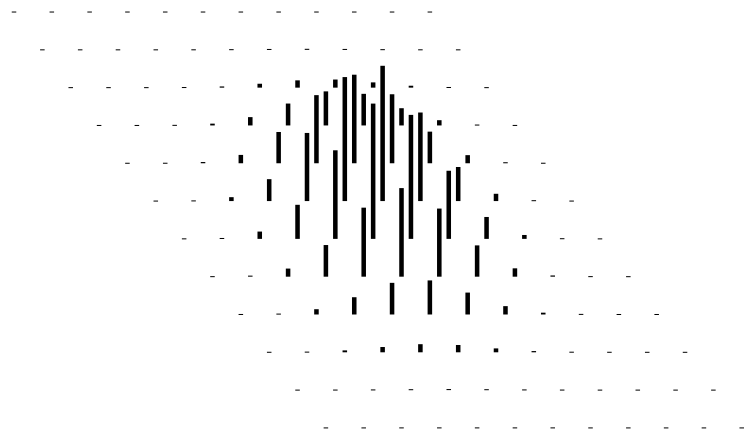
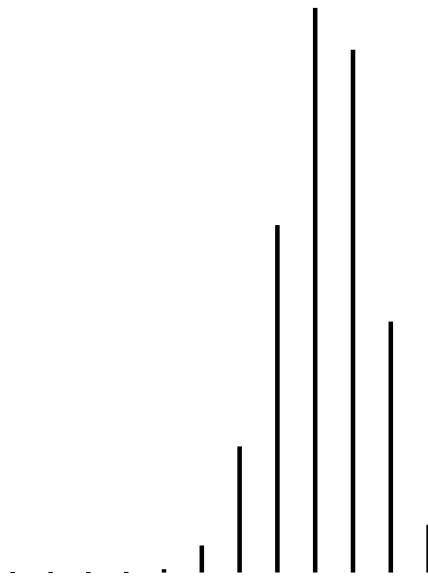


FISHER
TESTING INDEPENDENCE
IN $r \times c$ TABLES



FISHER 3.10
TESTING INDEPENDENCE
IN $r \times c$ TABLES

(2nd revised printing)

Albert Verbeek & Pieter M. Kroonenberg



-

This is the User's Manual of FISHER 3.11. The program itself is available for IBM PCs and compatibles, and can be made available for most other computer systems with an ISO Fortran compiler. Copies of this manual and the program can be ordered from

The Three-Mode Company - P. M. Kroonenberg

Mail: Department of Education and Child Studies,
Leiden University, Wassenaarseweg 52
2333 AK Leiden, The Netherlands
Email: kroonenb/@/fsw.leidenuniv.nl
Tel: +31 71 5273446/3434
Fax: +31 71 5273945

© iec ProGAMMA, 1990, 1993; The Three-Mode Company, 2004

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

ISBN 90-5144-010-3

The Three-Mode Company is a non-profit institution supported by the buyers of its software and by the Leiden University. Sole representative: Pieter M. Kroonenberg

Manual date: August 1, 2005

Acknowledgements. The FISHER team consisted of Albert Verbeek († 1970), Pieter Kroonenberg, Siep Kroonenberg, and Bill Barit, with financial support from

- the University of Utrecht,
- Leiden University,
- the Netherlands Organization for the Advancement for Pure Research, ZWO, Grant 56-185,
- the Netherlands Society for Road Safety Research, SWOV,
- iec ProGAMMA.

This book has been prepared with L^AT_EX and has been reproduced from camera-ready copy supplied by the authors.

Contents

Preface	ix
Part I. Statistical Background	2
1 Introduction	2
1.1 Purpose	2
1.2 Large sample methods	2
1.3 A word of warning	4
2 Notation and terminology	5
2.1 The p-value, observed values, and expected values	5
2.2 Scores or values of the categories	6
2.3 The isomarginal family	6
3 Distributional assumptions	6
3.1 Sampling schemes	6
3.2 The hypergeometric distribution as null hypothesis	7
4 Sketch of the algorithm	8
4.1 Exact computation	8
4.2 Monte Carlo estimation	12
5 Test statistics	13
5.1 Pearson's X^2 (X2)	13
5.2 The log likelihood ratio G^2 (LR)	13
5.3 The Freeman-Tukey test statistic $F-T$ (F-T)	13
5.4 The exact probability test statistic (EPT)	13
5.5 Kruskal-Wallis' test statistic H (K-W)	14
5.6 The correlation ratio η^2 (ETA2)	14
5.7 Kendall's τ_b (TAU)	14
5.8 Spearman's rank correlation coefficient R_S (RS)	15
5.9 Pearson's correlation coefficient R (R)	15
5.10 Tests for two columns = two groups	16
5.11 Which test statistic?	16
5.12 Scores for numerical variables	18
5.13 Midranks for ordinal variables	18
6 Large sample tests	20
6.1 Limit distributions	20
6.2 Should most expected values be larger than about 1?	21

7	Small sample tests	22
7.1	The calculation of the probability mass at the observed value	22
7.2	One-sided tests	23
7.3	Two-sided tests	24
8	Criterion robustness versus inference robustness	25
9	More on Pearson's X^2	26
9.1	X^2 terms and residuals	26
9.2	Collapsing and grouping	27
9.3	Quick methods	29
	 Part II. Example of a FISHER job	 33
10	Input	33
	 Part III. Reference Manual	 37
11	The user interface	37
11.1	General points	37
11.2	Files	38
11.3	Writing the entire distribution to file	38
11.4	Numbers	39
12	FISHER command syntax	39
12.1	General rules	39
12.2	Command descriptions	40
12.3	Data entry commands	40
12.4	Analysis commands	43
12.5	Control commands	45
12.6	File commands	47
	 Part IV. Appendices	 51
A	Installing and running FISHER on IBM PCs and compatibles	52
A.1	System requirements	52
A.2	The distribution disk	52
A.3	Installation	52
A.4	Running FISHER	52
A.5	Running the demo	53

B	Additional technical information	54
B.1	Marginal zeroes	54
B.2	Conditioning	54
B.3	Monte Carlo estimation	57
C	Annotated output	60
	References	71
	FISHER command summary	74

