BONFERRONI'S INEQUALITIES WITH APPLICATIONS TO TESTS OF STATISTICAL HYPOTHESES

by

RAYMOND NIEL CARR

B. A., Southwestern College, 1963

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1965

Approved by:

Major Professor

TABLE OF CONTENTS

INTRODUCTION

The problem of testing a statistical hypothesis was formulated by Neyman and Pearson (1933) as is given below.

A random variable X is known to be distributed over a space S according to some member of a family $\Gamma = \{F(X|\theta),\ \theta \in \Omega\}$ of probability distributions. A statistical hypothesis, H_ω, specifies a subset ω of the parameter space, Ω, and states that the distribution of X is $F(X|\theta)$ where $\theta \in \omega$. Any subset, s, of S may be considered a test of H_ω with the convention that H_ω is rejected if x, the observed value of X, is in s. Otherwise H_ω is accepted. The test is selected in the following manner:

A number α $(0 < \alpha < 1)$, called the level of significance of the test, is selected and s must be such that

$$P(X \in s \mid \theta = \theta_o) \leq \alpha \quad \text{for all} \quad \theta_o \in \omega. \tag{1}$$

Subject to this restriction, it is desired to maximize

$$P(X \in s \mid \theta = \theta_1) \quad \text{for all} \quad \theta_1 \in \bar{\omega}(\Omega). \tag{2}$$

The interpretation of these conditions is straightforward. Since $P(X \in s \mid \theta)$ is the probability of rejecting H_ω under the assumption that $F(X \mid \theta)$ is the distribution of X, condition (1) states that the probability of rejecting H_ω when in fact H_ω is true is to be at most α. Likewise condition (2) is that H_ω is to be rejected with high probability, called the power of the test, when in fact H_ω is false.

In practice, however, the test s is generally transformed to a test

$$s' = \{t\colon t \geq c\} \tag{3}$$

where $t = t(x_1, \ldots, x_n)$ is a function of the variates, and possibly of known parameters, and c is the critical value of the statistic t, i.e.

$$P(t \geq c) \leq \alpha \tag{4}$$

For many tests, say for outliers, the statistics involved are extreme statistics and the exact value of c for a given α or of α for a given c is difficult to obtain. Useful bounds, in either case, may be found by application of Bonferroni's Inequalities.

BONFERRONI'S INEQUALITIES

If $A_1, A_2, \ldots, A_N$ are N events then $A = \bigcup_{\alpha=1}^{N} A_\alpha$ denotes the event that at least one of the events A_α $(\alpha = 1, \ldots, N)$ occur. Let $p_{\alpha,\beta\ldots\gamma}$ denote the joint probability of j $(j \leq N)$ events $A_\alpha, A_\beta, \ldots, A_\gamma$; and S_j the sum of the $\binom{N}{j}$ p's with j subscripts. Then P(A) and the probability, $P_{[m]}$, that exactly m $(i \leq m \leq N)$ of the N events $A_\alpha(\alpha=1,\ldots,N)$ occur simultaneously are (Feller, 1957)

$$P(A) = \sum_{j=1}^{N} (-1)^{j-1} S_j \tag{5}$$

and

$$P_{[m]} = \sum_{j=m}^{N} (-1)^{j-m} \binom{j}{m} S_j, \tag{6}$$

respectively.

<u>Theorem I</u>: For any integer m $(1 \leq m \leq N)$ the probability, P_m, that at least m of the events $A_\alpha(\alpha = 1, \ldots, N)$ occur simultaneously is given by

$$P_m = \sum_{j=m}^{N} (-1)^{j-m} \binom{j-1}{m-1} S_j. \tag{7}$$

Proof: Consider the relationship

$$P_{m+1} = P_m - P_{[m]}. \tag{8}$$

Now, if $m = 1$, (7) becomes

$$P_1 = \sum_{j=1}^{N} (-1)^{j-1} \binom{j-1}{0} S_j$$

$$= \sum_{j=1}^{N} (-1)^{j-1} S_j$$

which is (5). Thus (7) holds for $m = 1$.

If $m = 2$, by (8)

$$P_2 = P_1 - P_{[1]}$$

$$= \sum_{j=1}^{N} (-1)^{j-1} S_j - \sum_{j=1}^{N} (-1)^{j-1} \binom{j}{1} S_j$$

$$= \sum_{j=1}^{N} (-1)^{j-1} (1-j) S_j$$

$$= \sum_{j=2}^{N} (-1)^{j-2} (j-1) S_j$$

$$= \sum_{j=2}^{N} (-1)^{j-2} \binom{j-1}{1} S_j$$

which is (7) with $m = 2$.

Now, assume (7) holds for $m = m - 1$.

Then

$$P_{m-1} = \sum_{j=m-1}^{N} (-1)^{j-m+1} \binom{j-1}{m-2} S_j .$$

Applying (8) to obtain P_m gives

$$P_m = P_{m-1} - P_{[m-1]}$$

from which, with $m = m - 1$ in (6) and (7),

$$P_m = \sum_{j=m-1}^{N} (-1)^{j-m+1} \binom{j-1}{m-2} S_j - \sum_{j=m-1}^{N} (-1)^{j-m+1} \binom{j}{m-1} S_j$$

$$= \sum_{j=m-1}^{N} (-1)^{j-m+1} \left[\binom{j-1}{m-2} - \binom{j}{m-1}\right] S_j$$

$$= \sum_{j=m-1}^{N} (-1)^{j-m} \binom{j-1}{m-1} S_j$$

$$= \sum_{j=m}^{N} (-1)^{j-m} \binom{j-1}{m-1} S_j$$

which is (7) for m = m.

