Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 225–226, January 1994
Statistics

Extending nondirectional heterogeneity tests to evaluate simply ordered alternative hypotheses

(data analysis/ordered test/biostatistics)

WILLIAM R. RICE* AND STEVEN D. GAINES†

*Biology Department, University of California, Santa Cruz, CA 95064; and †Program in Ecology and Evolutionary Biology, Box G-W202, Brown University, Providence, RI 02912

Communicated by Thomas W. Schoener, September 21, 1993

ABSTRACT **Biologists frequently use nondirectional heterogeneity tests when comparing three or more populations because a suitable directional test is unavailable or is not practical to implement. Here we describe a test, the ordered heterogeneity test, that permits testing against simply ordered alternative hypotheses in the context of almost any nondirectional test. The test has a wide range of parametric and nonparametric applications. Graphs are developed for calculating exact *P* values.**

Biologists commonly prefer to test a null hypothesis (H_o) against a one-sided (i.e., directional) alternative hypothesis (H_A). This preference is motivated by increased power of one-sided tests when appropriately applied. In the context of two populations, many one-sided tests are widely used and available in most statistical computer packages. But when three or more populations are classified by category, techniques for carrying out directional tests are absent or poorly represented in most textbooks of basic statistics as well as most statistical computer packages. We believe that this has led biologists to routinely forego the power advantage of directional testing by carrying out a nondirectional test when a directional test is appropriate.

Statisticians have not been remiss in developing a diversity of tests against a directional H_A for comparisons of three or more populations. These techniques have been reviewed (1) and the power of various alternative tests has been compared (1). Isotonic regression and its extensions have demonstrated high power over a wide range of possible deviations from H_o, and we recommend that these be used whenever possible. There are many testing circumstances, however, where specific tests are either not yet available or impractical to implement, especially in the context of statistical computer packages.

Here we describe a generic test that permits one to easily convert almost any nondirectional heterogeneity test into a directional test for the context of simple ordering. The procedure is not meant to replace specialized tests, such as isotonic regression, but is for use when these techniques cannot be practically applied. Statistical details of our test will be presented elsewhere (2). Here the logic and implementation of our test are described for practitioners.

To introduce the test, consider the following experiment. Three populations of fruit flies were established from the same base population: one large and sexual (LS), one small and sexual (SS), and the third small and asexual (SA). There is a clear prediction from theory that the levels of additive genetic variance for net fitness should diverge such that $\sigma^2_{LS} > \sigma^2_{SS} > \sigma^2_{SA}$. When the experiment is completed, the H_o will be tested: $\sigma^2_{LS} = \sigma^2_{SS} = \sigma^2_{SA}$. The preferred H_A is $\sigma^2_{LS} \geq \sigma^2_{SS} \geq \sigma^2_{SA}$, with at least one inequality strict, yet the available tests known to us for comparing population variances only test against the nondirectional H_A, $\sigma^2_{LS} \neq \sigma^2_{SS} \neq \sigma^2_{SA}$, with at least one inequality strict. How can we test H_o against the more appropriate directional (i.e., simply ordered) H_A and, thereby, gain statistical power?

The best solution is to identify the most appropriate variance heterogeneity test and then solve for, or find in the literature or an appropriate text, an extension of this test to the context of directional alternative hypotheses. When this cannot be done, or when such a test is impractical to implement, a useful alternative is the ordered heterogeneity test (OH test) described below.

The logical basis of the OH test is quite simple. Suppose we compared the levels of additive genetic variance among the three populations with a Bartlett test and the *P* value was 0.25 despite the fact that the ordering of the parameter estimates was $S^2_{LS} > S^2_{SS} > S^2_{SA}$, where S^2 denotes the sample estimate of σ^2. The data contain two types of statistically independent information: the magnitude of variation among the S^2 values from the three populations, and the ordering of these parameter estimates. Tests such as the Bartlett, ANOVA, analysis of covariance (ANCOVA), multivaiate analysis of variance (MANOVA), and contingency analysis are examples of nondirectional heterogeneity tests. All evaluate the magnitude of variation among the parameter estimates from the populations irrespective of any directional pattern of this variation. By ignoring the ordering of the data, nondirectional heterogeneity tests are ineffective in testing directional alternative hypotheses as has been more formally developed elsewhere (1, 3) in the context of ANOVA.

A straightforward way to extend these nondirectional heterogeneity tests so that they evaluate simply ordered alternatives is to incorporate a second independent measure that is based solely on the ordering information. Spearman's rank correlation (r_s; Fig. 1) between the observed and expected rankings of the groups fulfills this criterion. In the above example $r_s = 1$, because the observed and expected ranks are identical.

A simple composite test statistic that combines the magnitude and ordering measures is the product r_sP_c, where P_c is the complement of the *P* value from the nondirectional heterogeneity test (e.g., $P_c = 1 - P_{Bartlett}$). The P_c statistic extracts the magnitude information in the sample and the r_s statistic extracts the independent ordering information. P_c is used instead of the *P* value itself so that larger values imply more evidence against H_o. The product r_sP_c becomes increasingly large as the data increasingly refute the null hypothesis in the direction of the alternative hypothesis. In this example, $r_sP_c = [1 \times (1 - 0.25)] = 0.75$ and the probability of obtaining a value of r_sP_c at least this large by chance, under H_o, is <0.05 (see Fig. 1 and below).

