KM 2(1981)2
pag 66-86

FISHER,

een programma voor de analyse van rxc-tabellen bij kleine steekproeven

Pieter M. Kroonenberg, vakgroep WEP, Schuttersveld 9, Rijksuniversiteit Leiden
& Albert Verbeek, Inst v Math Stat en Sociol Inst, Rijksuniversiteit Utrecht

Samenvatting

Dit artikel komt vrijwel overeen met de voorlopige versie van de gebruikershandleiding van FISHER, een programma voor de analyse van twee-dimensionale tabellen, vooral bij kleine steekproeven. Het voornaamste verschil is dat de inputbeschrijving hier is weggelaten. FISHER is vooral geschreven voor het efficient berekenen van "exacte" (in tegenstelling tot asymptotische) overschrijdingskansen voor zes toetsen op onafhankelijkheid of homogeniteit. De zes toetsgrootheden zijn: Pearsons X^2, Kruskal-Wallis, de correlatieratio, Kendalls tau_b, Spearmans rangcorrelatie, en de gewone correlatiecoefficient.

In de laatste twee paragrafen wordt een voorbeeld van output gegeven en een vergelijking gemaakt met andere programma's. Het is vanuit sociologische optiek leerzaam om te zien, hoe rommelige en ongedefinieerde output van SPSS voor zoiets alledaags als 2x2-tabellen tien jaar lang gebruikt kan worden, zonder dat dit tot opstandjes of gemor onder de gebruikers aanleiding geeft. (Tussen haakjes, onze beste vrienden gebruiken SPSS, en wijzelf ook, en al jarenlang ...)

Beschikbaarheid van het programma

Een beperkte testversie met alleen Pearsons X^2 is momenteel beschikbaar op de academische rekencentra in Amsterdam, Groningen, Leiden en Utrecht. Een versie met de zes bovengenoemde toetsen komt binnenkort gereed. Deze wordt tegen verzendkosten geleverd door het Centrum voor Data-Analyse, Varkenmarkt 2, Utrecht, tel 030 323[illegible].

CONTENTS

INTRODUCTION

FISHER is a program to compute the distributions and/or descriptive levels of significance (p-values) of statistics used for testing independence or homogeneity in an r×c contingency table with fixed marginal totals.

The 'exact' distributions of the test statistics themselves are discrete, and difficult to treat analytically. So usually their continuous limit distributions are employed to obtain an approximation to the descriptive levels of significance. Until the advent of very fast computers and efficient algorithms the computational task to obtain the exact distributions of the test statistics themselves was prohibitive. In addition, no convenient way is available to tabulate them due to the large number of parameters, and the erratic behaviour of the descriptive levels of significance caused by the discreteness of the distributions. The asymptotic distributions have the advantage of easy tabulation and fast computation, but in cases where the number of observations (or more precisely the expectations) are small asymptotic methods are not accurate enough. However, not too much is known about what is meant by 'small'.

Cochran (1952, 1954) gave the following widely applied rule of thumb for the use of Pearson's X^2 in tables with more than one degree of freedom: if all expected frequencies are larger than 1, and at least 80% of them are larger than 5, the χ^2 critical region of 5 percent (1 percent) will really be at least 3 and at most 7 percent (.5 and 1.5 percent). For other statistics such general rules of thumb are not available.

For more than one degree of freedom exact calculation can be very much more expensive than the calculation of the asymptotic approximation; its cost still may be prohibitive. Moreover the asymptotic approximations prove adequate in many cases with remarkably small samples. Therefore one should not turn to exact calculation automatically. Through the program the necessity and the cost of exact calculations can be judged in advance, or interactively.

The exact tests used here are permutation tests and are a direct generalization of Fisher's exact test for the 2×2 table (which is also efficiently handled by this program). Put another way: tests are conditional on the observed marginals with the hypergeometric distribution as null hypothesis.

The alternative hypothesis should determine which test statistic is most appropriate. In addition to the usually employed X^2 for the most general alternative hypothesis the following exact tests are available for more specialized alternatives:

Kruskal-Wallis' one-way analysis of variance, the correlation ratio, Kendall's τ_b, Spearman's rank correlation coefficient, and Pearson's ordinary correlation coefficient.

A number of algorithms and some programs to perform the calculations necessary for computing the exact significance levels have been published: Boulton & Wallace (1973), Baker(1977), Agresti & Wackerly (1977), Klotz & Teng (1977), and Hancock (1975), IMSL Version 8, subroutine CTPR (1980).

The basic algorithm of the present program evolved form Agresti & Wackerly's implementation of Boulton & Wallace's algorithm. It has, however, a number of features which distinguishes it from its predecessors :

1. it is more efficient and faster than any of the previous programs;
2. it can be used for up to 12 x 12 tables with marginal totals up to 200 - be it that in combination one might compute forever -;
3. it can compute exact and approximate significance levels for six different statistics;
4. it also prints tables of expected values and residuals;
5. it can be used interactively;
6. it contains complete input checking, and performs also a number of internal consistency checks;
7. it is written in transportable ANSI-FORTRAN IV, chekced by PFORT (Ryder, (1974));
8. it is written in a structured way, so as to ensure efficient compiling, easy modification or addition of statistics, and source legibility.

By giving descriptions at increasing levels of detail, we hope to ease the problem of checking correctness for others as well as for ourselves.

A. User's Guide
   global description, definition of statistics, outline of statistical theory, input and output specifications, examples of input and output.
B. Programmer's guide
   detailed description, including rewritten definitions and calculations in a form that is better suited for computation, flowcharts.
C. Program Listing
   comments in source text
D. FORTRAN source text itself

In this way it is possible to check one level by comparing it with the next coarser level which involves only a small step, and which makes it more easy to check correctness.

GLOBAL DESCRIPTION

The main problem for computing the exact significance level of the test statistics is that, given the marginal totals of the observed contingency table, one needs to generate all possible tables with the same marginals in order to obtain the distributions of the test statistics. The table of observed frequencies from which we derive the fixed marginals will be called 'the observed table'. It must be clearly distinguished from the tables composing the set of all possible tables with the same marginals, which we will call the 'homo-marginal family'. Each element of this family will be called a 'member' or 'possible table'.