It has been shown that if (7) holds for $m = m - 1$ it holds for $m = m$ and since it holds for $m = 1, 2$ it holds for $m = 3, 4, \ldots, N$.

It should be noted that evaluation of P_m requires knowledge of the $N - m + 1$ sums $S_m, \ldots, S_N$, which in turn require knowledge of the probabilities of all possible occurrences of $m, m + 1, \ldots, N$ of the events A_α ($\alpha = 1, \ldots, N$). This knowledge is not always readily available to the statistician. In view of this fact, the following theorem is very useful.

Theorem II: For an approximation

$$\hat{P}_m^{(r)} = \sum_{j=m}^{m+r-1} (-1)^{j-m} \binom{j-1}{m-1} S_j \tag{9}$$

of P_m involving only the r $(1 \leq r \leq N - m + 1)$ sums $S_m, S_{m+1}, \ldots, S_{m+r-1}$, the error $(P_m - \hat{P}_m^{(r)})$ is

$$\varepsilon_r = \sum_{j=m+r}^{N} (-1)^{j-m} \binom{j-1}{m-1} S_j \tag{10}$$

which has the sign of the first term omitted and is less in absolute value. Thus the sign of ε_r is $(-1)^r$ and

$$|\varepsilon_r| \leq \binom{m+r-1}{m-1} S_{m+r} \tag{11}$$

Proof:

$$\varepsilon_r = P_m - \hat{P}_m^{(r)}$$

$$= \sum_{j=m}^{N} (-1)^{j-m} \binom{j-1}{m-1} S_j - \sum_{j=m}^{m+r-1} (-1)^{j-m} \binom{j-1}{m-1} S_j$$

$$= \sum_{j=m+r}^{N} (-1)^{j-m} \binom{j-1}{m-1} S_j$$

which proves (10).

To prove that ε_r has the properties given in Theorem II, the following lemma is needed.

Lemma: The S_j can be expressed in terms of the P_m as

$$S_j = \sum_{m=j}^{N} \binom{m-1}{j-1} P_m. \tag{12}$$

Proof: By (7)

$$\underline{P} = C\,\underline{S}$$

where

$$\underline{P} = [P_1, P_2, \ldots, P_N]'$$

$$\underline{S} = [S_1, S_2, \ldots, S_N]'$$

and

$$C = \left[(-1)^{i+j} \binom{j-1}{i-1}\right]$$

is an NxN upper triangular matrix with 1's on the diagonal. Since $|C| \neq 0$, C is nonsingular and thus has an inverse, C^{-1}. Thus

$$\underline{S} = C^{-1}\underline{P}$$

where

$$C^{-1} = \left[\binom{j-1}{i-1}\right]$$

is also an upper triangular matrix with 1's as diagonal elements. Now, equating the j^{th} elements of $\underline{S}$ and $C^{-1}\underline{P}$ gives

$$S_j = \sum_{m=j}^{N} \binom{m-1}{j-1} P_m$$

which is (12).

Proof of Theorem II (continued): By direct substitution for the S_j in terms of the P_k, (10) becomes

$$\varepsilon_r = \sum_{j=m+r}^{N} \left\{(-1)^{j-m} \binom{j-1}{m-1}\left[\sum_{k=j}^{N} \binom{k-1}{j-1} P_k\right]\right\} \quad . \qquad (13)$$

For any k $(m + r \leq k \leq N)$, the coefficient of P_k is

$$\sum_{j=m+r}^{N} (-1)^{j-m} \binom{j-1}{m-1}\binom{k-1}{j-1} = \sum_{j=m+r}^{k} (-1)^{j-m} \binom{j-1}{m-1}\binom{k-1}{j-1}$$

$$= \binom{k-1}{m-1} \sum_{j=m+r}^{k} (-1)^{j-m} \binom{k-m}{j-m} \tag{14}$$

But,

$$\sum_{j=m+r}^{k} (-1)^{j-m} \binom{k-m}{j-m} = \sum_{j=m}^{k} (-1)^{j-m} \binom{k-m}{j-m} - \sum_{j=m}^{m+r-1} (-1)^{j-m} \binom{k-m}{j-m}$$

$$= (1-1)^{k-m} + \sum_{j=m}^{m+r-1} (-1)^{j-m+1} \binom{k-m}{j-m}$$

$$= \sum_{i=0}^{r-1} (-1)^{i+1} \binom{k-m}{i} \tag{15}$$

where $i = j - m$.

Applying the combinatorial result

$$\sum_{i=0}^{n} (-1)^{i} \binom{a}{i} = (-1)^{n} \binom{a-1}{n}$$

to the right hand side of (15)

$$\sum_{j=m+r}^{k} (-1)^{j-m} \binom{k-m}{j-m} = (-1)^{r} \binom{k-m-1}{r-1} \; . \tag{16}$$

Note that $\binom{k-m-1}{r-1} > 0$ since $r \leq k - m$, and thus

$$\sum_{j=m+r}^{N} (-1)^{j-m-r} \binom{j-1}{m-1}\binom{k-1}{j-1} > 0 \tag{17}$$

since $\binom{k-1}{m-1} > 0$ in (14). Therefore $(-1)^{r} \varepsilon_{r} \geq 0$ by (13) and it follows that the sign of ε_r is $(-1)^r$, proving the first properly of ε_r in Theorem II.

