PROBABILITY OF DIRECTIONAL ERRORS WITH DISORDINAL (QUALITATIVE) INTERACTION

By

Juliet Popper Shaffer
University of California

Technical Report No. 232
December 1989

Requests for reprints should be sent to Juliet P. Shaffer,
Department of Statistics, University of California, Berkeley, CA 94720.

Department of Statistics
University of California
Berkeley, California

PROBABILITY OF DIRECTIONAL ERRORS WITH DISORDINAL (QUALITATIVE) INTERACTION

Juliet Popper Shaffer

University of California

Requests for reprints should be sent to Juliet P. Shaffer, Department of Statistics, University of California, Berkeley, CA 94720.

Abstract

In a factorial design with two or more factors, there is nonzero interaction when the differences among the levels of one factor vary with levels of other factors. The interaction is disordinal or qualitative with respect to a specific factor, say A, if the difference between at least two levels of A is positive for some and negative for some levels of the other factors. Using standard methods of analysis, there is a potentially large probability of drawing incorrect conclusions about the signs of differences in the presence of disordinal interaction. The maximum probability of such incorrect conclusions, or directional errors, is derived for two-factor designs in which the factor of interest has two levels and the number of levels of the other factor varies.

Key words: factorial design, interaction, disordinal interaction, qualitative interaction, directional errors.

Introduction

Consider a factorial design in which the main interest is in the effects of one factor, say Factor A, under various conditions represented by the other factors. For example, in a two-factor design, we may be comparing two teaching methods (Factor A) in effects on achievement for J ethnic groups (Factor B). Or we may be comparing two methods of treating an illness (Factor A) for J different types of patients (Factor B), varying in age, sex, and so on. Let μ_{ij} be the expected value in the cell with level i of Factor A and level j of Factor B (i.e., the expected achievement in that cell, in the first example above), and define $\delta_j = \mu_{1j} - \mu_{2j}$. If the δ_js are not all equal, we say there is interaction between A and B. The interaction is disordinal with respect to Factor A if δ_j is positive for at least one j and negative for at least one j; otherwise, it is ordinal with respect to A. (Note that interaction can be ordinal with respect to one factor but disordinal with respect to others.) In a general factorial design with an arbitrary number of factors, interaction is disordinal with respect to a factor, say A, if there exist two levels of A for which the difference is positive for some combinations of levels of Factors B, C,..., and negative for other combinations of levels of those factors. In the biological sciences, the term qualitative or crossover interaction is used for the same concept.

Although the definition of disordinal interaction can easily be extended to a wider range of models, at least to all generalized linear models, and the issues it raises generalize similarly, the present paper will deal in detail only with a two-factor design to which a fixed-effects analysis of variance model applies. The standard way of writing that model is:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \tag{1}$$

$i = 1, \ldots, I$; $j = 1, \ldots, J$; $k = 1, \ldots, K_{ij}$; where the random error vector $\underset{\sim}{\varepsilon} \sim N(0, \sigma^2 I)$, and with the side conditions $\sum_i \alpha_i = \sum \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$. Designating the expected value of Y_{ijk} as μ_{ij}, where $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, the side conditions imply that μ is the unweighted mean of the μ_{ij}. Let H_A be the hypothesis that the main effects of A (the α_i s) are 0, and let H_{AB} be the hypothesis that the AB interactions (γ_{ij} s) are 0. If $I = 2$, as in the educational and medical examples discussed above, H_{AB} is equivalent to the hypothesis that the δ_js are equal for all j, and H_A is equivalent to the hypothesis that the average $\delta_j = 0$. The quantities $\mu_{ij} - \bar{\mu}_j$ are the simple effects of Factor A, where $\bar{\mu}_j = \sum_i \mu_{ij} / I$, and the average over j of the simple effects are the α_is, the main effects of A.

In analysis of variance of factorial designs, it is customary to test for interactions and, if no interactions involving a specified factor are significant, to assume there is no interaction with respect to that factor--to assume, in the example above, that all δ_js are equal to the average difference $\bar{\delta}$. Of course, it is recognized that acceptance of the null hypothesis of zero interaction doesn't mean that interaction is really equal to zero. In many cases, a small or moderate interaction, if not detected, would result in some bias in the estimated effects of the relevant factor, but that might not make much difference as far as the inferences or decisions based on the analysis are concerned. However, the consequences of an undetected disordinal interaction are likely to be more serious than the consequences of an undetected ordinal interaction of the same size. In the examples given above, if we assumed Teaching Method 1 was better than Method 2 for all ethnic groups, while in fact it was worse for some, or assumed Treatment 1 was better than Treatment 2 for all patient types when this was not the case, the consequences could be highly undesirable. Such disordinal interactions have been noted in both the biological-medical and social science-educational literatures (Simon, 1980; Cronbach & Snow, 1977). In deciding whether or not to use the standard approach, it would be helpful to know how large the resulting probability of an erroneous conclusion about the direction of an effect can be. Such erroneous conclusions will be called directional errors. In other words, a directional error is made if it is concluded that a simple effect of some level of A is positive when it is really negative, or vice versa.