List of Tables

1	Violations of Cochran's rule of thumb	4
2	Example of one- and two-sided tests in a 2×2 table	8
3	Example of exact testing in a 2×3 table	10
4	The null distribution of X^2	10
5	Choosing a statistic based on the alternative hypothesis	16
6	The calculation of midranks	18
7	Asymptotic approximations to the null distributions	20
8	Example of collapsing	28

List of Figures

	Hypergeometric and generalized hypergeometric probabilities	ii
1	Hypergeometric distributions with 1 and 2 df	9
2	3D Plots of Pearson's X^2 and the hypergeometric probabilities	11
3	3D Plots of p , X^2 , η^2 , τ , and R	19

Preface

FISHER is a program to test independence in $r \times c$ contingency tables, especially (but not only) when the sample size is small. In all tests we will condition on the margins. We hope that the program will be useful to applied statisticians and to users of statistics from other disciplines. As minimal prerequisites we assume a basic understanding of statistical tests in general, and of the usual test statistics for $r \times c$ tables in particular.

Originally FISHER has been developed to compute exact p-values (as opposed to the usual asymptotic approximations). This is necessary for small samples. Unfortunately no one knows what ‘small’ really means. Notwithstanding the emphasis on small sample tests, the program is so easy to use, that many people use it routinely for the computation of statistics in $r \times c$ contingency tables with or without any testing. However, one should not use *exact testing* routinely, as this may be prohibitively costly in terms of computer time. Remember always to take a good look at your data first. For many tables a little ingenuity can give results faster than the time required for keying the data into a computer program. Some quick and dirty methods are described in section 9.3.

The statistics built in are:

- Pearson’s X^2 and some asymptotically equivalent χ^2 statistics, such as $G^2 = -2 \log$ likelihood ratio, and the Freeman-Tukey statistic,
- the correlation ratio η^2 from one-way analysis of variance, which is equal to explained SS/total SS,
- Kruskal-Wallis’ H , the nonparametric counterpart of η^2 ,
- Kendall’s τ_b , which as a test statistic is equivalent to the other τ -s and to Kruskal-Goodman’s γ ,
- Spearman’s rank correlation coefficient R_S , and
- Pearson’s (ordinary) correlation coefficient R .

Implicitly included are:

- Fisher’s exact test ($r = c = 2$, in which case all one-sided test statistics are equivalent),
- Mann-Whitney-Wilcoxon’s test (= two-sided Kruskal-Wallis with $c = 2$), and
- the two-sample t-test (equivalent to the correlation coefficient R with $c = 2$).

For definitions, references and other details see section 5. For these statistics FISHER provides the observed value and

- the popular, large sample approximations of the p-values (default)
- the exact p-values (optionally, if the table and the total number of observations are not too large, if `ANAL EXACT` is specified)
- a Monte Carlo estimate of the exact p-values (optionally, if `ANAL MC` is specified)

Presto

For the user who wants to use FISHER straightway: First make sure FISHER is installed on your computer system. For a PC, Appendix A contains a summary

of the installation procedure, and indicates how FISHER is run. For installation on other systems consult the README file on the distribution tape. For PCs the distribution diskette also contains the file DEMO.BAT that invokes FISHER with the input file DEMO.

Now, before you get started, there are six more things you should know.

- One exits FISHER by specifying STOP, BYE, END, or QUIT
- All key words can be abbreviated to the first 3 characters.
- Almost all parameters in FISHER are optional. Almost all parameters are order independent. Almost all unspecified parameters retain the previous values, and all parameters are initialized. There are a few exceptions, but these are very natural.
- Data are entered after the command

DATA *nr nc*

where *nr* is the number of rows, and *nc* is the number of columns. Thus, for a 3×4 table, type DATA 3 4

- Subsequently an analysis starts with ANA. To explore the program, first run ANA without any options, and then compare ANA ALL with ANA X2 and ANA X2+. Next compare ANA EXACT, ANA APPROX, and ANA MC. For a list of all (optional) parameters see HELP ANA. Finally, try TABLE to see the observed table, or EXRE to inspect the EXpected values and the RESiduals.

Adagio

This book consists of four parts. Part I gives the statistical background. It explains how exact tests in $r \times c$ contingency tables work, it gives the formulas of the statistics incorporated in FISHER, and it explains which statistic should be used for which alternative hypothesis. Most sections are short, so the table of contents gives a convenient survey.

Part II gives an example of the use of FISHER. It gives the input of a job with some typical analyses, and it gives an annotated listing of the output.

Part III is the Reference Manual.

Part IV consists of assorted appendices. The main topics are the installation of FISHER on a PC, limitations and restrictions, and an extensive bibliography. A summary of the command syntax is printed on the last page for easy reference.

For **teachers of statistics** we have the following suggestion. Rather than providing only the p-values, FISHER can also write the entire distribution (or a Monte Carlo sample thereof) of the statistic(s) to a file (which easily becomes enormous; take care). For details, see section 11.3. Small examples of exact distributions as provided in section 4 are very helpful to explain the basic notions of statistical testing in a classroom even to statistical laymen. The output provides an excellent illustration for the introduction of the very idea of statistical testing, but also for more specific notions such as critical regions, p-values, the difference between different statistics, and equivalence of statistics (especially in 2×2 tables). Also Monte Carlo sampling, simulation, and empirical cumulative distribution functions can be introduced this way.

Availability, support & fair play: from us to you, and from you to us. FISHER Version 3.10 is available for IBM PCs and compatibles, and also for many other computer systems. Distribution of the program plus this manual is handled by the Interuniversity Expertise Centre ProGAMMA. ProGAMMA is a non-profit institution, but for its continuation it depends on the success of its sales. FISHER is not copy protected, illegal copying is prohibited, and the price of FISHER is small. Hence, if you are frequently using an illegal copy of the program, or of this manual, please be decent enough to buy a legal copy.

If the program contributes to your research, please make an acknowledgement and include this booklet in your list of references. We would be pleased to receive a reprint.

If you have any suggestions, questions, or if you should observe any errors, let us know. The second author¹ can be contacted via ProGAMMA.

¹Note to the second printing. Albert Verbeek died on 9 September 1990 only weeks after finishing the final version of FISHER 3.0. This second revised printing of the manual had to be reconstructed from his original files, some minor errors have been corrected, some new ones may have unintentionally crept in, and references to papers he still intended to write have now been removed. New technical appendices have been included as well.