In principle we must do the following to obtain the exact level of significance p of the statistic S:

- set p=0; compute the observed value of S: obsS
- generate all members of the homo-marginal family, and for each member:
    - compute the probability, pcurrent;
    - compute for the value of the test statistic S, Scurrent;
    - compare the value of the statistic with that of the observed table and accumulate the probabilities of the more extreme members:

      IF Scurrent ≥ obsS THEN p = p + pcurrent
- print p.

From an algorithmic point of view the interesting part of the program is the subroutine generating all members of the homo-marginal family. This can be done in some 200 lines (as has also been done in IMSL). An elegant recursive algorithm is given by Boulton & Wallace (1973), but we implemented a faster method based on the simulation of a dynamic number -viz (r-1)x(c-1)- of nested DO loops. The largest part of the program (in total over 4000 lines) deals with I/O. The computer time needed to compute exact significances is roughly proportional to the size of the homo-marginal family. There is no simple, explicit formula for this family size, but several approximations are available. FISHER computes the approximatimation ingeniously derived by Gail & Mantel (1977). It usually seems to be off by far less than a factor two, which is sufficiently precise for our purpose.

NOTATION and TERMINOLOGY

The random variable that we observe is

$T = (T_{ij}),\ i = 1,\dots,r;\ j = 1,\dots,c,$

where T_{ij} is the frequency or count in the i-j-cell of an rxc classification. The observed values will be denoted by t and t_{ij}. Summation over the first or second index will be indicated the by a "+"; the observed marginal totals are $t_{1+}, t_{2+}, \dots, t_{r+}$ and $t_{+1}, t_{+2}, \dots, t_{+c}$, while N ($=t_{++}$) is the total number of observations. As in the present program we only treat rxc tables with fixed marginal totals, and $r \geq 2$ and $c \geq 2$, the degrees of freedom of such tables will be $(r-1)\times(c-1)$ if no marginal is zero. Finally the variable giving rise to the row classification or <u>'row variable'</u> will be denoted by X, and the <u>'column variable'</u> by Y. If, say X, has been measured on an ordinal scale, the <u>'values'</u> of X for each row will be the midranks of the observations of that row. The <u>midrank</u> RX_i is the average of the ranks of the observations in that row, i.e.

table	row marginal	row ranks	midranks of X
$t_{11}\ t_{12}\dots t_{1c}$	t_{1+}	$1,2,\dots,t_{1+}$	$RX_1 = (1+t_{1+})/2$
$t_{21}\ t_{22}\dots t_{2c}$	t_{2+}	$t_{1+}+1,\dots,t_{1+}+t_{2+}$	$RX_2 = t_{1+}+(1+t_{2+})/2$
...	...	...	...

For example

table			row marginal	row ranks	midranks
2	3	1	6	1,2,3,4,5,6	$RX_1 = 3\frac{1}{2}$
1	4	0	5	7,8,9,10,11	$RX_2 = 9$
3	7	1	11		$RY_1 = 2,\ RY_2 = 7,\ RY_3 = 11$

Recall that Kendall's tau only depends on the order of rows and columns. The midranks are used in Kruskal-Wallis and Spearman, and are computed by the program. The values of X and Y in the interval variable cases (correlation ratio and correlation coefficient) of course have to be supplied by the user (default values 1,2,3,... may be obtained by entering all zeros or blanks). These values are denoted by X_i for the i^{th} row and Y_j for the j^{th} column. In order to avoid confusion, the values must be strictly increasing:

$X_1 < X_2 < \dots < X_r$, and

$Y_1 < Y_2 < \dots < Y_c$.

Note that Kruskal-Wallis coincides with the correlation ratio if $X_i = RX_i$, and that Spearman coincides with the correlation coefficient if $X_i = RX_i$ and $Y_j = RY_j$ (cf. def). So generally it seems useful to compare Kruskal-Wallis with the correlation ratio and to compare Kendall's tau, Spearman's rank correlation and the ordinary correlation coefficient.

DISTRIBUTION of the POSSIBLE TABLES

FISHER handles contingency tables with fixed marginal totals. This limits its use to those problems for which the marginal totals are fixed by design or for which conditioning on the marginal totals is appropriate. This is however not at all a serious limitation (see e.g. Kempthorne (1979), Verbeek & Kroonenberg (1980), and their references).

In accordance with the fixed marginal totals we assume that the distribution of T (= the frequencies of the observed table) under the null hypothesis is the (generalized) hypergeometric one

$$\Pr(T=t) = \frac{\prod_{i=1}^{c} t_{+j}! \prod_{i=1}^{r} t_{i+}!}{N! \prod_{i=1}^{r} \prod_{j=1}^{c} t_{ij}!} = (\text{constant}) \times \frac{1}{\prod_{i,j} t_{ij}!} .$$

For details and derivations see, for instance, Mood, Graybill & Boes (1974, section 5.4).

TEST STATISTICS

The exact significance levels will be computed for the following test statistics, which are expressed here by the definition formulas rather than by the computational ones. We will refer the reader for details to two easily accesible textbooks rather than to the original references. In brackets [ ], we give the mnemonic used in the output. (Compare Dixon & Brown (1977) p 778, and Norušis (1979) p 12).

· Pearson's X^2 [X2] - generally called the 'chi-square' statistic

$$X^2 = \sum_{\text{all cells}} \frac{(\text{observed-expected})^2}{\text{expected}} = \sum_{i,j} \frac{(T_{ij} - t_{i+} t_{+j}/N)^2}{t_{i+} t_{+j}/N}$$

References: Hays (1973, sections 17.4-17.7, 17.9, 17.10);
Siegel (1956, pp. 104-111, 175-179).

· Kruskal-Wallis' one way analysis of variance K [K-W]
(grouped by the column variable Y)

$$K = \frac{\text{between sum of squares for midranks}}{\text{total sum of squares for midranks}} = \frac{\sum_j \{\sum_j T_{ij} (RX_i - \overline{RX}_j)^2\}}{\sum_i t_{i+} (RX_i - \overline{RX})^2}$$

where the RX_i are midranks, $\overline{RX}_j$ = average midrank of the j^{th} column = $\sum_j T_{ij} X_i / t_{+j}$, $\overline{RX}$ = overall average midrank = $\sum_i t_{i+} RX_i / N$.
The form given here is such that it is comparable to η^2: if $X_i = RX_i$ the two are equal.