Neyman, in a comment on a paper on factorial design presented before the Royal Statistical Society by Yates (Neyman, 1935; Yates, 1935), considered this problem. Yates gave as an example a $2 \times 2 \times 2$ factorial experiment that had been carried out to assess the effects of all combinations of three fertilizers on yield of peas. Because interactions were small, Yates felt that the direction of the significant main effects of two of the fertilizers could be assumed to hold under all combinations of the others. Neyman questioned this assumption, and gave an example of possible true yields for which Yates' analysis could lead with high probability to recommending a procedure that was actually harmful. Traxler (1976) verified and extended Neyman's results, and they were further generalized by Bohrer and Sheft (1979).

Of course, the problem can be avoided by analyzing the simple effects of the crucial factor, that is, the factor's separate effects for each set of values of the other factors, and using an appropriate multiple comparison method to control the overall error rate. However, with that approach, the factorial design loses its advantage in efficiency over one-factor-at-a-time designs, since the power for testing simple effects is much lower than power for testing the average effect, particularly when the multiple comparison nature of the problem is taken into account.

The three studies mentioned (Neyman, Traxler, and Bohrer & Sheft) all considered a three factor, 2^3 design as first investigated by Neyman. The results are rather complicated and difficult to grasp intuitively, since there are three possible interactions of Factor A with the other factors--AB, AC, and ABC--to be considered. In contrast, in a two-factor design, the situation is much simpler, and the problem can be considered in greater detail. Some results on error probabilities for a two-factor design will be given in this paper.

The analysis will be restricted to a balanced two-factor design: $K_{ij} = K$ for all i,j, with two levels of Factor A, analyzed under the standard analysis of variance model (1). Suppose the hypothesis H_{AB} is tested at some level α'. If the AB interaction is not significant, H_A is then tested at level α. If the main effect is significant, it is concluded that the difference between the two levels of A is nonzero and is in the same direction for all levels of B. Given this sequence, what is the maximum probability of a directional error, and how does it vary with α'?

The Supremum of the Probability of a Directional Error

Suppose the main effect of level 1 of Factor A is $\underline{a}$, which without loss of generality will be taken as nonnegative. Then SS_A / σ^2 is distributed as χ^2_{1,λ_A} (where SS_A is the sum of squares for the main effects of Factor A, and $\chi^2_{v,w}$ is the chi-square distribution with degrees of freedom v and noncentrality parameter w), with the noncentrality parameter $\lambda_A = (2JK\underline{a}^2)$, or $\underline{a} = (\lambda_A / (2JK))^{1/2}$. What is desired, then, is

$$\sup_{di} \mathrm{Prob}_{\lambda_A} [\text{ Accepting } H_{AB} \text{ at } \alpha' \text{ and rejecting } H_A \text{ at } \alpha],$$

or

$$\sup_{di} \mathrm{Prob}_{\lambda_A, \alpha', \alpha} [\, T_{AB} < T_{AB;\alpha'} \text{ and } T_A > T_{A;\alpha} \,], \tag{2}$$

where T_{AB} and T_A are the test statistics for testing H_{AB} and H_A with level-α' and level-α critical values $T_{AB;\alpha'}$ and $T_{A;\alpha}$, respectively, and the supremum is taken over all configurations of the μ_{ij} for which there is disordinal interaction (di).

