FORTRAN IV functions for calculating exact probabilities associated with z , χ^2 , t , and F values

WILLIAM P. DUNLAP and JOHN A. DUFFY

*Tulane University
New Orleans, Louisiana 70118*

Veldman (1965) provided the most widely used computer function to determine probabilities associated with the basic statistics, χ^2 , t , F , and the normal deviate, z . This function, called PRBF, was adapted by Veldman from an approximate procedure first published by Jaspens (1965). When degrees of freedom (df) are large, this approximate procedure is sufficiently accurate, however, when df are small (for example F ratios on 2 and 2 df), Jaspens (1965) has shown that the function is inaccurate in the third decimal place which can lead to erroneous decisions regarding the null hypothesis at standard levels of significance. The Jaspens procedure transforms t , χ^2 , or F to an approximate normal deviate. The probability of the normal deviate is then approximated by a polynomial expression given in Zelen and Severs (1964). The polynomial chosen is such that normal deviates evaluated by this function lead to inaccuracy in the fourth decimal place. It appeared useful, therefore, to write FORTRAN IV functions that would calculate exact probabilities of z , t , χ^2 , and F with at least the accuracy of commonly available tabled values, and if possible, to the limits of accuracy of the single precision computer.

The first function, ZPRB, uses an extremely accurate polynomial approximation from Zelen and Severs (1964) that calculates the probability of a normal deviate with an absolute error of less than 7.5×10^{-8} ; therefore, this function is accurate to about 7 decimal places, which is compatible with the accuracy of the standard single precision floating point word used by the IBM 360. The polynomial approximation was chosen over several infinite series approximations (also in Zelen & Severs, 1964), since empirical tests showed the former to have greater accuracy in single precision as well as greater speed. ZPRB computes the integral of the normal distribution from z to infinity; therefore, the probability generated can be used directly to access the one-tailed significance of a positive z value. For a two-tailed significance test the probability should be doubled. To evaluate the significance of negative z scores, subtract the probability generated from one.

The second function, CPRB, calculates the probability associated with a χ^2 value, given its degrees of freedom. Thus CPRB computes the integral of the chi square distribution from χ^2 to infinity. The computational procedures for the exact integration of the χ^2 distribution were published by Fisher (1935) and also appear in Zelen and Severs (1964). The particular function chosen was the one which showed the greatest accuracy and speed in empirical tests. When the df are odd, function CPRB requires the calculation of a cumulative normal deviate, so that the function ZPRB is called by function CPRB, thereby limiting its accuracy to 7 digits. Other than this limitation, the accuracy of the probability calculated by CPRB is limited only by

Preparation and testing of these programs was supported by the Tulane University Computer Center.

```

C
C      FUNCTION ZPRB(Z)
C      COMPUTES THE CUMULATIVE PROBABILITY ASSOCIATED WITH NORMAL DEViate Z
C      (THE INTEGRAL OF THE NORMAL DISTRIBUTION FROM Z TO INFINITY)
C
      X=Z
      IF (Z.LT.0.0) X=1-Z
      ZPRB=1.0
      IF (X.GT.5.612) GOTO 1
      ZX=.39894228*EXP(-X*X/2.0)
      T=1.0/(1.0+.2316419*X)
      B1=.319381530*T
      B2=(-.356563782)*T**2
      B3=1.781477937*T**3
      B4=(-1.821255978)*T**4
      B5=1.330274429*T**5
      ZPRB=1.0-ZX*(B1+B2+B3+B4+B5)
      IF (Z.GT.0.0) ZPRB=1.0-ZPRB
      RETURN
      END

```

the computational accuracy of the particular computer used.

The third function, FPRB, evaluates the probability of an F value, given the numerator and denominator degrees of freedom. FPRB computes the integral of the F distribution from F to infinity. Again, the computational procedures for this integration were published by Fisher (1935) and appear in Zelen and Severs (1964). Since the procedure is a direct solution rather than an approximation the accuracy is determined by the computational accuracy of the computer used. Since the square of a t -value on C degrees of freedom is equal to an F value on 1 and C degrees of freedom, two-tailed probabilities of t -values are found by referring to function FPRB as follows:

$$P = 2 \times \text{FPRB}(T^2, 1, C)$$

One-tailed t probabilities are one half of this computed value.

Time requirements. Since ZPRB uses a highly accurate polynomial approximation the computational time is not determined by the value of z . Since its accuracy is no greater than 7 digits, the probabilities of z -scores greater than 5.612 are returned as zero because the actual probability is less than 10^{-8} . The computation of both χ^2 and F values is done by solving series equations, the number of terms of which depend upon the number of df. As df become larger the number of computations required increases arithmetically. Since this is true it is difficult to consider the time required to compute the probability associated with an "average" F ratio; however, when FPRB was used to calculate the probabilities of all F ratios having noninfinite df appearing in the 5% F table in Fisher and Yates (1963, p. 53), approximately 3 sec of computing time was required per 100 F ratios.

Program Availability. Although developed on an IBM 7044 these functions should operate with any computer programmable in standard FORTRAN IV. The accuracy of the calculations is, of course, limited by computer word length. There are no special peripheral requirements and memory size need only be sufficient to load the function desired and a main routine to call it. A listing of these functions may be obtained at no cost from William P. Dunlap, Department of Psychology, Tulane University, New Orleans, Louisiana, 70118.

REFERENCES

- Fisher, R. A. The mathematical distributions used in the common tests of significance, *Econometrica*, 1935, 3, 353-365. Reprinted in: Tukey, J. (Ed.) Fisher, R. A. *Contributions to mathematical statistics*, New York: Wiley, 1950.
- Fisher, R. A. & Yates, F. *Statistical tables for biological, agricultural and medical research*, (6th ed.) New York: Hafner, 1963.

Jaspen, N. The calculation of probabilities corresponding to values of z , t , F , and chi-square. *Educational and Psychological Measurement*, 1965, 25, 877-880.

Veldman, D. J. *Fortran programming for the behavioral sciences*.

New York: Hold, Rinehart, & Winston, 1967.

Zelen, M. & Severs, N. C. "Probability functions" In: *Handbook of mathematical functions* (National Bureau of Standards Applied Mathematics Series No. 55). Washington: U.S. Government Printing Office, 1964, 927-990.