Now, multiplying both sides of (10) by $(-1)^r$ gives

$$\sum_{j=m+r}^{N} (-1)^{j-m+r} \binom{j-1}{m-1} S_j \geq 0$$

for all r $(1 \leq r \leq N - m)$.

Thus for $r = r + 1$,

$$\sum_{j=m+r+1}^{N} (-1)^{j-m+r+1} \binom{j-1}{m-1} S_j \geq 0$$

so that

$$\sum_{j=m+r+1}^{N} (-1)^{j-m+r} \binom{j-1}{m-1} S_j \leq 0 .$$

Hence

$$\sum_{j=m+r}^{N} (-1)^{j-m+r} \binom{j-1}{m-1} S_j \leq \binom{m+r-1}{m-1} S_{m+r} . \tag{18}$$

But the left hand side of (18) is

$$(-1)^r \varepsilon = |\varepsilon|$$

and thus

$$|\varepsilon| \leq \binom{m+r-1}{m-1} S_{m+r}$$

which is (11), proving the second property of ε_r in Theorem II.

By theorem II a set of inequalities which give bounds on P_m may be obtained, namely

$$\sum_{j=m}^{m+r}(-1)^{j-m}\binom{j-1}{m-1}S_j \le P_m \le \sum_{j=m}^{m+r-1}(-1)^{j-m}\binom{j-1}{m-1}S_j \qquad (r \text{ odd})$$

(19)

$$\sum_{j=m}^{m+r-1}(-1)^{j-m}\binom{j-1}{m-1}S_j \le P_m \le \sum_{j=m}^{m+r}(-1)^{j-m}\binom{j-1}{m-1}S_j \qquad (r \text{ even})$$

for $r = 1, 2, \ldots, N - m + 1$.

The inequalities given by (19) are known as Bonferroni's Inequalities. The special case for $m = 1$ is particularly useful in testing hypotheses (or in confidence region procedures) as many important statistics can be expressed as maxima (or minima).

APPLICATIONS OF BONFERRONI'S INEQUALITIES

Application to the Extreme Deviate from the Sample Mean

Let $x_i (i = 1, \ldots, n)$ be n independent variates drawn from normal populations with means $\mu + \lambda_i$ and common unit variance. Suppose it is desired to test the following hypotheses:

H_o: $\lambda_i = 0$ for all i

H_A: $\lambda_i > 0$ for one (possibly a few), but

$\lambda_i = 0$ for the others.

A test given in terms of the extreme deviate from the sample mean, d, as

$$s = \{d : d \geq c\}$$

where

$$d = \max_{1 \leq i \leq n} d_i = x_{max} - \bar{x}$$

and c is the upper 100α - percentage point of the distribution of d has a level of significance α .

The problem encountered here is the determination of c given α or, equivalently, of α given c . One possible solution is to choose c such that

$$P(d_i \geq c \mid H_o) = \frac{\alpha}{n} \qquad \text{for } i = 1, \ldots, n \text{ .} \tag{20}$$

This simplifies the problem considerably since, under H_o, the d_i are normally distributed with

i) $E(d_i) = 0$

ii) $\sigma^2(d_i) = (n-1)/n$

iii) $Cov(d_i, d_j) = -1/n, \quad i \neq j$.

Thus c may be found as a solution to

$$\alpha = \frac{n}{\sqrt{2\pi}} \int_{c(\frac{n}{n-1})^{1/2}}^{\infty} e^{-\frac{1}{2}t^2} \, dt \tag{21}$$

which is extensively tabled. McKay (1935) suggested (20) as a first approximation to the critical value of this test. That this approximation of the actual critical value is useful however, it must be shown that the error level of the test using c as a critical value is close to α. This may be accomplished by obtaining bounds on $P(d \geq c \mid H_o)$ with Bonferroni's inequalities. Now, $P(d \geq c) = P(\text{at least one } d_i \geq c)$ so that, applying Bonferroni's inequality with $m = 1$ and $r = 1$ where A_i is the event $d_i \geq c$ $(i = 1, \ldots, n)$,

$$S_1 - S_2 \leq P(d \geq c) \leq S_1 \tag{22}$$

where

$$S_1 = \sum_{i=1}^{n} P(d_i \geq c) \tag{23}$$

and

$$S_2 = \sum_{i<j} P(d_i \geq c,\ d_j \geq c)\ . \tag{24}$$

Under the assumption that the null hypothesis is true, (23) becomes

$$\begin{aligned} S_1 &= nP(d_i \geq c) \\ &= n\left(\frac{\alpha}{n}\right) \\ &= \alpha \end{aligned}$$

and (24) becomes

$$S_2 = \binom{n}{2} P(d_i \geq c,\ d_j \geq c) \tag{25}$$

since the d_i are identically distributed and c was chosen to satisfy (20).