References: Hays (1973, section 18.9);
Siegel (1956, pp. 184-193).

• Correlation ratio η^2 [ETA2], or one-way analysis of variance (grouped by the column variable Y)

$$\eta^2 = \frac{\text{between sum of squares for scores}}{\text{total sum of squares for scores}} = \frac{\sum_i\{\sum_j T_{ij}(X_i-\overline{X}_j)^2\}}{\sum_i t_{i+}(X_i-\overline{X})^2},$$

where X_i = value (or score) for the ith category of X, and $\overline{X}_j$ = average value of X in the j^{th} column = $\Sigma_i T_{ij}X_i/t_{+j}$, $\overline{X}$ = overall average value of X = $\Sigma_i t_{i+}X_i/N$.

Reference: Hays (1973, section 16.6).

• Kendall's τ_b [TAU B]

$$\tau_b = \frac{\text{number of concordant pairs of obs} - \text{number of discordant pairs of obs}}{\sqrt{\text{number of pairs, not tied on X}}\ \sqrt{\text{number of pairs, not tied on Y}}}$$

$$= \frac{\sum_{i',j'}\left(\sum_{i<i',j<j'} T_{ij}\right)T_{i'j'} - \sum_{i',j'}\left(\sum_{i<i',j>j'} T_{ij}\right)T_{i'j'}}{\sqrt{\tfrac{1}{2}N(N-1)-\Sigma\,\tfrac{1}{2}t_{i+}(t_{i+}-1)}\ \sqrt{\tfrac{1}{2}N(N-1)-\Sigma\,\tfrac{1}{2}t_{+j}(t_{+j}-1)}}$$

Here a 'pair (of obs)' is one of the $\frac{1}{2}N(N-1)$ unordered pairs of observations $\{(X,Y),(X',Y')\}$.

References: Hays (1973, sections 18.13-18.15)
Siegel (1956, pp. 213-223)
Kendall (1948, 1975, section 3.4)

• Spearman's rank correlation coefficient r_s [RS]

$$r_s = \frac{\text{covariance }(X,Y)}{\text{st.dev.}(X)\ \text{st.dev.}(Y)} = \frac{\Sigma_{ij}\, T_{ij}(RX_i-\overline{RX})(RY_j-\overline{RY})}{\sqrt{\sum_i t_{i+}(RX_i-\overline{RX})^2}\ \sqrt{\sum_j t_{+j}(RY_j-\overline{RY})^2}}$$

where the values of X and Y are midranks, as with Kruskal-Wallis.

References: Hays (1973, sections 18.12, 18.14)
Siegel (1956, pp. 202-213)

• Pearson's correlation coefficient r [R]

$$r = \frac{\text{covariance }(X,Y)}{\text{st.dev.}(X)\ \text{st.dev.}(Y)} = \frac{\sum_{i,j} T_{ij}(X_i-\overline{X})(Y_j-\overline{Y})}{\sqrt{\sum_i t_{i+}(X_i-\overline{X})^2}\ \sqrt{\sum_j t_{+j}(Y_j-\overline{Y})^2}}$$

where the values of X and Y are the values (or scores) of the row and column variable respectively. If $X_i = RX_i$ and $Y_j = RY_j$ then $r = r_s$.

References: Hays (1973, chapter 15)
Kendall & Stuart (1973, vol. II section 31.19 The permutation distribution of r)

The choice of a test statistic depends on the type of alternative hypothesis one has in mind. Display 1 gives an overview of the relation between the statistics available in the program and the corresponding alternative hypothesis. One can, of course, always use a statistic with a more general alternative hypothesis, but this will result in a loss of power.

DISPLAY 1

	Alternative hypothesis	Statistic	Mnemonic	Comment
1.	general	X^2	X2	
2.	row variable ordered; columns are groups; one-way anova	K	K-W	if values of row variable are (mid)ranks; e.g. the row variable is ordinal.
		η^2	ETA2	if values of row variable are scores
3.	both variables are ordered	τ_b	TAU B	if values of both variables are (mid)ranks; e.g. both variables are ordinal
		r_S	RS	(as for TAU B)
		r	R	if values of both variables are scores and the alternative assumes linear dependency

ASYMPTOTIC APPROXIMATIONS

If the expected values are not too small, then asymptotic limit distributions can be used to approximate the exact distributions of the statistics with sufficient accuracy. How large 'small' is, is still a matter of investigation. Some papers with results for X^2 are Agresti & Wackerly (1977), Larntz (1978), and Kroonenberg & Verbeek (1980). Klotz & Teng (1977) present some results for Kruskal & Wallis' K.

In Display 2 we present the asymptotic limit distributions used in FISHER. Some of the statistics have to be transformed before the approximating distribution can be used.

DISPLAY 2

Statistic	Mnemonic	Transformation	Limit distribution	Degrees of freedom	References
X^2	X2		χ^2	$(r-1)(c-1)$	Cf. Mood et al (1974)
K	K-W	$(N-1)K$	χ^2	c-1	Lehmann(1975)pag 396 ex 31
η^2	ETA2	$(N-1)\eta^2$	χ^2	c-1	Lehmann(1975)pag 396 ex 31
τ_b	TAU B	$\tau_b/\text{st.dev.}(\tau_b)$	normal	-	Kendall(1975) sections 5.6 & 4.9
r_s	RS	$r_s\sqrt{N-2}/(1-r_s^2)$	student's t i.e. standard normal, as $N\to\infty$	N-2	Kendall&Stuart(1973) section 31.19 The permutation distribution of r
r	R	$r\sqrt{N-2}/(1-r^2)$	student's t i.e. standard normal, as $N\to\infty$	N-2	Kendall&Stuart(1973) section 31.19 The permutation distribution of r

ONE and TWO-SIDED TESTS

Tests of independence based on τ_b, r_S, or r , and also tests in a 2x2-table can be against a two-sided alternative or against a one-sided alternative. Let us call the test statistic S.

