ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 196, 2006

*Dariusz Parys**

A MODIFIED HOLM'S STEPWISE REJECTIVE MULTIPLE TEST PROCEDURE

Abstract. In the case of Holm's stepwise rejective procedure we consider the multiple test problem where there are n hypotheses $H_1, H_2, ..., H_0$ and corresponding p-values $R_1, ..., R_n$. The procedure is said to control strongly the familywise error rate when the property $P(H_s s \in I,$ are accepted $|(H_s, s \in I$ true$) \geqslant 1 - \alpha$ holds. In this paper the modification of this procedure is presented. The refinement retains strong control of familywise error rate. There is a cost in calcalational simplicity, but a substantial improvement in actual error rate, according to simulations.

Key words: multiple test procedure, stepwise procedure, familywise error rate.

1. INTRODUCTION

We consider the multiple test problem where are n hypotheses $H_1, H_2, ..., H_n$ and corresponding p-values $R_1, R_2, ..., R_n$, assuming the test statistics $X_1, ..., X_n$ are from a continous distribution. Suppose that in a multiple test procedure the property

$$P(H_s, s \in I, \text{ are accepted } |H_s, s \in I \text{ true}) \geqslant 1 - \alpha \tag{1}$$

holds, for prespecified size of test (familywise error rate) α, where I is any non-null subset of $\{1, 2, ..., n\}$, and thus contains m items, $1 \leqslant m \leqslant n$. Then the procedure is said to control strongly the familywise error rate (e.g. Hochberg, Tamhane, 1987).

Let $R_{(1)}, R_{(2)}, ..., R_{(n)}$ be the ordered p-values, and $H_{(1)}, H_{(2)}, ..., H_{(n)}$ the corresponding hypotheses. The "Bonferroni" multiple test procedure rejects the composite hypothesis $\{H_{(1)}, H_{(2)}, ..., H_{(n)}\}$ if $R_{(1)} \leqslant \alpha/n$, and accepts it otherwise. This procedure was refined by Holm (1979) as follows.

* Ph.D., Chair of Statistical Methods, University of Łódź.

Examine whether $R_{(1)} \leqslant \alpha/n$: if not, accept $H_{(i)}$, $i = 1, ..., n$ as with Bonferroni; if so, reject $H_{(1)}$ and examine whether $R_{(2)} \leqslant \alpha/(n-1)$. If the inequality is not satisfied accept $H_{(2)}, ..., H_{(n)}$; otherwise reject $H_{(2)}$. Continue in this way.

To summarize: $R_{(i)} \leqslant \alpha/(n-i+1)$, $i \leqslant j-1$ then at step j the remaining hypotheses are $H_{(j)}, ..., H_{(n)}$ and the inequality next to check is $R_{(j)} \leqslant \alpha/(n-j+1)$. The process may run at most until a decision is made on the basis of whether $R_{(n)} \leqslant \alpha$ or not. Holm showed that his procedure strongly controls the familywise error rate. Inasmuch as it essentially depends on Boole's (first Bonferroni) inequality, which is a degree 1 bound (e.g. Seneta 1997), Holm's procedure retains an elegant simplicity.

There have been a number of improvements on the Boonferroni-Holm degree 1 procedures, all of which are aimed at increasing power while relaining a simple structure of critical points (such as $\alpha/(n-j+1)$ above).

In Seneta and Chen (1997), a degree 2 step-down procedure is proposed which retains familywise control of error rate. This procedure is adaptive in that calculation at each step is determined by the joint outcome of all pairs of statistics in the experiment involved until the procedure stops. In view of the continuing interest in a general procedure with famylywise control of error rate, we present here a substantial refinement of this procedure, procedure M in a from which resembles Holm's. Specifically, the values $\sigma/(n-j+1)$, $j \geqslant 1$, are replaced by large ones, thus increasing the power. We present, using simulation, a crude power comparison with the Bonferroni/Holm procedure and with the Hochberg procedure in the setting of multivariate t.

2. PROCEDURE M

Write for the moment $R_{(i)} = R_{t_i}$, t_i, t_i is a random variable from the set $\{1, 2, ..., n\}$. Using the ordered p-values R_{t_i}, $i = 1, ..., n$ observed, define the index sets $K(\cdot)$ by

$$\gamma(p) = \max_{j \in K(p)} \sum_{i \in K(p)-(j)} P\left(R_i \leqslant \frac{a}{n-p+1},\ R_j \leqslant \frac{a}{n-p+1} \,|\, H_s, s \in K(p), \text{ true}\right)$$

$$\text{for } 1 \leqslant p \leqslant n-1, \text{ with } \gamma(n) = 0. \tag{2}$$

These may be calculated for successive p as far as required in what follows.

