming of this approach is that the anticipated ordering of the responses is not taken into consideration in the analysis. All ANOVA tests are 'two-tailed,' where the hypotheses are set up simply to detect differences between means, as opposed to a specific ordering of means. The assumption that the responses tend to have a certain order suggests that more specific alternatives could be tested. In turn, proper conclusions made from ANOVA are not as specific as they could be.

There are different types of order that can be assumed on the treatment means for an experiment, depending on the nature of the data. For the experiments previously described, the assumption is of *lattice-ordered means* (Strand 2000, Higgins and Bain, 1999). To formally define this for a two-factor experiment, let $\mu_{\mathbf{i}} = \mu_{(i_1,i_2)}$ be the mean for treatment $\mathbf{i} = (i_1, i_2)$, where i_1 denotes level of factor A, $i_1 = 1, \ldots, m_A$, and i_2 denotes level of factor

B, $i_2 = 1, \ldots, m_B$. (Usually the factor levels are defined so that the higher the level, the higher the dosage or amount of that factor.) Let $\mathbf{j} = (j_1, j_2)$ denote another treatment. The relation $\mathbf{i} \leq \mathbf{j}$ means that $i_1 \leq i_2$ and $j_1 \leq j_2$, while $\mathbf{i} < \mathbf{j}$ means the same, but with a strict inequality holding for at least one pair. The treatment means are *lattice ordered increasing* if

$$\mu_{\mathbf{i}} \leq \mu_{\mathbf{j}}, \qquad \mathbf{i} \leq \mathbf{j}. \tag{1}$$

Similarly, the treatment means are *lattice ordered decreasing* if the '$\leq$' between the means in (1) is changed to '$\geq$'. For the previous examples, the assumption would be that of lattice-ordered-increasing means. Figure 1 gives a graphical representation of lattice order.

Figure 1: Illustration of lattice-ordered-increasing means for a 3x4 factorial experiment, using treatment (2,2) as a reference. Shaded means above and/or to the left are no greater than $\mu_{(2,2)}$; shaded means below and/or to the right are no less than $\mu_{(2,2)}$. Unshaded means ($\mu_{(1,3)}$, $\mu_{(1,4)}$ and $\mu_{(3,1)}$) have no hypothesized order with $\mu_{(2,2)}$. The pattern is similar when considering another treatment as a reference.

$\mu_{(1,1)}$	$\mu_{(1,2)}$	$\mu_{(1,3)}$	$\mu_{(1,4)}$
$\mu_{(2,1)}$	$\boldsymbol{\mu_{(2,2)}}$	$\mu_{(2,3)}$	$\mu_{(2,4)}$
$\mu_{(3,1)}$	$\mu_{(3,2)}$	$\mu_{(3,3)}$	$\mu_{(3,4)}$

The following real experiments illustrate when the assumption of lattice-ordered means may be very reasonable. The data for these experiments are analyzed in Section 6.

Application 1: lattice-ordered-decreasing means. Myostatin is a muscle-specific protein that regulates muscle mass in cattle and experimental animals. Cattle and mice with mutations of the myostatin gene have a marked increase in muscularity, suggesting that myostatin is an inhibitor of skeletal muscle mass. These observations have stimulated enormous pharmaceutical and agricultural interest in myostatin because of its potential as a target for the development of drugs that might increase meat production in cattle and improve muscle mass and function in human disease states characterized by muscle wasting such as AIDS wasting syndrome, end stage renal disease, chronic obstructive lung disease, and many types of cancer. However, the mechanisms by which myostatin inhibits muscle mass are unknown.

Taylor, et al. (2001) carried out several experiments to test hypotheses that myostatin inhibits muscle mass by its effects on muscle protein synthesis or degradation. Pure, recombinant myostatin protein was generated in an *in vitro* expression system and its effects on protein synthesis and degradation were examined using $L - [1 -^{14} C]$ leucine pulse labeling

of muscle cells. C2C12 muscle cells in culture were used because this skeletal muscle cell line has been extensively used to characterize the effects of a number of muscle growth factors.

