

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1990 - 2nd Annual Conference Proceedings

ORDERED ALTERNATIVES: A MEANS OF IMPROVING POWER

John E. Boyer Jr.

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Boyer, John E. Jr. (1990). "ORDERED ALTERNATIVES: A MEANS OF IMPROVING POWER," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1444>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

ORDERED ALTERNATIVES: A MEANS OF IMPROVING POWER

by

John E. Boyer, Jr.
Department of Statistics
Kansas State University
Manhattan, Kansas

ABSTRACT

In analysis of variance settings it is often known that if there are any differences among the means, those differences will fall in a particular order. The usual F test used to look for the existence of differences is not sensitive to the particular order. This paper presents two procedures from the nonparametric literature which have sensitivity to the suggested ordering. The analogy is drawn between these procedures and the two-sample t test. The paper concludes with a simulation study which investigates the power properties of the proposed tests and makes comparisons with the F test.

Key words: Ordered alternatives, Jonckheere statistic

1. INTRODUCTION

In the two sample problem, one often tests the hypothesis $H_0: \mu_1 = \mu_2$ against a one-tailed alternative $H_a: \mu_1 < \mu_2$, knowing (or at least believing) that if the two means are not equal, the first treatment will produce a smaller mean than the second. In this case one would not hesitate to use the one-tailed version of the t test. In the analysis of variance problem where one is testing for the equality of the means of t treatments however, the hypotheses that are tested are invariably $H_0: \mu_1 = \mu_2 = \dots = \mu_t$ versus H_a : at least two means not equal. The test statistic used is, of course, the usual F statistic.

There are situations in which the researcher knows that if the null hypothesis specified above is not true, the alternative will be in the form $\mu_1 \leq \mu_2 \leq \dots \leq \mu_t$ with at least one of the inequalities strict. Examples that quickly come to mind include those in which the subscripts 1, 2, ... t reflect ordered amounts of some quantity (but for which the usual regression models do not seem like a reasonable choice). For instance, a biologist, in studying the desirability of freshness as a quality of food for rodents, presents them with sources of food that are three, six, nine, twelve and fifteen days old and compares the amounts consumed. If freshness of the particular food provided is a criterion for the animals, the experimenter expects to see larger amounts of the freshest foods consumed. On the other hand, if freshness is not a criterion, then relatively equal amounts of the sources of food would be expected to be used.

The classical statistical literature does not provide a procedure for the analysis of variance which is analogous to the one-tailed test in the two sample problem. The nonparametric literature does provide several such tests. Two of the simplest are presented in the next section, with motivation as to why they are sensitive to the alternatives of interest. An example showing the type of data which might fail to be rejected under the F test, but for which the nonparametric tests find significance is given in the following section. A power study showing the advantages of these procedures is described in the fourth section of the paper. The final section reports the conclusions and some recommendations.

2. CONSTRUCTION AND MOTIVATION

Alternative hypotheses of the form $H_a: \mu_1 \leq \mu_2 \leq \dots \leq \mu_t$ are referred to in the nonparametric literature as "ordered alternatives". There has been considerable work done on the problem for the usual parametric case (see Barlow et al (1972) or Robertson et al (1988)) and the resulting procedures are termed isotonic regression. It has been demonstrated that the resulting techniques yield significant gains in power over the F test. (See, for example, Gaines and Rice (1990).) Unfortunately the methods presented there are computationally quite involved.

Jonckheere (1954) and Terpstra (1952) presented the test procedure now known in the nonparametric literature as Jonckheere's statistic and henceforth denoted in this paper as J. If W_{ij} is the Mann-Whitney statistic for testing for differences between groups i and j then J is defined to be

$$\sum_{i < j} W_{ij} \quad (1)$$

Clearly this statistic should be sensitive to alternatives of the type described above, as each of the terms in the sum ought to tend to be large (or small, depending on how the statistic is calculated) when the treatments are aligned in the order specified by the alternative. Notice also that if the null hypothesis is not true, but the treatment means follow an order other than that specified by the alternative, some of the terms in the sum will be large and some will be small. Thus the resulting test statistic will not take on an extreme value, and consequently the null hypothesis will not be rejected. Obviously, an equivalent test can be constructed from the corresponding Wilcoxon rank sum statistic.

It should be pointed out that under the null hypothesis, J is distribution free and that some tables of the values of J exist for equal numbers of observations in each treatment group and for small numbers of treatments (see Hollander and Wolfe, 1973). In most circumstances, however, a normal approximation to the distribution of J gives probabilities which are sufficiently close to the exact values. For the case of equal numbers of observations in each treatment group (where t denotes the number of treatments, o denotes the number of observations per treatment, and $N = t * o$ is the total number of observations) the mean of J under the null hypothesis is given by

$$E(J) = N(N - o)/4 \quad (2)$$

and the variance is given by

$$\text{Var}(J) = N o (t(2N + 3) - (2o + 3)) / 72. \quad (3)$$

The corresponding formulas for the case in which the number of observations varies with the treatments can be found in Hettmansperger (1984).