If σ^2 were known, T_A (T_{AB}) would be the test statistic SS_A / σ^2 (SS_{AB} / σ^2), distributed as χ^2_{1, λ_A} ($\chi^2_{J-1, \lambda_{AB}}$). If σ^2 were not known, T_A (T_{AB}) would be the test statistic MS_A / MS_W (MS_{AB} / MS_W), distributed as F_{1, v, λ_A} ($F_{J-1, v, \lambda_{AB}}$), where $F_{u,v,w}$ is the F distribution with numerator degrees of freedom u, denominator degrees of freedom v, and noncentrality parameter w, and the variance is estimated with df = v. Since for known σ^2 or fixed MS_W, T_{AB} and T_A are independent, and T_{AB} is stochastically increasing in λ_{AB}, the supremum (2) will be attained for the disordinal interaction that results in the minimum noncentrality parameter λ_{AB}. Disordinal interaction requires $\min_j \delta_j < 0$, under the constraint that the average $\delta_j = 2a$. A set of cell means satisfying these conditions is: all J cell means for Level 2 of Factor A are 0; one cell mean for Level 1 of Factor A is barely smaller than 0; and the other J cell means for Level 1 of Factor A are equal to $2aJ / (J - 1)$. For the derivation of this set of means, see the Appendix.

Since the noncentrality parameter for A, λ_A, equals ($2JK\underline{a}^2$), it follows from the configuration described above that the noncentrality parameter λ_{AB}, which equals ($2JK\underline{a}^2 / (J - 1)$), is equal to $\lambda_A / (J - 1)$.

The implications of this result for the probability of directional errors will be given first in detail for J = 2, followed by a briefer description of results for J = 3. It will initially be assumed that σ^2 is known, since the probabilities of directional errors for known σ^2 should be good approximations to those that obtain with large degrees of freedom ν.

J = 2

Variance known. Since T_A and T_{AB} (in (2)) are independent with σ^2 known, the required supremum is given, for fixed λ_A, by

$$\mathrm{Prob}_{\lambda_A}[T_A > \chi^2_{1;\alpha}]\ \mathrm{Prob}_{\lambda_A}[T_{AB} < \chi^2_{1;\alpha'}], \tag{3}$$

where T_A and T_{AB} are iid. noncentral chi-square with 1 df and noncentrality parameter λ_A, and $\chi^2_{\nu;\gamma}$ is the level-γ critical value of a central chi-square distribution with df of ν. Note that as $\lambda_A \to 0$, the product (3) approaches $\alpha(1-\alpha')$, while as $\lambda_A \to \infty$, (3) approaches zero, and these same limiting values would apply for all J. Furthermore, for any λ_A, the probability will clearly be minimum over J for J = 2. From the results for J = 2, described in this section, it then follows that the maximum is achieved for some nonzero finite value of λ_A for all values of J.

Results were computed for $\alpha = .05$ and α' varying from .05 to .95. In each case, the supremum of the probability of a directional error appears to be a unimodal function of λ_A, and thus, has only a single relative maximum. Although an analytic proof of unimodality has not been found for arbitrary α', it is easily proved for $\alpha' = \alpha$. The probability of a directional error can then be expressed as

$$\pi_{\lambda_A}(1 - \pi_{\lambda_A}), \tag{4}$$

where π_{λ_A} is the probability that a random variable distributed as noncentral χ^2_{1,λ_A} is greater than $\chi^2_{1;\alpha}$, that is, π_{λ_A} is the power of the tests of H_A and H_{AB}. By (4), the supremum of the probability of a directional error as a function of π_{λ_A} when $\alpha' = \alpha$ is therefore unimodal, reaching a maximum of .25 when $\pi_{\lambda_A} = .50$. Since π_{λ_A} is a monotonic increasing function of λ_A, unimodality with respect to λ_A follows.

Using (3), the supremum of the probability of directional errors was calculated for different values of α' and λ_A using the GAUSS statistical software system on an IBM PC. Table 1 gives the supremum over all λ_A of the probability of a directional error for $\alpha = .05$, as α' varies from .05 (.05) .95, and the values of $\beta_A = (\lambda_A)^{1/2}$ for which the supremum is achieved. (Results are reported in terms of β_A, the noncentrality parameter used by Fisher, because it is proportional to the main effect of Factor A.) In Figure 1, the supremum of the probability of a directional error is plotted for varying β_A, for $\alpha = .05$ and $\alpha' = .05, .25$, and .50.

Insert Table 1 and Figure 1 about here

Note that, as mentioned above, all the curves in Figure 1 are unimodal. The value of the noncentrality parameter for which the supremum is achieved appears to decrease with α'. The supremum, of course, becomes smaller as α' increases, but remains above .05 even when H_{AB} is tested at level .50. The supremum over all λ_A at $\alpha' = .50$ is .062; it reaches .05 only when α' becomes as large as .58. The overall supremum for $\alpha' \geq \alpha$ is at $\alpha' = \alpha$ and, as pointed out above, is .25, a result which is independent of the value of α.