One of the factorial experiments involved measuring protein degradation over three times (24 hr, 48 hr, and 72 hr), and in the absence and presence of myostatin. The rate of decay of $L - [1 -^{14} C]$ leucine served as a marker of the rate of protein degradation. (After cells were incubated and labelled with $L - [1 -^{14} C]$ leucine, they were treated so that no more proliferation of this leucine was possible.) For each treatment, muscle cells were grown in four separate tissue culture wells, thus providing a balanced experiment with $n = 4$. Myostatin and time were labelled as the A and B factors, respectively, so that $m_A = 2$ and $m_B = 3$.

It was anticipated that the $L - [1 -^{14} C]$ leucine levels would decrease over time in both the myostatin and control (no-myostatin) groups. In addition, the myostatin group was expected to have greater protein degradation (i.e. lower $L - [1 -^{14} C]$ leucine levels) than the no-myostatin group, at each time level. In other words, it was assumed that the means would be lattice ordered decreasing. Table 1 contains the data for the experiment and does indeed show a decreasing trend for the values, moving between treatments from left to right and top to bottom.

Table 1: Protein degradation data from the Myostatin experiment (Taylor, et. al., 2001). Degradation was measured by amount of $L - [1 -^{14} C]$ leucine (cpm x 10^3) remaining in muscle cells. Experiment-wide ranks are given in parentheses.

		Time	
Group	24 hours	48 hours	72 hours
Control	6568 (20)	4992 (8)	4092 (4)
	6802 (22)	5242 (11)	4331 (5)
	7198 (23)	5285 (12)	5135 (9)
	7280 (24)	6284 (17)	6087 (16)
Myostatin	5516 (14)	4512 (6)	3076 (1)
	6023 (15)	4706 (7)	3209 (2)
	6334 (18)	5175 (10)	3462 (3)
	6400 (19)	6612 (21)	5364 (13)

Application 2: lattice-ordered-increasing means. Bhasin, et. al. (1996) conducted a 2x2 experiment to determine the effects of anabolic-androgenic steroids (no-placebo / yes-600mg) and exercise (no/yes) on quadriceps muscle volume. It was expected that either exercise or steroid use would increase muscle mass, and that the combination of the two

would be no worse than having just one. However, it was not as certain how the steroid-only group would compare with the exercise-only group. Thus, it was assumed that the means would be lattice ordered increasing. The ranks for percent change in quadriceps muscle volume over the 10-week treatment period (1 was lowest increase, 30 was highest increase) for the subjects in the experiment follow. Placebo, no exercise: 1, 2, 5, 6, 7, 8, 11; placebo, exercise: 3, 4, 9, 10 , 12, 13, 15, 18, 22; testosterone, no exercise: 14, 16, 17, 19, 21, 24; testosterone, exercise: 20, 23, 25, 26, 27, 28, 29, 30. (See also Strand, 2000.) As will be seen forthcoming, ranks are sufficient for the lattice-ordered tests.

The methodology for lattice ordered tests are discussed in Sections 2 through 5, which are then applied to data from the previously described experiments in Section 6, using the associated computer programs. The relationship between interaction and lattice order is discussed in Section 7, among other issues.

2 Measuring the Degree of Lattice Order in the Data

The response for treatment $\mathbf{i}$ and replicate r will be denoted by $Y_{\mathbf{i}r}$, for $r = 1, \ldots, n_{\mathbf{i}}$. The Kendall-type statistic that can be used to measure the degree of lattice order in the data is

$$
\begin{aligned}
L &= \frac{2}{N} \sum_{\mathbf{i}} \sum_{\mathbf{j} > \mathbf{i}} \left(\sum_{r=1}^{n_{\mathbf{i}}} \sum_{s=1}^{n_{\mathbf{j}}} I(Y_{\mathbf{i}r} \leq Y_{\mathbf{j}s}) \right) - 1 \\
&= \frac{1}{N} \sum_{\mathbf{i}} \sum_{\mathbf{j} > \mathbf{i}} \left(\sum_{r=1}^{n_{\mathbf{i}}} \sum_{s=1}^{n_{\mathbf{j}}} [I(Y_{\mathbf{i}r} \leq Y_{\mathbf{j}s}) - I(Y_{\mathbf{i}r} > Y_{\mathbf{j}s})] \right),
\end{aligned} \tag{2}
$$

where

$$
N = \sum_{\mathbf{i}} \sum_{\mathbf{j} > \mathbf{i}} n_{\mathbf{i}} n_{\mathbf{j}} . \tag{3}
$$

The statistic L can be computed if one only has ranks of the original data. The closer L is to $+1$, the more the responses are lattice ordered and increasing, while the closer L is to -1, the more the responses are lattice ordered and decreasing. A value that is near zero indicates no real trend. Thus the scale for L is much like that of other common measures of correlation. In some cases the researcher might expect increasing responses as one factor is increased but the other is decreased. In such a case, measuring lattice order may still be practical and can be done if the levels of one factor are arranged so that the anticipated order of responses is that of either lattice-ordered-increasing or decreasing means.