Another statistic provided to test for ordered alternatives is given by

$$H^* = \sum_{i=1}^t (i - (t + 1) / 2) (\bar{R}_i - (N + 1) / 2). \quad (4)$$

Here the \bar{R}_i denotes the average rank for the ith group when all the items are ranked (as in the Kruskal-Wallis statistic). It can be shown that under

$$H_0, E(H^*) = 0 \quad (5)$$

and

$$\text{Var}(H^*) = \frac{t^2(t^2 - 1)}{144} N + 1 \quad (6)$$

This is one of a class of tests proposed by Tryon and Hettmansperger (1973). In Hettmansperger

(1984) it is shown that the statistic H^* is equivalent to $H = \sum_{i < j} (j - i) W_{ij}$ and that is the form

that has been utilized in this paper for computations. It follows that H is really a generalization of J, in that it is a weighted sum of the Mann-Whitney statistics which go into J. In this form, it can be shown that the mean of H under the null hypothesis is given by

$$E(H^*) = N o (t^2 - 1) / 12 \quad (8)$$

and the variance is given by

$$\text{Var}(H) = N^2 (t^2 - 1) (N + 1) / 144. \quad (9)$$

From the H^* version of H , it is clear that if the ranks of the groups tend to increase as the subscripts on the groups increase, the statistic H will be large. In fact, this expression for H looks very similar to a correlation coefficient, computed between the observed ranks and the hypothesized ordering of the treatments. As with J , the usual way of performing the test is to standardize the test statistic and compare to critical points of the normal distribution.

3. AN EXAMPLE

An entomologist collects data on four treatments for which it is expected that the means, if not equal, will be in an order which increases with the subscript. The data are given in the table below.

<u>Group 1</u>	<u>Group 2</u>	<u>Group 3</u>	<u>Group 4</u>
36.2	44.0	57.1	58.7
43.3	49.1	57.3	67.2
56.8	62.0	64.1	68.0
62.6	69.6	71.0	74.6

When the usual analysis of variance is computed for this data set, an observed F value of 2.48, with a p -value of .1140 is computed. On the other hand, using the Jonckheere statistic, an observed value of 75 is found for J . When standardized this gives a z value of 2.52, which corresponds to a p -value of .0058. (It is pertinent to note here that the exact p -value for this set of values, from tables available in Hollander and Wolfe (1973) is .0056) The same analysis for H yields a computed value of 131 (H^* is 12.75), a standardized value of 2.40 and a p -value of .0082.

The important thing to notice is that even though the F test was not sensitive enough to find treatment differences, at least at the .05 level, both nonparametric procedures do find them and report very similar p -values.

4. POWER STUDY

A computer simulation study was performed to demonstrate the differences in power between the nonparametric statistics defined above and the F test. An additional aim was to discover whether there are significant differences between the two proposed nonparametric tests. The simulation generated 1000 sets of data for each parameter configuration and computed the empirical power for each of the F , J , and H tests, when using .05 level critical points. The results are not tabled for this paper but shown in graphical form, where the power functions are graphed with the appropriate letter as the plotting symbol.

The primary alternative studied is what might be termed the "stairstep" alternative, denoted in some other sources as the "simple order" alternative; that is, the consecutive means differ by a constant amount. In the graphs that follow the text of the paper, the value on the horizontal scale is the difference between any two consecutive means; hence in each case the value at zero corresponds to the null hypothesis and should yield a value of .05 for each of the tests. This alternative was studied for all combinations of $t = 3, 4$ and 5 and $o = 6$ and 10 , where the error distribution is the standard normal. Comparisons were also made for the case where $t = 4$ and $o = 6$ with errors having the uniform and Cauchy distributions.

Some additional runs were made at alternatives where the means of the distribution did not follow the "stairstep" pattern; the pattern used is identified on the appropriate graph. Each of these comparisons was also made with $t = 4$ and $o = 6$.

5. CONCLUSIONS AND RECOMMENDATIONS

In each case for the staircase alternative under the assumption of normal errors the nonparametric procedures were considerably more powerful than the F test, at all sample sizes. In the case of nonnormal errors the evidence is even more convincing. In fact, when the underlying errors have the Cauchy distribution, the F test has problems even maintaining its nominal level of .05 at the null hypothesis. Consequently, one must be very cautious in using the F test in this situation.

When the alternatives are not in the staircase pattern, the nonparametric procedures still show a strong advantage over the F test. That advantage diminishes somewhat when two or more of the means are actually equal, but it is still significant.