Part I

Statistical Background

1 Introduction

1.1 Purpose

FISHER is a program for testing independence or homogeneity in an $r \times c$ contingency table. It calculates descriptive levels of significance (= p-values) or, optionally, entire distributions of nine popular statistics. Usually the significance levels are calculated by large sample methods (and the program provides these too). In 2×2 tables large sample methods are satisfactory only if all expected values exceed 25. (Although several computer programs suggest otherwise, it is not the overall sample size that matters in 2×2 tables, but the smallest expected value.) For smaller samples, Fisher's exact test (Fisher, 1925, 1970) is a popular and superior alternative. However, also for larger tables, it is theoretically straightforward to compute the exact distributions of test statistics; for details see section 4.

Calculating the exact distribution is straightforward in theory, but in tables with more than 1 degree of freedom it may be a lot of work in practice. The exact distributions of the test statistics are discrete and difficult to treat analytically. In addition, there is no convenient way to tabulate exact distributions due to the large number of parameters. Therefore, they have to be calculated by rather involved combinatoric methods. Actually, until the advent of sufficiently fast computers and of efficient algorithms (around 1975) this computation was prohibitively laborious for almost all samples of any interest. Since then many cases of practical interest can be handled, but even with FISHER and computers built in 1990, exact calculations are too time consuming in larger tables and/or with larger samples, and say 12 df or a few hundred observations is pretty large. If computation of the exact distribution is too laborious, Monte Carlo estimation is an alternative to the standard asymptotic approach. The advantage of Monte Carlo estimation over asymptotic methods is, that its accuracy is both known and under control. The advantage of Monte Carlo estimation over exact computation is that the time it takes hardly depends on the sample size. FISHER computes p-values by the following three methods:

- **Large sample approximation.** This is the usual method, it is very fast, but its accuracy is unknown.
- **Exact calculation by combinatorial enumeration.** This is the best method, but it may require large amounts of computer time.
- **Monte Carlo approximation of the exact p-value.** This takes a moderate amount of computer time, that increases only slightly with the total number of observations, and is proportional to the number of cells and to the Monte Carlo sample size. Its accuracy is known, and depends on the size of the Monte Carlo sample.

1.2 Large sample methods

What are large sample methods? Large sample methods give an approximation to the distribution of a statistic, by letting the sample size go to infinity, while all (population) proportions are kept fixed. This limiting distribution is called the

asymptotic distribution. Note, that with a finite sample, there are only finitely many possible values for the statistic, so its ‘real’ distribution is discrete. Commonly, the asymptotic distribution is continuous, and it is used as an approximation to the real distribution.

Why are large sample methods so popular? While the exact p-values or Monte Carlo estimates thereof are hard to obtain without specialized software, the continuous limit distributions of the most popular statistics are simple and well-tabulated families such as normal distributions, t-distributions, and χ^2 distributions. Moreover, FISHER shows that in many cases their accuracy is remarkable.

Then, what is the problem with large sample methods? The asymptotic distributions have the advantage of easy tabulation and fast computation, but except for Pearson’s X^2 , little is known about the accuracy, especially for ‘small samples’. For the use of Pearson’s X^2 Cochran (1952, 1954) gave the following, famous rule of thumb:

If in a contingency table with more than 1 degree of freedom, 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).

We will call a table satisfying these conditions a **Cochran table**. Often Cochran’s rule is quoted in a washed down version, such as “In Cochran tables the χ^2 approximation ‘may be used’.” Note how much more precise Cochran’s rule is, and how remarkable the precision of his statement is in those pre-computer days. Cochran’s papers did not contain a proof. A complete search of all Cochran tables with at most 10 cells and at most 70 observations, has revealed very few counterexamples, and has shown that Cochran’s limits cannot be made narrower, at least not in 2×3 tables. Rather, based on extensive computer simulations, we would recommend the following slightly amended version.

If in a contingency table with more than 1 degree of freedom, 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 2.5 and at most 7.5 percent (.5 and 2 percent), with four minor exceptions. All violation are 2×3 -tables. They are given in Table 1.

Hence, for Pearson’s X^2 we know when the sample size is large enough to apply asymptotic methods. If the conditions from Cochran’s rule are violated however, the literature tells us virtually nothing about the accuracy of asymptotic methods. Many have proposed ‘corrections’ to X^2 or to its approximating χ^2 distribution, but with remarkably little success. Yet for example a table in which all expected values are around 3 may contain enough information to make the question of independence a sensible one. In such a case, only the methods implemented in programs like FISHER provide a reliable answer.

For χ^2 -statistics other than Pearson’s X^2 there are no rules of thumb such as Cochran’s. Amongst the statistics that test against the general alternative, with $(r - 1)(c - 1)$ df, Pearson’s X^2 seems to be the one with the best χ^2 approximation

Table 1: Violations of Cochran’s rule of thumb. The p-value indicated is the exact probability of the 5% χ^2 region. That is the exact probability that $X^2 > 5.991$. Note that the smallest expected values are not very small, and that the margins are remarkably ‘nice’.

	row	col		smallest
n	margin	margin	p-value	expected value
44	22,22	11,11,22	.084	5.5
42	21,21	11,11,20	.079	5.5
48	24,24	16,16,16	.077	8
54	18,36	18,18,18	.077	6

for its null distribution. For example, there is ample evidence, that the $-2 \log$ likelihood ratio G^2 is quite inferior in this respect, cf. Cressie & Read (1984). Other possibilities are Freeman-Tukey’s statistic $F-T$ (Freeman & Tukey, 1950), which was especially designed to handle small samples, and the hypergeometric probability of the table itself, which is often suggested in analogy with Fisher’s exact test for 2×2 tables. The quality of the χ^2 approximation of these two statistics is even worse than that of G^2 . Nevertheless, we have incorporated them in the option X2+ for the command ANAL, because the accuracy of the χ^2 approximation for the null distribution is only one property discriminating between these statistics, section 8: ‘Criterion robustness versus inference robustness’.

