Biometrika (1986), 73, 3, *pp.* 751–4
Printed in Great Britain

An improved Bonferroni procedure for multiple tests of significance

BY R. J. SIMES
Ludwig Institute for Cancer Research, University of Sydney, Sydney, N.S.W. 2006, Australia

SUMMARY

A modification of the Bonferroni procedure for testing multiple hypotheses is presented. The method, based on the ordered p-values of the individual tests, is less conservative than the classical Bonferroni procedure but is still simple to apply. A simulation study shows that the probability of a type I error of the procedure does not exceed the nominal significance level, α, for a variety of multivariate normal and multivariate gamma test statistics. For independent tests the procedure has type I error probability equal to α. The method appears particularly advantageous over the classical Bonferroni procedure when several highly-correlated test statistics are involved.

Some key words: Bonferroni inequality; Multiple comparisons; Simultaneous test procedures.

1. INTRODUCTION

The Bonferroni inequality is often used when conducting multiple tests of significance to set an upper bound on the overall significance level α (Miller, 1981, pp. 67–70). If $T_1, \ldots, T_n$ is a set of n statistics with corresponding p-values $P_1, \ldots, P_n$ for testing hypotheses $H_1, \ldots, H_n$, the classical Bonferroni multiple test procedure is usually performed by rejecting $H_0 = \{H_1, \ldots, H_n\}$ if any p-value is less than α/n. Furthermore the specific hypothesis H_i is rejected for each $P_i \leq \alpha/n$ $(i = 1, \ldots, n)$. The Bonferroni inequality,

$$\text{pr}\left\{\bigcup_{i=1}^{n} (P_i \leq \alpha/n)\right\} \leq \alpha \quad (0 \leq \alpha \leq 1),$$

ensures that the probability of rejecting at least one hypothesis when all are true is no greater than α.

Although several multivariate methods have been developed for multiple statistical inference, the Bonferroni procedure is still valuable, being simple to use, requiring no distributional assumptions and enabling individual alternative hypotheses to be identified. Nevertheless, the procedure is conservative and lacks power if several highly correlated tests are undertaken.

This paper introduces a modified Bonferroni procedure, based on the ordered p-values of the individual tests, which has an actual significance level closer to the nominal level in a wide range of circumstances and which has a lower type II error rate for a given nominal significance level than the classical procedure. Section 2 describes the procedure and shows that the probability of a type I error for the test procedure equals α for independent test stastistics. Simulation studies in § 3 show that α is an upper bound on the type I error probability for a variety of multivariate normal and chi-squared distributions. The powers of the classical and modified procedures are compared for some alternative hypotheses in § 4.

2. MODIFIED BONFERRONI PROCEDURE

Let $P_{(1)}, \ldots, P_{(n)}$ be the ordered p-values for testing hypotheses $H_0 = \{H_{(1)}, \ldots, H_{(n)}\}$. Then H_0 is rejected if $P_{(j)} \leq j\alpha/n$ for any $j = 1, \ldots, n$.

This test procedure has type I error probability equal to α for independent tests as shown by the following result.

THEOREM. *Let $P_{(1)}, \ldots, P_{(n)}$ be the order statistics of n independent uniform* (0, 1) *random variables and let $A_n(\alpha) = \mathrm{pr}\{P_{(j)} > j\alpha/n; j = 1, \ldots, n\}$ $(0 \leq \alpha \leq 1)$. Then $A_n(\alpha) = 1 - \alpha$.*

Proof. The result is clearly true for $n = 1$. For $n > 1$, $\{P_{(1)}/P_{(n)}, \ldots, P_{(n-1)}/P_{(n)}\}$ are the order statistics of $n-1$ independent uniform random variables on (0, 1), independent of $P_{(n)}$, and $P_{(n)}$ has distribution function p^n $(0 < p < 1)$. Hence

$$A_n(\alpha) = \int_\alpha^1 A_{n-1}\left\{\frac{\alpha(n-1)}{pn}\right\} np^{n-1}\, dp.$$

If $A_{n-1}(\alpha) = 1 - \alpha$ then $A_n(\alpha) = 1 - \alpha$ follows. Hence the result is proved by induction. □

The modified test procedure is conservative provided

$$\mathrm{pr}\left\{\bigcup_{j=1}^n P_{(j)} \leq j\alpha/n\right\} \leq \alpha.$$

This inequality is not true in general as counterexamples, albeit pathological, can be found. Nevertheless, it may well be true for a large family of multivariate distributions as suggested by the simulation studies below.