Variance unknown. Since the variance known case provides a good approximation to the supremum of the probability for large sample sizes, the supremum with unknown variance was considered in addition for the *very small* sample size K = 3, giving degrees of freedom 8 for the estimates of σ^2. Figure 2 shows the relations between the supremum with known variance and the supremum with variance estimated with 8 df, for $\alpha' = .05$. The comparison was also carried out for $\alpha' = .25$ and $\alpha' = .50$. For $\alpha' = \alpha$ it follows from the positive quadrant dependence of T_{AB} and T_A (Lehmann, 1966, Example 1, Part (iv)) that the supremum is maximum when the variance is known (though not necessarily for a fixed λ_A as can be seen). This conclusion follows from the fact that the joint probability (2) is bounded above by the product of the probabilities $\pi_1 = \text{Prob}\,(T_{AB} < T_{AB;\alpha'})$ and $\pi_2 = \text{Prob}\,(T_A > T_{A;a})$, where T_{AB} and T_A are identically distributed and therefore $T_{AB;\alpha'} = T_{A;\alpha}$ for $\alpha' = \alpha$. Since π_2 then equals $1 - \pi_1$, the product is $\pi_1\,(1 - \pi_1)$ and is bounded above by .25, the supremum for known variance.

Insert Figure 2 about here

Despite the very small value of 8 for the df, the results are reasonably close to those in the asymptotic case of known variance. The supremum for $\alpha' = .05$ is .21, as compared with .25 when σ^2 is known, and in each case the supremum appears to be moderately smaller for unknown than for known variance. It seems clear that with sample sizes of 5 or above (df=16 or higher), the values would be very close to those for the known variance case.

J = 3

For J = 3, the supremum of the probability of directional error was investigated only for σ^2 known. Table 2 gives the supremum as a function of λ_A for $\alpha = .05$ and $\alpha' = .05\ (.05)\ .95$, with the $\beta_A = (\lambda_A)^{1/2}$ for which the supremum is achieved. The overall supremum, achieved at $\alpha' = \alpha$, is .476, considerably higher than the .25 for $J = 2$. The supremum is .128 for $\alpha' = .50$, and doesn't become as small as .05 until α' reaches .77. Figure 3 shows the supremum as a function of β_A for $\alpha' = .05$, .25, and .50. Comparison with Figure 1 for $J = 2$ shows that, although the numerical values are different, the shapes of the curves are remarkably similar for $J = 2$ and $J = 3$.

Insert Table 2 and Figure 3 about here

Conclusion

The results given here indicate that the standard approach to the treatment of interaction in analysis of variance designs can result in probabilities of directional errors of a magnitude too large to be acceptable in many situations, even in the best case in which $J = 2$. (It is noteworthy that in a recent investigation of a somewhat analogous issue in estimation, Fabian (1989) came to similarly pessimistic conclusions about standard analysis of variance practice.) When $\alpha' = \alpha$, the probability can be as large as .25 for $J = 2$ and .48 for $J = 3$. Increasing the level (α') of the test for interaction reduces the probability of error, but is of limited effectiveness for two reasons: (i) the reduction to acceptable levels requires a drastic increase in α', and (ii) the procedure that has been investigated is, of course, only part of a strategy, since if the main effect and interaction are both significant, further testing must be carried out, adding to the probability of directional errors, and the probability that such further tests will be needed increases as α' increases.

A more direct attack on the problem would be to test the hypothesis that the interaction is ordinal, rather than testing the hypothesis of no interaction. A maximum likelihood test for this hypothesis was developed by Gail and Simon (1985), who propose a strategy of assuming all effects are in the same direction (assessed by the main effect) if the hypothesis is accepted. Unfortunately, the probability of directional errors is even higher under this strategy than under the standard strategy, so that it should be used only if it seems highly unlikely that disordinal interaction is present. An intermediate approach is taken by Azzalini and Cox (1984), who developed a test for the hypothesis of zero interaction with power higher under disordinal than under ordinal interaction. The probability of directional error using their test, while smaller than that under the Gail and Simon procedure, is still much higher than under the standard strategy. Berger (1984) developed a likelihood ratio test for the hypothesis that the interaction is disordinal, and for $J = 2$ the Berger and Gail-Simon tests have been improved by Berger (1989) and Zelterman (1989). Further work to develop a satisfactory strategy for minimizing the probability of directional errors, without undue sacrifice of power, is necessary, both in the specific designs investigated here and in the much wider class of designs for which these considerations are relevant.