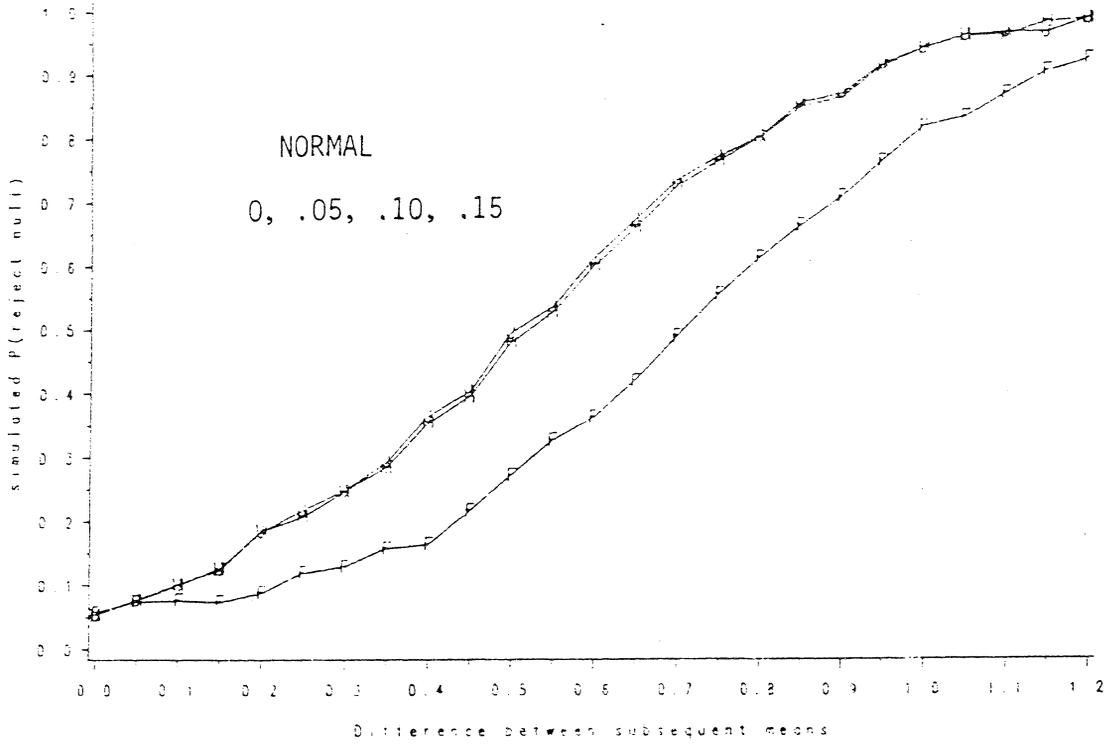
The last graph shown is one in which the true means were not equal, but were unequal in a way not consistent with the ordered alternative hypothesis. As one would well expect, the two ordered alternative procedures never manage to pick up the differences in the means, but the F test, since the means are consistent with its alternative, does. It is clear from this example that a misspecification of the alternative can be disastrous.

The evidence seems compelling that the procedures presented will increase the power over a wide variety of parameter configurations, when the alternatives are ordered in the way specified by the alternative hypothesis. There seems to be very little difference between the two procedures, thus leading the statistician to use the one which seems simplest to compute.

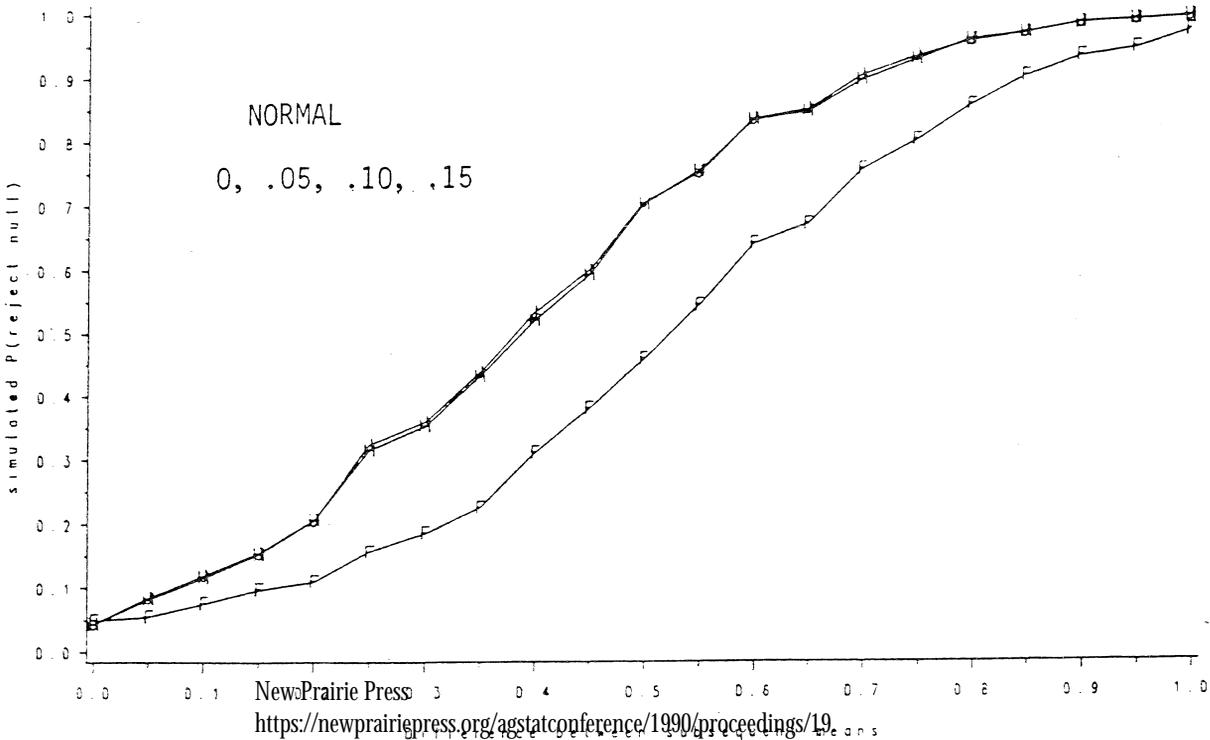
REFERENCES

- Barlow, R. E., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk (1972). *Statistical Inference under Order Restrictions*, John Wiley and Sons, London.
- Gaines, S. D., and W. R. Rice (1990). Analysis of biological data when there are ordered expectations. *Am. Nat.* 135:310-317.
- Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks*, John Wiley and Sons, London.
- Hollander, M. and D. A. Wolfe (1973). *Nonparametric Statistical Methods*, John Wiley and Sons, London.
- Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika* 41:133-145.
- Robertson, T., F.T. Wright, and R.L. Dykstra (1988). *Order Restricted Statistical Inference*, John Wiley and Sons, London.
- Terpstra, T.J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indag. Math.* 14:327-333.
- Tryon, P. V. and T. P. Hettmansperger (1974). A class of nonparametric tests for homogeneity against ordered alternatives. *Ann. Stat.* 1:1061-1070.

