The Annals of Statistics
1980, Vol. 8, No. 6, 1342–1347

CONTROL OF DIRECTIONAL ERRORS WITH STAGEWISE MULTIPLE TEST PROCEDURES[1]

BY JULIET POPPER SHAFFER

University of California, Berkeley

This paper is concerned with jointly testing the hypotheses $\theta_i = \theta_{i0}$, $i = 1, \cdots, s$, using tests based on independent statistics T_i with distributions $P(T_i \leq t) = F_i(t, \theta_i)$ nonincreasing in θ_i. Holm proposed a sequentially rejective test procedure, applicable to this problem, for which, for fixed α ($0 \leq \alpha \leq 1$), the probability that the joint conclusion contains no false rejections is $\geq 1 - \alpha$ for all possible values of the θ_i.

Suppose, however, that if the hypothesis $\theta_i = \theta_{i0}$ is rejected, it is desired to conclude not only that $\theta_i \neq \theta_{i0}$ but also either that it is greater than θ_{i0} or smaller than θ_{i0}. Usually one then requires a probability $\geq 1 - \alpha$ that the joint conclusion contains neither false rejections nor false directional statements. This paper considers the use of Holm's nondirectional procedure for rejecting hypotheses, supplemented by decisions on direction based on the values of the T_i. It is shown that this procedure does not in general provide the required control over error probabilities, but that it does so under specified conditions on the distributions of the T_i.

1. Introduction. Assume s hypotheses of the form H_i^0: $\theta_i = \theta_{i0}$, $i = 1, \cdots, s$, which will be tested against two-sided alternatives using independent statistics T_i with continuous or discrete distributions

$$P_i(T_i \leq t) = F_i(t, \theta_i), \qquad F_i \text{ nonincreasing in } \theta_i. \tag{1.1}$$

It will be assumed that any combination of true values of the θ_i's is possible; while not necessary for proof of the theorems, the comments on the favorable power properties of the procedure considered in this paper would not otherwise apply. Rejection of a hypothesis will be interpreted, at this point, as being equivalent to the conclusion only that $\theta_i \neq \theta_{i0}$, without specification of the direction of the difference between actual and hypothesized parameter values. Acceptance of a hypothesis will be interpreted as reaching no conclusion about the value of the corresponding parameter, so that errors occur only when true hypotheses are rejected. The set of s acceptances or rejections resulting from the application of a joint testing procedure will be called a correct joint decision if it does not contain any errors.

Holm (1977, 1979) proposed a sequentially rejective test procedure which can be applied to this problem, and for which the probability of a correct joint decision, to be denoted P_c, is $\geq 1 - \alpha$ for all possible values of the θ_i. Essentially the same procedure was proposed in a more general context by Marcus, Peritz, and Gabriel (1976). Holm's procedure involves testing in stages; any hypothesis rejected at one stage is eliminated from the set considered in setting critical values for subsequent stages. More formally, the procedure is as follows. For $i = 1, \cdots, s$

STAGE 1. If $T_i < C_{is}$ or $T_i > C'_{is}$, reject H_i^0. If at least one hypothesis is rejected, proceed to Stage 2.

Received October 1978; revised September 1979.

[1] This paper was prepared with support of U.S. Office of Naval Research contract number N00014-75-C-0444/NR042-036, and with the support of National Science Foundation grant number MCS76-10238-A02.

AMS 1970 *subject classifications.* Primary 62F99, 62J15; secondary 62F05.

Key words and phrases. Hypothesis testing, multiple tests, multiple inference, simultaneous inference, stagewise multiple tests, experimentwise error rate, directional errors.

Let r_{j-1} be the number of hypotheses H_i^0 rejected at stage $j-1, j-1 = 1, \cdots, s-1$. If $r_{j-1} = 0$, accept all hypotheses which have not been rejected at an earlier stage. If $r_{j-1} > 0$, proceed to stage j.

STAGE j. Consider all i for which H_i^0 has not been rejected. If $T_i < C_{i,m}$ or $T_i > C'_{i,m}$ reject H_i^0, where $m = s - \sum_{k=1}^{j-1} r_k$.

Continue in this way until each hypothesis has been either accepted or rejected.