Thus bounds on the error level are found to be

$$\alpha - \binom{n}{2} P(d_i \geq c, d_j \geq c \mid H_o) \leq P(d \geq c \mid H_o) \leq \alpha \tag{26}$$

where

$$P(d_i \geq c, d_j \geq c \mid H_o) = \int_c^\infty \int_c^\infty n(\underline{0}, \frac{n-1}{n} V)\, dd_i dd_j$$

with

$$V = \begin{bmatrix} 1 & -\frac{1}{n-1} \\ -\frac{1}{n-1} & 1 \end{bmatrix}$$

is tabled for various values of c and p (K. Pearson, 1931).

However, due to the negative correlation, between d_i and d_j,

$$P(d_i \leq c, d_j \geq c \mid H_o) < \left[P(d_i \geq c \mid H_o)\right]^2 = \frac{\alpha^2}{n^2} \tag{27}$$

so that

$$\alpha - \frac{1}{2}(n-1)\alpha^2/n \leq P(d \geq c \mid H_o) \leq \alpha \quad . \tag{28}$$

Also, due to the nature of the coefficient of $\alpha^2/2$ in the left hand inequality of (28), namely that

$$\lim_{n\to\infty} \left(\frac{n-1}{n}\right) = 1 \text{ and } \frac{n-1}{n} < \frac{n}{n+1} \quad ,$$

$$\alpha - \frac{1}{2}\alpha^2 \leq P(d \geq c \mid H_o) \leq \alpha \tag{29}$$

for all n.

From (29) it is evident that while the test is conservative the error level is quite close to the level of significance. The following table has been prepared to illustrate the closeness of the bounds.

Table 1. Bounds on $P(d \geq c \mid H_o)$ with lower bounds as entries.

Upper Bound (α)	.01	.05	.10
n = 10	.009955	.048875	.095500
15	.009954	.048835	.095334
20	.009952	.048812	.095250
∞	.009950	.048750	.095000

Entries for n = 10, 15, 20 were computed from (28).
Entries for last line was computed from (29).

From Table 1 it is seen that with level of significance .01 (.05,.10) the bounds on the error level of the test described in this section are

$$\left.\begin{matrix} .009950 \\ .048750 \\ .095000 \end{matrix}\right\} \leq P(d \geq c \mid H_o) \leq \left\{\begin{matrix} .01 \\ .05 \\ .10 \end{matrix}\right. ,$$

for all n. That improvement of these bounds by using (28) rather than (29) is slight is evident from Table 1 - although the bounds do become better as n decreases as was explained in deriving (29) from (28). It is also evident that due to the nature of the correlation between d_i and d_j (namely, $-\frac{1}{n-1}$)

$$P(d_i \geq c, d_j \geq c \mid H_o) = P(d_i \geq c \mid H_o)^2 \qquad (30)$$

for all practical purposes if $n \geq 10$, say. Thus, unless the sample is quite small, the improvement in bounds obtained from (26) over those obtained from (28) is of no consequence.

Bonferroni's inequalities could also be used to obtain quite good bounds on the upper percentage points of the distribution of d under the assumption of a true null hypothesis. To see this, recall that under H_o

$$nP(d_i \geq c) - \binom{n}{2} P(d_i \geq c,\ d_j \geq c) \leq P(d \geq c) \leq nP(d_i \geq c)\ . \tag{31}$$

An upper bound, c_1, and a lower bound, c_2, on the upper 100α - percentage point, c, may be obtained in the following manner:

Given α and n, solve for c_1 in

$$nP(d_i \geq c_1) = \alpha \tag{32}$$

and for c_2 in

$$nP(d_i \geq c_2) - \binom{n}{2} P(d_i \geq c_2,\ d_j \geq c_2) = \alpha\ . \tag{33}$$

Solving (33) for c_2, however, is not very convenient, but $\binom{n}{2} P(d_i \geq c_2,\ d_j \geq c_2)$ may be replaced by $\frac{1}{2}(n-1)\alpha^2/n$ to obtain a lower bound as the solution of

$$nP(d_i \geq c_2) = \alpha + \frac{1}{2}(n-1)\ \alpha^2/n\ . \tag{34}$$

Actually, the distribution of d, under the assumption of H_o, has been tabled by Thigpen and David (1961) for $\alpha = .10, .05, .025, .01$ and $.005$ and $n = 2\ (1)\ 10$. Nair (1948) computed the probability integral to six decimal places in increments of 0.01 for $n = 3\ (1)\ 9$ and Grubbs (1950) computed the c.d.f. to five decimal places in increments of 0.05 for $n = 2\ (1)\ 25$.

Application to the Maximum Absolute Deviate

Let $x_i (i = 1, \ldots, n)$ be n independent normally distributed variates with means $\mu + \lambda_i$ and common variance of unity. Suppose it is desired to test the hypotheses

$$H_o: \lambda_i = 0 \text{ for all } i$$

$$H_A: \lambda_i \neq 0 \text{ for one (possibly a few) } i \text{ but } \lambda_j = 0 \text{ for all } j \neq i.$$

The statistic commonly used for this test is the maximum absolute deviate

$$d = \max_{1 \leq i \leq n} |d_i|$$

where $d_i = x_i - \bar{x}$, and the test is given by $s = \{d \geq c\}$ where c is determined by

$$P(d \geq c) = \alpha \ .$$

An approximation, c_1, for c may be obtained by solving for c_1 in

$$P(d_i \geq c_1 \mid H_o) = \frac{\alpha}{2n} \quad . \tag{35}$$

Bounds on the error level of test determined by the critical value c_1 may then be obtained from Bonferroni's inequality with $m = 1$ and $r = 1$ where A_i is the event that $|d_i| \geq c_1$.