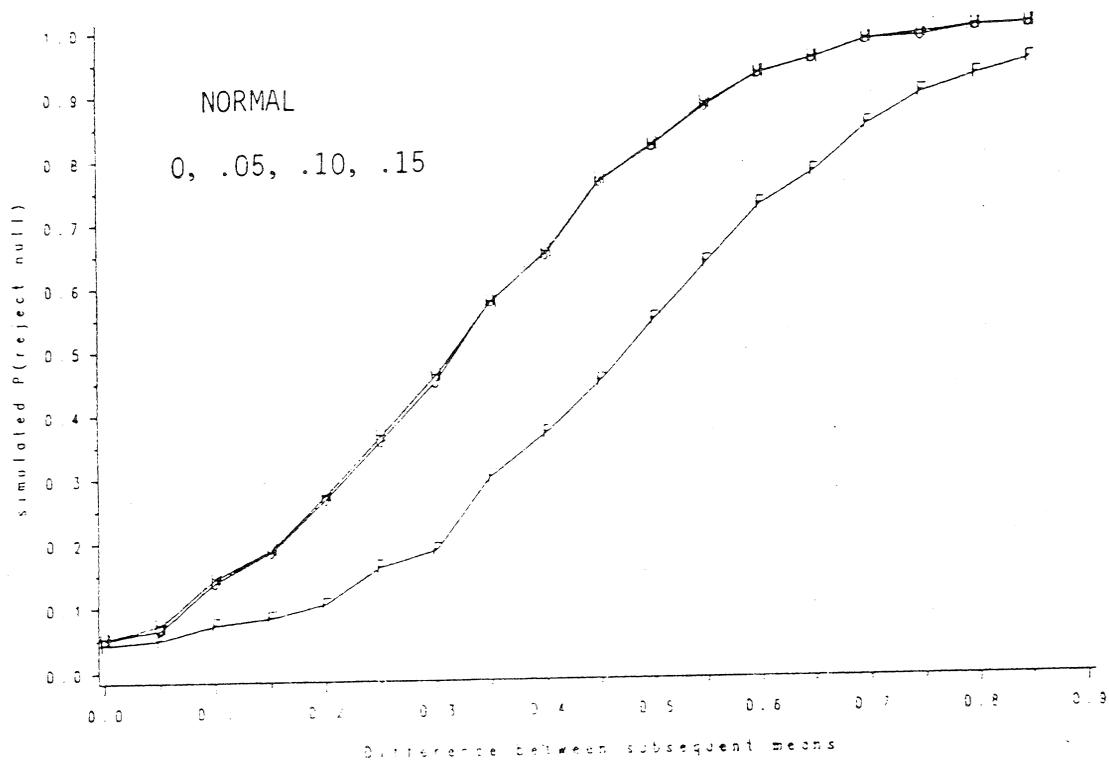
Power Comparison of J, H and F Tests
3 trts, 6 obs/trt (simulated)



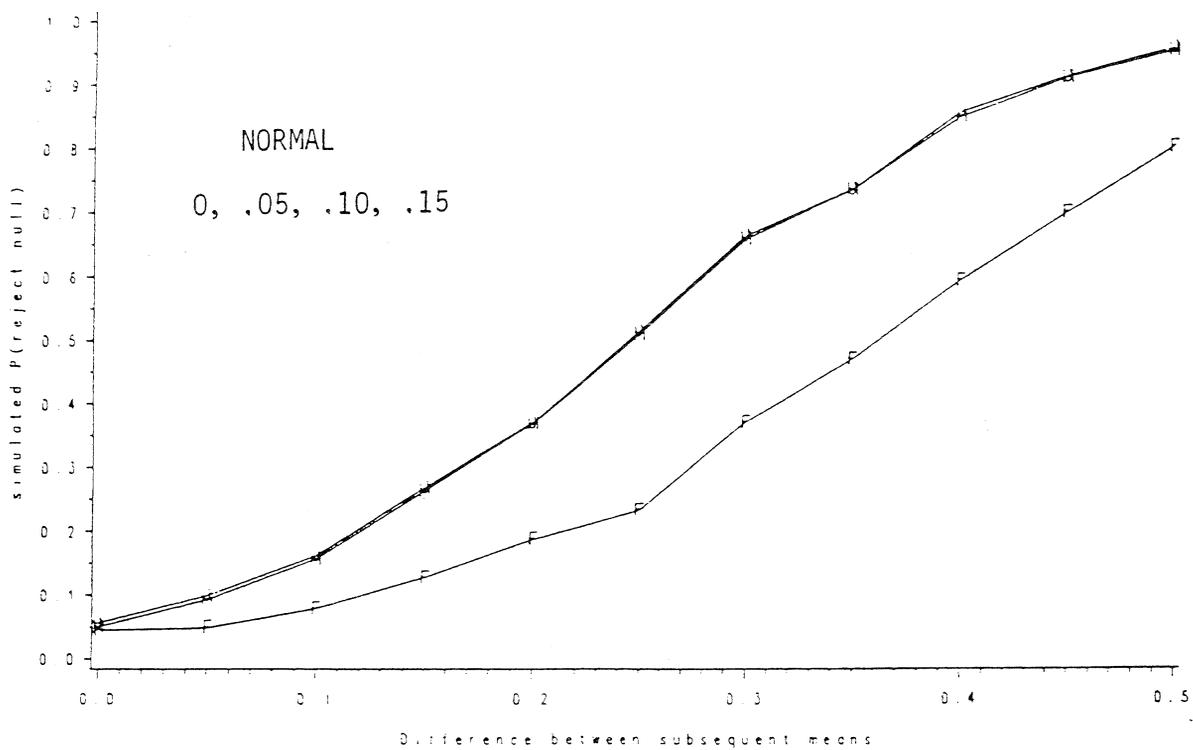
Power Comparison of J, H and F Tests
3 trts, 10 obs/trt (simulated)