Next, consider statistics which test against restricted alternatives. FISHER supplies the correlation ratio η^2 of one-way Anova, its nonparametric counterpart Kruskal-Wallis’ H , Kendall’s τ_b , and Spearman’s and Kendall’s correlation coefficients R_S and R , respectively. Some experience with FISHER suggests, that for tables with more than 1 degree of freedom, the asymptotic p-values of the χ^2 distribution usually are sufficiently accurate for remarkably small samples. But it takes a program like FISHER to be sure.

1.3 A word of warning

For more than one degree of freedom exact calculation can be very time consuming. Moreover, the asymptotic approximations prove adequate in many cases with remarkably small samples. Therefore, one should not turn to exact calculation automatically or routinely. The program can estimate the cost of exact calculations in advance, while Monte Carlo estimation may provide a better compromise between accuracy and costs. Moreover some quick methods for Pearson’s X^2 are discussed in section 9.3.

Let us conclude this section with a horror story. Once, many years ago now, in a remote computing centre there was a user spending *hours* of CPU time on a large, fast, and expensive Cyber main frame, doing exact calculations for a 15×2 table,

only because Verbeek & Kroonenberg made the exact method available. In this run so many probabilities were computed and added together, that FISHER produced a serious warning that the accumulated roundoff error in the sum of all probabilities (this sum should be 1) exceeded a built in threshold. So, finally, this user came to us: throughout several identical trials, each costing several hours of CPU time, the warning persisted. From the data, a trained statistician could tell at a glance that the hypothesis of independence would be rejected. . .

2 Notation and terminology

2.1 The p-value, observed values, and expected values

Random variables will be denoted by italic capitals, such as X^2 , or by small greek letters, such as τ . The row variable of our tables will be denoted by X , and the column variable by Y . We use $\Pr(A)$ for the probability of an event A , S for an arbitrary statistic, and S_{obs} for the observed value. The descriptive level of significance or p-value of S_{obs} is defined as usual by

$$\text{p-value} = \Pr(S \geq S_{\text{obs}}) \quad (1)$$

Non-random variables are denoted by small letters.

We will always **condition on the margins**. That is, we treat the margins as fixed, and hence we denote them by small letters. For the heated discussion on conditioning, see Appendix B.2.

The **random table** of observed frequencies is denoted by

$$T = (T_{ij})_{i=1}^r_{j=1}^c,$$

where T_{ij} is the frequency or count in the i, j -cell of an $r \times c$ classification. Summation over the first or second index will be indicated by a '+' sign. Hence, the observed marginal totals are denoted by $t_{1+}, t_{2+}, \dots, t_{r+}$ and $t_{+1}, t_{+2}, \dots, t_{+c}$, while

$$n = t_{++}$$

is the total number of observations. Occasionally, we will denote the random **observed values** by

$$\text{obs} = T_{ij},$$

and the fixed **expected values** by

$$\text{exp} = t_{i+}t_{+j}/n,$$

Since we only treat $r \times c$ tables with fixed marginal totals, we assume (in Part I) that all marginal totals are positive (that is, any column or row with a zero margin has been removed), and that $r \geq 2$ and $c \geq 2$. Hence the number of degrees of freedom will be $(r-1)(c-1)$. Adding zero rows or columns does not affect the degrees of freedom, nor the distribution of any statistic. In FISHER, and in Part III, we allow zero rows and columns, as long as there are at least two non-zero rows, and at least two non-zero columns.

2.2 Scores or values of the categories

For the computation of η^2 the categories of the row variable have to have numeric values. One must compute overall mean, group means, sums of squares etc. Similarly, for R the categories of both variables have to have numeric values. Default values are 1, 2, 3, ... These values can be inspected and changed with the commands XVAL, YVAL, and DVAL. In this manual we call these numeric values ‘scores’ or ‘values’.

In Part I we assume that deletion of a row or column does not affect the scores of the other rows or columns. In FISHER and in Part III, however, the *default* scores are 1, 2, 3, ... Hence, if the default scores are used, removal of a row or column *by hand* leads to a change in the scores of subsequent rows or columns. If, for example, the second row is removed by hand, the scores of the remaining rows change from 1, 3, 4, ... to 1, 2, 3, ... This affects only η^2 (for row deletion) and R . Usually the effects are very small.

On the other hand, if for example the second row is not removed by hand, but by FISHER, because it contains only zero frequencies, the remaining rows keep the scores they had: 1, 3, 4, ...

Deletion of rows or columns in which all frequencies are zero is only relevant if FISHER is applied more or less mechanically to a series of tables. For if one enters tables by hand, one will generally simply skip such rows and columns.

2.3 The isomarginal family

An important concept for exact testing is the family of all tables with the same dimensions and the same margins as the observed table. This is called **the isomarginal family**. Because we condition on the margins, the isomarginal family is our sample space, that is, the set of possible outcomes.

3 Distributional assumptions

3.1 Sampling schemes

We consider four possible sampling schemes. These cover most practical cases in which the usual asymptotic or exact tests are appropriate. Recall X and Y are the row and column variable respectively.

- The n observations of the pair (X, Y) are independently, identically distributed. Hence the frequencies in the $X \times Y$ table have a multinomial distribution. This situation is typical for a survey, if the sample is a simple random sample with replacement.
- For each value of X (that is, for each stratum) we have a different population (that is a different distribution of Y), and an independent random sample from that population. Thus our frequencies have a product-multinomial distribution. An important special case is the comparison of two independent multinomial or binomial samples, for example a treatment group and a control group, or a ‘case group’ and a control group.

- Similarly to the situation above with the roles of X and Y interchanged.
- The frequency in each cell follows a Poisson distribution, and the frequencies in different cells are independent. Hence we have a product-Poisson distribution for our table. This situation is typical for an experiment where one observes and cross-classified random ‘occurrences’ in a preset time interval.

For more complex sampling schemes, such as sampling without replacement from a finite population, multi-stage sampling, and cluster sampling, it is quite difficult to calculate the exact null distribution of almost any statistic. In general, if the above assumptions about the sampling scheme are violated, the ‘exact methods’ discussed here only provide an approximation. Unfortunately it is quite difficult to assess the quality of this approximation. For complex sampling schemes it is not even certain that ‘exact methods’ always provide better approximations than asymptotic methods.