The C_{ij}, C'_{ij}, $i, j = 1, \cdots, s$, are the maximum and minimum values, respectively, for which

$$P_{\theta_{i0}}(T_i < C_{ij}) \le (1 - c_i)\alpha_j; \qquad P_{\theta_{i0}}(T_i > C'_{ij}) \le c_i\alpha_j \tag{1.2}$$

where $0 \le c_i \le 1$ and

$$(1 - \alpha_j)^j = 1 - \alpha. \tag{1.3}$$

Thus, for any subset $k, \cdots, n$ containing j of the integers $1, \cdots, s$

$$P_{\theta_{k0},\cdots,\theta_{n0}}(C_{kj} < T_k \le C'_{kj}, \cdots, C_{nj} < T_n < C'_{nj}) \ge 1 - \alpha, \tag{1.4}$$

with equality if all the T_i are continuous.

A joint testing procedure δ will be said to be more powerful than a joint testing procedure δ' if the probability of rejecting any false hypothesis under δ is $\ge$ the probability of rejecting that hypothesis under δ', and if it is greater for at least one false hypothesis for some values of the θ_i. A single stage test procedure with critical values C_{is}, C'_{is} clearly cannot be more powerful than Holm's procedure, since the critical regions for each H_i^0 for the single stage procedure are included in the critical regions given by Holm's procedure.

Under an additional assumption, it is also easily shown that the power of the procedure cannot be improved, if all T_i are continuous, by increasing any C_{ij} or decreasing any C'_{ij} without violating the requirement

$$P_c \ge 1 - \alpha \qquad \text{for all values of the } \theta_i. \tag{1.5}$$

(The choice of the c_i in (1.2), which determines the joint selection of C_{ij} and C'_{ij}, affects power but also involves other considerations outside the scope of this paper.) To state this assumption, let $[\underline{t}_{i,\theta_i}, \bar{t}_{i,\theta_i}]$ be the convex support of $F_i(t, \theta_i)$. Suppose that the possible values of θ_i are $\underline{\theta}_i < \theta_i < \bar{\theta}_i$ $(-\infty \le \underline{\theta}_i, \bar{\theta}_i \le \infty)$. Then the additional assumption is

$$\lim_{\theta_i \to \underline{\theta}_i} F_i(t, \theta_i) = 1, \qquad \lim_{\theta_i \to \bar{\theta}_i} F_i(t, \theta_i) = 0; \qquad \underline{t}_i, \theta_{i0} < t < \bar{t}_{i,\theta_{i0}}. \tag{1.6}$$

(In the discrete case, the lower endpoint of the interval for t would be included, provided it is finite.)

To show that, given (1.5), (1.6), and all T_i continuous, the procedure cannot be made more powerful by increasing any C_{ij} or decreasing any C'_{ij}, consider, for any j, the configuration in which $\theta_i = \theta_{i0}$, $i = 1, \cdots, j$, and θ_i approaches $\underline{\theta}_i$ or $\bar{\theta}_i$, $i = j + 1, \cdots, s$. Then, by (1.6), the probability of rejecting all hypotheses $H_{j+1}^0, \cdots, H_s^0$ at Stage 1 can be made arbitrarily close to one, in which case P_c approaches the probability that $(C_{1j} < T_1 < C'_{1j}, \cdots, C_{jj} < T_j < C'_{jj})$ which equals $1 - \alpha$ by (1.4). An increase in C_{ij} or reduction in C'_{ij} would result in P_c being smaller than $1 - \alpha$ for the configuration given above.

It seems likely that, in most applications of such procedures, when H_i^0 is rejected one would want to decide either $\theta_i > \theta_{i0}$ or $\theta_i < \theta_{i0}$, not simply $\theta_i \ne \theta_{i0}$. For this purpose it seems natural, in view of (1.1), to augment the Holm procedure by concluding, after rejecting H_i^0, either $\theta_i > \theta_{i0}$ or $\theta_i < \theta_{i0}$, depending upon whether T_i is large or small. Under this augmented procedure, additional errors, sometimes referred to as Type III errors, may arise by concluding $\theta_i > \theta_{i0}$ when $\theta_i < \theta_{i0}$ or vice versa. These will be called directional errors, in distinction to the errors of falsely rejecting true hypotheses H_i^0, which will be called nondirectional.

The augmented procedure will be referred to as the directional test procedure. It seems plausible that this procedure would still satisfy (1.5), with P_c now the probability of a joint decision containing neither directional nor nondirectional errors. This is easily seen to be true when $s = 1$, since P_c under the directional test procedure is at least as large when $\theta \neq \theta_0$ as when $\theta = \theta_0$, assuming (1.1). One might then expect that, for $s > 1$, P_c would be minimum when $\theta_i = \theta_{i0}$ for all i, and thus $\geq 1 - \alpha$ for all possible θ_i. (If so, then among procedures satisfying (1.5) with all T_i continuous, no more powerful procedure could be obtained by increasing any C_{ij} or reducing any C'_{ij}, as shown by the argument used with respect to Holm's procedure.) However, as pointed out by Marcus, Peritz, and Gabriel (1976), neither the directional test procedure, nor any other procedure within the more general class of "closed testing procedures" which they consider, has been shown to have this property.