For a one-sided alternative there is little disagreement about the definition of significance (but cf. the last paragraph in the section on the probability jump).

For a right-sided alternative significance = $Pr(S \geq obsS)$, and

for a left-sided alternative significance = $Pr(S \leq obsS)$.

Note that these probabilities add up to

$$1 + Pr(S = obsS).$$

As explained in the section on the probability jump, Pr(S = obsS) is hard to determine for r. In that case the significances produced by FISHER add up to

$$1 + JUMP.$$

For two-sided alternatives there are various methods to distribute the critical region over both tails. For example:

- A. Equally sized tails: take a left tail critical region of size $\alpha/2$ (under H_o) and take a right tail of the same size. However, due to the discreteness this usually can only be done approximately. Then one creates a (discrete) nested family of critical regions, and the significance of an observation is defined as the size of the smallest critical region containing the observation. Due to the discreteness this significance can be substantially less than twice the one-sided significance (cf the Example below).
- B. Take S^2 as test statistic: make critical regions of the form "$S \leq -c$ or $S \geq c$". A problem with this approach is that e.g. X^2 and -2 ln LR are equivalent as one-sided test statistics in a 2x2-table, but not as two-sided test statistics: they may interlace both tails differently, as in the Example below.
- C. Choose the critical values c_L and c_R in such a way that the critical region "$S \leq c_L$ or $S \geq c_R$" is unbiassed or at least locally unbiassed. For discrete distributions this usually requires randomization, but it is rather obvious how to define "nearest" non-randomized tests.

We have (not yet) given serious thought about the implementation of two-sided significances for the statistics mentioned above. As yet the program simply only produces the left-sided and the right-sided significances.

Example. Suppose we observe the following 2x2-table.

A	5-A	5
5-A	16+A	21
5	21	26

A	-2 ln LR	X^2	Pr(A=a)
0	2.4	1.5	.309
1	.0	.0	.455
2	1.5	1.7	.202
3	5.5	6.6	.032
4	12.4	14.7	.0016
5	25.5	26.0	.000015

This table has a distribution virtually without left tail. So left-sided or "really" two-sided testing is impossible. Also note that -2 ln LR and X^2 order A=0 and A=2 differently; according to -2 ln LR the significance of A=2 is .545, according to X^2

the significance is .236 .

RESIDUALS

In most cases significance testing should not be the end-point of an analysis, but only a 'matter of hygiene', i.e. performed without much ado.*) If dependency is found then one should look at residuals. These residuals are defined as

residual = observed - expected.

Examining the residuals is essential for building or refining any model. In FISHER two types of residuals have been included:

- standardized residual $= \dfrac{\text{residual}}{\sqrt{\text{expected}}} = \dfrac{t_{ij}-t_{i+}t_{+j}/N}{\sqrt{t_{i+}t_{+j}/N}}$,

- adjusted residual $= \dfrac{\text{residual}}{\text{estimated st.dev.}} = \dfrac{t_{ij}-t_{i+}t_{+j}/N}{\sqrt{t_{i+}t_{+j}\,(1-t_{i+}/N)(1-t_{+j}/N)/N}}$

In the output of the program the square of the standardized residual is printed, as this is the contribution of a cell to Pearson's X^2.

Note that the standardized residual of a cell will always be smaller than the adjusted residual. The difference will be negligible when t_{+j}/N and t_{i+}/N are small. The adjusted residuals have the advantage that they have asymptotically standard normal distributions.

References: Everitt (1977, section 3.4.3) and Haberman(1973).

PROBABILTY JUMP

A peculiar problem with real valued, discrete statistics, like X^2, is the following. Two different members of a homo-marginal family (= the collection of all tables with the same marginals) may have the same value of X^2, e.g. the tables are equal up to a permutation of rows or columns. However, due to a different order of the calculation,rounding effects may lead to two slightly different values. Therefore, when two members with nearly the same X^2 values occur, it is almost impossible to say which of the two values 'really' is larger. This makes it very hard to determine the significance $Pr(X^2 \geq X^2\text{observed})$ and the size of the 'jump' $Pr(X^2 = X^2\text{observed})$ in the exact, discrete distribution function at X^2observed.

*)This refreshing comparison is borrowed from Ehrenberg (1975) page 323.

In FISHER we determine

$$\text{significance} = \Pr(X^2 \geq X^2\text{(observed)} - .0005)^*, \text{ and}$$

$$\text{jump} = \Pr(X^2\text{observed} - .0005 \leq X^2 < X^2\text{observed} + .0005)^*.$$

The implicit assumption is that two X^2 values are likely to be equal if their absolute difference is smaller than .0005. If in reality the two values are different, then the significance indicated is too large (i.e. too insignificant), but at most the size of the jump. If a borderline case with a large jump occurs - and an extensive analysis is justified - one should (and can) print the entire distribution to check the members with an X^2 close to X^2observed. However, discriminating between 'X^2observed' and 'X^2observed - .0005' almost invariably presupposes a precision not available in the data.

The cited peculiarity affects X^2, Kruskal-Wallis K, η^2, and r, but not τ_b, r_S, and tests in the 2x2-table, as these can be done by integer arithmetic (see the Programmers Guide for details).

Two other possible approaches to the 'jump problem' are:

- A. Ignore the problem altogether.
- B. Don't mention it to the user, but do subtract some small $\varepsilon > 0$ from the observed value(s). This guarantees that the significance reported is not too small (i.e. not exaggerated).

Admittedly these approaches are more user friendly (in the sense that they arouse less discomfort), and often they are quite sufficient. The main problems are: one cannot tell whether they are sufficiently precise, and the computed significances are more machine dependent than just the last few bits. However, different results produced by different machines do not differ more than the jump (and actually less, but it is hard to give an efficient algorithm for a better upperbound).

On the whole we feel, that if one is careful enough to compute the exact distribution rather than its asymptotic approximation, then one should also pay tribute to the inherent discreteness of the exact distribution by glancing at the jump. However, we would be pleased to hear different opinions or other suggestions.