Tests based on the r_sP_c statistic are called OH tests (2), have a broad range of application, and can be used in both

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: OH test, ordered heterogeneity test.

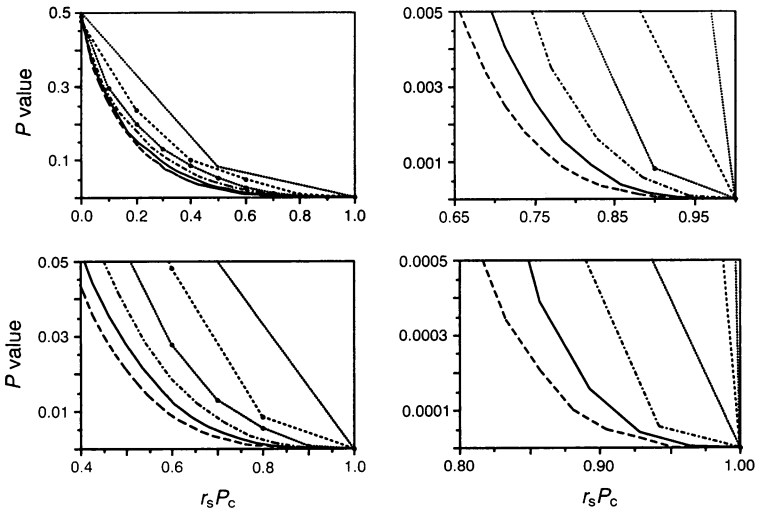

FIG. 1. Critical values for the r_sP_c statistic, where r_s is Spearman's rank correlation coefficient $(1 - \{\Sigma (\text{rank}_{obs} - \text{rank}_{exp})^2/[k(k^2 - 1)/6]\}$, k being the number of groups), $P_c = 1 - P_{HT}$, and P_{HT} is the P value from a nondirectional heterogeneity test. To use the graphs begin with the full-range plot (*Upper Left*) (the other plots are identical except the axes are expanded to improve resolution for small P values) and select the curve with the appropriate number of populations (k, see below), locate the observed r_sP_c value on the abscissa, and then read the corresponding P value from the ordinate. For negative values of r_sP_c, enter the figure with the absolute value of r_sP_c and substitute the corresponding P value from the ordinate with its complement (1 − P value). For two-tailed tests, enter the figure with the absolute value of r_sP_c and double the P value, except for the case of $r_sP_c = 0$ for which the P value = 1.0. The curves are arranged in descending order; upper-most curve for $k = 3$, second from top for $k = 4, \ldots$, and the bottom curve for $k = 8$.

parametric and nonparametric contexts. They are particularly useful in complex or "messy" applications such as the evaluation of single-factor effects in multiway ANOVA and contingency tables. For example, a simple order test for any single factor in a multiway ANOVA can be obtained by calculating P_c from the appropriate F statistic (contained in the ANOVA table) and r_s from the observed and expected rank order of the appropriate means (corrected for covariates if necessary). If a parametric test is inappropriate in evaluating the magnitude variation, a nonparametric alternative can be used to obtain P_c. In fact any test can be used to obtain P_c with the constraint that the P values from the test be both independent of the ordering of the data and distributed, at least approximately, as a Uniform{0,1} variate under the null hypothesis. This last condition will be met in most potential applications, except for those where the test statistic can assume a small number of discrete values—e.g., a 2 × 2 contingency test with very small sample sizes.

The broad application of the r_sP_c test derives from the fact that the magnitude information is incorporated into the composite test statistic via the corresponding P_c value rather than the specific test statistic (χ^2, F, etc.) itself. As a last fact concerning the flexibility of the r_sP_c test, we point out that while the ordering of groups must be specified in the alternative hypothesis, the direction of the response need not; i.e., one- and two-tailed tests are possible (see Fig. 1).

Critical values for the r_sP_c statistic depend on the pattern of expected ordering of the group parameters being tested. The most common application will be the case of simple ordering [i.e., $H_o(\text{PAR}_i = \text{PAR}_j)$ vs. $H_A(\text{PAR}_i \leq \text{PAR}_j)$, for $i < j$, at least one inequality strict and PAR denoting some population parameter of interest]. The distribution of the r_sP_c statistic for this case has been described elsewhere (2) and was used to generate the P value vs. r_sP_c curves shown in Fig. 1.

The OH test fills a void that has previously hampered efficient statistical testing. There are many applications where specific directional tests are not yet developed, unknown to the researcher, or impractical to implement, forcing biologists to use nondirectional tests when inappropriate. The simplicity and broad application of the OH test should permit simple order testing in virtually all cases where it is needed.

This research was supported by grants from the National Science Foundation to W.R.R. and S.D.G.

1. Robertson, T., Wright, F. T. & Dykstra, R. L. (1989) *Order Restricted Statistical Inference* (Wiley, London).
2. Rice, W. R. & Gaines, S. D. (1994) *Biometrics*, in press.
3. Gaines, S. D. & Rice, W. R. (1990) *Am. Nat.* **135**, 310–317.