5

3   The Overall Test for Lattice Order

The statistic L can be used carry out a test for means that are lattice ordered and either increasing or decreasing. For the increasing case, the null hypothesis is that of equal treatment means, and the alternative is that of lattice-ordered-increasing means, minus the case of equal means. It is assumed that all responses are independent and that the response $Y_{\mathbf{i}r}$ has a continuous c.d.f. $F_{\mathbf{i}}$ for all $\mathbf{i}$ and all r. The test is appropriate if

$$F_{\mathbf{i}}(x) \geq F_{\mathbf{j}}(x), \qquad \mathbf{i} \leq \mathbf{j} \tag{4}$$

for all real x. If (4) is satisfied, then it follows that the means are lattice ordered increasing and that all responses have a common distribution under the null hypothesis of equal means. A special case of (4) is when the treatment means follow (1) and there exists some c.d.f. F such that

$$F_{\mathbf{i}}(x) = F(x - \mu_{\mathbf{i}})$$

for all $\mathbf{i}$ and all real x. In this case, all distributions differ only by location.

For tests involving means that are lattice ordered decreasing, the assumption for distributions is the same as (4), but replacing '$\geq$' with '$\leq$'. The null hypothesis is that of equal means, and the alternative hypothesis is that of lattice-ordered-decreasing means, minus the case of equal means.

Under the null hypothesis, all responses are *iid*, in which case the mean and variance of L will be denoted by $E_0(L)$ and $Var_0(L)$, respectively. (Note: when the response variables have continuous c.d.f.'s, as is considered here, ties are not possible and one may replace the '$\leq$' appearing in (2) with '$<$'.) It is easy to verify that $E_0(L) = 0$. The variance of L under H_0 is

$$Var_0(L) = \frac{N + Q}{3N^2}, \tag{5}$$

where

$$Q = \sum_{\mathbf{i}} n_{\mathbf{i}} \left(\sum_{\mathbf{j} < \mathbf{i}} n_{\mathbf{j}} - \sum_{\mathbf{j} > \mathbf{i}} n_{\mathbf{j}} \right)^2$$

and N is as given in (3). When all treatments have n replicates (i.e. the balanced data case), the formulas for N and Q can be expressed in closed form. They are

$$N = n^2 \left(\binom{m_A + 1}{2} \binom{m_B + 1}{2} - m_A m_B \right)$$

and

$$Q = \frac{2n^3}{3}(m_A m_B - 1)\binom{m_A + 1}{2}\binom{m_B + 1}{2}.$$

More general results for $Var_0(L)$ are derived in Strand (2000), for factorial experiments with any number of factors, for both balanced and unbalanced data cases.

If one believes that the means are lattice ordered increasing (minus the equal means case), then the null hypothesis may be rejected for values of L that are sufficiently close to 1. Similarly, one may reject the null for sufficiently small values of L, those close enough to -1, if the alternative hypothesis involves lattice-ordered-decreasing means.

4 Tests for Individual Factors

If the overall test for lattice order is performed, tests for individual factors may still be desired. It should be stressed that even if the null hypothesis for the overall test of lattice order is rejected, that does not imply that all of the individual factors are significant. The statistic L is computed by comparing all unique pairs of responses between treatments that have an expected order on means. The pairs can be classified into three types: pairs between levels of A but within a level of B, pairs between levels of B but within a level of A, and pairs between levels of both A and B. The first two types can be used to construct Kendall-type statistics to test for individual effects of factors A and B. For the rest of this article, results will be derived for the factor A test, and results for B can be derived similarly.