References

Azzalini, R., & Cox, D.R. (1984). Two new tests associated with analysis of variance. *Journal of the Royal Statistical Society, 46*, 335-343.

Berger, R. (1984). Testing for the same ordering in several groups of means. In T.J. Santner, & A.C. Tamhane (Eds.), *Design of experiments: Ranking and selection* (pp.241-249). New York: Dekker.

Berger, R. (1989). Uniformly more powerful tests for hypothesis concerning linear inequalities and normal means. *Journal of the American Statistical Association, 84*, 192-199.

Bohrer, R., & Sheft, J. (1979). Misclassification probabilities in 2^3 factorial experiments. *Journal of Statistical Planning and Inference, 3*, 79-84.

Cronbach, L.J., & Snow, R.E. (1977). *Aptitudes and instructional methods.* New York: John Wiley.

Fabian, V. (1989). On the problem of interactions in the analysis of variance. *Journal of the American Statistical Association*, in press.

Gail, M., & Simon, R. (1985). Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics, 41*, 361-372.

Lehmann, E.L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics, 37*, 1137-1153.

Neyman, J. (1935). Complex experiments: Comments by Dr. Neyman on a paper by F. Yates. *Journal of the Royal Statistical Society Supplement, 2*, 235-242.

Simon, R. (1980). Patient heterogeneity in clinical trials. *Cancer Treatment Reports, 64*, 405-410.

Tong, Y.L. (1980). *Probability inequalities in multivariate distributions.* New York: Academic Press.

Traxler, R.H. (1976). A snag in the history of factorial experiments. In D.B. Owen (Ed.), *On the history of statistics and probability* (pp.281-295). New York: Dekker.

Yates, F. (1935). Complex experiments. *Journal of the Royal Statistical Society Supplement, 2*, 181-223.

Zelterman, D. (1989). On tests for qualitative interactions. *Statistics and Probability Letters*, in press.

Appendix

Determining a Configuration of Cell Means which Maximizes the Probability of a Directional Error

Since T_A and T_{AB} are independent with σ^2 known, and $\mathrm{Prob}_{\lambda_A}([T_A > T_{A;\alpha}]$ is not affected by the interaction, we must find

$$\sup_{di} \mathrm{Prob}_{\lambda_A}[T_{AB} < T_{AB,\alpha'}], \tag{5}$$

where T_{AB} is distributed as noncentral $\chi^2_{1,\lambda_{AB}}$, so the supremum in (5) will be attained for the disordinal interaction that results in the minimum noncentrality parameter λ_{AB}, regardless of the value of α'. Disordinal interaction requires $\min_j \delta_j < 0$, under the constraint that the average $\delta_j = 2\underline{a}$ (for the main effect of Level 1 of Factor A to equal $\underline{a}$), and we want to find a configuration with these properties that minimizes the interaction noncentrality parameter. Such a configuration is found as follows.

Start with any set of means μ_{ij}. Since the effect of B is irrelevant, constants can be added to the columns as desired, so that without loss of generality, set $\mu_{1j} = 0$ for all j. Since $\bar{\delta}$ is assumed positive, for disordinal interaction δ_j must be < 0 for some j; without loss of generality, make that j = 1. To obtain the infimum of the disordinal interaction, set $\delta_1 = 0$ (i.e., $\mu_{21} = 0$); if it were any other value, the interaction could be decreased by increasing δ_1 towards zero and reducing all other δ's proportionately. Under these conditions, by use of Lagrange multipliers, it is easily seen that the infimum of the interaction occurs when $\mu_{2j} = 2\underline{a}J/(J-1))$, $j = 2, \ldots, J$. Alternatively, the result follows from the fact that (a) interaction is a Schur-convex function of the difference vector $\tilde{\delta}$, and (b) the vector $(0, \delta, \delta, \ldots, \delta)$ is majorized by all other vectors $\tilde{\delta}$ satisfying the constraint that the minimum component must be < 0.

Figure captions

Figure 1. Supremum of the probability of a directional error with known variance, $J = 2$, $\alpha = .05$, α' varying from .05 (upper curve) to .50 (lower curve).

Figure 2. Supremum of the probability of a directional error with known variance (solid line) and estimated variance (dashed line), $\alpha = .05$, $\alpha' = .05$.

Figure 3. Supremum of the probability of a directional error with known variance, $J = 3$, $\alpha = .05$, α' varying from .05 (upper curve) to .50 (lower curve).