Now,

$$P(d \geq c_1) = P(\text{at least one } |d_i| \geq c_1) .$$

Thus applying Bonferroni's inequality under the assumption of a true null hypothesis,

$$S_1 - S_2 \leq P(d \geq c_1) \leq S_1$$

where

$$S_1 = \sum_{i=1}^{n} P(|d_i| \geq c_1)$$

$$= nP(|d_i| \geq c_1)$$

$$= 2n\ P(d_i \geq c_1) = \alpha \qquad (36)$$

and

$$S_2 = \sum_{i<j} P(|d_i| \geq c_1,\ |d_j| \geq c_1)$$

$$= \binom{n}{2}\ h(c_1,\ c_1,\ -\frac{1}{n-1}) \qquad (37)$$

The quantity $h(c_1,\ c_1,\ -\frac{1}{n-1})$ may be computed from Pearson's tables (1931) as

$$h(a,b,\ -\frac{1}{n-1}) = 2 \int_a^\infty \int_b^\infty n\ (\underline{0},\ \frac{n-1}{n}\ V)\ dt_1\ dt_2$$

$$+ 2 \int_a^\infty \int_b^\infty n\ (\underline{0},\ \frac{n-1}{n}\ V_1)\ dt_1\ dt_2$$

where

$$V = \begin{bmatrix} 1 & -\frac{1}{n-1} \\ & \\ -\frac{1}{n-1} & 1 \end{bmatrix} \qquad \text{and} \quad V_1 = \begin{bmatrix} 1 & \frac{1}{n-1} \\ & \\ \frac{1}{n-1} & 1 \end{bmatrix} .$$

and $a = b = c_1$

However, for $n \geq 10$, say, $\rho \gtrsim 0$ and (37) becomes

$$S_2 = \binom{n}{2} \left(\frac{\alpha}{n}\right)^2$$

$$= \frac{1}{2}(n - 1)\ \alpha^2/n \quad . \tag{38}$$

Thus the error level of the test $s = \{d : d \geq c\}$

is bounded by $\frac{1}{2}(n - 1)\ \alpha^2/n$ and α where

$$\alpha = \frac{2n}{\sqrt{2\pi}} \int_{c(\frac{n}{n-1})^{1/2}}^{\infty} e^{-\frac{1}{2}t^2}\, dt$$

for n sufficiently large, as indicated above.

The upper 10.0, 5.0, 2.5, 1.0 and 0.5 percentage points of the maximum absolute standard normal deviate for $n = 2\ (1)\ 10$ were tabled by Thigpen and David (1961).

For $n > 10$, bounds on the percentage points could be obtained by application of Bonferroni's Inequalities as indicated in the previous section.

Application to the Studentized Extreme Deviate

If $x_i(i = 1, \ldots, n)$ are n independent normal variates with common mean μ and unknown variance σ^2, Dunnett and Sobel (1954) have shown that the variates defined by

$$d_i = \frac{x_i - \bar{x}}{s}$$

where s^2 is an estimate of σ^2 (obtained independent of the x_i) have a joint distribution which is an n-variate generalization of the Student t-distribution with the degrees of freedom of s^2, say ν, and correlation matrix $[\rho_{ij}]$ of the associated n-variate normal.

Thus given a set of variates, say $x_i (i = 1, \ldots, n)$ which are independently obtained from normal populations with means $\mu + \lambda_i$ and common, but unknown, variance σ^2, a test for the hypotheses

$$H_o: \lambda_i = 0 \text{ for all } i$$

$$H_A: \lambda_j > 0 \text{ for one (possibly a few) } j \text{ but}$$

$$\lambda_i = 0 \text{ for all } i \neq j$$

is given by $\quad s' = \{d \geq c\}$

where

$$d = \max_{1 \leq i \leq n} d_i = \frac{x_{max} - \bar{x}}{s}$$

and c is such that

$$\alpha = n \int_{c(\frac{n}{n-1})^{1/2}}^{\infty} f(t)\, dt$$

has an error level bounded by $\alpha - \frac{1}{2}(n - 1)\,\alpha^2/n$ and α. Here $f(t)$ is the ordinary Student-t density with ν degrees of freedom.

Applying Bonferroni's inequality with $m = 1$, $r = 1$ and A_i as the event that $d_i \geq c$ gives as bounds on the error level

$$S_1 - S_2 \leq P(d \geq c \mid H_o) \leq S_2$$

where

$$S_1 = \sum_{i=1}^{n} P(d_i \geq c \mid H_o)$$

$$= n\, P(d_i \geq c \mid H_o)$$

$$= \alpha$$

and

$$S_2 = \sum_{i<j} P(d_i \geq c,\ d_j \geq c \mid H_o)$$

$$= \binom{n}{2}\, d_\nu \left(c\left(\frac{n}{n-1}\right)^{1/2}, c\left(\frac{n}{n-1}\right)^{1/2}, -\frac{1}{n-1}\right).$$