Another view relevant to the jump (or rather to Pr(S = obsS)) is the following. Some have argued that it is more reasonable to define significance as $\Pr(S > \text{obs}S) + c\cdot\Pr(S = \text{obs}S)$, with $c=\frac{1}{2}$, than with c=1 (as usually done). After all, refining the order induced by S, by randomly ordering ("breaking") S's ties, approximately and on average gives this decrease in significance "for free".

---

*Here X^2 and X^2 (observed) ± .0005 should be interpreted as their <u>rounded</u> machine representations. The FORTRAN text contains expressions like: IF(X2.GE.(X2OBS-.0005))...

DESIGN GOALS of the PROGRAM

FISHER is written in transportable ANSI-FORTRAN with special care for the following aspects:

- effiency and speed;
- legibility; it is relatively easy to modify for special purposes (e.g. high speed theoretical simulations; relaxing restrictions; including other statistics)
- user friendliness of input
  - it has both a batch and on interactive mode
  - it is largely self-explaining, it can produce both information about the program and an input summary either upon request or automatically in case of input errors.
  - it checks all input parameters
  - the input format has been designed to allow extensive checking of the input data
  - cooperative error handling
- user friendliness of output
  - care has been taken to include the information necessary to analyse two-dimensional tables: two kinds of residuals, expected values, observed values of the test statistics, their approximate significances, and the exact ones, the hypergeometric probability of the observed table, and the jump;
  - various parts of the output can be written in condensed or extended form to user specified files for further inspection and analysis;
  - in default mode the program does not embark on - possibly time consuming and unnecessary - exact calculations, but it does provide both the asymptotic approximations to the significance(s), and information about the necessity and the cost of exact calculations; viz. the table of expected values and an estimate (or exact calculation) of the number of tables with the same marginals.

OUTPUT EXAMPLES

The next pages contain some annotated examples of output from FISHER. All alphabetic input is in lower case; moreover all input is encircled. At this point, the User Guide has discussed all input parameters for FISHER. In this paper we skip this discussion, hoping that the most important parameters are self-evident from the examples given.

all input is encircled like this

```
COMMAND- connect,input,output  <- required on our CDC computer for an
                                  interactive session
COMMAND- fisher <- loading and executing FISHER

X X2  FISHER R*C EXACT TESTS
 X    C(1980) AV/PMK
X X   TEST VERSION  CDC-0.78SP.
      TODAY IS 04/03/81  09.53

      ia  <- we want to work in the interactive mode
           <- a "task" can be: - a request for help or information (see below);
                               - the analysis of one table; this starts by
                                 answering "TITL" followed by a one line header;
                               - one of a few odd jobs (see HELP, I)
1° TASK:help

HELP - SYNTAX:
  INFO    GIVES INFORMATION ABOUT THE PROGRAM
  HELP    GIVES THIS SYNTAX OF HELP-
  HELP,A  GIVES ALL SECTIONS OF THE INPUT SUMMARY
  HELP,F  GIVES INFORMATION ON THE INPUT FORMATS
  HELP,I  GIVES AN EXPLANATION FOR INOPT (=INPUT OPTION)
  HELP,O  GIVES AN EXPLANATION FOR OUTOPT (=OUTPUT OPTION)
  HELP,S  GIVES AN EXPLANATION FOR ISTAT (=STATISTICS)
  HELP,T  GIVES AN EXPLANATION FOR THRESH (=THRESHOLD)

2° TASK:help,o

   OUTOPT (=OUTPUT OPTION) SPECIFIES WHICH SELECTION OF A-E BELOW IS REQUESTED:
POSSIBLE OUTPUT : (A) INPUT TABLE,
                      OBSERVED VALUES OF THE STATISTIC(S),
                      APPROXIMATE SIGNIFICANCES,
                      APPROXIMATE OR EXACT NUMBER OF TABLES WITH THE SAME
                      MARGINALS,
                      AND (IF AVAILABLE) THE SUM OF THEIR PROBABILITIES (= 1.),
                      AND THE HYPERGEOMETRIC PROBABILITY OF THE OBSERVED TABLE.
                  (B) TABLES OF EXPECTED VALUES AND OF RESIDUALS.
                  (C) (FAST) CALCULATION OF THE NUMBER OF TABLES WITH THE SAME
                      MARGINALS.
                  (D) EXACT SIGINIFICANCE(S).
                  (E) WRITE WHOLE DISTRIBUTION ON UNIT  1.

OUTOPT = 0        AB
         1        ABC       OUTOPT = -1       A C
         2        AB D               -2       A  D
         3        ABCD               -3       A CD
         4        ABCDE              -4       A CDE
```

3° TASK: titl everitt(1977) p 46 placebo versus drug 5 ← user defined header
keyword signalling analysis of a new table

```
FISHER   C(1980) AV/PMK   TEST VERSION  CDC-0.78SP.            04/03/81  09.56

ANALYSIS NUMBER     1
 EVERITT(1977) P 46 PLACEBO VERSUS DRUG 5
```

outopt=-3, see above at "help,o"

```
NR NC INOPT OUTOPT ISTAT THRESH (5I2,F4.2) :
```

← mnemonics for the input expected

```
  2 2 1-3
(INOPT= 1.OUTOPT=-3.ISTAT= 0.)  ( THRESH IS INACTIVE.)
N OF ROWS    =     2
N OF COLUMNS=      2
INPUT : CELL FREQUENCIES ONLY.
ROW  1:   8  15
ROW  2:  22  15
```

echo + effect of the input parameters

the expected format is multiple I4; after conversion to FORTRAN V this may be changed to free format

```
INPUT TABLE + MARGINS:

    8      15  :     23
   22      15  :     37
..............:........
   30      30  :     60

OBSERVED X SQUARED =    3.45 WITH   1 DEGREE(S) OF FREEDOM.

*********************************************************************************

         1 DEGREE OF FREEDOM.
         LEFT  TAIL SIGNIFICANCE =  .0551
         RIGHT TAIL SIGNIFICANCE =  .9837
```