TABLE 1

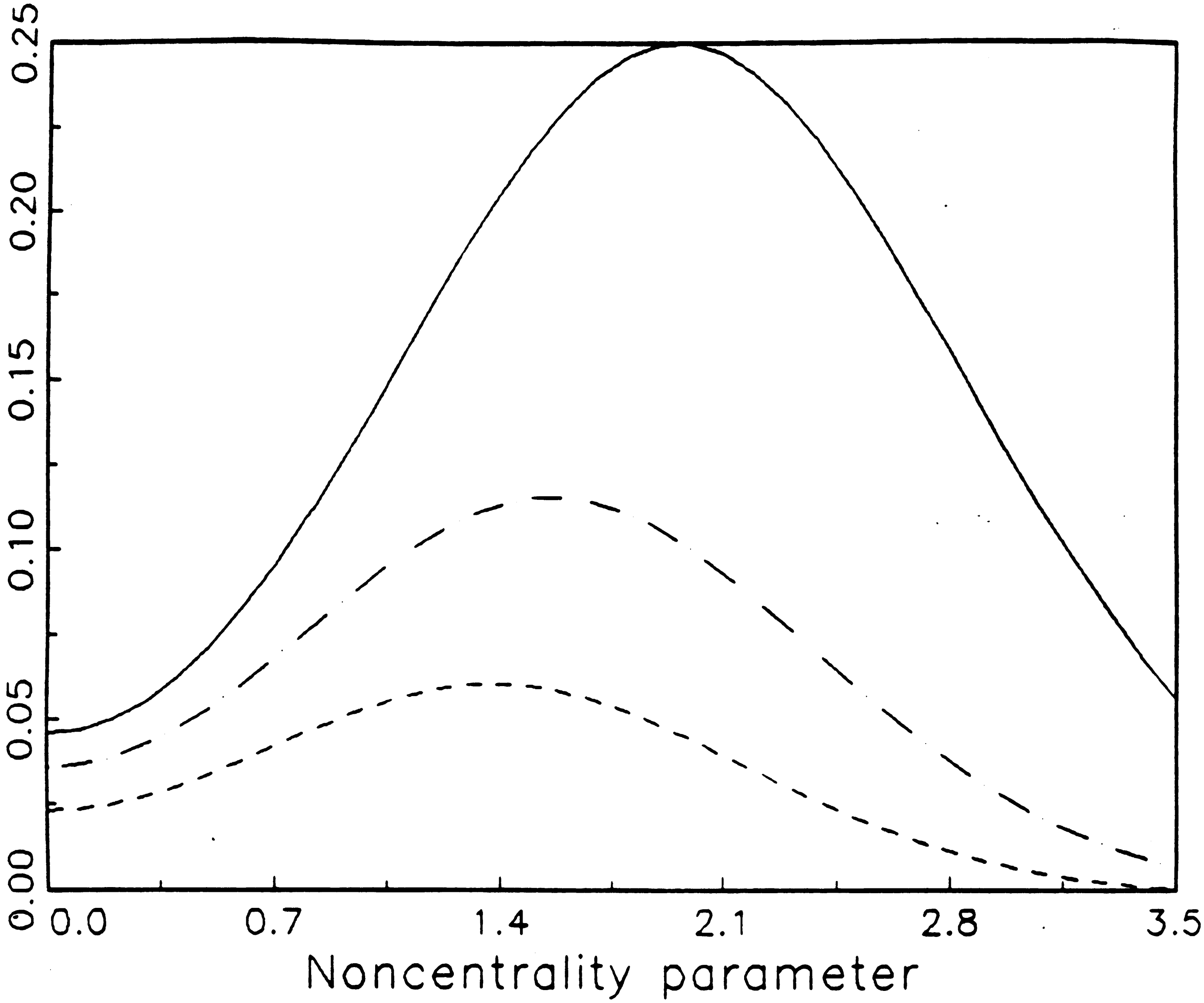
Supremum of Probability of Directional Error as a Function of α' for $\alpha = .05$, $J = 2$.

