A BASIC program for Tukey's multiple comparison procedure

ALFRED L. BROPHY

Behavioral Science Associates, West Chester, Pennsylvania

Tukey's honestly significant difference (HSD) test may be the best method for making pairwise multiple comparisons between means of groups with homogeneous variances (Dunnnett, 1980; Einot & Gabriel, 1975; Stoline, 1981; Zwick & Marascuilo, 1984). Although less powerful than some other multiple comparison procedures, it effectively controls the experiment-wise Type I error rate, it works with samples of equal or unequal size, and it provides simultaneous confidence intervals as well as significance tests.

Table 1 presents a BASIC program that performs Tukey's HSD test and displays the results. The program prompts the user to enter the number of samples to be compared (k), the mean and size of each sample, the variance as estimated by the mean square error (s^2), and the critical value of the studentized range statistic (q) at the relevant alpha level (α) for k and for the degrees of freedom associated with s^2 .

If the sample sizes (n) are equal, the minimum significant difference between means is the same for all comparisons, and the program calculates that quantity [$qs/(n)^{1/2}$]. For each pair of means, the program calculates the difference between the means ($d = \bar{x}_i - \bar{x}_j$) and the observed q ($q' = |d|/r$), where $r = s[(1/n_i + 1/n_j)/2]^{1/2}$, $1 \leq i < k$, and $j = i + 1$. Finally, the $100(1 - \alpha)\%$ confidence limits ($d \pm rq$) are calculated for each comparison. In displaying the results, the program prints an asterisk to signify a significant difference if $q' > q$. (A significant result also is indicated by confidence limits that do not include a value of zero.)

The program evaluates only nonredundant pairs of means. The results of differences calculated by subtracting $\bar{x}_i$ from $\bar{x}_j$, rather than the reverse, would be the same as the results yielded by the program, except for a change in the sign of each difference and of each confidence limit. The program is set for $\alpha = .05$, but this can be changed easily by assigning a different value to A in line 30 of the program.

The correct value of q that is required as input can be obtained from the extensive tables of Harter, Clemm, and Guthrie (1959) or from more accessible tables in Harter (1960), Pearson and Hartley (1966, Table 29), and some statistics texts. An approximation routine could be included to estimate the q value so that the user would not have to provide that value. Dunlap, Powell, and Konnerth (1977) noted that the Statistical

Table 1
A BASIC Program to Perform Tukey's HSD Test

```
10 DEFINT I-N
20 INPUT"NUMBER OF SAMPLES "; K
30 DIM N(K), X(K): L=1: A=.05
40 INPUT"ARE THE SAMPLES OF EQUAL SIZE (Y/N) "; Z$
50 IF Z$="Y" L=0: INPUT"SIZE OF EACH SAMPLE "; N
60 INPUT"MEAN SQUARE ERROR (MSE)"; S: S=SQR(S)
70 PRINT"CRITICAL VALUE OF STUDENTIZED RANGE STATISTIC FOR" A
  "LEVEL,"
80 PRINT" DF FOR MSE, AND" K "SAMPLES ";: INPUT Q
90 PRINT"ENTER THE DATA FOR EACH SAMPLE:"
100 FOR I=1 TO K: PRINT"SAMPLE" I
110 INPUT" MEAN "; X(I): IF L INPUT" N "; N(I): GOTO 130
120 N(I)=N
130 NEXT I
140 CLS: IF L THEN 180
150 PRINT"MINIMUM SIGNIFICANT DIFFERENCE BETWEEN MEANS (P ="
  A ")"
160 PRINTTAB(22) Q*S/SQR(N)
170 PRINT: GOSUB 320: CLS
180 PRINT"AN ASTERISK INDICATES A SIGNIFICANT DIFFERENCE (P <"
  A ")"
190 PRINT: M=2
200 PRINT, "DIFFERENCE", (1-A)*100 "% CONFIDENCE LIMITS"
210 PRINT"SAMPLES", "BET. MEANS", " LOWER", " UPPER"
220 FOR I=1 TO K-1: FOR J=I+1 TO K
230 D=X(I)-X(J): R=S*SQR((1/N(I)+1/N(J))/2): Q1=ABS(D)/R: W=R*Q
240 M=M+1: IF M=14 M=0: GOSUB 320
250 PRINT I "-" J, D,: IF Q1>Q PRINTTAB(26)"*";
260 PRINT, D-W, D+W
270 NEXT J,I
280 IF L AND K<6 THEN 310
290 Z$="": PRINTTAB(15)"REVIEW RESULTS AGAIN (Y/N) ";: INPUT Z$
300 IF Z$<"N" THEN 140
310 END
320 PRINTTAB(15)"PRESS <ENTER> TO CONTINUE ";: INPUT Z$: RETURN
```