3.2 The hypergeometric distribution as null hypothesis

All tests considered here will be conditional on the observed margins (see also Appendix B.2). The null hypothesis is either ‘independence of the row and column variable’, or ‘homogeneity of independent samples’. These two notions differ a little in meaning, and refer to different sampling schemes, but after conditioning on the observed margins, both lead to the same null distribution, namely the (generalized) **hypergeometric** distribution

$$\Pr(T = t) = \frac{\prod_{j=1}^c t_{+j}! \prod_{i=1}^r t_{i+}!}{n! \prod_{i=1}^r \prod_{j=1}^c t_{ij}!} = \text{constant} / \prod_{i,j} t_{ij}! \quad (2)$$

Here ‘constant’ means that this factor only depends on the margins. For details and derivation see, for instance, Mood, Graybill & Boes (1974, section 5.4). One of the major technical advantages of conditioning on the margins is that the conditional null hypothesis consists of a single distribution.

Some authors reserve the name ‘hypergeometric’ for the case of 2×2 tables, and use ‘generalized hypergeometric’ or something similar for general $r \times c$ tables. We will call both ‘hypergeometric’, unless there is a special reason to emphasize the fact that we are dealing with more than one degree of freedom.

Figure 1 shows two hypergeometric distributions. The first distribution is the isomarginal family with 1 df, and with margins 20, 60, and 20, 60. In this table the smallest expected value \exp_{11} is 5, and the skewness is approximately $1/\sqrt{5} = .45$, which is only just visible. It looks very much like a discretized normal distribution, but the eye can not discriminate very well between normal and nonnormal distributions. The second distribution is the isomarginal family with 2 df and with margins 22, 22, and 22, 11, 11. It looks very much like a discretized binormal distribution.

4 Sketch of the algorithm

4.1 Exact computation

It was mentioned that the distribution of the table under the null hypothesis is the (generalized) hypergeometric distribution. The resulting exact tests are also called permutation tests. Exact tests in $r \times c$ tables are a direct generalization of Fisher's exact test for the 2×2 table (which is also efficiently handled by `FISHER`). Similarly as for 2×2 tables, we first have to determine the sample space, the set of all possible outcomes. This is the isomarginal family, which was introduced in section 2.3. For two-mini examples, see Tables 2 and 3, and Figure 2, in which the isomarginal families contain only 6 and 16 tables.

Table 2: Example of one- and two-sided tests in a 2×2 table. This table has a distribution virtually without a left tail. Hence, left-sided and 'really two-sided' testing are impossible. Also note that G^2 and X^2 order $A = 0$ and $A = 2$ differently; according to G^2 the exact significance level (p-value) of $A = 2$ is 0.545, according to X^2 it is 0.236.

The observed table

2	3	5
3	18	21
5	21	26

The isomarginal family

T_{11}	$5 - T_{11}$	5
$5 - T_{11}$	$16 + T_{11}$	21
5	21	26

The distribution of X^2 and G^2

t_{11}	G^2	X^2	$\Pr(T_{11} = t_{11})$
0	2.4	1.5	.309
1	.0	.0	.455
2	1.5	1.7	.202
3	5.5	6.7	.032
4	12.4	14.7	.0016
5	25.5	26.0	.000015

Figure 1: Hypergeometric distributions with 1 and 2 degrees of freedom. Generalization from Fisher's exact test for 1 degree of freedom to higher dimensions is what FISHER is all about. The top shows the hypergeometric probabilities of a 2×2 table. The second figure shows the generalized hypergeometric probabilities of a 2×3 table.