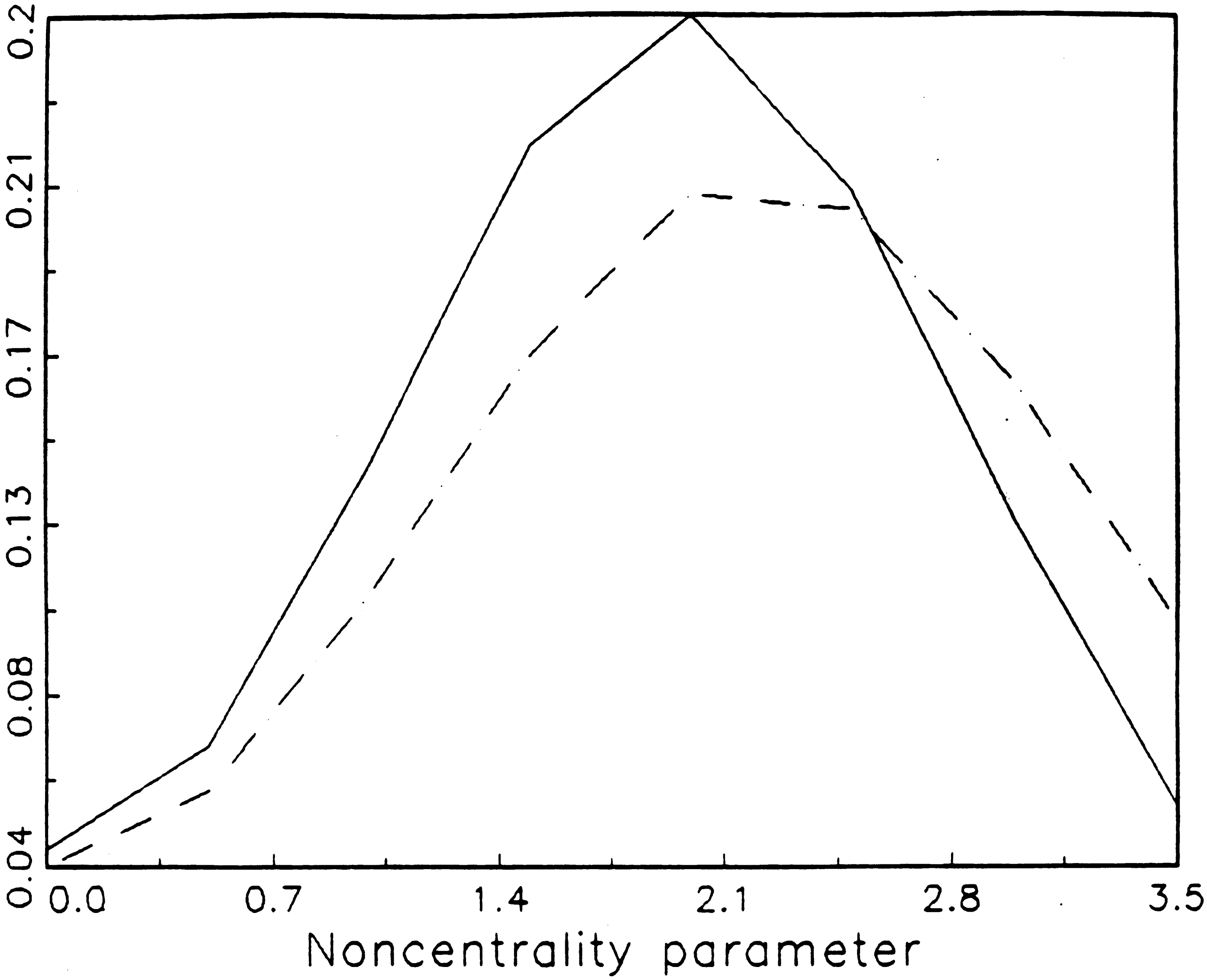
α'	Supremum	Value of β_A for Which Supremum is Attained
.05	.250	1.96
.10	.191	1.80
.15	.158	1.70
.20	.134	1.63
.25	.116	1.57
.30	.102	1.52
.35	.090	1.47
.40	.080	1.44
.45	.070	1.41
.50	.062	1.38
.55	.054	1.36
.60	.047	1.34
.65	.041	1.32
.70	.034	1.31
.75	.028	1.30
.80	.022	1.29
.85	.017	1.28
.90	.011	1.27
.95	.005	1.27

