Behavior Research Methods & Instrumentation
1975, Vol. 7 (5), 482-483

An algorithm for calculating probabilities for the F distribution (including the chi square, t, and normal distributions as special cases)

S. SELVIN and J. T. WONG
University of California, Berkeley, California 94720

A number of numerical methods exist for computing probabilities associated with variables that have F distributions. Most of these procedures depend on approximations and yield rather poor accuracy when the degrees of freedom associated with either the numerator (n_1) or denominator (n_2) are small. The approach presented here is to combine an approximation (Paulson, 1942) which is sufficiently accurate for a large number of situations with a computation algorithm based on the incomplete Beta distribution. The incomplete Beta distribution is simply related to the F distribution and can be evaluated exactly by means of a recurrence relationship (Abramowitz & Stegun, 1964). Since the chi-square and t distributions are special cases of the F distribution, probabilities associated with these distributions are easily calculated from slight modifications of the same recurrence formula. A FORTRAN function subroutine is provided to compute these statistical probabilities with a known degree of accuracy. Also included is the option of calculating probabilities for variables with standard normal distributions (mean = 0 and variance = 1.0).

Method. The exact right-hand tail F distribution probabilities are calculated by a recurrence relationship unless the approximate formula yields a relative error rate[1] of less than .5% for probabilities greater than .01.

The chi-square distribution is related to the F distribution ($F = \chi^2/n$ has F distribution with n_1 = degrees of freedom and $n_2 = \infty$) and it is, therefore, easy to calculate chi-square probabilities from a procedure that computes F distribution probabilities. Probabilities associated with the t distribution can be similarly calculated from the F distribution ($F = t^2$ has the F distribution with $n_1 = 1$ and n_2 = degrees of freedom). Normal distribution probabilities are necessary for the computational algorithm and are also included as a separate option. The calculated normal probabilities have absolute error rates never exceeding 10^{-8}.

Algorithm. The FORTRAN function subroutine entitled PCTNF (Table 1) is based on the incomplete Beta function recurrence relationship and a closed form approximation. The computational strategy sacrifices a small and known amount of accuracy for efficiency.

The function subroutine returns the value PCTNF which is the probability area to the right of a specific value (X) for the chi square, the t, the normal, or the F distributions. The input values are:

X: the value of the test statistic

and

chi square: ID = 1, n_1 = degrees of freedom and $n_2 = 0$
t distribution: ID = 2, $n_1 = 1$, and n_2 = degrees of freedom
normal: ID = 3, $n_1 = 0$, and $n_2 = 0$
F distribution: ID = 4, n_1 and n_2 = degrees of freedom.

Notice that this algorithm assumes that the system subroutines for inverse tangent (ATAN) and for the square root (SQRT) are available.

Hardware. The algorithm PCNTF is written in a version of FORTRAN that is compatible with the IBM 1130 computer and most more advanced systems. The subroutine requires 720 words of core storage. This computational algorithm has been tested for execution time and accuracy on both the IBM

Table 1
Subroutine for Calculating Probabilities Associated With the Right-Hand Tail of Four Statistical Distributions (Error Rate ⩽ .5%)

```
      FUNCTION PCTNF(ID,X,N1,N2)
C     ID=1 CHI SQUARED, IC=2 STUDENT T, ID=3 STANDARD NORMAL, ID=4 F
C     X=VALUE OF TEST STATISTIC
C     IF ID=1 THEN SET N1=D.F. AND N2=0
C     IF ID=2 THEN SET N1=1 AND N2=D.F.
C     IF ID=3 THEN SET N1=0 AND N2=0
C     IF ID=4 THEN SET N1=C.F. OF NUMERATOR AND N2=D.F. OF DENOMINATOR

      DIMENSION Q(2),GA(2),GAB(2)
      DATA PI,BON1,BON2,IPAR/3.1415927,10.,15.,69/
      AN1=N1
      AN2=N2
      SPI=SQRT(PI)
      U=X
      GO TO (1,2,10,4),ID
    1 U=SQRT(U)
      FF=X/AN1
      AN2=1.0E10
      GOTO 5
    2 QU = 1.0
      FF = U*U
      IF(FF) 3,28,3
    3 IF (N2-IPAR) 12,12,10
    4 FF=U
    5 PCTNF=1.
      IF (FF) 6,34,6
    6 IF (N1+N2-IPAR) 7,7,9
    7 IF (AN1-BON1) 12,12,8
    8 IF (AN2-BON2) 12,12,9
    9 A1=2./(9.*AN1)
      A2=2./(9.*AN2)
      U=((1.-A2)*FF**.3333333-1.+A1)/(SQRT(FF**.6666667*A2+A1))
   10 T=1./(1.+.2316419*ABS(U))
      ZU=EXP(-U*U/2.)*.3989423
      QU=T*(.31938153-T*(.35656378-T*(1.7814779-T*(1.82125598
     1 -1.33027443*T))))*ZU
      IF (ID-1) 29,11,29
   11 IF (AN1-BON1) 13,13,29
   12 NN=N1
      IF (ID-1) 10,10,14
   13 Q(1)=QU*2.
      UI=X*0.5
      Y=EXP(-UI)
      Q(2)=Y
      MM=1
      IF (NN-2) 26,26,15
   14 MM=2
      F=AN2/(AN2+AN1*FF)
      TEMP=1.-2.*F
      Q(1)=ATAN(TEMP/SQRT(1.-TEMP*TEMP))/PI+0.5
      Q(2)=1.-SQRT(F)
      GB=SPI
      GAB(1)=1.
      GAB(2)=GB
      B=0.5
      Y=1.-F
   15 DO 25 JJ=1,MM
      GA(1)=SPI
      GA(2)=1.
      IF (NN-2) 22,22,16
   16 DO 21 I=3,NN
      J=I-2*((I-1)/2)
      A=(I-2.)*0.5
      GA(J)=A*GA(J)
      IF (ID-1) 17,20,17
   17 GAB(J)=(A+B-1.)*GAB(J)
      IF (GAB(J)) 18,18,19
   18 GAB(J)=1.
   19 Q(J)=Q(J)-GAB(J)/(GA(J)*GB)*(Y**A)*((1.-Y)**B)
      GOTO 21
   20 Q(J)=Q(J)+(UI**A)*Y/GA(J)
   21 CONTINUE
      IF (JJ-MM) 23,27,27
   22 J=NN
   23 B=NN*0.5
      Q(1)=1.-Q(J)
      Q(2)=1.-(Y**B)
      NN=N2
      IF (NN-2) 26,26,24
   24 GB=GA(J)
      GAB(1)=GAB(J)
      GAB(2)=GB
   25 Y=F
      GOTO 27
   26 J=NN
   27 QU=Q(J)
      IF (ID-2) 29,28,29
   28 QU=QU*0.5
   29 IF (U) 30,31,31
   30 QU=1.-QU
   31 IF (QU-1.) 33,33,32
   32 QU=1.
   33 PCTNF=QU
   34 RETURN
      END
```