Denoting response $Y_{\mathbf{i}r}$ as $Y_{(i_1,i_2)r}$ and treatment sample size $n_{\mathbf{i}}$ as $n_{(i_1,i_2)}$, a statistic that is built upon all unique pairs of responses between levels of A but within levels of B is

$$L^A = \frac{2}{N^A} \sum_{h=1}^{m_B} \left(\sum_{i_1} \sum_{j_1 > i_1} \left[\sum_r \sum_s I(Y_{(i_1,h)r} \le Y_{(j_1,h)s}) \right] \right) - 1, \tag{6}$$

where

$$N^A = \sum_{h=1}^{m_B} \left(\sum_{i_1} \sum_{j_1 > i_1} n_{(i_1,h)} n_{(j_1,h)} \right). \tag{7}$$

As before, the '$\le$' appearing in (6) can be replaced with '$<$' when the response variables are continuous.

Defining the marginal means for factor A as $\overline{\mu}_{i_1\cdot} = (\sum_{i_2=1}^{m_B} \mu_{(i_1,i_2)})/m_B$, $1 \leq i_1 \leq m_A$, the statistic L^A can be used to test H_0 versus $H_A : H_1 - H_0$, where

$$H_0 : \overline{\mu}_{1\cdot} = \overline{\mu}_{2\cdot} = \ldots = \overline{\mu}_{m_A\cdot}.$$

and

$$H_1 : \overline{\mu}_{1\cdot} \leq \overline{\mu}_{2\cdot} \leq \ldots \leq \overline{\mu}_{m_A\cdot}.$$

(This test would be of interest when (1) is assumed. Similarly, one can use L^A to perform a test for marginal means in the lattice-ordered-decreasing case, for which the inequalities in H_1 above would be reversed.) At first glance, these order-restricted main-effect hypotheses may not seem to be consistent with the statistic that does not directly involve marginal means. However, note that if the treatment means are lattice ordered increasing as in (1), then for any i_1, j_1 and h within the levels tested,

$$\mu_{(i_1,h)} < \mu_{(j_1,h)} \Rightarrow \overline{\mu}_{i_1\cdot} < \overline{\mu}_{j_1\cdot}.$$

Also, if the means are lattice ordered increasing and $\overline{\mu}_{i_1\cdot} < \overline{\mu}_{j_1\cdot}$, then there exists at least one set of i_1, j_1 and h such that $\mu_{(i_1,h)} < \mu_{(j_1,h)}$. Without the lattice order assumption, these clearly would not be true.

As before, it is easy to verify that $E_0(L^A) = 0$. Calculating the variance can be done by employing $Var_0(L)$. Specifically, note that (6) can be rewritten as

$$L^A = \frac{\sum_{j=1}^{m_B} N_j^A L_j^A}{\sum_{j=1}^{m_B} N_j^A},$$

where L_j^A and N_j^A are calculated as in (2) and (3), respectively, but for the $(m_A \text{x} 1)$ data for level j of factor B only. (Note: one-way data can be considered a special case of two-way data, where one of the factors has just one level.)
Now

$$Var_0(L^A) = \frac{\sum_{j=1}^{m_B} (N_j^A)^2 Var_0(L_j^A)}{(\sum_{j=1}^{m_B} N_j^A)^2},$$

where $Var_0(L_j^A)$ is calculated as in (5), but for the $(m_A \text{ x } 1)$ data at the jth level of B only. This result follows because $Var_0(L_j^A), j = 1, \ldots, m_B$ are independent. In the balanced data case where all treatments have n replicates, the variance reduces to

$$Var_0(L^A) = \frac{4n(m_A + 1) + 6}{9m_A(m_A - 1)n^2 m_B}.$$

5 Determining observed levels of significance

There are at least three ways that p-values can be obtained for the statistic L (or L^A or L^B): (i) Obtain the exact permutation distribution of L and find the proportion of permutations that result in values of L that are at least as large as the observed L (for the lattice ordered increasing case), (ii) Same as (i), but using a random sample of permutations, and (iii) use the fact that L is asymptotically normal for increasing N to approximate the p-value. Ager and Brent (1978) show that a generalization of L is asymptotically normally distributed. The fact that L^A (or L^B) is asymptotically normal follows from the fact that it is a linear combination of independent U-statistics, each of which converge to normal distributions.

