

A FORTRAN IV function for calculating probabilities associated with the studentized range statistic

WILLIAM P. DUNLAP, RICHARD S. POWELL,
and TED K. KONNERTH
Tulane University, New Orleans, Louisiana 70118

Computer programs are available for calculating probabilities associated with most of the common parametric statistics, either by approximation (Jaspens, 1965; Veldman, 1967) or by direct solution (Dunlap & Duffy, 1975). Programs for probability levels associated with the studentized range statistic, on the other hand, are not readily available. The only such program that we have been able to find is the ONEWAY subroutine from the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975), and this program provides only an approximation via the fit of a polynomial surface to the 10%, 5%, and 1% tabled critical values of the studentized range. Although this program allows decisions regarding the null hypothesis, it does not provide probability values for particular studentized range statistics.

Although limited tabulations of the integral of the studentized range have been available for some time (e.g., Newman, 1939; Pearson & Hartley, 1943), the most accurate and extensive tables were compiled by Harter, Clemm, and Guthrie (1959). The Harter, Clemm, and Guthrie tables were the result of a multistage numerical integration procedure, but the probabilities reported were checked and are claimed to be accurate to at least six decimal places; this provides values against which to check the output of our program. The Harter et al. (1959) tables are reprinted in part by most statistical textbooks that describe range tests. The disadvantage of any tabular presentation of a statistical distribution, no matter how complete, is that only specific discrete values of the statistic are tabled. A computer function, on the other hand, provides a probability for any value of the statistic.

The program discussed in the present paper solves the studentized range integral in the following form:

$$1 - F(q) = c_v \int_0^\infty [4(t/q) z(t/q)]^\nu [1 - P_n(t)] \cdot \frac{1}{t} dt,$$

where

$$c_v = 2[\sqrt{\pi\nu}/4]^\nu / \Gamma(\nu/2)$$

$$z(x) = 1/\sqrt{2\pi} e^{-x^2/2}$$

$$P_n(t) = n \int_{-\infty}^\infty z(u) [E(u+t) - E(u)]^{n-1} du$$

$$E(t) = \int_{-\infty}^t z(x) dx.$$

The inner integral, $E(t)$, is the cumulative probability of a standard normal deviate which is solved by the probability function provided in Dunlap and Duffy (1975). Since extreme normal deviates are encountered often, this function is augmented by another approximation for extreme scores provided by Zelen and Severs (1964, p. 933, Function 26.2.24).

The next integral, $P_n(t)$, is solved numerically using Simpson's rule for 32 points in the range -7 to $+4$, which is found to encompass the useful portion of the integral. The outer integral, $1 - F(q)$, is also solved by a 32-point Simpson integration. However, the useful range for this integral depends upon q and ν and is determined by the program depending on the values of these quantities.

The output of the present program agrees to at least four places with the tabled values in Harter et al. (1959) and should suffice for most practical situations. The integral with degrees of freedom (df) of 1 or 2 has unique problems; since 1 or 2 df are seldom encountered in practice, the extra steps necessary to fit these points accurately are not included in the program.

The program provides the probability for a given q , n , and ν . The value of the studentized range statistic, q , is usually defined as:

$$q = (\bar{x}_a - \bar{x}_b) / \sqrt{MS(E) / m},$$

where \bar{x}_a and \bar{x}_b are two sample means, $MS(E)$ is the mean square error, and m is the sample size for each mean. Along with q the user must provide ν , which is the df error, and n , the number of means from which the two means in question were selected. To perform the Tukey HSD test, the user sets n equal to the total number of means involved regardless of the particular pair of means compared. For the Newman-Keuls test, n is set equal to the number of means falling between \bar{x}_a and \bar{x}_b , inclusively. For Duncan's test one follows the Newman-Keuls procedure, but compares the probability obtained to the quantity $[1 - (1 - \alpha)^n]$, rather than to α , to decide significance. A more complete discussion of the range tests can be found in Winer (1971, p. 185-201) or Games (1971).

Time Requirements. Since the program must calculate 1,024 (32×32) points in the numerical integration, the time required to find each probability value is rather long; however, run time is essentially constant regardless of the parameters. On a DECsystem PDP-10 computer the average time for each probability calculation is 1.15 sec.

Program Availability. The program is written as a function subprogram in standard single-precision FORTRAN IV, and runs on a DECsystem PDP-10, an IBM 7044, and an Interdata Model 70 computer. Since the language is standard, the program should run on any

machine that has a FORTRAN IV compiler. Since the computations are done in log form and converted back to arithmetic form only if such conversion is possible without underflowing, a quantity, E, is set equal to the largest value whose negative exponent can be calculated by the user's computer [i.e., $X = \text{EXP}(-E)$] without underflowing.

There are no special peripheral requirements; memory size need only be sufficient to load the function and a main routine to call it. A listing of the function may be obtained at no cost from William P. Dunlap, Department of Psychology, Tulane University, New Orleans, Louisiana 70118.