$.0551 = Pr(\text{left upper corner cell} \leq 8)$

$.9837 = Pr(\text{left upper corner cell} \geq 8)$

$1.0388 = 1 + Pr(\text{left upper corner cell} = 8)$

```
*********************************************************************************
HYPERGEOMETRIC PROBABILITY OF THE OBSERVED TABLE = .0388

NUMBER OF TABLES WITH THE GIVEN MARGINS =          24 .

EXECUTION TIME FOR THIS ANALYSIS :     .05 SEC.
```

$= Pr(\text{left upper corner cell} = 8) = \frac{23!\,37!\,30!\,30!}{60!\,8!\,15!\,15!\,22!}$

4° TASK: titl voorbeeld met meerdan 1 vrijheidsgraad

```
FISHER   C(1980) AV/PMK   TEST VERSION  CDC-0.78SP.            04/03/81  09.56

ANALYSIS NUMBER     2
 VOORBEELD MET MEERDAN 1 VRIJHEIDSGRAAD
```

80

```
NR NC INOPT OUTOPT ISTAT THRESH (5I2,F4.2) :
   2 3 1 4
(INOPT= 1.OUTOPT= 4.ISTAT= 0.)  ( THRESH IS INACTIVE.)
N OF ROWS    =     2
N OF COLUMNS=     3
INPUT : CELL FREQUENCIES ONLY.
ROW  1:    0    1    5
ROW  2:    3    2    1

INPUT TABLE + MARGINS:

    0      1      5  :    6
    3      2      1  :    6
........................:........
    3      3      6  :    12

THE DISTRIBUTION WILL BE WRITTEN ON UNIT  1 (BECAUSE OUTOPT =4 OR -4).


EXPECTED VALUES + MARGINS :
(SMALL EXPECTED FREQUENCIES (< 5.0) ARE FLAGGED BY A # )

   1.5#   1.5#   3.0#:      6
   1.5#   1.5#   3.0#:      6
........................:........
    3      3      6  :    12

RESIDUALS:
1ST ENTRY: ADJUSTED RESIDUALS = (OBS-EXP) / ESTIMATED ST. DEV.
           (LARGE ABSOLUTE RESIDUALS (> 3.0) ARE FLAGGED BY A # )
2ND ENTRY: X2 TERMS = (OBS-EXP)**2 / EXP

  -2.0    -.7    2.3
   1.5     .2    1.3

   2.0     .7   -2.3
   1.5     .2    1.3



OBSERVED X SQUARED =    6.00 WITH    2 DEGREE(S) OF FREEDOM.
SIGNIFICANCE ACCORDING TO CHI2 APPROXIMATION  .050
*****************************************************************************

        EXACT SIGNIFICANCE =  .1234

PROBABILITY JUMP =  .1212
*****************************************************************************
HYPERGEOMETRIC PROBABILITY OF THE OBSERVED TABLE = .0195

NUMBER OF TABLES WITH THE GIVEN MARGINS =        16 .
SUM PROBABLITY OF ALL TABLES =1.00000000000012 ,WHICH SHOULD BE 1., EXCEPT
FOR ROUNDING ERRORS.


EXECUTION TIME FOR THIS ANALYSIS :    .10 SEC.
```

outopt = 4 : maximal output

istat = 0 : X^2 only

see help,t

a rather discomforting difference

Large! Note that many permutations of rows and columns lead to the same X^2. From unit 1 we may learn that the distribution of X^2 is very discrete:

X^2	probability	cumulative probability
12	.002	.002
6	.12	.12
4	.19	.31
1.33	.29	.61
.67	.39	1.00

5° no "titl"

TASK: maxwell(1977) table 10.3

```
!! SEARCHING FOR A TASK CARD, THE FOLLOWING NON-TASK
   CARD(S) WERE FOUND, AND IGNORED:
MAXWELL(1977) TABLE 10.3
```

6° TASK: titl maxwell(1977) table 10.3

```
FISHER  C(1980) AV/PMK   TEST VERSION  CDC-0.78SP.              04/03/81  09.56

ANALYSIS NUMBER     3
 MAXWELL(1977) TABLE 10.3

NR NC INOPT OUTOPT ISTAT THRESH (5I2,F4.2) :
  5 5 1 0 1
(INOPT= 1.OUTOPT= 0.ISTAT= 1.)  ( THRESH IS INACTIVE.)
N OF ROWS    =    5
N OF COLUMNS=     5
INPUT : CELL FREQUENCIES ONLY.
ROW  1:   3   2   5  10  11
ROW  2:  11   8  16  35  19
ROW  3:  28  13  23  33   6
ROW  4:  27  11  23  12   5
ROW  5:  63  10   9   4   0
```

default output option

istat=1 : all 6 statistics; values are 1,2,3,...

population: 387 patients of a psychiatrist, each rated on two 5 point scale
x (down) = "tension"
y (across) = "worry"

INPUT TABLE + MARGINS:

3	2	5	10	11	:	31
11	8	16	35	19	:	89
28	13	23	33	6	:	103
27	11	23	12	5	:	78
63	10	9	4	0	:	86
132	44	76	94	41	:	387

ROW NUMBER	:	1	2	3	4	5
VALUES (ETA2 & R)	:	1.00	2.00	3.00	4.00	5.00
MIDRANKS (K-W & RS)	:	16.0	76.0	172.0	262.5	344.5
COLUMN NUMBER	:	1	2	3	4	5
VALUES (R)	:	1.00	2.00	3.00	4.00	5.00
MIDRANKS (RS)	:	66.5	154.5	214.5	299.5	367.0

← default values

EXPECTED VALUES + MARGINS :
(SMALL EXPECTED FREQUENCIES (< 5.0) ARE FLAGGED BY A #)

10.6	3.5#	6.1	7.5	3.3#	:	31
30.4	10.1	17.5	21.6	9.4	:	89
35.1	11.7	20.2	25.0	10.9	:	103
26.6	8.9	15.3	18.9	8.3	:	78
29.3	9.8	16.9	20.9	9.1	:	86
132	44	76	94	41	:	387

Cochran's rule allows the χ^2-approximation to the X^2-distribution

```
RESIDUALS:
1ST ENTRY: ADJUSTED RESIDUALS = (OBS-EXP) / ESTIMATED ST. DEV.
           (LARGE ABSOLUTE RESIDUALS (> 3.0) ARE FLAGGED BY A # )
2ND ENTRY: X2 TERMS = (OBS-EXP)**2 / EXP

  -3.0     -.9     -.5     1.1     4.7#
   5.4      .7      .2      .8    18.1

  -4.9#    -.8     -.4     3.8#    3.8#
  12.3      .4      .1     8.3     9.7

  -1.7      .5      .8     2.1    -1.8
   1.4      .1      .4     2.5     2.2

    .1      .9     2.5    -2.1    -1.3
    .0      .5     3.9     2.5     1.3

   8.7#     .1    -2.4    -4.8#   -3.6#
  38.6      .0     3.7    13.7     9.1
```

positive residuals

A splendid pattern of residuals:
- large in the corners
- positive along one diagonal.