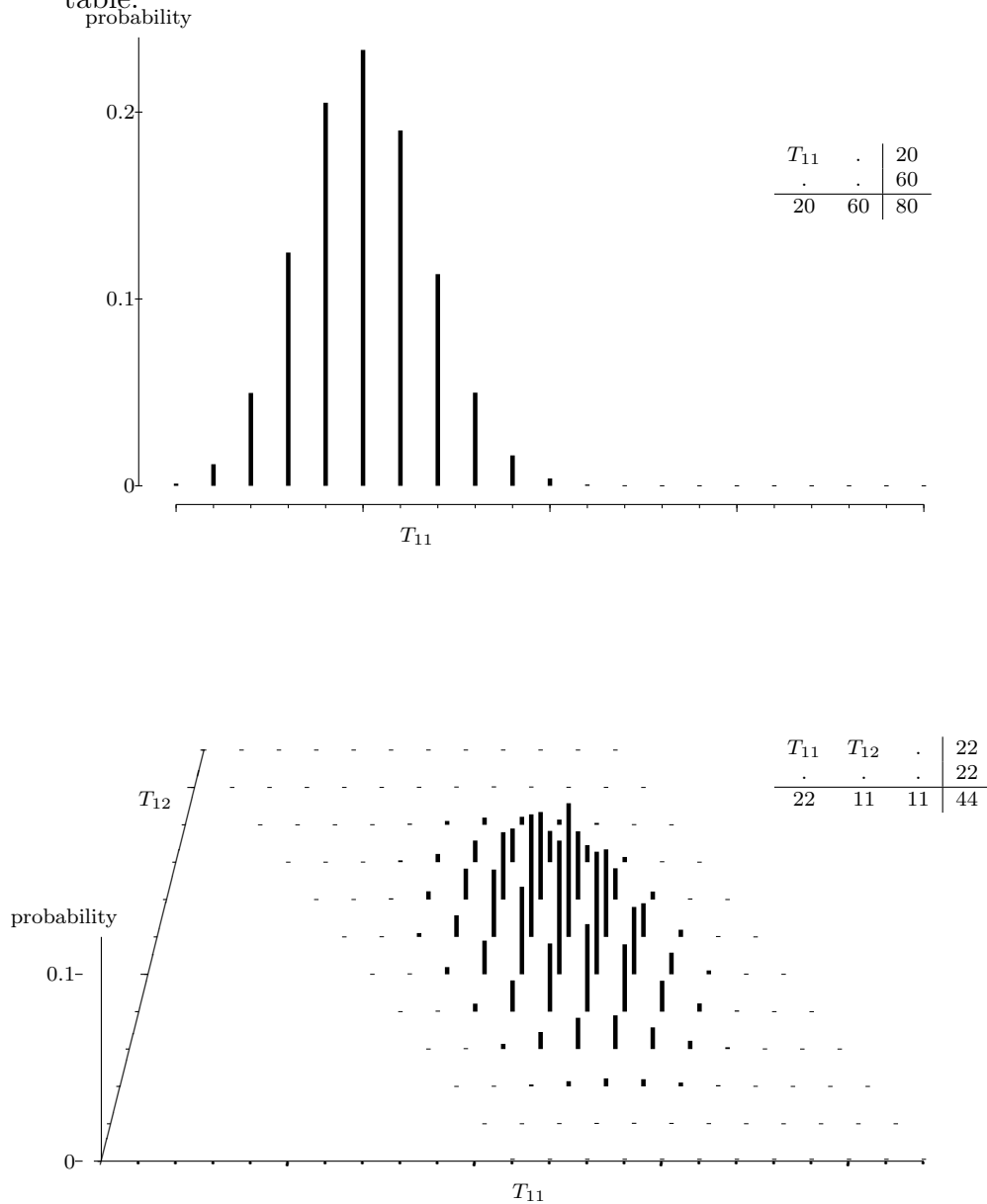


Table 3: Example of exact testing in a 2×3 table.

The observed table					The isomarginal family				
	treatment					treatment			
	A	B	C			A	B	C	
survived	5	1	0	6	survived	T_{11}	T_{12}	.	6
deceased	1	2	3	6	deceased	.	.	.	6
	6	3	3	12		6	3	3	12
Is treatment A really better?					$0 \leq T_{11} \leq 6$				
$X^2 = 6 > 5.991 = \chi^2_{.95}$					$0 \leq T_{12} \leq 3$				
Hence $p_{\text{asyp}} < .05 \dots$					$3 \leq T_{11} + T_{12} \leq 6$				

Table 4: The null distribution of X^2 in the isomarginal family of Table 3. Figure 2 gives the visual impression of the distribution.

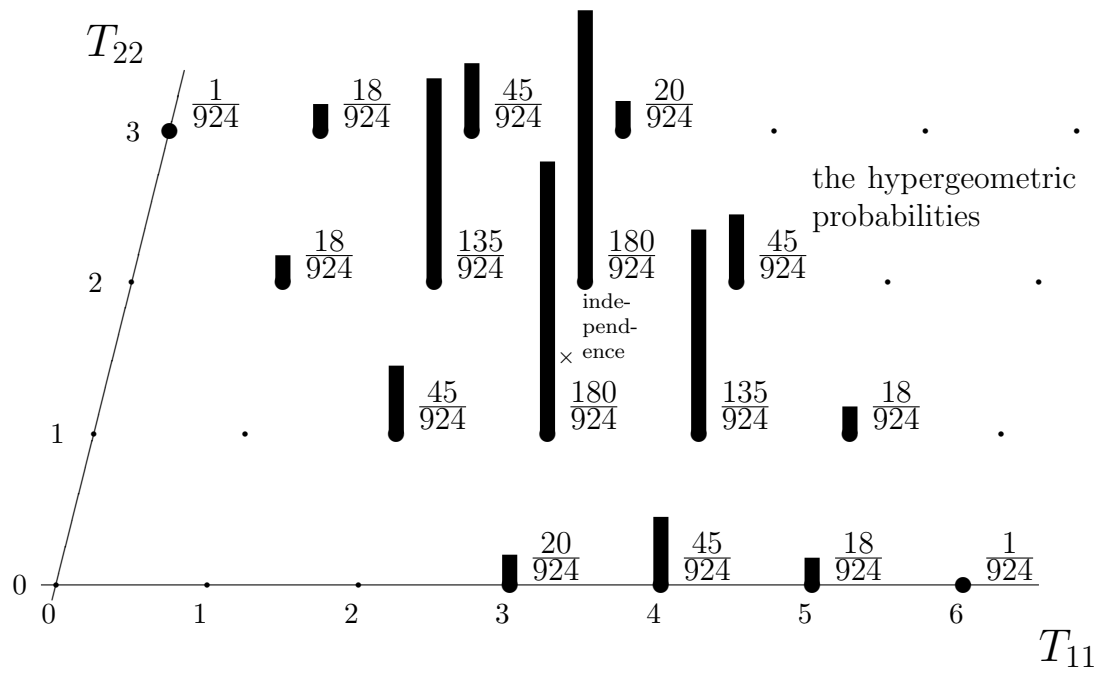
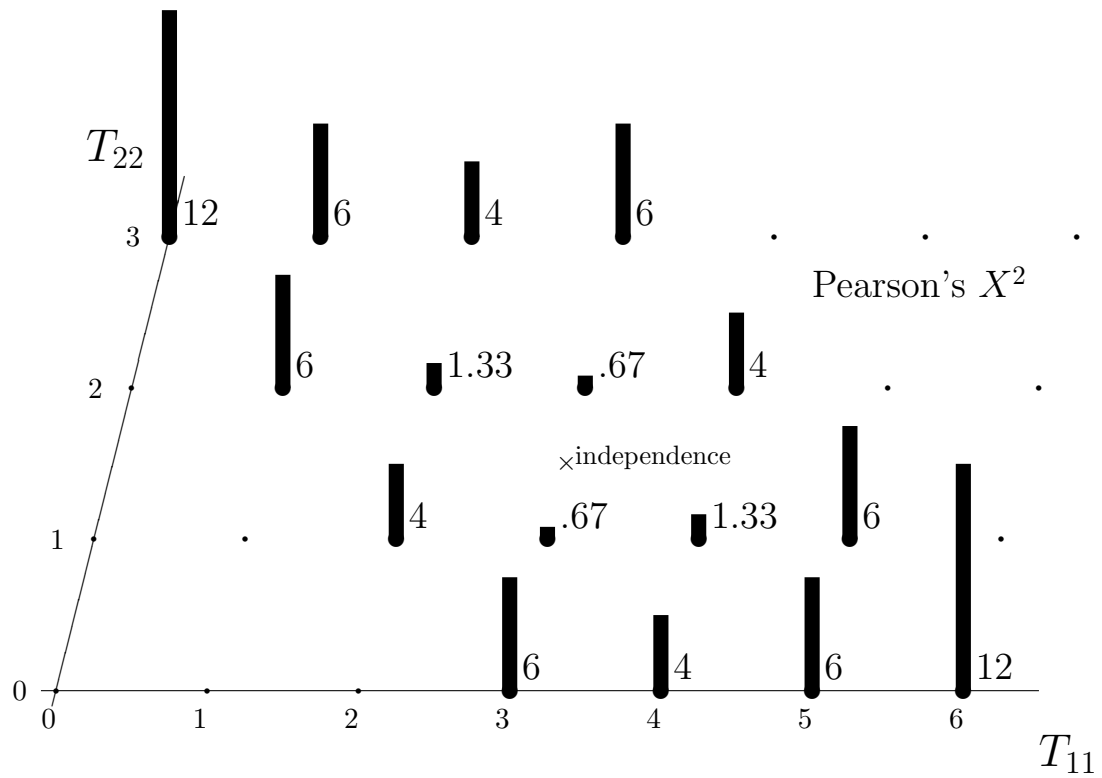
X^2	probability	p-value	$\Pr(\chi^2_2 > X^2)$
.67	.390	1	.717
1.33	.292	.610	.513
4	.195	.318	.135
6	.121	.123	.050
12	.002	.002	.002