The symbol $d_\nu(a,b,\rho)$ is defined by

$$d_\nu(a,b,\rho) = \int_a^\infty \int_b^\infty f(t_1,\ t_2,\ \rho)\ dt_1\ dt_2$$

with

$$f(t_1,\ t_2,\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \left[1 + \frac{t_1^{\,2} - 2\rho t_1 t_2 + t_2^{\,2}}{\nu(1-\rho^2)}\right]^{-\left(\frac{\nu}{2}+1\right)}$$

being the bivariate generalization of the Student t-distribution for which the probability intergal was evaluated by Dunnett and Sobel (1954). Here, $a = b = c\ \left(\frac{n}{n-1}\right)^{1/2}$ and $\rho = -\frac{1}{n-1}$. Thus bounds on $P(d \geq c \mid H_o)$ are,

letting $c' = c\left(\frac{n}{n-1}\right)^{1/2}$,

$$\alpha - \binom{n}{2} d_\nu \left(c', c', -\frac{1}{n-1}\right) \leq P(d \geq c \mid H_o) \leq \alpha . \tag{39}$$

However, since $\rho = -\frac{1}{n-1}$,

$$d_\nu \left(c', c', -\frac{1}{n-1}\right) < \left(\frac{\alpha}{n}\right)^2 \qquad \text{for all } n$$

giving bounds on the error level as

$$\alpha - \frac{1}{2}(n-1)\alpha^2/n \leq P(d \geq c \mid H_o) \leq \alpha . \tag{40}$$

And, since $\frac{1}{2}(n-1)\alpha^2/n < \frac{\alpha^2}{2}$ for all n, (40) becomes

$$\alpha - \frac{\alpha^2}{2} < P(d \geq c \mid H_o) \leq \alpha . \tag{41}$$

It should be noted that (40) and (41) are the bounds given in table 1. The distinction is in the value of c. For the extreme deviate from the sample mean (actually standardized) the $c\left(\frac{n}{n-1}\right)^{1/2}$ was taken as the upper 100α - percentage point of the standard normal distribution; for the studentized extreme deviate $c\left(\frac{n}{n-1}\right)^{1/2}$ was taken as the upper 100α - percentage point of the Student t-distribution with ν degrees of freedom.

Application to the Studentized Maximum Absolute Deviate

Halperin et al. (1955) computed upper and lower limits for percentage points of

$$d = \max |d_i| = \max \frac{|x_i - \bar{x}|}{s_\nu}$$

by use of Bonferroni's Inequalities, where the $x_i (i = 1, \ldots, n)$ are n independent normal variates with common mean μ and common, but unknown, variance σ^2 and s_ν^2 is an estimate of σ^2 made independently of the

x_i $(i = 1, \ldots, n)$ having ν degrees of freedom.

By Bonferroni's inequality with $m = 1$, $r = 1$, and A_i being the event that $|d_i| \geq c$, bounds or $P(d \geq c)$ are obtained as

$$nP(|d_i| \geq c) - \binom{n}{2} h_\nu \left(c', c', -\frac{1}{n-1}\right) \leq P(d \geq c) \leq nP(|d_i| \geq c) \tag{42}$$

where

$$nP(|d_i| \geq c) = 2n \int_{c(\frac{n}{n-1})^{1/2}}^{\infty} f(t)\, dt$$

and

$$c' = c\left(\frac{n}{n-1}\right)^{1/2}$$

with $f(t)$ being the Student t-distribution with ν degrees of freedom and

$$h_\nu\left(c', c', -\frac{1}{n-1}\right) = P(|d_i| \geq c,\ |d_i| \geq c)$$

$$= 2d_\nu\left(c', c', -\frac{1}{n-1}\right) + 2d_\nu\left(c', c', \frac{1}{n-1}\right)$$

with the notation $d_\nu\left(c', c', -\frac{1}{n-1}\right)$ being previously defined.

By solving

$$\int_{c(\frac{n}{n-1})^{1/2}}^{\infty} f(t)\, dt = \alpha$$

and

$$\int_{c(\frac{n}{n-1})^{1/2}}^{\infty} f(t)\, dt - \binom{n}{2} h_\nu \left(c', c', -\frac{1}{n-1}\right) = \alpha$$

for c, with $\alpha = .05$ and $.01$, lower and upper limits were obtained for the 5% and 1% points of the distribution of d. The tables were prepared for $n = 3\ (1)\ 10, 15, 20, 30, 40, 60$ and $\nu = 3\ (1)\ 10, 15, 20, 30, 40, 60, 120$ and ∞ in both cases.

Application to the Computation of Percentage Points
for the Studentized Extreme and Maximum Absolute Deviates

Quesenberry and David (1961) tabled the 1% and 5% points of the distribution of

$$b = \max_{1 \le i \le n} b_i = \frac{x_{max} - \bar{x}}{S}$$

and

$$b^* = \max_{1 \le i \le n} |b_i|$$

where

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{\nu+1} (y_i - \bar{y})^2,$$

by application of Bonferroni's inequalities. Here it is assumed that the x_i ($i = 1, \ldots, n$) are n independent normal variates with common mean μ and common variance σ^2 and the y_i ($i = 1, \ldots, \nu + 1$) are $\nu + 1$ variates, independent of the x_i, which may be used to obtain an estimate s_ν^2 of the variance σ^2. The values of the lower and upper bounds tabled are for $\alpha = .01, .05$, $n = 3\ (1)\ 10, 12, 15, 20$ and $\nu = 0\ (1)\ 10, 12, 15, 20, 25, 30, 40, 50$ for both b and b^*.