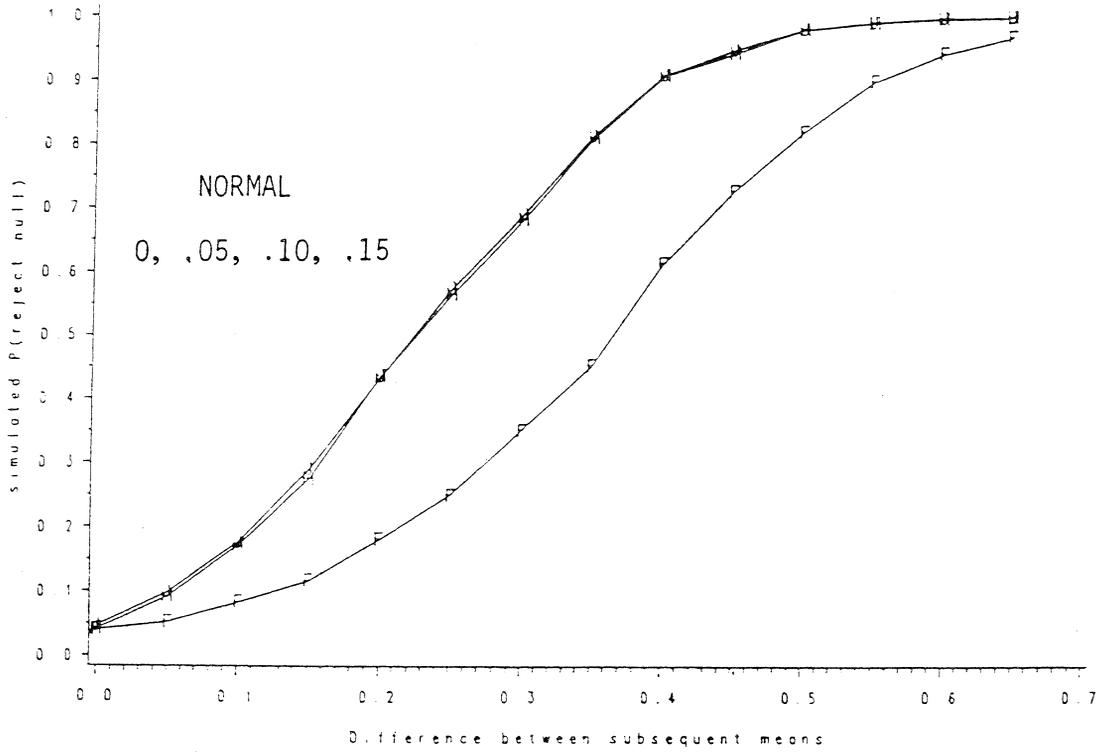
Power Comparison of J, H and F Tests
4 trts, 8 obs/trt (simulated)



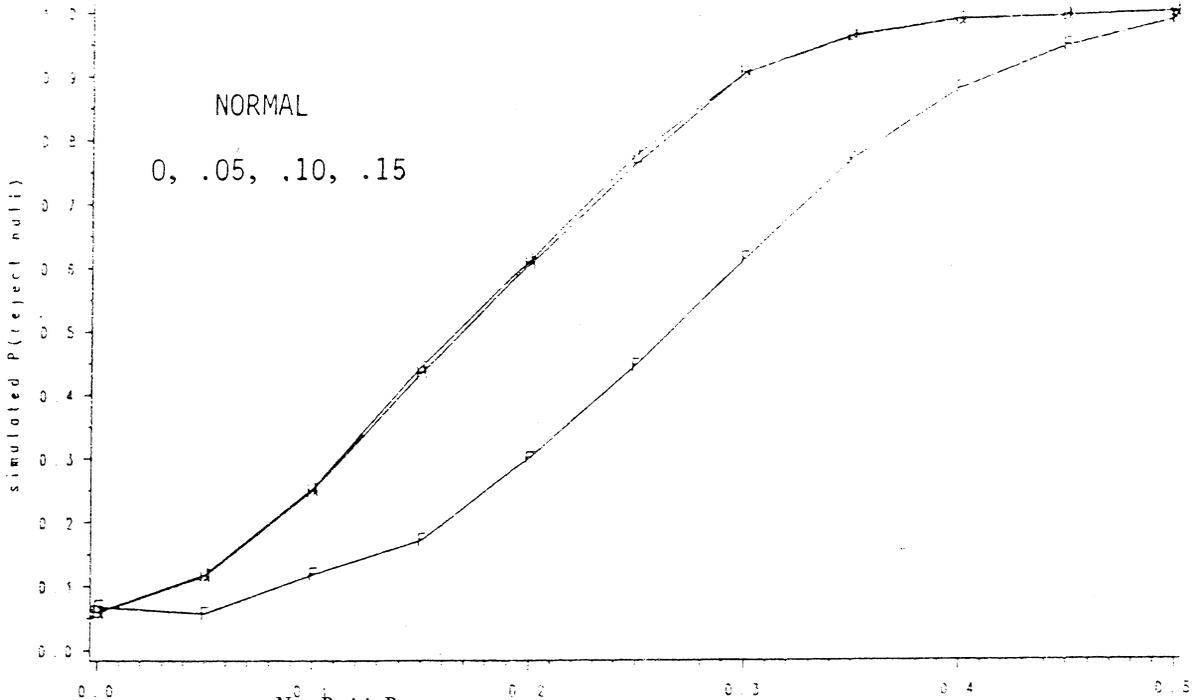
Power Comparison of J, H and F Tests
4 trts, 10 obs/trt (simulated)