Step 1:

$$R_{(1)} \leqslant \min\left(\frac{\alpha}{n-1}, \frac{\alpha+\gamma(1)}{n}\right)?$$

If yes, reject $H_{(1)}$ and go to Step 2. If no, accept $H_{(1)}, H_{(2)}, \ldots, H_{(n)}$ and top. Continue in this way.

Step i:

$$R_i \leqslant \min\left(\frac{\alpha}{n-i}, \frac{\alpha+\gamma(i)}{n-i+1}\right)?$$

If yes, reject $H_{(i)}$ and go to Step $i+1$. If no, accept $H_{(i)}, H_{i+1}, \ldots, H_{(n)}$ and stop. If the n-th Step is reached:

Step n:

$$R_{(n)} \leqslant \alpha.$$

If yes, reject $H_{(n)}$ and stop. In no, accept $H_{(n)}$ and stop.

3. STRONG CONTROL OF FAMILYWISE ERROR RATE

A key feature of the proof of the theorem is the use of the inequality (from which (2) derives) of Kounias (1968)

$$P\left(\bigcup_{i=1}^{I} A_i\right) \leqslant \sum_{i=1}^{k} P(A_i) - \max_{j=1,\ldots,A} \sum_{j \neq j}^{k} P(A_i \cap A_j)$$

a second-degree inequality.

Lemma. Let I, of fixed size m, ony non-null subset of $\{1, \ldots, n\}$. Define

$$\gamma = \max_{I} \sum_{j \in I-\{i\}} P(R_i \leqslant \frac{\alpha}{m}, R_j \leqslant \frac{\alpha}{m} \mid H_s, s \in I, \text{ true}), \tag{3}$$

where $\gamma = 0$ when $m = 1$. Then

$$\bigcap_{i \in I}\left\{R_i > \min\left(\frac{(\alpha+\gamma)}{m}, \frac{\alpha}{m-1}\right)\right\} \subseteq \{H_s, s \in I, \text{ are all accepted}\}.$$

Corollary. For any given I, $\gamma \geqslant \alpha/(m-1)$, then

$$P\{H_s, s \in I, \text{ are accepted } |A\} \geqslant 1 - m\alpha(m-1) + \gamma.$$

The quantity γ defined in the lemma above is required for the proof of the theorem following it, which establishes strong control of the adaptive test procedure, but is not needed in the adaptive test procedure itself. Note that $\gamma \leqslant (m-1)\alpha/m \leqslant \alpha$.

4. EXAMPLE AND SIMULATIONS

We shall measure power by

$$P(\text{Reject at least one } H_i,\ i = 1, \ldots, n).$$

This has the advantage that when all of the H_i, $i = 1, \ldots, n$ hold, from (1) this value will be $\leqslant \alpha$, and its closeness to the nominal error α will measure the actual conservativeness of the error rate. According to the corollary above, if we take $I = \{1, 2, \ldots, n\}$ then $\gamma > \alpha/(n-1)$ results in a bound $\leqslant \alpha$. This suggests that the degree of conservativeness of the procedure CS is related to the strength of positive association between the R_i 's (and hence of X_i 's) from the definition of γ. This is confirmed by Table 1 below.

We take the test statistics to be exchangeable under corresponding null hypotheses, so from (2)

$$\gamma(p) = (n-p)P(R_1 \leqslant \frac{\alpha}{n-p-1},\ R_2 \leqslant \frac{\alpha}{n-p+1}), \tag{4}$$

which is thus non-random in this special setting. More specifically we consider uppertail tests where $X_i = |T_i|$, $i = 1, 2, \ldots, n$ with $T_1, T_2, \ldots, T_n$ defined by $T_i = W_i/\sqrt{\chi^2/v}$, $i = 1, 2, \ldots, n$ where the W_i's are multivariate normal with $EW_i = \mu_i$, $\text{Var}(W_i) = 1$, $i = 1, 2, \ldots, n$ Corr $(W_i, W_j)\rho$ $i \neq j$, and are independently distributed of χ^2_v. Thus under $H_i : \mu_i = 0$, $i = 1, \ldots, n$ the T_i, $i = 1, \ldots, n$ have jointly a multivariate exchangeable t distribution with parameters n, $\rho(\rho \geqslant -1/(n-1)$, v as in Dunnett's tests. We take $n = 3$, $v = 16$, $\alpha = 0{,}005$, and consider $0 \leqslant \rho \leqslant 1$. We can calculate from tables giving upper-tail values $P(T_1 \leqslant a,\ T_2 \leqslant a)$ for various a and $\rho = 0$, $\pm 0.1, \ldots, \pm 0.9$ our values $\Delta(1) = \alpha + \gamma(1)/n$. Some of these are shown in Table 1. Notice that in our setting $\Delta(1) = (\alpha + \gamma(1)/n$. Some of these are shown in Table 1. Notice that in our setting $\Delta(1) > \alpha/2 = 0.025$ at $\rho = 1$, but is < 0.025 for $\rho < 0.9$. Our measure of power when $\rho < 0.9$ (in fact for ρ upto approximately 0.95) is thus $P(R_{(1)} \leqslant \alpha/n)$, which is smaller; and remains smaller than our measure of power for ρ very close to 1 viz.