For all but very small sample sizes, (i) is not practical. Even for a 2x2 experiment with only $n = 3$ per cell, $N = 369600$. Approach (ii) can be used regardless of the size of the experiment, using a fixed number of permutations. The standard error on the estimated p-value in this case would be $\sqrt{\frac{p(1-p)}{n}}$, where n is the number of permutations used and p is the true p-value. The easiest approach is (iii), which yields approximate p-values that are quite accurate for even relatively small sample sizes.

To carry out (iii), one can create a z-statistic that includes a continuity correction, and then find $P(Z \geq z)$. The continuity correction based on the usual approach would be $1/N$. To illustrate why, consider the 2x2 experiment with $n = 1$ in each cell. In this case, $N = 5$ and possible values of L are -1, -0.6, -0.2, 0.2, 0.6 and 1. If the observed value of L is 0.6, then $L - (1/N) = 0.6\text{-}0.2 = 0.4$, which is halfway between possible values. Thus, the Z-statistic would be $Z = [L - (1/N)]/\sqrt{(Var(L))}$. Ager and Brent (1978) mention such a statistic and use $'1/(2N)'$ as the continuity correction (after changing notation). Whether their correction choice was an error or due to other reasons, using either $1/N$ or $1/(2N)$ will yield fairly accurate answers and both are an improvement over not using any continuity correction.

Another approach to (iii) is to simply compute

$$p_z = \sum_{i=c}^{N} f(i) \tag{8}$$

where $f(\cdot)$ is the probability density function of the normal distribution with mean $N/2$ and variance $(N^2/4)Var(L)$, and c is an observed value of $C = N(L+1)/2$, which is the number of pairs of responses that are lattice ordered, among those that can be compared under lattice order. Similar results can be derived for L^A and L^B. The accuracy of this approach is generally at least as good as the approach involving the standardized test statistic, and it is computationally much easier. This approach is like creating a normal probability histogram that approximates C (and hence L), and then adding the bars over the values of c that are at least as big as that which was observed. Note that the width of each bar is 1, and hence

$P(C = i) \approx 1 * f(i) = f(i)$, for $i = 0, \ldots, N$. For left tailed p-values, bars are added over values of c that are no greater than that which was observed. This approach is used in the computer algorithm.

The rates of convergence to normal distributions occur very quickly for these statistics, particularly for L. For a simple 2x2 experiment, with only $n = 3$ in each cell, Table 2 shows exact upper tail p-values with those based on normal approximations using (8), for both L and L_A. For most practitioners, these approximations would be quite sufficient, and they would only get better for experiments involving more replicates.

Table 2a: Comparison of exact and normally approximated p-values for L in a 2x2 experiment, with $n = 3$ for each treatment.

	$P(L \geq x)$ under null	
x	Exact	Normal Approximation
0.333	0.150	0.146
0.422	0.091	0.088
0.511	0.049	0.049
0.600	0.024	0.025
0.689	0.010	0.012
0.778	0.004	0.005

Table 2b: Comparison of exact and normally approximated p-values for L^A in a 2x2 experiment, with $n = 3$ for each treatment.

	$P(L^A \geq x)$ under null	
x	Exact	Normal Approximation
0.333	0.228	0.218
0.444	0.145	0.137
0.555	0.085	0.080
0.666	0.045	0.043
0.777	0.020	0.020
0.888	0.008	0.008
1.000	0.003	0.003

6 Analysis of data from real experiments

Application 1: The lattice order test results for the Myostatin data presented in Table 1 are given in Table 3. The overall test indicates that the treatment means are lattice ordered and decreasing, but not equal ($p < 0.0001$). The individual tests indicate that both factors are associated with this decrease. In other words, mean $L - [1 - {}^{14}C]$ leucine counts decrease over time ($p < 0.0001$), and cells with myostatin exhibited greater protein loss than for the control group, on average ($p = 0.0049$). These test results can be obtained using either the associated R or FORTRAN programs. The raw input and output for this application using either program are given in the Appendix.

Table 3: Results of lattice order tests for the Myostatin experiment. Note: For factor A, 'N' and 'L' represent N^A and L^A, respectively. Similar for B.