The main problem in computing the exact significance level of a test statistic is the efficient enumeration of the isomarginal family of the observed table. How this is done, will not be discussed here; details and references are given in Verbeek and Kroonenberg (1985). Let us just suppose that it can be done. Then, to obtain the exact level of significance p of the statistic S we simply must do the following.

- Set $p = 0$.
- Compute the observed value of S : S_{obs} .
- Generate all members of the isomarginal family, and for each member:
 - . Compute the probability, p_{current} .
 - . Compute for the value of the test statistic S , S_{current} .
 - . Accumulate the probabilities of the tables more extreme than the observed table:
 - IF $(S_{\text{current}} \geq S_{\text{obs}})$, $p = p + p_{\text{current}}$
- Print p .

A more sophisticated and faster algorithm for computing exact significances is given in Mehta & Patel (1983), and it is contained in their program StatXact. For comparison with our algorithm see Verbeek & Kroonenberg (1985).

Figure 2: 3D Plots of Pearson's X^2 and the hypergeometric probabilities in the isomarginal family of Table 4.



The exact size of the isomarginal family can only be calculated by a complete enumeration. A simple, rough approximation is provided by Gail & Mantel (1977). FISHER prints this approximation before it embarks on the enumeration, so in interactive use the user has some idea how long the computation will take.

4.2 Monte Carlo estimation

For ‘small’ tables and ‘small’ sample sizes, complete enumeration of the isomarginal family can be done in a reasonable time. For ‘large’ sample sizes, asymptotic methods give good approximations. But for large sample sizes with a few small marginal frequencies, and for large tables with sample that are not really very large, both methods may fail. Monte Carlo estimation is a third possibility. One simply draws m random tables from the generalized hypergeometric distribution; for each table one computes the statistic, and one estimates the p-value by the fraction of tables for which the statistic is equal to or larger than the observed value. The execution time is proportional to the Monte Carlo sample size m , and about proportional to the number of cells in the table, but it varies only little with the number of observations. The accuracy of the estimate is known and proportional to $1/\sqrt{m}$. This is a great advantage over the asymptotic methods, of which the accuracy is unknown.

Here is a sketch algorithm that uses Monte Carlo estimation of the p-value.

- Monte Carlo sample size $:= m$ (initial value: 2000)
- $k := 0$
- Repeat m times:
 - . Generate a random table from the hypergeometric distribution, with the given margins (Patefield, 1981).
 - . Compute the statistic S_{current} for this table.
 - . Compute the number k of tables for which the value of the statistic is greater than or equal to the observed value.
 - . IF $(S_{\text{current}} \geq S_{\text{obs}})$, $k = k + 1$
- estimated p-value is k/m

The confidence interval for this estimate: $k/m \pm 1.96 \cdot$ the ‘asymptotic standard error’, where the latter is defined as

$$\frac{|1 - 2p| + \sqrt{p(1 - p)m}}{m + 4}.$$

There is one important refinement. Rather than drawing tables at random from the null distribution, one may draw tables from any other distribution, and correct the estimated p-value by weighting. One may choose another distribution to reduce the variance of the estimator (‘importance sampling’) and/or to reduce the computational complexity. In the current version of FISHER, however, we simply draw from the hypergeometric distribution. A motivation for the asymptotic standard error formula and some other technical details are presented in Appendix B.3.

5 Test statistics

In this section we give the definition formulas and references for statistics incorporated in FISHER. In computations (by hand or machine) other, mathematically equivalent formulas may be more convenient. Whenever possible, we will refer the reader for details to some easily accessible textbooks, mostly in addition to the original references. One may also consult the BMDP manual: Dixon & Brown (1977, 1981; Appendix A5), or the SPSS Algorithm Manual: Norušis (1979, p. 12). A very nice unification of many χ^2 statistics, including X^2 , G^2 , and a simplified version of $F - T$ is provided by Cressie & Read (1984), and in their book (Read & Cressie, 1988). They also give a very informative discussion of small sample properties.

In brackets we give the mnemonic used in the output of FISHER.

5.1 Pearson's X^2 (X2)

Pearson's X^2 is also known as 'the chi-square statistic'. This name is less appropriate however, because the null distribution of X^2 is only asymptotically χ^2 . Moreover also G^2 , $F-T$, and $-2\ln(EPT) + \text{constant}$, which are mentioned below, and many other statistics are asymptotically χ^2 distributed.

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

References: Pearson (1900); Siegel (1956, pp. 104 – 111, 175 – 179); Fienberg (1980, p.40); Hays (1988, sections 18.3 – 18.7).

5.2 The log likelihood ratio G^2 (LR)

$$G^2 = 2 \sum_{\text{all cells}} \text{obs} \ln(\text{obs}/\text{exp})$$

References: Wilks (1938); Fienberg (1980, p. 40).

5.3 The Freeman-Tukey test statistic $F-T$ (F-T)

$$F-T = \sum_{\text{all cells}} \left(\sqrt{\text{obs}} + \sqrt{\text{obs} + 1} - \sqrt{4 * \text{exp} + 1} \right)^2$$

Reference: Freeman & Tukey (1950).