3. SIMULATION STUDIES

Test statistics $T_1, \ldots, T_n$ were simulated from an n-variate normal distribution, $N(0, \Omega)$ with unit variances and common correlation coefficients ρ $(0 < \rho < 1)$. Two-sided p-values were obtained from each univariate normal statistic as $P_i = 2\min(Y_i, 1 - Y_i)$, where $Y_i = \Phi(T_i)$, and Φ is the standard normal distribution function. Then the classical and modified Bonferroni test procedures were applied to each set of simulated p-values.

Random variables with a multivariate gamma distribution were constructed from linear combinations of independent gamma variables. The degree of dependence between the resulting random variables was determined by the number of gamma variables used in common for each sum. Let $\{X_{ij}\}$ be a set of independent gamma (θ_1, θ_2) variables. Then $T_1, \ldots, T_n$, defined by

$$T_i = \sum_{j=1}^{l} X_{0j} + \sum_{j=1}^{m-l} X_{ij} \quad (i = 1, \ldots, n),$$

is a set of gamma $(m\theta_1, \theta_2)$ variables with common correlation coefficient $\rho = l/m$. Individual chi-squared test statistics with 1 and 5 degrees of freedom were constructed by choosing $m = 10$ and $\theta_1 = 0{\cdot}05$ and $0{\cdot}25$ respectively. Then P-values corresponding to the right-hand tail of each gamma variate were obtained: $P_i = 1 - G(Y_i)$ for $i = 1, \ldots, n$, where G is the gamma $(m\theta_1, \theta_2)$ distribution function.

Simulations were carried out on a VAX 11/780 computer using IMSL subroutines GGNSM for multivariate normal and GGAMR for gamma variables. The results of the type I error rates for the modified and classical Bonferroni procedures are shown in Table 1 using $n = 5$ and 10 simultaneous

Table 1. *Type I error rates* for modified,* M, *and classical,* C, *Bonferroni test procedures; ρ, correlation coefficient*

	Distribution						Distribution					
ρ	Normal		χ^2_1		χ^2_5		Normal		χ^2_1		χ^2_5	
	M	C	M	C	M	C	M	C	M	C	M	C
	Number of tests = 5						Number of tests = 10					
0·0	0·049	0·048	0·050	0·049	0·049	0·048	0·049	0·048	0·049	0·048	0·049	0·048
0·3	0·049	0·048	0·045	0·040	0·044	0·040	0·047	0·045	0·043	0·037	0·042	0·039
0·6	0·043	0·039	0·044	0·029	0·039	0·033	0·039	0·034	0·042	0·026	0·035	0·029
0·9	0·033	0·024	0·048	0·016	0·041	0·019	0·028	0·017	0·047	0·012	0·039	0·014

* Based on 100000 simulations each; estimated standard error $\leq 0{\cdot}007$

tests, $\rho = 0\cdot0$, $0\cdot3$, $0\cdot6$ and $0\cdot9$ and $\alpha = 0\cdot05$. The results for the modified procedure, based on 100000 simulations in each case, are consistent with an upper bound on the type I error probability of 0·05. The estimated error rate drops as low as 0·028 for highly correlated multivariate normal statistics but is in the range 0·04–0·05 for most conditions simulated. The results for the classical Bonferroni procedure demonstrate that it has a similar type I error rate for independent tests but is appreciably more conservative than the modified procedure for highly correlated tests. This is particularly so for the χ_1^2 distribution.

4. Power comparisons

Since the modified Bonferroni test procedure contains the classical Bonferroni procedure it is clear that the power of the modified procedure is greater than the classical procedure at the same nominal significance level. A simulation study, undertaken to evaluate the relative power of the procedures for a range of alternative hypotheses, is illustrated in Table 2 for the multivariate normal case with 10 simultaneous tests. Alternative hypotheses examined were of the form H_A: $\mu_i = \mu$ $(i = 1, \ldots, k \leq m)$, $\mu_i = 0$ otherwise, with choices of $\delta = \frac{1}{2}$, 1 or $1\frac{1}{2}$. The results are expressed as the ratio of the powers of the classical to modified procedures for each alternative.