Typical for strong (negative) dependency. With so many observations, and such a strong dependency significance testing trivially leads to extreme significance; it is non-informative.

```
****************************************************************************
TYPE OF       : GENERAL  : ROW VARIABLE ORDERED:   BOTH VARIABLES ARE ORDERED   *
ALTERNATIVE   :          : ONE WAY ANOVA       :                                *
HYPOTHESIS H1:           : COLUMNS ARE GROUPS  :  H1 IS A ONE-SIDED HYPOTHESIS  *
              :          :           :         :             :             *
STATISTIC     :PEARSON'S :KRUSKAL-   :CORR RATIO:KENDALL'S :SPEARMAN'S:PEARSON'S *
              :   X2     :  WALLIS   :  ETA2   :  TAU B    :   RS     :   R      *
OBSERVED      :          :           :         :           :          :          *
VALUE         :  136.16  :    .28    :   .28   :  -.45     :  -.53    :  -.53    *
              :          :           :         :           :          :          *
              :          :           :         :           :          :          *
APPROXIMATION: CHI2 WITH:   CHI2/(N-1) WITH    :  NORMAL   : T DISTRIBUTION FOR  *
              :  16 DF.  :        4 DF.        :           :R*SQRT((N-2)/(1-R*R))*
APPROXIMATE   :          :           :         :           : WITH 385 DF.        *
SIGNIFICANCE :   .000    :   .000    :  .000   :           :          :          *
RGHT-SIDED H1:           :           :         :  1.000    :  1.000   :  1.000   *
LEFT-SIDED H1:           :           :         :  0.000    :   .000   :   .000   *
              :          :           :         :           :          :          *
****************************************************************************
HYPERGEOMETRIC PROBABILITY OF THE OBSERVED TABLE = .0000

GAIL-MANTEL ESTIMATE OF THE NUMBER OF TABLES WITH THE SAME MARGINALS*********
```

estimate $\geq 10^9$, format overflow; this has been repaired

```
EXECUTION TIME FOR THIS ANALYSIS :     .14 SEC.
```

7° TASK: 0

```
END OF INPUT FILE REACHED.
NUMBER OF ANALYSES:      3 .
EXECUTION TIME :        .32 SEC.
     STOP
     050500 MAXIMUM EXECUTION FL.
       .331 CP SECONDS EXECUTION TIME.
```

```
GAIL-MANTEL ESTIMATE OF THE NUMBER OF
TABLES WITH THE SAME MARGINALS .127E+23
```

In other words: we advise to bypass the exact calculations, which at the Cyber would require some 10 000 000 000 CPyears. Moreover the approximations seem adequate, and significance is guaranteed

Finally, an example showing what happens to a user unfamiliar to the input expected by FISHER: the program is self-documenting.

```
COMMAND- fisher


X X2  FISHER R*C EXACT TESTS
 X    C(1980) AV/PMK
X X   TEST VERSION  CDC-0.78SP.
      TODAY IS 04/03/81  10.04
```

8° nonsense input, omdat de programmagebruiker fisher nog niet kent

```
PROGRAM INFORMATION

*************************************************************************

                               FISHER

    INTRODUCTION

      THIS PROGRAM IS DESIGNED FOR THE ANALYSIS OF TWO DIMENSIONAL
    CONTINGENCY TABLES. ITS NAME IS DERIVED FROM ITS COMPUTATION OF THE
    "EXACT" SIGNIFICANCE OF PEARSON'S X-SQUARED=X2 FOR AN R*C CONTINGENCY
    TABLE, CONDITIONAL ON THE MARGINS.
      HERE, SIGNIFI[illegible] IS DEFINED AS [illegible] X2 >= X2OBSERVED
                       [illegible]BSERVED-EXPE[illegible]  2
      [illegible]SON'S [illegible]   -------
                                   EX[illegible]
```

There follows a synopsis of the User Guide, including the Input Summary.

COMPARISON with BMDP, IMSL and SPSS

The BMDP program for measures of association, BMDP1F, gives significances only for X^2. For other statistics they provide the following information:

- The "t-value", which is asymptotically t-distributed under H_0. So this can easily be used for one-or two-sided testing; two-sided tests are equal tails tests.
- An estimate of the asymptotic standard error under H_0 (which is used as denominator for the t-value).
- An estimate of the asymptotic standard error under the alternative H_1. This is convenient for constructing approximate confidence intervals.

Moreover BMDP1F reports the minimal expected value in the table, as a signal warning for a possible breakdown of asymptotic methods. It can also generate tables of expected values and of various types of residuals. Finally, for 2x2-tables always the exact, hypergeometric probabilities are used.

IMSL version 8 has a new subroutine CTPR, which is claimed to produce "exact significance tests in rxc-contingency tables". However it is not stated what test statistic is used. The source text reveals that they use the probability of the table itself as test statistic. This is all right for 2x2-tables, but the generalization to rxc-tables seems to lack any justification beyond a superficial intuition that has been critized e.g. by Fisher (1950) and by Gibbons & Pratt (1975). We know of three properties in favour of this test:

- It is admissable.
- It is asymptotically equivalent to X^2, -2 ln LR etc.
- The significance can be calculated much easier (i.e. faster) than for any other statistic.

