Behavior Research Methods, Instruments, & Computers
1988, 20 (4), 416-418

PROGRAM ABSTRACTS/ALGORITHMS

A BASIC program for the multistage Bonferroni procedure for many correlations

ALFRED L. BROPHY
Behavioral Science Associates
West Chester, Pennsylvania

As is generally true with multiple analyses of data sets, significance tests of more than one correlation coefficient from a single sample pose a risk of an inflated Type I error. Larzelere and Mulaik (1977) devised a multistage Bonferroni procedure to test the significance of a set of correlations that successfully controls the familywise Type I error rate and that is more powerful than the customary single-stage Bonferroni method. Crosbie (1986) prepared a Pascal program to perform the multistage procedure.

Larzelere and Mulaik (1977) and Crosbie (1986) gave clear, concise descriptions of the multistage Bonferroni procedure. In brief, the procedure divides the desired familywise Type I error rate (a_f) equally among the number of tests to be made at each stage. Thus, in the first stage, each of the m correlation coefficients is tested at the $a_t = a_f/m$ level of significance, a_t being the per-test Type I error rate. The usual Bonferroni procedure terminates at that point, but the multistage method continues if at least one coefficient has been found to be significant. In the next stage, $m-k$ coefficients are tested at the $a_t = a_f/(m-k)$ level, where k is the number of coefficients earlier deemed significant. The process continues until no additional coefficients are found significant at a given stage.

The Appendix provides a listing of MANYCORR, a BASIC program that performs the multistage Bonferroni procedure with a set of correlation coefficients. Output from the program is highly similar to that of Crosbie's (1986) program. MANYCORR does, however, possess some advantages: it is relatively short, it employs a very accurate method of approximating critical values of the Pearsonian correlation coefficient (r), and it yields two supplementary statistics that Larzelere and Mulaik (1977) recommended for use in research reports. MANYCORR performs two-tailed tests, considering the size but not the sign of the coefficients. It determines whether a coefficient is significantly different from zero by comparing the coefficient with the critical value of r at the relevant a_t level with $\nu = (n-2)$ degrees of freedom, where n is the size of the sample.

The author's mailing address is Behavioral Science Associates, P.O. Box 748, West Chester, PA 19381.

Input and Output. MANYCORR prompts the user to enter m, n, a_f, the method of entering the correlation coefficients (data statements, sequential disk file, or keyboard), and the device to which output will be directed (monitor or printer). At the completion of each run, the user has the option of requesting another analysis at a different a_f level.

MANYCORR outputs the results at each stage, including the number of correlations tested, the critical value of r at the appropriate a_t level, the corresponding values of the standard normal deviate (z) and Student's t, and a list of correlations that have been found significant. As suggested by Larzelere and Mulaik (1977), at the end of each run, MANYCORR prints the value of a_t that is equivalent to the a_f level used in the analysis and the critical value of r at the $a_t = a_f$ level. Larzelere and Mulaik noted that the last statistic can identify coefficients of "borderline" (p. 566) significance, pending confirmatory evidence from other samples. The researcher is cautioned not to overinterpret such results, however.

The 15 coefficients contained in line 600 of the program listing were taken from a numerical example used by Larzelere and Mulaik (1977) and Crosbie (1986). Those sample data can be used in conjunction with the printout in Crosbie (pp. 328–329) to check the operation of the program.

Estimation of Critical Values. MANYCORR computes a different critical value of r at each stage and at the end of the analysis, using the relation $r = t/(t^2+\nu)^{1/2}$, where t is the critical value of the t distribution at the relevant a_t level with ν degrees of freedom. The inverse form of an approximation by Bailey (1980, Equation 7) estimates the critical value of t as

$$t = \left\{\nu \exp\left[\frac{z^2}{\nu}\left(\frac{4\nu^2+\nu+(4z^2+9)/12}{4\nu^2+5(2z^2+3)/24}\right)^2\right] - \nu\right\}^{1/2}, \quad (1)$$

where z is the standard normal deviate at the relevant a_t level. The value of z in Equation 1 is found by the approximation of Odeh and Evans (1974), which is accurate to five decimal places (Brophy, 1985). For $\nu \leq 2$ (corresponding to an improbably low n of 3 or 4), t is computed exactly by equations derived from Student (1908). In those cases z is not necessary, but it is approximated for display with the results. With parameters that are likely to occur in MANYCORR, the approximated t values usually have at least three-decimal-place accuracy, although they sometimes are correct to only one figure. The approximations are one-tailed, so the program halves a_t before branching to the approximation routine.