In Section 2 of this paper, it will be shown that (1.5) does hold provided that the distributions F_i satisfy (1.1), (1.6), and

$$(1.7) \qquad F'_i(t', \theta)/F'_i(t, \theta) \leq F'_i(t', \theta')/F'_i(t, \theta') \qquad \text{for } t' > t, \theta' > \theta, i = 1, \cdots, s$$

where $F'_i(t, \theta)$ is the derivative of $F_i(t, \theta)$ with respect to θ. On the other hand, it will be shown by an example in Section 3 that (1.5) is not necessarily true if the F_i satisfy (1.1) and (1.6) but not (1.7).

In treating the directional test procedure, it will be convenient to replace H_i^0, $i = 1, \cdots, s$, by the hypothesis-pair H_i, H'_i defined as

$$(1.8) \qquad H_i\colon \theta_i \leq \theta_{i0}; \qquad H'_i\colon \theta_i \geq \theta_{i0}.$$

The directional test procedure can then be described as follows.

Directional test procedure. In the adaptation of Holm's sequentially rejective test procedure described above, replace each statement "If $T_i < C_{ij}$ or $T_i > C'_{ij}$ reject H_i^0" by "If $T_i > C'_{ij}$ reject H_i; if $T_i < C_{ij}$ reject H'_i" where $j = s$ or m.

Note that rejection of either H_i or H'_i implies rejection of H_i^0, and rejection of H_i^0 implies rejection of either H_i or H'_i.

THEOREM 1. *Assume that the hypotheses* (1.8), $i = 1, \cdots, s$, *are being tested using the directional test procedure defined above, that each* T_i *is continuous or discrete, that the* T_i *are independently distributed as* $F_i(t, \theta_i)$, *and that the distributions* F_i *satisfy* (1.1), (1.6), *and* (1.7). *Then* (1.5) *holds.*

In Section 4 it will be shown that the conditions of Theorem 1 are satisfied by wide classes of distributions.

2. Proof of Theorem 1. A hypothesis H_i is false if $\theta_i > \theta_{i0}$; H'_i is false if $\theta_i < \theta_{i0}$. Let k be the number of false hypotheses. Since H'_i is true if H_i is false, and vice versa, the possible values of k are $0, 1, \cdots, s$.

Without loss of generality, we can assume that the false hypotheses are $H_1, \cdots, H_k$. For any fixed values of $\theta_i > \theta_{i0}$; $i = 1, \cdots, k - 1$, it will be shown that $P_c \geq 1 - \alpha$ for all $\theta_k > \theta_{k0}$.

The proof will be carried out in two stages.

(A) P_c will be shown to approach limits $\geq 1 - \alpha$ as (i) $\theta_k \rightarrow \theta_{k0+}$ and as (ii) $\theta_k \rightarrow \bar{\theta}_k$.

(B) It will be proved that P_c has no local minima between $\theta_k = \theta_{k0+}$ and $\theta_k = \bar{\theta}_k$.

Together, these two facts imply the desired result.

PROOF OF (A). The proof will proceed by induction. If $k = 0$, $P_c = P$(all hypotheses are accepted) $\geq 1 - \alpha$ by (1.4). Thus, (A) holds for $k = 0$. It will now be assumed to hold for $k - 1$, and proved for k false hypotheses.

(i) $\underline{\theta_k \rightarrow \theta_{k0+}}$. Let $P_c = P_1 + P_2$, where $P_1 = P$(correct joint decision with H_k accepted) and $P_2 = P$ (correct joint decision with H_k rejected). As $\theta_k \rightarrow \theta_{k0+}$, $P_1 \rightarrow P$(correct joint decision with s hypothesis-pairs and $k - 1$ false hypotheses) $\geq 1 - \alpha$ by the induction hypothesis, and part (i) of (A) follows.

(ii) $\underline{\theta_k \to \bar{\theta}_k}$. Put $P_c = P_3 + P_4$, where $P_3 = P$(correct joint decision with $T_k < C'_{ks}$) and $P_4 = P$(correct joint decision with $T_k > C'_{ks}$). As $\theta_k \to \bar{\theta}_k$, $P_4 \to P$(correct joint decision with $s - 1$ hypothesis-pairs and $k - 1$ false hypotheses) $\geq 1 - \alpha$ by the induction hypothesis, and part (ii) of (A) follows.