REFERENCES

- DUNLAP, W. P., & DUFFY, J. A. FORTRAN IV functions for calculating exact probabilities associated with z, chi-square, t, and F values. *Behavior Research Methods & Instrumentation*, 1975, 7, 59-60.
- GAMES, P. A. Multiple comparisons of means. *American Educational Research Journal*, 1971, 8, 531-565.
- HARTER, H. L., CLEMM, D. S., & GUTHRIE, E. H. *The probability integrals of the range and of the studentized range* (WADC Technical Report 58-484, Vol. II). Dayton, Ohio: Wright Air Development Center, 1959.
- JASPEN, N. The calculation of probabilities corresponding to values of z, t, F, and chi-square. *Educational and Psychological Measurement*, 1965, 25, 877-880.
- NEWMAN, D. The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika*, 1939, 31, 20-30.
- NIE, N. H., HULL, C. H., JENKINS, J. G., STEINBRENNER, K., & BENT, D. H. *Statistical package for the social sciences* (2nd ed.). New York: McGraw-Hill, 1975.
- PEARSON, E. S., & HARTLEY, H. O. Tables of the probability integral of the "studentized" range. *Biometrika*, 1943, 33, 89-99.
- VELDMAN, D. J. *FORTRAN programming for the behavioral sciences*. New York: Holt, Rinehart, & Winston, 1967.
- WINER, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.
- ZELEN, M., & SEVERS, N. C. "Probability functions." In M. Abramowitz & I. A. Stegun (Eds.), *Handbook of mathematical functions* (National Bureau of Standards Applied Mathematics Series No. 55). Washington, D. C.: U. S. Government Printing Office, 1964. Pp. 927-990.

(Accepted for publication March 3, 1977.)

APPENDIX

```

FUNCTION RPRB (R, S, DF)
C      R = STUDENTIZED RANGE STATISTIC = (ML-MS)/SQRT(MS(E)/N)
C      WHERE:  ML = LARGEST MEAN IN RANGE TESTED
C              MS = SMALLEST MEAN IN RANGE TESTED
C              MS(E) = MEAN SQUARE ERROR
C              N = SAMPLE SIZE
C      S = NUMBER OF SAMPLES BETWEEN ML AND MS, INCLUSIVELY
C      DF = DEGREES OF FREEDOM FOR MEAN SQUARE ERROR
C
C      DIMENSION ZP(2),COEF(33)
C      SET EMAX SUCH THAT EXP(-EMAX) DOES NOT UNDERFLOW
      EMAX=89.
      DATA COEF/1.,4.,2.,4.,2.,4.,2.,4.,2.,4.,2.,4.,2.,4.,2.,4.,
*2.,4.,2.,4.,2.,4.,2.,4.,2.,4.,2.,4.,2.,4.,2.,4.,1./
C      SOLVE FOR CONSTANT OF INTEGRAL
      IDF=DF
      DG=DF/2.-1.
      IDG=DG
      G=0.
      IF(DG) 1,5,2
1     G=.572364943
      GO TO 5
2     IF (DG.EQ.,5) GO TO 4
      DO 3 I=1,IDG
      XI=I
3     G=G+ALOG(DG-XI+1.)
      IF(AMOD(DF,2.).EQ.0.) GO TO 5
4     G=G-.120782238
5     C=EXP(.693147181+ALOG(.196349541*DF)*DF/2.-G)
C      COMPUTE RANGE OF INTEGRATION FOR OUTER INTEGRAL
      DT=ALOG(R+1.)*100.
      IF(DF.GT.DT) GO TO 6
      DMAX=DF
      GO TO 7
6     DMAX=DT
7     RI=5.2+R*.5-ALOG(DMAX)-1./SQRT(R)
      IF(RI.LT..15) RI=.15
      BI=R-RI
      IF(BI.LT.0.) BI=0.

```

```

UI=R+RI
SI=(UI-BI)/32.
C      STUDENTIZED RANGE INTEGRAL (SIMPSON'S RULE)
SRA=0.
DO 11 I=2,33
XI=I-1
BR=BI+SI*XI
C      SR=(-.918938533-BR*BR/(R*R*2.)+ALOG(4.*BR/R))*DF
RANGE INTEGRAL (SIMPSON'S RULE)
RA=0.
DO 10 J=1,33
XJ=J-1
BZ=XJ*.34375-7.
C      NORMAL DISTRIBUTION INTEGRAL
Z=BR+BZ
DO 9 K=1,2
IF(K.EQ.2) Z=BZ
X=ABS(Z)
Q=.39894228*EXP(-X*X/2.)
IF(X.GT.3.7) GO TO 8
T=1./(1.+2316419*X)
P=.31938153*T
P=P-.356563782*T**2
P=P+1.78147937*T**3
P=P-1.821255978*T**4
P=P+1.330274429*T**5
ZP(K)=Q*P
GO TO 9
8 ZP(K) = Q*(SQRT(4.+X*X)-X)/2.
9 IF(Z.GT.0.) ZP(K)=1.-ZP(K)
RH=0.
D=ZP(1)-ZP(2)
IF(D.LE.0.) GO TO 10
H=ALOG(D)
PR=-.918938533-BZ*BZ/2.+H*(S-1.)+ALOG(S)
IF(PR+EMAX.GT.0.) RH=EXP(PR)
10 RA=RA+COEF(J)*RH
RA=.34375*RA/3.
SRH=0.
D=(1.-RA)/BR
IF(D.LE.0.) GO TO 11
SR=SR+ALOG(D)
IF(SR+EMAX.GT.0.) SRH=EXP(SR)
11 SRA=SRA+COEF(I)*SRH
RPRB=SI*SRA/3.*C
RETURN
END

```