Table 2. *Power of classical Bonferroni test procedure relative to modified procedure*; multivariate normal, $n = 10$; ρ, correlation coefficient*

No. of correct alternatives	ρ	$\mu = \frac{1}{2}$	$\mu = 1$	$\mu = 1\frac{1}{2}$	No. of correct alternatives	$\mu = \frac{1}{2}$	$\mu = 1$	$\mu = 1\frac{1}{2}$
5	0·0	0·97	0·97	0·96	10	0·97	0·95	0·94
	0·3	0·96	0·95	0·96		0·94	0·92	0·94
	0·6	0·90	0·92	0·94		0·87	0·90	0·91
	0·9	0·73	0·81	0·85		0·64	0·71	0·75

* Based on 15000 simulations; each estimate of relative power has standard error less than $\frac{1}{2}$%.

The results for multivariate normal tests indicate little advantage to the modified test procedure over the classical method when the test statistics are independent or poorly correlated. However, the modified procedure is considerably more powerful when the test statistics are highly correlated and several alternative hypotheses are correct, particularly when the magnitude of the alternatives is small. Presumably, this is due mainly to the unduly small true type I error of the classical procedure.

5. Discussion

A criticism of the classical Bonferroni test procedure is that it is too conservative for highly-correlated test statistics. Then the modified procedure should be advantageous by having an actual significance level much closer to the nominal level and a consequent lower type II error probability. Even when the advantage is small, the only disadvantage seems to be a slight increase in computation.

Since the Bonferroni inequality leads to a conservative test procedure, there have been several attempts to improve on the method. Šidák (1968, 1971) has shown that the significance level for each test α/n can be improved by using $1-(1-\alpha)^{1/n}$ under certain conditions, although the degree of improvement for $n < 10$ and $\alpha = 0\cdot05$ is slight. Worsley (1982) found an upper bound on the probability of a type I error which is an improvement over both the Bonferroni and Šidák upper bounds, but it requires knowledge of the joint probabilities of pairs of events and thus is not directly applicable here.

If the overall null hypothesis H_0 is rejected, statements about individual hypotheses can be made using the classical Bonferroni test procedure: any individual null hypothesis H_i can be rejected provided $P_i \leq \alpha/n$. An improvement on this method suggested by Holm (1979) is the sequentially rejective Bonferroni test. This procedure rejects the specific hypothesis $H_{(i)}$ for $i = 1, \ldots, n$, provided both $P_{(i)} \leq \alpha/(n-i+1)$ and $H_{(1)}, \ldots, H_{(i-1)}$ have all been rejected. The sequential test procedure has multiple level of significance α for free combinations of null hypotheses. A question arises as to what statements about individual hypotheses can be made using the modified Bonferroni test procedure. One possiblity is to reject the individual hypotheses $H_{(1)}, \ldots, H_{(j)}$, where $j = \max\{j: P_{(j)} \leq j\alpha/n\}$. However, since there is no formal basis for rejecting the individual hypotheses $H_{(i+1)}, \ldots, H_{(j)}$ not rejected by the sequentially rejective test, statements about these latter hypotheses should be considered exploratory and preferably confirmed in subsequent studies.

ACKNOWLEDGEMENTS

The work was supported by an NH&MRC Applied Health Sciences Research Fellowship and by a National Cancer Institute grant and undertaken in part at the Harvard School of Public Health. The author would like to thank Malcolm Hudson, Marvin Zelen and a referee for constructive suggestions and a shorter proof of the theorem.

REFERENCES

HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* **6**, 65–70.
MILLER, R. G. (1981). *Simultaneous Statistical Inference*, 2nd ed. New York: Springer-Verlag.
ŠIDÁK, Z. (1968). On multivariate normal probabilities of rectangles: their dependence on correlations. *Ann. Math. Statist.* **39**, 1425–34.
ŠIDÁK, Z. (1971). On probabilities of rectangles in multivariate Student distributions: their dependence on correlations. *Ann. Math. Statist.* **42**, 169–75.
WORSLEY, K. J. (1982). An improved Bonferroni inequality and applications. *Biometrika* **69**, 297–302.

[*Received June* 1985. *Revised February* 1986]