1130 and the CDC 6600 computers. The time required on the much slower IBM 1130 is .5 sec per call. There seem to be no problems in computational accuracy; and single precision mode was adequate for all cases tested.

Availability. A limited number of program decks can be obtained by writing:

S. Selvin
School of Public Health
University of California
Berkeley, California 94720

There will be no charge for the present supply.

REFERENCE NOTE

1. Abramowitz, M., & Stegun, I. A. (Eds.). *Handbook of Mathetmatical functions.* National Bureau of Standards Applied Mathematical Series 55. U. S. Government Publication, 1964.

REFERENCE

Paulson, E. An approximation of the analysis of variance distribution. *Annals of Mathematical Statistics*, 1942, **10**, 233-235.

NOTE

1. Relative error rate = (approximate value—actual value/actual value) × 100, where these values are the probabilities for the right-hand tail of the F distribution.

Behavior Research Methods & Instrumentation
1975, Vol. 7 (5), 483

VALUES: A program for the analysis of Milton Rokeach's "value surveys"

ERIC TANENBAUM
The University of Durham
Social Science Advisory Service, Computer Services
Durham, England DH1 3LE

and

SHARON SUTHERLAND
Dalhousie University, Halifax, Nova Scotia, Canada

In numerous articles, Milton Rokeach (e.g., 1968; 1969) argues that social psychologists might find the study of individual value structures to be more fruitful than the more familiar examination of attitudes. To this end, he devised his "value survey." Briefly, this consists of two lists of 18 value statements each, one representing instrumental (or "style") modes of behavior and the other pertaining to terminal (or "end") states of existence. Subjects are asked to rank value statements in each list in terms of personal importance.

A growing body of literature reporting research with the value survey attests to their potential (see Rokeach, 1973). However, for several reasons, analyzing data collected with them is awkward with available general-purpose packages. The researcher must manipulate 18 variables simultaneously, where these variables are measured at the ordinal level. Moreover, while individual ranks are interesting, it is important to consider relative ranks; that is, to evaluate the entire pattern of values. Finally, the analyst is often interested in comparing value profiles among several groups. It was to meet these needs that VALUES was written.

Input. VALUES accepts raw data, including an optional group identifier. Provision is made for declaration of missing values.

Output. Total and group medians, modes, and semi-interquartile ranges are calculated per value. Two measures, based on the median and semi-interquartile range, which indicate relative importance and relative consensus (see Sutherland & Tanenbaum, Note 1) are then calculated. Tables output include the above statistics for each set of values organized alphabetically within sets, as well as arranged by importance and consensus rank. A lineprinter scattergram of the latter two measures is also available. Further options include a Kruskal-Wallis approximation to one-way analysis of variance using ranked data and Kolmogorov-Smirnov's test for two samples (Hays, 1965).

Computer and language. VALUES is written in "machine-independent" Fortran IV. It has run successfully on an ICL-1909, a DEC-PDP-10, and an IBM 360/67. Core requirements depend upon the maximum number of observations. One scratch device is required.

Availability. A source listing, user documentation, and a test setup with output is available from Eric Tanenbaum, Social Science Advisory Service, Computer Services, The University, Durham, England DH1 3LE.

REFERENCE NOTE

1. Sutherland, S. L., & Tanenbaum, E. J. Rokeach's value survey in use: Towards validation with criterion attitude scales. *Canadian Review of Sociology and Anthropology*, in press.

REFERENCES

Hays, W. L. *Statistics for psychologists.* Toronto: Holt, Rinehart, and Winston, 1965.

Rokeach, M. A. Theory of organization and change within value systems. *Journal of Social Issues*, 1968, **24**, 13-33.

Rokeach, M. The H. Paul Douglas Lectures for 1969. *Review of Religious Research*, 1969, **11**, 3-38.

Rokeach, M. *The nature of human values.* New York: The Free Press, 1973.

(Received for publication April 14, 1975; accepted April 17, 1975.)