Package for the Social Sciences used such a routine. The program might have applied the algorithm developed by Dunlap et al. to determine the probability corresponding to each q' . However, because, even after considerable modification to increase speed, a BASIC translation of the Dunlap et al. FORTRAN function required 9 min to evaluate a single q value on a TRS-80 Model I microcomputer, use of the routine in the program is impractical.

Numerical Example. Four treatment groups had means of 7.48, 6.35, 9.80, and 8.06, and sample sizes of 8, 17, 8, and 11, respectively. An analysis of variance (ANOVA) found that the means are significantly different [$F(3,40) = 7.58$, $p < .001$] and that the within-group mean square is 2.98. (A significant overall F test is not, however, required before using Tukey's procedure.) The means and sample sizes were entered into the program, together with the values of 2.98 for s^2 , taken from the ANOVA, and 3.791 for q , taken from Harter (1960, p. 1134). The results displayed by the program are shown in Table 2. Group 3 differs from Groups 1 and 2; none of the other comparisons are significant.

Language and Modifications. The program is written in TRS-80 Level II BASIC, a version of the popular Microsoft BASIC. It should run in most BASIC dialects with little or no change. Changes will be necessary for a properly formatted display on machines with a monitor width of less than 64 characters. If the display depth is greater than 16 lines, the value of the first M in the conditional statement in line 240 of the program can be

Table 2
Example of Results Displayed by the Program

Samples	Difference Between Means	95% Confidence Limits	
		Lower	Upper
1-2	1.13	-.854026	3.11403
1-3	-2.32*	-4.63375	-6.24681E-03
1-4	-.579999	-2.73022	1.57022
2-3	-3.45*	-5.43403	-1.46597
2-4	-1.71	-3.50063	.0806277
3-4	1.74	-.410217	3.89022

*Significant difference of $p < .05$.

increased to allow display of more results at one time. A printed copy of the results can be obtained by changing every PRINT (including those followed by TAB) in lines 150 through 260 to LPRINT.

The program uses 1,068 bytes, and the run-time memory requirement is only slightly greater. For example, processing the data of the example required an additional 127 bytes. The execution time for analysis of those data and display of the results on a TRS-80 Model I was 1.7 sec.

Availability. A listing of the program can be obtained without charge from Alfred L. Brophy, Behavioral Science Associates, P.O. Box 748, West Chester, Pennsylvania 19381.

REFERENCES

- DUNLAP, W. P., POWELL, R. S., & KONNERTH, T. K. (1977). A FORTRAN IV function for calculating probabilities associated with the studentized range statistic. *Behavior Research Methods & Instrumentation*, 9, 373-375.
- DUNNETT, C. W. (1980). Pairwise multiple comparisons in the homogeneous variance, unequal sample size case. *Journal of the American Statistical Association*, 75, 789-795.
- EINOT, I., & GABRIEL, K. R. (1975). A study of several methods of multiple comparisons. *Journal of the American Statistical Association*, 70, 574-583.
- HARTER, H. L. (1960). Tables of range and studentized range. *Annals of Mathematical Statistics*, 31, 1122-1147.
- HARTER, H. L., CLEMM, D. S., & GUTHRIE, E. H. (1959). *The probability integrals of the range and of the studentized range—Probability integral and percentage points of the studentized range; critical values for Duncan's new multiple range test* (Tech. Rep. No. 58-484, Vol. 2). Wright-Patterson Air Force Base, OH: Wright Air Development Center.
- PEARSON, E. S., & HARTLEY, H. O. (Eds.). (1966). *Biometrika tables for statisticians* (Vol. 1, 3rd ed.). Cambridge, England: Cambridge University Press.
- STOLINE, M. R. (1981). The status of multiple comparisons: Simultaneous estimation of all pairwise comparisons in one-way ANOVA designs. *American Statistician*, 35, 134-141.
- ZWICK, R., & MARASCUILO, L. A. (1984). Selection of pairwise multiple comparison procedures for parametric and nonparametric analysis of variance models. *Psychological Bulletin*, 95, 148-155.

(Manuscript accepted for publication March 9, 1984.)