Power Comparison of J, H and F Tests
5 trts, 6 obs/trt (simulated)

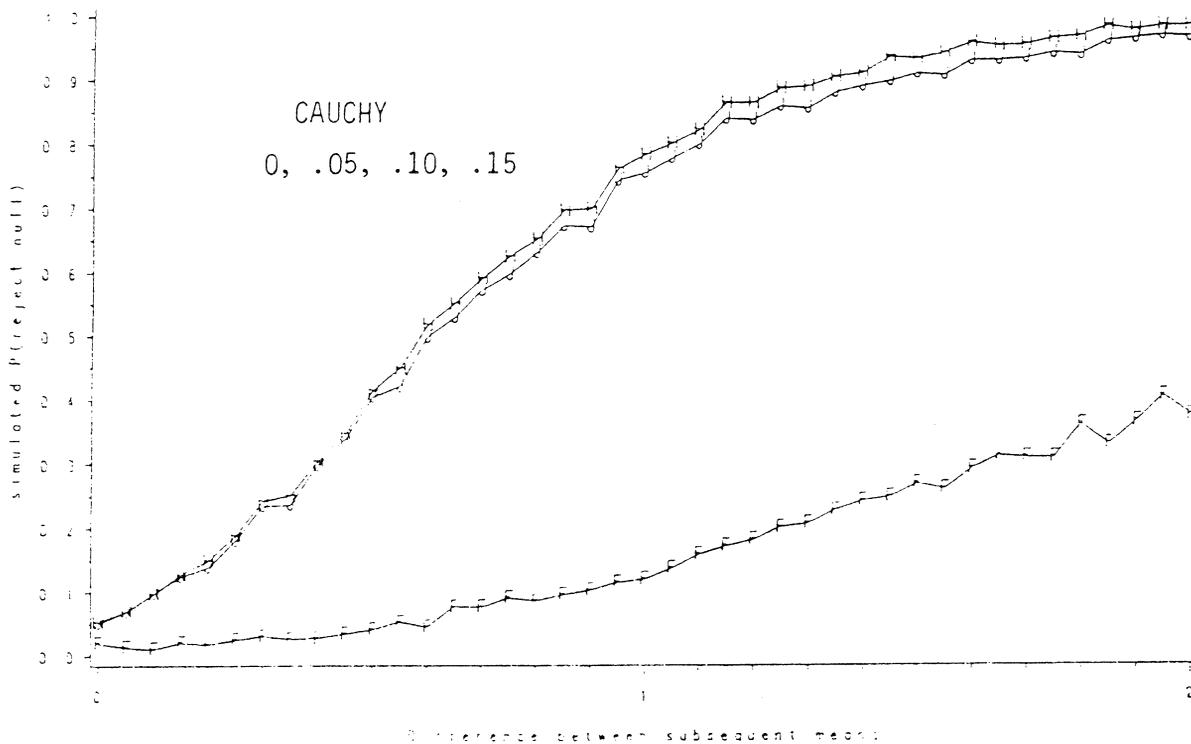


Power Comparison of J, H and F Tests
5 trts, 10 obs/trt (simulated)