PROOF OF (B). For fixed $\theta_i > \theta_{i0}$, $i = 1, \cdots, k - 1$, let θ_k vary from θ_{k0+} to $\bar{\theta}_k$. It will be shown that if the derivative of P_c with respect to θ_k is negative for any value $\theta_k = \theta^*$, it cannot be positive for $\theta_k > \theta^*$, and this implies (B).

It is convenient to define sets D_{ij}, $i = 1, 2, \cdots, s$; $j = -s, -(s - 1), \cdots, (s - 1), s$ in the following way:

$$D_{is} = (C'_{is}, \infty)$$

$$D_{ij} = (C'_{ij}, C'_{i,j+1}], \qquad j = 1, 2, \cdots, s - 1$$

$$D_{i0} = D_{i,-0} = [C_{i,1}, C'_{i,1}]$$

$$D_{ij} = [C_{i,-(j-1)}, C_{i,-j}), \qquad j = -(s - 1), \cdots, -1$$

$$D_{i,-s} = (-\infty, C_{is}).$$

Then

$$P_c = \textstyle\sum_{j=-s}^{s} P(T_k \in D_{kj})\, p_j \tag{2.1}$$

where $p_j = P$(correct joint decision $| T_k \in D_{kj}$). By (1.1) and the independence of the T_i, the p_j do not depend on θ_k and

$$\begin{aligned} p_j &\leq p_{j-1} \\ p_{-j} &\leq p_{-(j-1)}, \end{aligned} \qquad j = 1, \cdots, s \tag{2.2}$$

since values of $T_1, T_2, \cdots, T_{k-1}, T_{k+1}, \cdots, T_s$ resulting in a correct joint decision for $T_k \in D_{kj}$ also result in a correct joint decision for $T_k \in D_{k,j-1}$; and similarly for $T_k \in D_{k,-j}$ and $T_k \in D_{k,-(j-1)}$.

In terms of $F_k(t, \theta_k)$ we can rewrite (2.1) as

$$\begin{aligned} P_c = p_s &- \textstyle\sum_{j=1}^{s} F_k(C'_{kj}, \theta_k)(p_j - p_{(j-1)}) \\ &+ \textstyle\sum_{j=1}^{s} F_k(C_{kj-}, \theta_k)(p_{-j} - p_{-(j-1)}), \end{aligned} \tag{2.3}$$

where $F_k(C_{kj-}, \theta_k)$ is the limit of $F_k(t, \theta_k)$ as t approaches C_{kj} from the left. (If T_k is continuous, $F_k(C_{kj-}, \theta_k) = F_k(C_{kj}, \theta_k)$, while if T_k is discrete, $F_k(C_{kj-}, \theta_k) = F_k(C^*_{kj}, \theta_k)$ for some $C^*_{kj} < C_{kj}$.)

Let $F'_k(t, \theta_k)$ be the derivative of $F_k(t, \theta_k)$ with respect to θ_k. Differentiating P_c in (2.3) with respect to θ_k, denoting the derivative by P'_c, and multiplying and dividing by $F'_k(C, \theta_k)$ for some C such that $C_{k,1} \leq C \leq C'_{k,1}$, gives

$$P'_c = F'_k(C, \theta_k) G(\theta_k) \tag{2.4}$$

where

$$\begin{aligned} G(\theta_k) = &\textstyle\sum_{j=1}^{s} [F'_k(C'_{kj}, \theta_k)/F'_k(C, \theta_k)](p_{j-1} - p_j) \\ &- \textstyle\sum_{j=1}^{s} [F'_k(C_{kj-}, \theta_k)/F'_k(C, \theta_k)](p_{-(j-1)} - p_{-j}). \end{aligned} \tag{2.5}$$

By (1.7), the ratios of derivatives in the first sum in (2.5) are nondecreasing in θ_k, and the ratios of derivatives in the second sum are nonincreasing in θ_k. Since, by (2.2), all coefficients of these ratios are nonnegative, $G(\theta_k)$ is nondecreasing in θ_k. By (1.1), $F'_k(C, \theta_k)$ in (2.4) is nonpositive, and therefore if P'_c is negative for any $\theta_k = \theta^*_k$ ($> \theta_{k0}$) it is nonpositive for all $\theta_k > \theta^*_k$, as was to be proved.

This completes the proof of Theorem 1.

3. A counterexample. The following example shows that the directional test procedure described in Section 1 need not satisfy (1.5) when (1.7) is not satisfied.