TABLE 2

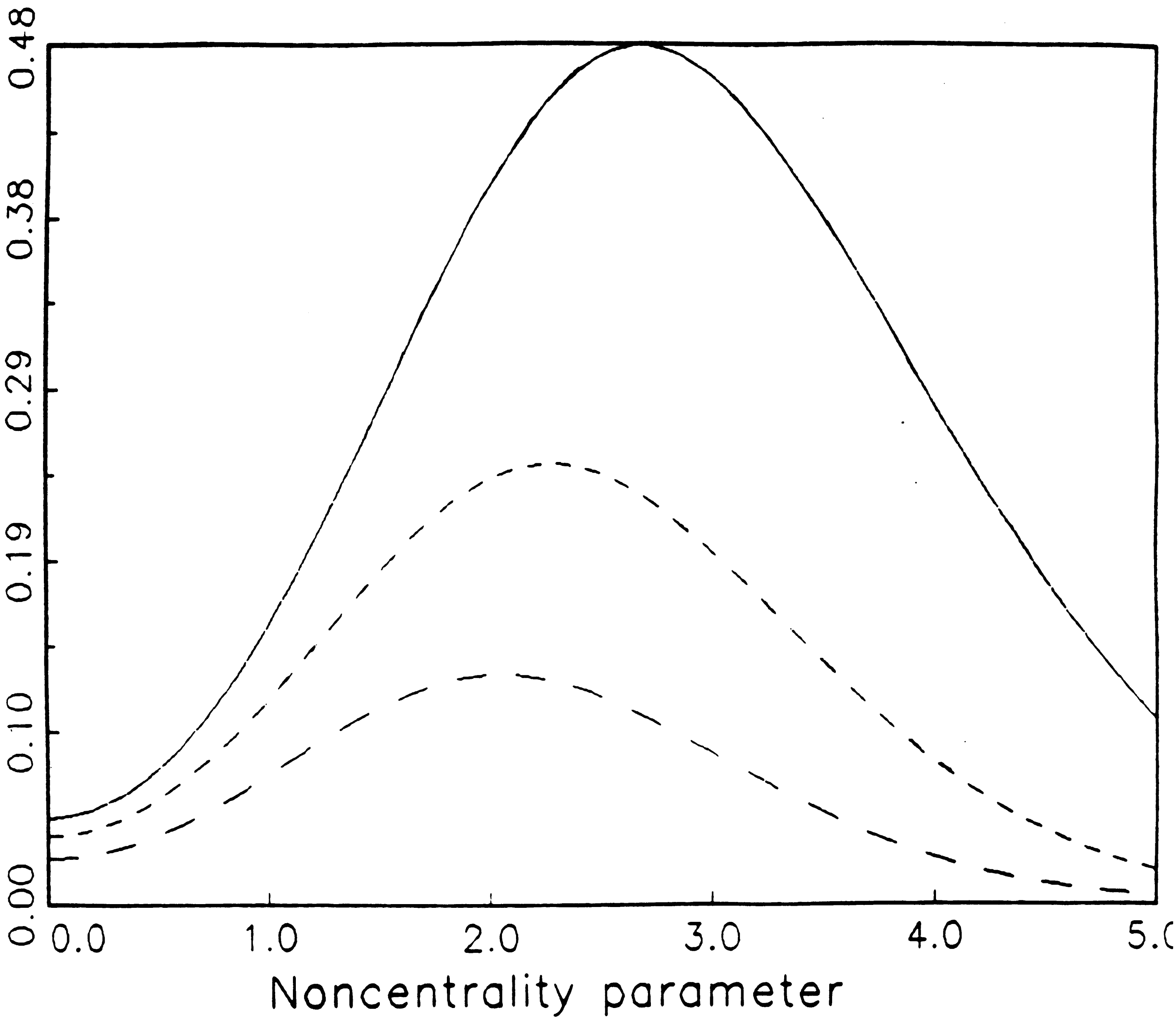
Supremum of Probability of Directional Error as a Function of α' for $\alpha = .05$, $J = 3$.

α'	Supremum	Value of β_A for Which Supremum is Attained
.05	.476	2.68
.10	.383	2.52
.15	.324	2.42
.20	.280	2.34
.25	.244	2.27
.30	.215	2.22
.35	.189	2.17
.40	.166	2.13
.45	.146	2.09
.50	.128	2.05
.55	.111	2.02
.60	.095	1.99
.65	.081	1.96
.70	.067	1.94
.75	.054	1.91
.80	.042	1.89
.85	.031	1.86
.90	.020	1.84
.95	.010	1.82


0.00
0.05
0.10
0.15
0.20
0.25
0.0
0.7
1.4
2.1
2.8
3.5
Noncentrality parameter



0.25
0.21
0.17
0.13
0.08
0.04
0.0
0.7
1.4
2.1
2.8
3.5
Noncentrality parameter


0.00
0.10
0.19
0.29
0.38
0.48
0.0
1.0
2.0
3.0
4.0
Noncentrality parameter