$P(R_{(1)} \leqslant \alpha/(n-1))$. (A more sensitive measure of power would separate out Holm from Bonferroni.). Table 2 displays the power at $\rho = 0.9$ when $\mu_1 = \delta$, $\mu_2 = 2|\delta|$, $\mu_3 = 3|\delta|$ for the Bonferroni-Holm, Hochberg and CS procedures.

Table 1. Value of $\Delta(1)$ and error rate (ER) ($n = 3$, $\nu = 16$, $\alpha = 0.05$)

ρ	0	0.5	0.8	0.9	1
$\Delta(1)$	0.0171	0.0183	0.0211	0.0228	0.0278
ER (α/n)	0.049	0.042	0.033	0.028	
ER ($\Delta(1)$	0.050	0.046	0.042	0.040	

Table 2. Power at $\rho = 0.09 (n = 3,\ \nu = 16,\ \alpha = 0.05)$

δ	-1	-0.5	0	0.5	1
(α/n)	0.685	0.184	0.028	0.159	0.633
Hochberg	0.689	0.187	0.034	0.164	0.634
$\Delta(1)$	0.749	0.227	0.040	0.193	0.685

The error rate (ER) entries in Table 1 were produced from a simulation of 20 000 independent sets of values of the triple T_i, $i = 1, 2, 3$ at each ρ. These values for $\rho = 0.9$ are given again in Table 2 at $\delta = 0$. The other values of Table 2 were also produced from 20 000 triples.

Overall, the simulations support a conclusion that our proposed procedure is most effective as regards power when test statistics are strongly positively dependent. The error rate is closer to the nominal value α irrespective of degree of dependence, and is not much affected by it. The indication is that procedure CS controls error rate well, and has significantly better power than Hochberg.

While our procedure may be useful only for small n(1) control strongly the FER holds without any restriction on the continuous joint distribution of test statistics.

Finally, our computational results on ER are consistent with those of Sarkar and Chang (1997, Table 2), inasmuch as

$$P\left(R_{(i)} \geqslant \frac{\alpha}{n-i+1},\ 1 \leqslant i \leqslant n\right) \geqslant P\left(R_i \geqslant \frac{i\alpha}{n},\ 1 \leqslant i \leqslant n\right).$$

REFERENCES

Hochberg Y. (1988), "A Sharper Bonferroni Procedure for Multiple Tests of Significance", *Biometrics*, **75**, 800–802.
Hochberg Y., Tamhane A. C. (1987), *Multiple Comparison Procedures*, Wiley, New York.
Holm S. (1979), "A Simple Sequentially Rejective Multiple Test Procedure", *Scandinavian Journal of Statistics*, **6**, 65–70.
Hommel G. (1988), "A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test, *Biometrica*, **75**, 383–386.
Sarkar S. K., Chang C. K. (1997), "The Simes Method for Multiple Hypothesis Testing with Positively Dependent Test Statistics", *Journal of American Statistical Association*, **92**, 1601–1608.
Seneta E., Chen T. H. (1997), "A Sequentially Rejective Test Procedure", *Theory of Statistic Processes*, **3**(19), 393–402.

Dariusz Parys

ZMODYFIKOWANA WIELOKROTNA PROCEDURA TESTOWA KROCZĄCEGO ODRZUCENIA HOLMA

(Streszczenie)

Rozważamy przypadek testowania wielokrotnego, w którym istnieje n hipotez $H_1, H_2, ..., H_n$ i odpowiadające im p-wartości $R_1, ..., R_n$. Mówimy, że procedura ma mocną kontrolę nad błędem na rodzinę, jeżeli prawdopodobieństwo nie odrzucenia hipotezy prawdziwej, pod warunkiem że jest nie mniejsze niż α.

W artykule przedstawiono modyfikację kroczącej wielokrotnej procedury Holma. Wprowadzone zmiany zapewniają silniejszą kontrolę nad błędem na rodzinę.