Quesenberry and David suggest using b and b^* as statistics in testing for outliers in a normal sample and in the treatment of the slippage problem for normal samples.

The density of b_i is

$$f(b_i) = \left(\frac{n}{n-1}\right)^{1/2} \frac{\Gamma\left[\frac{1}{2}(n + \nu - 1)\right]}{\sqrt{\pi}\ \Gamma\left[\frac{1}{2}(n + \nu - 2)\right]} \left[1 - n\, b_i^2 \;/\; (n-1)\right]^{\frac{1}{2}(n+\nu-4)}$$

$$\text{for } -\left(\frac{n-1}{n}\right)^{1/2} \leq b_i \leq \left(\frac{n-1}{n}\right)^{1/2} \tag{43}$$

and $f(b_i) = 0$ otherwise.

The joint density of b_i, b_j $(i \neq j)$ is

$$f(b_i, b_j) = \left(\frac{n}{n-2}\right)^{1/2} \frac{n+\nu-3}{2\pi} \left[1 - \frac{n-1}{n-2}\, b_i^2 - \frac{2b_i b_j}{n-2} - \frac{n-1}{n-2}\, b_j^2\right]^{\frac{1}{2}(n+\nu-5)}$$

over the ellipse

$$\frac{n-1}{n-2}\, b_i^2 - \frac{2b_i b_j}{n-2} + \frac{n-1}{n-2}\, b_j^2 \leq 1 \ . \tag{44}$$

By Bonferroni's inequality with $m = 1$, $r = 1$ and A_i being the event $b \geq c$

$$nP(b_i \geq c) - \binom{n}{2} P(b_i \geq c,\ b_j \geq c) \leq P(b \geq c)$$

$$\leq nP(b_i \geq c) \quad . \tag{45}$$

The upper bound was obtained by solving for c_1 in

$$nP(b_i \geq c_1) = \alpha \tag{46}$$

The lower bound was obtained by solving for c_2 in

$$nP(b_i \geq c_2) - \binom{n}{2} P(b_i \geq c_2,\ b_j \geq c_2) = \alpha \quad . \tag{47}$$

by the following iterative technique.

A first approximation to c_2, say $c_{2,0}$, is given by

$$nP(b_i \geq c_{2,0}) = \alpha + \binom{n}{2} P(b_i \geq c_1, b_j \geq c_1) . \tag{48}$$

On replacing c_1 by $c_{2,0}$ in (48) a second approximation $c_{2,1}$ is obtained. This process was continued until $c_{2,t+1}$ and $c_{2,t}$ agreed to three decimal places. It was found however that $c_{2,0}$ was sufficiently accurate in all but a few cases. The lower and upper bounds on b (for $\alpha = 0.05, 0.01$) agreed so well that only one value was tabled.

Essentially the same procedure was used to obtain the percentage points of b^*. The Bonferroni inequality gives

$$nP(|b_i| \geq c) - \binom{n}{2} P(|b_i| \geq c, |b_j| \geq c) \leq P(b^* \geq c)$$

$$\leq nP(|b_i| \geq c) . \tag{49}$$

But, from the symmetry of $f(b_i)$,

$$P(|b_i| \geq c) = 2P(b_i \geq c)$$

thus an upper bound c_1 can be obtained by solving (46) with α replaced by $1/2\ \alpha$. An upper bound can be obtained by use of the same iterative technique as was applied to (48) on

$$nP(b_i \geq c_{2,0}) = \frac{1}{2}\alpha + \binom{n}{2} P(|b_i| \geq c_1, |b_j| \geq c_1) \tag{50}$$

The bounds on the percentage points of b^* (for $\alpha = 0.05, 0.01$) did not agree as well as did those for the percentage points of b, but were very close.

Applications to Various Maxima Statistics

David (1956) has applied what is essentially Bonferroni's inequality in the evaluation of the probability of rejecting the largest of n observations by the use of $x_{max} - \bar{x}$ at the 5% level of significance when all observations are normal with unit variance, $n - 1$ having mean μ and one having mean $\mu + \lambda$.

The lower bounds on this probability (a power function) are tabled for $\lambda = 1, 2, 3, 4$ and $n = 3\ (1)\ 10, 12, 15, 20, 25$.

Wallace (1958) uses Bonferroni's inequalities to establish bounds on the error level of intersection confidence region procedures, based on the use of maxima statistics.

It should be understood that the various procedures reported in this paper have applications in situations dealing with minima (eg, $x_{min} - \bar{x}$) as well as maxima.

SUMMARY

Bonferroni's Inequalities on the probability that at least m of n events occur simultaneously may be used to give either bounds on the percentage points of the distribution of statistics involving extreme (maxima, minima or absolute) values or bounds on the error level of tests based on these statistics. In some situations it also permits the evaluation of a power function. In the case of tests for outliers in normal samples the error level is shown to be bounded by $\alpha - \frac{1}{2}(n - 1)\alpha^2/n$ and α, where α is the nominal error level. These bounds may be improved only slightly due to the nature of the correlation involved in evaluating the probabilities of joint occurrences. In fact bounds independent of n (namely $\alpha - \frac{1}{2}\alpha^2$ and α) can be obtained since the correlation is negative.