Test	N	L	$Var_0(L)$	p-value (normal app.)
Overall	192	-0.76	0.0365	< 0.0001
A=Myostatin	48	-0.67	0.0625	0.0049
B=Time	96	-0.77	0.0405	< 0.0001

For comparison, the results of the ANOVA analysis are given in Table 4. The overall test for some difference in treatment means yielded $p = 0.0004$. The effect of time was also strong ($p < 0.0001$), but the effect of myostatin was shown to be only marginally significant ($p = 0.028$). Since the ANOVA and lattice order tests are based on different hypotheses, one cannot expect the p-values to be the same. In particular, the greater significance for the lattice order tests is likely due to the more restrictive hypotheses, which the data tend to follow.

Table 4: ANOVA test results for the Myostatin experiment.

Test	df	F	p-value
Overall (Model)	5	8.02	0.0004
Myostatin	1	5.74	0.0277
Time	2	16.38	< 0.0001
Myostatin*Time	2	0.82	0.4577

Application 2: Results for the lattice-ordered tests were also significant for the Testosterone and Exercise experiment presented in the Introduction. Table 5 summarizes these results.

Table 5: Results of lattice order tests for the Testosterone and Exercise experiment (Bhasin, et. al., 1996). Note: For factor A, 'N' and 'L' represent N^A and L^A, respectively. Similar for B.

Test	N	L	$Var_0(L)$	p-value (normal app.)
Overall	281	0.89	0.0332	< 0.0001
A=Exercise	111	0.73	0.0485	0.0005
B=Steroids	114	0.98	0.0483	< 0.0001

7 Discussion

While ANOVA is still the most common method of carrying out tests for data from factorial experiments, researchers and analysts should be aware that there are alternatives in certain cases. Not only do tests involving a particular order on the treatment means allow the researcher to make more specific conclusions, more significant results will often occur with such tests relative to ANOVA. The implication is that, in many cases, researchers may be able to use smaller sample sizes in order to detect differences in means. For experiments in which replicates are very costly, this is a very crucial issue. But a word of caution: a good statistician will warn researchers not to look at the data before determining hypotheses for tests on that same data. Otherwise, significant results will be claimed more often than they should be.

There is a partial relationship between lattice order and interaction. Specifically, if the means are lattice ordered, then certain types of interaction cannot occur. In a two-factor experiment, no interaction can be expressed mathematically as $\mu_{(i_1,i_2)} - \mu_{(i_1,j_2)} - \mu_{(j_1,i_2)} + \mu_{(j_1,j_2)} = 0$ for all (i_1, i_2) and (j_1, j_2). Figure 2 illustrates means from hypothetical 2x2 experiments. Panel (d) shows a type of interaction that could occur if the means are lattice ordered. However, the types of interaction shown in panels (a) and (b) could not occur if the means are lattice ordered. Panel (c) shows a case where the means are lattice ordered and there is no interation. (Note: no interaction implies lattice order for 2x2 experiments, but this is not necessarily true for larger experiments.)

Figure 2: Examples of 2x2 experiments illustrating the relationship between lattice order and interaction for treatment means. The values of the means are displayed vertically.
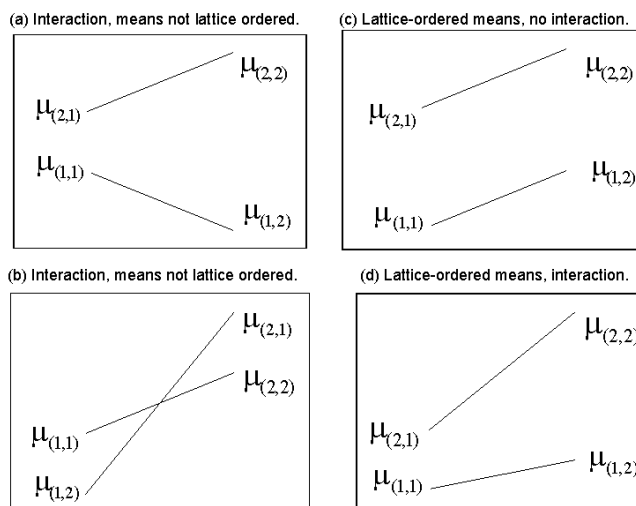
The lattice order assumption will be plausible in many cases, but researchers should use their expertise and discretion in knowing exactly when it is appropriate. Certain types of interaction might spoil the lattice order assumption. To help illustrate, consider a hypothetical 2x2 experiment involving two drugs, both designed to reduce hypertension. Drug A alone may reduce hypertension, as may drug B. However, taking both drugs might have a detrimental effect on subjects. Even if taking both drugs reduces hypertension, compared with taking no drugs, it is possible that taking two drugs is not as effective as taking the best of drug A alone or drug B alone, and in turn, the means are not lattice ordered. If the researcher does have some doubt about the mean response for some factor combinations, then using a testing procedure with fewer order restrictions is probably more appropriate. (When the control group is expected to have lower/higher responses than all of the remaining groups, then a *tree order* is assumed. Tree order has fewer assumed orders among treatments than lattice order.)

