UNIVERSITY OF ILLINOIS

DIGITAL COMPUTER

LIBRARY ROUTINE K 6 - 185

TITLE	Chi-Squared
TYPE	Closed with two program parameters
NUMBER OF WORDS	23
TEMPORARY STORAGE	0, 1, and 2
DURATION	(2.4m + 3) milliseconds
ACCURACY	$\pm 2^{-40}$
ENTRY	When this routine is located at q, entry is made by

p	-- mF
	50 pF
p + 1	26 qF
	00 nF
p + 2	any

At the end of this routine control is transferred to the left hand order at location p + 2.

PRESET PARAMETERS

When this routine is read in the following preset parameters must have been stored in locations 3, 4 and 5, respectively.

S3	00 F 00 sF	s is a scaling factor and is normally chosen to be 1, 10, 100, or 1000 depending on whether none, one, two, or three decimal places of accuracy are desired after the decimal point.
S4	00 F 00 aF	a is the location at which the quantities p_i, i = 0, 1, ..., m - 1, are stored prior to entering this routine.
S5	00 F 00 bF	b is the location at which the quantities f_i, i = 0, 1, ..., m - 1, are stored prior to entering this routine.

DESCRIPTION

In its most frequent application chi-squared is given by the formula

$$\chi^2 = \sum_{i=0}^{m-1} \frac{(E_i - \theta_i)^2}{E_i} \qquad 1)$$

where each E_i is the expected number of members in the i^{th} of m classes for a given sample size and the θ_i are the observed values. If we let

$$p_i = \frac{E_i}{n} \;; \quad f_i = \frac{\theta_i}{n} \qquad 2a, b)$$

and multiply both sides of equation 1) by a number s, the resulting equation is

$$s\chi^2 = sn \sum_{i=0}^{m-1} \frac{(p_i - f_i)^2}{p_i} \;. \qquad 3)$$

This last equation corresponds to the quantity computed by this routine. $s\chi^2$ is computed as an integer and is placed in the A register at the end of the routine. The quantities p_i and f_i are fractions and must be in the ranges

$$0 < p_i < 1$$
$$0 \leq f_i < 1 \qquad 4a, b)$$

A value of 1 for one of the f_i may be represented in the machine as -1. Each value p_i is stored at location a + i before this routine is entered, the first address a being given by preset parameter S4. Similarly the f_i are stored at b + i, b being specified by preset parameter S5. The number of values m of the p_i or f_i is specified by a program parameter (See Entry).

The number s is a positive integer specified by preset parameter S3 and serves as a scaling factor for χ^2. Normally s will be chosen to be 1, 10, 100 or 1000 depending upon the number of decimal places of accuracy required in the value of χ^2. For example if χ^2 is 2.531.... and s is 100, the number in the A register at the end of this routine will be 253×2^{-39}.*

The program parameter n is also a positive integer and is chosen so as to put the p_i and f_i in the required range given by 4a, b) as determined by equations 2a, b). A logical choice for n is the sample size. The p_i are then the predicted probabilities and the f_i are the corresponding observed values The requirements 4a, b) will then be satisfied automatically. If the values E_i and O_i are known directly it may be more convenient to chose n to be the smallest power of 10 which is greater than or equal to the sample size. In some applications a power of 2 may be more convenient. In any case the values p_i and f_i are determined by equations 2a, b), and n must be specified upon entering the routine (See Entry).

If it is much more convenient to produce the values p_i and f_i one at a time, it is possible to enter this routine m times with the new values of p_i and f_i in locations a and b, respectively, each time. If this procedure is used the program parameter m must always be 1, and the necessary summation must be carried on outside this routine. Other things being equal, this method will be slower and less accurate.

In addition to 4a, b) there are certain other moderate requirements on the quantities involved. The first of these is that the following inequality must hold:

$$\frac{\chi^2}{n} = \sum_{i=0}^{m-1} \frac{(p_i - f_i)^2}{p_i} < 256 = 2^8 \qquad 5)$$

---

* If Library Routine P 1 is used the results may be printed in such a way that the position of the decimal point is specified.

Since each term in this summation is non-negative, each term must also be less than 256. If the latter does not hold a division hang-up will occur. If the summation is too large the answer will be in error by a negative integral multiple of 512s. Any danger that the requirement 5) might be violated can usually be avoided by using a larger value of n and making the appropriate changes in the p_i and f_i as determined by 2a, b). Secondly, in order to obtain the stated accuracy of $\pm 2^{-40}$ the product mns should be small compared to $2^{30} \approx 10^9$. Although this generally means that values of χ^2 accurate to a large number of decimal places are obtainable, printing out results to greater accuracy than actually required or justified by the data should be assiduously avoided.

Rt: 7/22/59
RETYPED
DATE 6/2/55: 5/31/56
PROGRAMMED BY C.Farrington
APPROVED BY JPNash

CF/mge

K 6

LOCATION	ORDER		NOTES PAGE 1
0	41 F		
	K5 F		
1	42 5L		Plant address of n
	46 F		Store m
2	L5 19L		Set addresses
	40 8L		
3	46 11L		
	L4 F		Set test constant
4	46 20L		
	F5 5L		Plant link
5	42 18L		
	50 F		Extract and store n
6	00 1043F		
	01 20F		
7	40 F		
	41 1F		Clear Σ
8	L5 S4	from 15	$p_i - f_i$
	L0 S5		
9	40 2F		$\frac{(p_i - f_i)^2}{p_i} \times 2^{-8}$
	50 2F		
10	75 2F		
	10 8F		
11	66 S4		
	S5 F		Summation
12	L4 1F		
	40 1F		
13	L5 8L		Step addresses
	F4 21L		
14	40 8L		
	46 11L		
15	L0 20L		Test for end
	36 8L		
16	50 F		ns $\times 2^{-39}$ in Q
	75 22L		

LOCATION	ORDER		NOTES PAGE 2
17	L5 6L		sχ^2 rounded off
	74 1F		
18	00 8F		
	26 F		Exit
19	L5 S4		
	L0 S5		
20	75 F		
	L0 F		
21	00 1F		
	00 F		
22	00 F		
	00 S3		

K 6