The accuracy of the approximated r was assessed for 10 levels of (two-tailed) a_t between .0001 and .2 and for all values of v between 1 and 40, as well as for selected larger vs. Approximated r values were accurate to within .00026 for all vs, and to within .00004 for $v\text{s} \geq 9$. These results are more accurate than results from the approximation in Crosbie (1986), which is a five-term Cornish-Fisher expansion (Fisher & Cornish, 1960) that was reprinted in Zelen and Severo (1964, section 26.7.5), tested by Harris (1976), and referred to by both Crosbie and Harris as a four-term approximation. That approximation is accurate to within .00079 and .00021 for $v \geq 3$ and $v \geq 9$, respectively. Fisher and Cornish gave a six-term expansion that is about as accurate as the approximation used in MANYCORR, but it is comparatively long.

The critical values of r, with their three- or four-decimal-place accuracy, should be adequate for the purpose of the program. They are rounded to four decimal places before use in the significance tests. The program prints critical values of z, t, and r to five, three, and four decimal places, respectively; some printed figures for t and r are occasionally incorrect.

Language and Memory Requirements. MANYCORR is written in GW-BASIC, but all statements used are supported by most other popular BASIC dialects. The program, which was tested on a Tandy 1000 microcomputer, runs with less than 3.5K with the sample data in line 600. A shorter version of MANYCORR that directs output only to the monitor can be obtained by deleting all lines with line numbers that are not multiples of 10.

Availability. A listing of the program, as shown in the Appendix, can be obtained without charge from the author.

REFERENCES

Bailey, B. J. R. (1980). Accurate normalizing transformations of a Student's t variate. *Applied Statistics*, **29**, 304-306.

Brophy, A. L. (1985). Approximation of the inverse normal distribution function. *Behavior Research Methods, Instruments, & Computers*, **17**, 415-417.

Crosbie, J. (1986). A Pascal program to perform the Bonferroni multistage multiple-correlation procedure. *Behavior Research Methods, Instruments, & Computers*, **18**, 327-329.

Fisher, R. A., & Cornish, E. A. (1960). The percentile points of distributions having known cumulants. *Technometrics*, **2**, 209-225.

Harris, R. J. (1976). Hand computation of Bonferroni critical values. *Perceptual & Motor Skills*, **42**, 515-521.

Larzelere, R. E., & Mulaik, S. A. (1977). Single-sample tests for many correlations. *Psychological Bulletin*, **84**, 557-569.

Odeh, R. E., & Evans, J. O. (1974). Algorithm AS 70: The percentage points of the normal distribution. *Applied Statistics*, **23**, 96-97.

Student. (1908). The probable error of a mean. *Biometrika*, **6**, 1-25.

Zelen, M., & Severo, N. C. (1964). Probability functions. In M. Abramowitz & I. A. Stegun (Eds.), *Handbook of mathematical functions* (pp. 925-995). Washington, DC: U.S. Government Printing Office.

APPENDIX
A BASIC Program to Perform the Multistage Bonferroni Procedure with a Set of Correlation Coefficients

```
10 T$="Multistage Bonferroni Test of Correlation Coefficients":
   CLS: PRINT T$: PRINT
20 INPUT"Number of correlations "; M: DIM R(M)
30 INPUT"Sample size "; NN: N=NN-2
40 INPUT"Do you want to enter the coefficients from:
   (1) Data statements,  (2) Disk file,  (3) Keyboard "; K:
   IF K<>1 AND K<>2 AND K<>3 THEN 40
45 INPUT"Do you want to direct output to:
   (1) Monitor  or  (2) Printer "; L: IF L<>1 AND L<>2 THEN 45
50 PRINT: ON K GOTO 60,70,90
60 FOR I=1 TO M: READ R(I): NEXT I: GOTO 100
70 INPUT"Prepare disk drive, and enter File Name "; F$
80 OPEN"I",1,F$: FOR I=1 TO M: INPUT#1,R(I): NEXT I: CLOSE:
   GOTO 100
90 FOR I=1 TO M: PRINT"Correlation No." I;: INPUT R(I): NEXT I
100 PRINT
105 IF L=1 THEN 110
106 INPUT"Prepare printer, and press <ENTER>", Z$
107 LPRINT T$: FOR I=1 TO M: LPRINT R(I);: NEXT I: LPRINT:
    GOTO 120
110 CLS: PRINT T$: FOR I=1 TO M: PRINT R(I);: NEXT I: PRINT
120 PRINT: P$="  "+STRING$(25,"-"): Q$=STRING$(55,"*")
130 INPUT"Familywise alpha level (e.g., .05) "; AL
140 K1=0: M1=M: J=0
145 IF L=2 THEN LPRINT Q$: GOTO 160
150 PRINT Q$
160 K=0: J=J+1
165 IF L=1 THEN 170
166 LPRINT"STAGE" J: LPRINT"For familywise alpha level of" AL
167 LPRINT"  with N =" NN "and" M1 "correlations to be tested":
    GOTO 190
170 PRINT"STAGE" J: PRINT"For familywise alpha level of" AL
180 PRINT"  with N =" NN "and" M1 "correlations to be tested"
190 P=AL/M1/2: GOSUB 360
195 IF L=1 THEN 200
196 LPRINT: LPRINT"The following correlations are significant:":
    LPRINT: GOTO 210
```