Strand (2000) discusses the overall test for lattice-ordered means for factorial experiments involving any number of factors. Generalizations of the tests for individual factors to experiments with more than two factors are straightforward. In the three factor case, a test for factor A can be constructed by looking at the pairs of responses between levels of A but within levels of B and C.

Along with the Kendall-type statistic considered here, a test for lattice-ordered means has also been derived using a Spearman-type statistic (Higgins and Bain, 1999, and Bain, 1994). These nonparametric tests for lattice order are relatively easy to grasp, conceptually, and for which computations of test statistics and variances can be done by hand or with a

simple computer algorithm. Similar tests based on parametric theory (see Robertson, 1988) require more restrictive distributional assumptions and involve calculations that certainly require computer help. Distributions for the test statistics under the null hypothesis of *iid* responses have only been explicitly defined for special cases for these parametric tests. As the test statistics based on either parametric or nonparametric theory can be normalized for even relatively small experiments, the nonparametric tests have the advantage of simplicity.

References

Ager J, Brent S, 1978. An Index of Agreement Between a Hypothetical Partial Order and an Empirical Rank Order, *Journal of the American Statistical Association*; **73**: 827-830.

Bain P, 1994. *Tests of Hypotheses and Measures of Concordance for Lattice-Ordered Means*, Dissertation, Kansas State University, Manhattan, KS.

Bhasin S, Storer T, Berman N, et. al., 1996. The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. *The New England Journal of Medicine*; **335**: 1-7.

Higgins J, Bain P, 1999. Nonparametric Tests for Lattice-Ordered Alternatives in Unreplicated Two-Factor Experiments, *Nonparametric Statistics*; **11**: 307-318.

Robertson T, Wright F, Dykstra R, 1988. *Order-Restricted Statistical Inference.* New York: John Wiley and Sons.

Strand M, 2000. A Generalized Nonparametric Test for Lattice-Ordered Means, *Biometrics*; **56**: 1222-1226.

Taylor WE, Bhasin S, Artaza J, Byhower F, Azam M, Willard DH Jr, Kull FC Jr, Gonzalez-Cadavid N, 2001. Myostatin inhibits cell proliferation and protein synthesis in C2C12 muscle cells. *Am J Physiol Endocrinol Metab*; **280**(2):E221-8.

Appendix

This Appendix shows the Myostatin data input file to be used with the FORTRAN or R programs and the resulting output.

Input file (myostatin.txt):

```
Myostatin Time RankDeg
1 1 20
1 1 22
1 1 23
1 1 24
1 2 8
1 2 11
1 2 12
1 2 17
1 3 4
1 3 5
1 3 9
1 3 16
2 1 14
2 1 15
2 1 18
2 1 19
2 2 6
2 2 7
2 2 10
2 2 21
2 3 1
2 3 2
2 3 3
2 3 13
```

Output:

```
Outcome variable analyzed = RankDeg

----------------------------

Overall test of lattice order
N =  192
COUNT =  23
L =  -0.7604167
VAR(L) =   3.6458332E-02
PVALUE =   3.8157079E-05

----------------------------

Test for row factor (A)
Variable name = Myostatin
N^A =   48
COUNT =   8
L^A =  -0.6666667
VAR(L^A) =   6.2500000E-02
PVALUE =   4.8502577E-03

----------------------------

Test for column factor (B)
Variable name = Time
N^B =   96
COUNT =  11
L^B =  -0.7708333
VAR(L^B) =   4.0509257E-02
PVALUE =   7.8478726E-05
```