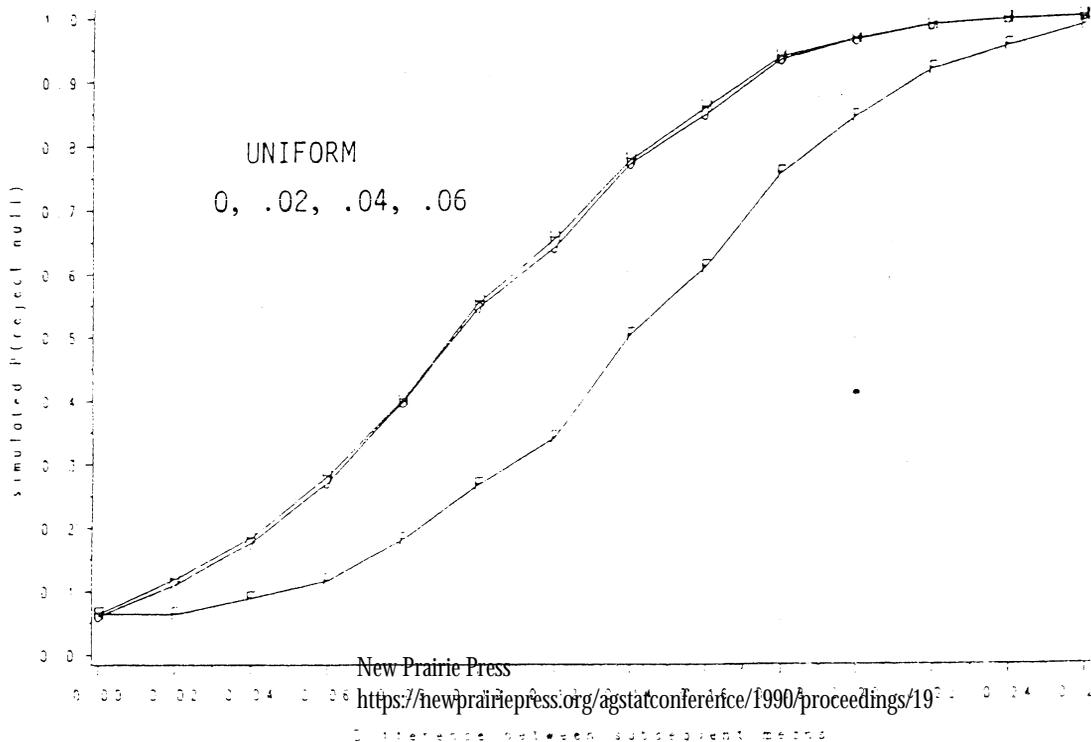
Power Comparison of J, H and F Tests

4 trts, 6 obs/trt (simulated)