APPENDIX (Continued)

```
200 PRINT: PRINT"The following correlations are significant:":
    PRINT
210 FOR I=1 TO M: IF ABS(R(I))<=RC THEN 240
215 IF L=2 THEN LPRINT"  Correlation" I ": " R(I): GOTO 230
220 PRINT"  Correlation" I ": " R(I)
230 K=K+1
240 NEXT I
245 IF L=1 THEN 250
246 IF K=0 THEN LPRINT"  None"
247 LPRINT P$: GOTO 270
250 IF K=0 THEN PRINT"  None"
260 PRINT P$;: PRINT TAB(54)"Press <ENTER> to continue";:
    INPUT"",Z$
270 IF K<>K1 AND K<>M THEN M1=M-K: K1=K: GOTO 160
275 IF L=1 THEN 280
276 LPRINT"The equivalent per-test alpha level =" P+P: LPRINT
277 LPRINT"For per-test alpha level of" AL: LPRINT"  with N ="
    NN: GOTO 300
280 PRINT"The equivalent per-test alpha level =" P+P: PRINT
290 PRINT"For per-test alpha level of" AL: PRINT"  with N =" NN
300 P=AL/2: GOSUB 360
305 IF L=2 THEN 320
310 PRINT Q$
320 INPUT"Repeat with another familywise alpha level (Y/N) "; Z$
330 IF Z$<>"N" AND Z$<>"n" THEN 130
335 IF L=2 THEN LPRINT Q$
340 END
350 'Approximate t corresponding to p
360    IF P>0 AND AL<=1 THEN 390
365    IF L=2 THEN LPRINT"  Unacceptable alpha": GOTO 130
370    PRINT"  Unacceptable alpha": GOTO 130
380 ' Exact calculation for df = 1,2
390    IF N>2 THEN 430
400    IF N=1 THEN T=1/TAN(3.141593*P) ELSE T=SQR(1/(2*P*(1-P))
       -2)
410    GOSUB 500: GOTO 460
420 ' Bailey, 1980, Equation 7 (inverted)
430    GOSUB 500: H=4*N*N: A=Z*(H+N+Z*Z/3+.75)/(H+Z*Z/2.4+.625)
440    T=SQR(N*EXP(A*A/N)-N)
450 'Compute critical value of r
460    RC=T/SQR(T*T+N)
470 Z=INT(Z*100000!+.5)/100000!: T=INT(T*1000+.5)/1000:
    RC=INT(RC *10000+.5)/10000
475 IF L=2 THEN LPRINT"  z =" Z " t =" T ", critical r =" RC:
    RETURN
480 PRINT"  z =" Z ", t =" T ", critical r =" RC: RETURN
490 'Approximate z corresponding to p (Odeh & Evans, 1974)
500    Y=SQR(-2*LOG(P)): Z=Y-((((4.536422E-05*Y+2.042312E-02)*Y
       +.3422421)*Y+1)*Y+.3222324)/((((3.85607E-03*Y+.1035378)*Y
       +.5311035)*Y+.5885816)*Y+9.934846E-02): RETURN
600 DATA -.29,-.32,-.42,-.22,-.28,.35,.23,.33,.29,.12,-.13,-.18,
    -.23,-.11,-.22
```

(Revision accepted for publication December 22, 1987.)