Let $s = 2$, $F_1(t, \theta_1) = G_1(t - \theta_1)$, where G_1 is a standard Cauchy distribution, and F_2 be continuous but otherwise arbitrary. It will be shown that P_c is $< 1 - \alpha$ for $\theta_2 = \theta_{20}$ and large values of θ_1 ($> \theta_{10}$).

From (2.1) we have

$$\begin{aligned} P_c &= [G_1(C'_{12} - \theta_1) - G_1(C_{12} - \theta_1)]P(C_{22} \le T_2 \le C'_{22}) \\ &\quad + [1 - G_1(C'_{12} - \theta_1)]P(C_{21} \le T_2 \le C'_{21}) \\ &= [G_1(C'_{12} - \theta_1) - G_1(C_{12} - \theta_1)](1 - \alpha)^{1/2} + [1 - G_1(C'_{12} - \theta_1)](1 - \alpha). \end{aligned}$$

Then $P_c \ge 1 - \alpha$ if and only if

$$G_1(C'_{12} - \theta_1)/G_1(C_{12} - \theta_1) \ge 1/[1 - (1 - \alpha)^{1/2}]. \tag{3.1}$$

Now, if g_1 is the density of G_1,

$$\lim_{\theta_1 \to \infty} G_1(C'_{12} - \theta_1)/G_1(C_{12} - \theta_1) = \lim_{\theta_1 \to \infty} g_1(C'_{12} - \theta_1)/g_1(C_{12} - \theta_1) = 1 \tag{3.2}$$

since G_1 is Cauchy. Then, for a sufficiently large value θ_1^*, (3.1) will be violated for $\theta_1 > \theta_1^*$, and therefore P_c will be $< 1 - \alpha$. (Holm (1979) has proposed a more general but less powerful class of directional procedures for which (1.5) is satisfied in the Cauchy case.)

4. Some families of distributions satisfying Theorem 1, and applications. Theorem 2 gives conditions under which Theorem 1 holds for some common classes of distributions.

THEOREM 2. *The conditions of Theorem* 1 *are satisfied for* (A) *location-parameter families with monotone likelihood ratio,* (B) *positive-valued scale-parameter families with monotone likelihood ratio, and* (C) *exponential families which are continuous or discrete and which satisfy* (1.6).

PROOF. The proof of (A) and (B) follows directly from the usual expressions for the cumulative distribution functions for such families. To prove (C), note that (1.6) is assumed and (1.1) is known to be satisfied; the inequality (1.7) is easily proved from the fact that $E_\theta(T)$ is an increasing function of θ, that the density (with respect to μ) expressed in natural form, $f(u, \theta) = a^\theta e^{\theta u}$, has monotone likelihood ratio, and that

$$F'(t, \theta) = \int_{u=-\infty}^{t} [u - E_\theta(T)]f(u, \theta)\, d\mu(u).$$

Making use of Theorem 2, it is easy to show that Theorem 1 applies to the usual tests, based on independent random samples, for single parameters of the following distributions (with any other parameters assumed known): means and variances of normal distributions, location and scale parameters of exponential distributions, scale parameters of uniform distributions and (for sample size 1) of Cauchy distributions, and the parameters of binomial and Poisson distributions. Tests for means of normal distributions with common unknown variance are not covered by Theorem 1, but when the variance is estimated with large degrees of freedom, the individual t-statistics are approximately normal and independent so that Theorem 1 holds approximately.

When the T_i are discrete, as in the binomial and Poisson cases, one may wish to use randomized tests in order to obtain values C_{ij} and C'_{ij} which result in equality in (1.4). It is easily shown that if the discrete T_i satisfy the conditions of Theorem 1, the same will be true for the randomized test statistics derived from them.

Note that nothing in Theorem 1 requires the hypotheses to refer to common families of distributions or parameters. Similarly, since Theorem 1 holds for arbitrary values of the c_i

used to define the C_{ij} and C'_{ij} in (1.2), any combination of one-sided and two-sided tests is covered. (If all tests are one-sided, Theorem 1 holds without any conditions; see Holm (1979).)

Acknowledgment. The author wishes to thank Sture Holm for suggesting a streamlined version of the original proof of Part (B) of Theorem 1.

REFERENCES

HOLM, STURE (1977). Sequentially rejective multiple test procedures. Statistical Research Report, Inst. Math. Statist. Univ. Umeå, Sweden.

HOLM, STURE (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* **6** 65–70.

MARCUS, R., PERITZ, E. and GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720