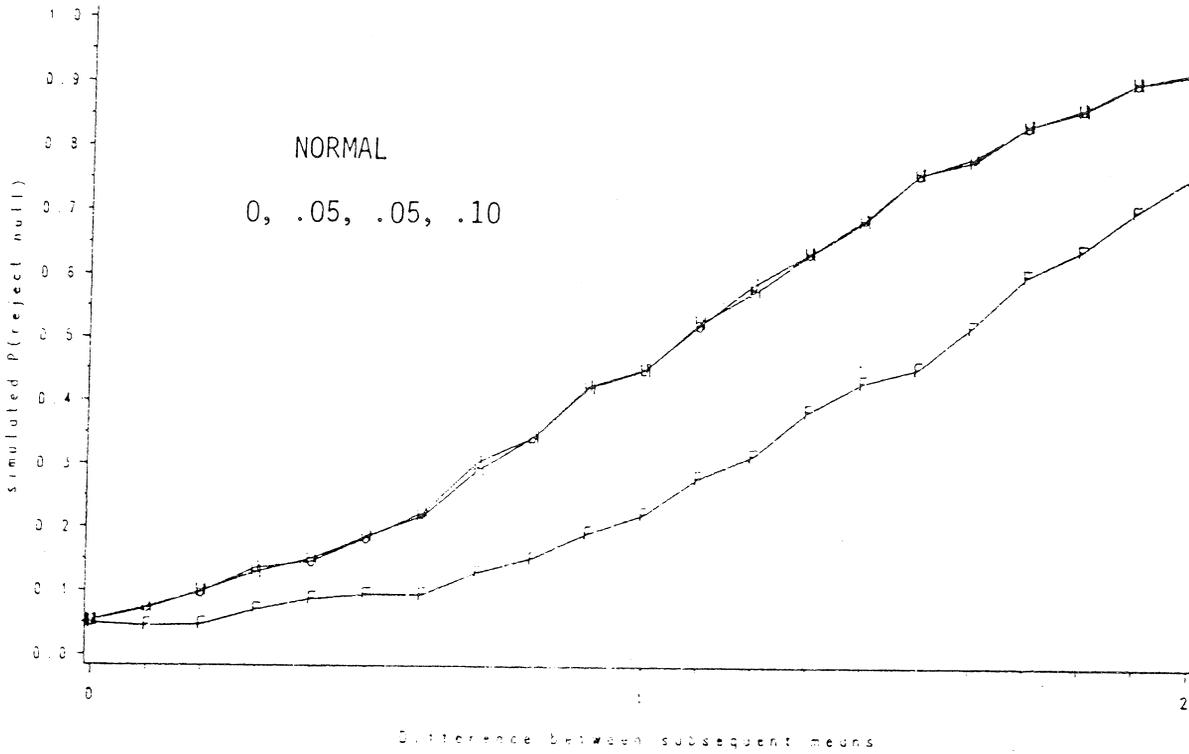


Power Comparison of J, H and F Tests

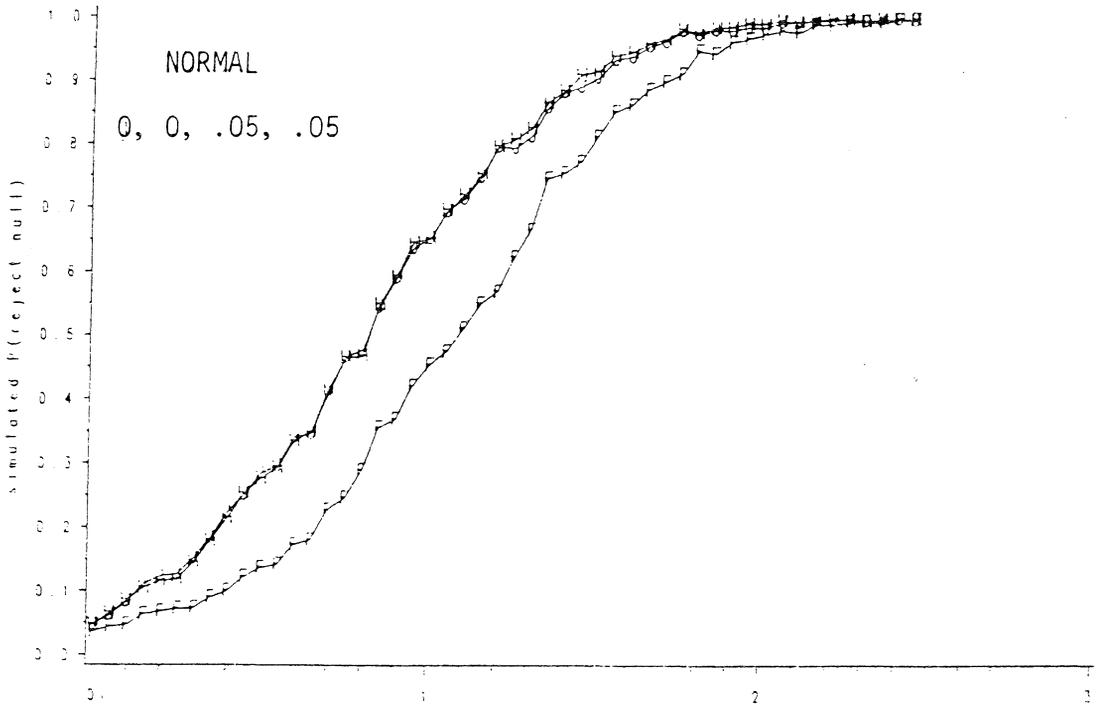
4 trts, 6 obs/trt (simulated)



Power Comparison of J, H and F Tests
 4 trts, 6 obs/trt (simulated)

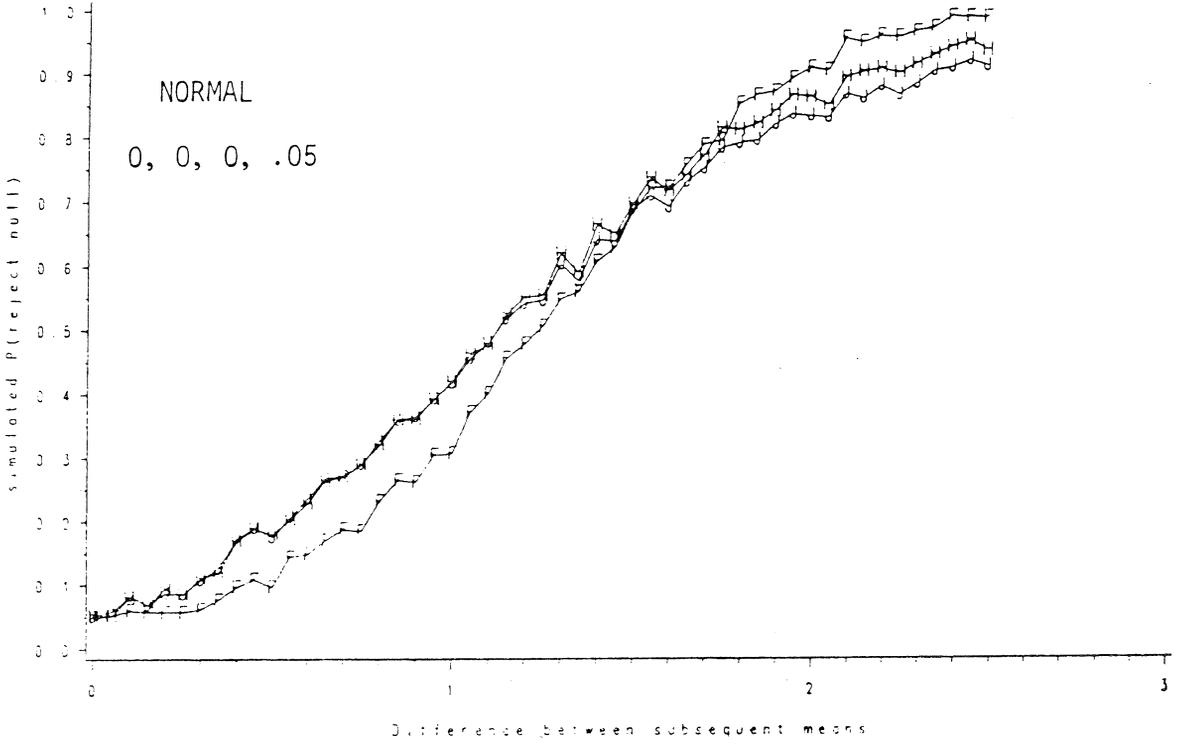


Power Comparison of J, H and F Tests
 4 trts, 6 obs/trt (simulated)



Power Comparison of J, H and F Tests

4 trts, 6 obs/trt (simulated)



Power Comparison of J, H and F Tests

4 trts, 6 obs/trt (simulated)