However none of these properties seems very relevant to the small sample situation, where one is interested in exact tests.

The SPSS subprogram CROSSTABS prints a bunch of statistics for rxc-tables. Significances are only given for X^2, τ_b, and τ_c (and in the CDC version also for r). Remarkably η (rather than η^2, the correlation ratio) is given, but without significance. In the manual by Nie et al. (1975) we could not find any reference to the fact that the significances are based on asymptotic methods, nor clues whether the significances for τ_b, τ_c, and r are one-sided or two-sided. In the SPSS Algorithm Manual by Norušis (1979) one reads on page 16 that the τ's are tests two-sidedly, while page 18 suggests that for r a one-sided test is applied. Checking the output of CDC-SPSS version 8.0 shows that for the 2x2-tables investigated both tests were applied one-sidedly.

Another curious Mumbo Jumbo in SPSS is the treatment of 2x2-tables. As if it had to be done by hand, no exact probabilities are calculated for more than 20 (twenty) observations. The one-sided tests for the τ's and for r are equivalent in 2x2-tables, but SPSS bravely computes and prints all slightly different asymptotic significances, next to the (two-sided) significances of X^2 with and without Yates' correction. This usually leads to three widely different (groups of) significances, with little help for the user how to interpret the differences.

The unorthodox habit of printing $\sqrt{\eta^2}$ and $\sqrt{\chi^2/N}$ rather than the squares may lead to some confusion in tables with negative τ's and r.

The most recent IBM version of SPSS prints as well the minimal expected value in the table, as the number and percentage of cells with an expectation less than 5.

REFERENCES

Agresti, A. and Wackerly, D.: Some exact conditional tests of independence for RxC cross-classification tables. Psychometrika, 1977, 42, 111-125.

Agresti, A., Wackerly, D. and Boyett, J.M.: Exact conditional tests for cross-classifications: approximation of attained significance levels. Psychometrika, 1979, 44, 75-83.

Baker, R.J.: Algorithm AS112. Exact distributions derived from two-way tables. Appl. Statist., 1977, 26, 199-206.

Boulton, D.M. and Wallace, C.S.: Occupancy of a rectangular array. The Computer Journal, 1973, 16, 57-63.

Boulton, D.M.: Remark on Algorithm 434. Communications of the ACM, 1974, 17, 326.

Boulton, D.M.: Remark on Algorithm 434. ACM Transactions on Mathematical Software, 1976, 2, 108.

Cantor, A.B.: A computer algorithm for testing significance in MxK contingency tables. 5th Proceedings of the Statistical Computing Section of the ASA, 1979.

Cochran, W.G.: The χ^2 test of goodness of fit. Ann. Math. Stat., 1952, 23, 315-345.

Cochran, W.G.: Some methods for strengthening the common χ^2 tests. Biometrics, 1954, 10, 417-451.

Dixon, W.J. and Brown, M.B.: BMDP-77. University of California Press, Berkeley, 1977. Esp. program BMDP1F and appendix A5.

Ehrenberg, A.S.C.: Data reduction. Wiley, 1975.

Everitt, B.S.: The analysis of contingency tables. Chapman and Hall, London, 1977.

Gail, M. and Mantel, N.: Counting the number of rxc contingency tables with fixed margins. Journal American Statistical Association, 1977, 72, 859-862.

Haberman, S.J.: The analysis of residuals in cross-classified tables. Biometrics, 1973, 29, 205-220.

Hancock, T.W.: Remark on algorithm 434. Communications of the ACM, 1975, 18, 117-119.

Hays, W.L.: Statistics for the social sciences. Holt, Rinehardt and Winston, London, 1973.

IMSL Edition 8, Houston, 1980.

Kempthorne, O.: In dispraise of the exact test: reactions. J. Statist. Planning Inf.,1979, 3, 199-213.

Kendall, M.G.: Rank correlation methods. Griffin, 1948, 4th ed.: 1975.

Kendall, M.G. and Stuart, A.: The advanced theory of statistics, vol. 2. Griffin, 3rd ed 1973. Esp. sections 31.18/31.49: tests of independence.

Klotz, J.N. and Teng, J.: One-way layout for counts and the exact enumeration of the Kruskal-Wallis H distribution with ties. Journal American Statistical Association, 1977, 72, 165-169.

Kroonenberg, P.M. and Verbeek, A.: Exact distributions of χ^2 tests in contingency tables with small numbers.(WEP Reeks, WR 80-18-EX). Leiden: Vakgroep Wijsgerige en Empirische Pedagogiek, Rijksuniversiteit Leiden, 1980 (poster session at the 4th COMPSTAT symposium, Edinburgh, August 1980.

Kroonenberg, P.M. and Verbeek, A.: Programmer's Guide to FISHER. (manuscript in preparation).

Larntz, K.: Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. Journal American Statistical Association, 1978, 73, 253-263.

Lehmann, E.L.: Non-parametrics: Statistical methods based on ranks. San Francisco, Holden-Day, 1975.

March, D.L.: Algorithm 434. Exact probabilities for rxc contingency tables. Communications of the ACM, 1972, 15, 991-992. (Rather inefficient, see Boulton 1974.)

Maxwell, A.E.: Multivariate Analysis in Behavioural Research. Chapman and Hall, London, 1977.

Mood, A.M., Graybill, F.A. and Boes, D.C.: Introduction to the theory of statistics. (3rd ed.) McGraw Hill, 1974.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K. and Bent, D.H.: SPSS. (2nd ed) McGraw Hill, 1975. Esp. CROSSTABS, p 222-230 and 243.

Norušis, M.J.: SPSS Statistical Algorithms. SPSS Inc, Chicago, 1979. Esp. p 12-18.

Ryder, B.G.: The PFORT Verifier, Software Practice and Experience 4 (1974), 359-377.

Siegel, S.: Nonparametric statistics: For the behavorial sciences. Tokyo, McGraw-Hill Kogakusha Ltd., 1956.

Verbeek, A. and Kroonenberg, P.M.: De X^2-toets voor kruistabellen: conditioneren of niet-vervolg. VVS-Bulletin, 1980, 13 (3), 22-25, 28-36.