Halperin et al. (1955) computed upper and lower limits for the 5% and 1% points of the distribution of $d = \max\limits_{1 \le i \le n} \frac{|x_i - \bar{x}|}{s_\nu}$ for $n = 3\ (1)\ 10, 15,$ $20, 30, 40, 60$ and $\nu = 3\ (1)\ 10, 15, 20, 30, 40, 60, 120, \infty$ under the assumption of a normal parent population and the availability of an estimate, s_ν^2, of σ^2 (independent of the sample and having ν degrees of freedom) by an application of Bonferroni's inequalities.

Quesenberry and David (1961), by use of Bonferroni's inequalities, tabled bounds on the 5% and 1% points of the distributions of

$$b = \max_{1 \le i \le n} b_i = \frac{x_{max} - \bar{x}}{S}$$

and

$$b^* = \max_{1 \le i \le n} |b_i| ,$$

where

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + \nu s_\nu^2 \quad .$$

They assumed a normal parent population and the availability of an estimate, s_ν^2, of σ^2 (independent of the sample and having ν degrees of freedom). Values are given for b and b^* with $n = 3\ (1)\ 10, 12, 15, 20$ and $\nu = 0\ (1)\ 10, 12, 15, 20, 25, 30, 40, 50$.

David (1956) has used the lower limit of a Bonferroni inequality to compute the lower bounds on a power function for the extreme standardized statistic for the particular alternative of a single outlier; while Wallace (1958) has demonstrated the applicability of Bonferroni's inequalities to intersection region confidence procedures.

Reference is also given to the tabulation of the percentage points of a bivariate generalization of the Student t-distribution by Dunnett and Sobel (1954).

ACKNOWLEDGEMENT

The writer wishes to express his sincere appreciation to his major professor, Dr. Gary F. Krause, for suggestion of this topic and for his advise and assistance during the preparation of this report.

REFERENCES

David, H. A.
"On the application to statistics of an elementary theorem in probability." Biometrika, 1956, vol. 43.

Dunnett, C. W. and Sobel, M.
"A bivariate generalization of Student's t-distribution, with tables for certain special cases." Biometrika, 1954, vol. 41.

Feller, W.
An Introduction to Probability Theory and its Applications, John Wiley & Sons, Inc., 1957.

Grubbs, F. E.
"Sample criteria for testing outlying observations." Ann. Math. Stat., 1950, vol. 21.

Halperin, M., Greenhouse, S. W., Cornfield, J. and Zalokar, J.
"Tables of percentage points for the studentized maximum absolute deviate in normal samples." Jour. of the Am. Stat. Assoc., 1955, vol. 50.

McKay, A. T.
"The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe." Biometrika, 1935, vol. 36.

Nair, K. R.
"The distribution of the extreme deviate from the sample mean and its Studentized form." Biometrika, 1948, vol. 35.

Neyman, J. and Pearson, E. S.
"On the problem of the most efficient tests of statistical hypotheses." Phil. Trans. Roy. Soc., (Series A), 1933, vol. 231.

Pearson, K.
Tables for Statisticians and Biometricians, Part II, Cambridge Univ. Press, 1931.

Quesenberry, C. P. and David, H. A.
"Some tests for Outliers." Biometrika, 1961, vol. 48.

Thigpen, C. C. and David, H. A.
"Distribution of extremes in normal samples when the variables are equally correlated with common, known mean and common variance." Technical Report No. 51. Virginia Polytechnic Institute, Blacksburg, Virginia, September 1961.

Wallace, David L.
"Intersection Region confidence procedures with application to the location of the maximum in quadratic regression." Ann. Math. Stat., 1958, vol. 29.

BONFERRONI'S INEQUALITIES WITH APPLICATIONS TO TESTS OF STATISTICAL HYPOTHESES

by

RAYMOND NIEL CARR

B. A., Southwestern College, 1963

AN ABSTRACT OF
A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1965

The inequalities of Bonferroni have useful applications in the testing of statistical hypotheses--particularily in the detection of outlying observations. These inequalities may be used either to obtain quite good bounds on the error level of these tests or to obtain percentage points of the distributions of the statistics involved.

In the case of the tests for outlying observations considered in this paper the approximate bounds on the error level are shown to be $\alpha - \frac{1}{2}(n-1)\alpha^2/n$ and α for remarkably small n, where α is the nominal error level of the test. This result is due entirely to the form of the correlation between the deviations of any two observations from the sample mean.

Formulae for obtaining bounds on the error levels of tests based on maximum deviations, absolute maximum deviations and Studentized maximum and absolute maximum deviations are given--as are references to tables for computing these bounds--if one does not wish to use the approximate bounds $\alpha - \frac{1}{2}(n-1)\alpha^2/n$ and α . These formulae may also be used to tabulate upper and lower bounds on the percentage points of the statistics involved in the above mentioned tests.

In particular, the computational procedure for obtaining bounds on the percentage points of the distributions of

$$b = \max_i b_i = (x_{max} - \bar{x})/S$$

and

$$b^* = \max_i |b_i|$$

is discussed. Here

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + \nu s_\nu^2$$

is computed from the independent normal variates $x_i (i = 1, \ldots, n)$ with common mean and common variance and from s_ν^2 - an estimate of the common variance with ν degrees of freedom, independent of the x_i. Reference is given to the tabulation of these bounds.