5.4 The exact probability test statistic (EPT)

In the exact probability test, the hypergeometric probability of the observed table, Pr from formula (2), is used as test statistic. Under the null hypothesis, the transformation $-2\ln \text{Pr} + \text{constant}$ is asymptotically χ^2 distributed. Some authors tacitly assume that this is the only generalization of Fisher's exact test for a 2×2 table.

This assumption is derived from the common use of the exact probability in the computation of the exact significance in a 2×2 table. However, this is only done for one-sided tests within the proper tail, and in that situation all test statistics are equivalent. So the choice of the probability as test statistic is only a matter of convenience. In tables with more degrees of freedom, the various test statistics are in general not equivalent, and each statistic has its own exact distribution. No statistic can validly claim to be the unique, natural extension of Fisher's exact test for 2×2 tables. We regard the exact distribution of any statistic as a valid generalization. *References:* Borel (1924); Freeman & Halton (1951); Mood, Graybill, & Boes (1974, section 9.5.4).

5.5 Kruskal-Wallis' test statistic H (K-W)

One-way analysis of variance by ranks. Groups are determined by the column variable Y .

$$H = \frac{\text{between sum of squares for midranks}}{\text{total sum of squares for midranks}} = \frac{\sum_j t_{+j}(\overline{R^X_j} - \overline{r^X})^2}{\sum_i t_{i+}(r_i^X - \overline{r^X})^2}$$

where r_i^X is the midrank of row i , $\overline{R^X_j}$ is the average row midrank of the j -th column $= \sum_i T_{ij} R_i^X / t_{+j}$, and $\overline{r^X}$ is the overall average midrank $= \sum_i t_{i+} r_i^X / n$. For midranks see section 5.13 below. Note that $\overline{R^X_j}$ is stochastic (and hence denoted by capitals) because it depends on T_{ij} . The form given here is such that it is comparable to the next statistic η^2 : if $x_i = r_i^X$ for all i , the two are equal.

References: Kruskal & Wallis (1952a, 1952b, 1953); Siegel (1956, pp. 184 - 193); Hays (1988, section 18.7).

5.6 The correlation ratio η^2 , (ETA2)

One-way analysis of variance. Groups are determined by the column variable Y .

$$\eta^2 = \frac{\text{between sum of squares for scores}}{\text{total sum of squares for scores}} = \frac{\sum_j t_{+j}(\overline{X_j} - \overline{x})^2}{\sum_i t_{i+}(x_i - \overline{x})^2}$$

Where x_i is the value (or score) for the i -th category of X , and $\overline{X_j}$ is average value of X in the j -th column $= \sum_i T_{ij} x_i / t_{+j}$, and \overline{x} is the overall average value of X , $= \sum_i t_{i+} x_i / n$.

Recall that random variables are denoted by capitals and constants (such as the margins) by small letters.

References: Kendall & Stuart (1973, section 26.21); Hays (1988, section 10.17).

5.7 Kendall's τ_b (TAU)

The three statistics τ_a , τ_b , and τ_c all have the same numerator. If

C = # of concordant pairs of observations, and
 D = # of discordant pairs of observations,

this numerator is $C - D$. Their denominators only depend on the (fixed) margins of the tables. Hence these three statistics are equivalent as test statistics: they lead to the same rejection regions with the same rejection probabilities, and hence to the same p-values. We have chosen to print τ_b in the output.

The same holds for **Kruskal-Goodman's** $\gamma = (C - D)/(C + D)$. After all, given the margins, both τ_b and γ are easily seen to be strictly increasing in C/D , and hence in each other. Some programs use different (but asymptotically equivalent) approximations to calculate their approximate p-values from the χ^2 distribution, and this may lead to (slightly) different outcomes! Some programs give a p-value for only one of these four statistics (without telling why).

$$\begin{aligned}\tau_b &= \frac{C - D}{\sqrt{\# \text{ of pairs not tied on } X} \sqrt{\# \text{ of pairs not tied on } Y}} \\ &= \frac{\sum_{i < i', j < j'} T_{ij} T_{i'j'} - \sum_{i < i', j > j'} T_{ij} T_{i'j'}}{\sqrt{\frac{1}{2}n(n-1) - \sum_i \frac{1}{2}t_{i+}(t_{i+} - 1)} \sqrt{\frac{1}{2}n(n-1) - \sum_j \frac{1}{2}t_{+j}(t_{+j} - 1)}}\end{aligned}$$

Here a ‘pair of observations’ is one of the $\frac{1}{2}n(n-1)$ unordered pairs of observations $\{(X, Y), (X', Y')\}$; such a pair is called concordant if $(X - X')$ and $(Y - Y')$ have the same sign, discordant if these two signs differ, and tied if at least one of $(X - X')$ and $(Y - Y')$ is zero.

References: Kendall (1948, 1975, section 3.4); Siegel (1956, pp. 213 – 223); Goodman & Kruskal (1979); Hays (1988, sections 19.11 – 19.13).

5.8 Spearman's rank correlation coefficient R_S (RS)

$$R_S = \frac{\text{cov}(R^X, R^Y)}{\sigma_{R^X} \sigma_{R^Y}} = \frac{\sum_{ij} T_{ij} (r_i^X - \bar{r}^X)(r_i^Y - \bar{r}^Y)}{\sqrt{\sum_i t_{i+} (r_i^X - \bar{r}^X)^2} \sqrt{\sum_j t_{+j} (r_j^Y - \bar{r}^Y)^2}},$$

where r_i^X and r_j^Y are midranks (see section 5.13 below).

References: Hays (1988, sections 19.10); Kendall (1948, 1975, section 3.4); Siegel (1956, pp. 213 – 223); Spearman (1902).

5.9 Pearson's correlation coefficient R (R)

$$R = \frac{\text{covar}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{ij} T_{ij} (x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_i t_{i+} (x_i - \bar{x})^2} \sqrt{\sum_j t_{+j} (y_j - \bar{y})^2}}.$$

where x_i is the value (or score) for the i -th category of X , and \bar{x} is the overall average value of X , and similarly for y_j and \bar{y} . If $x_i = r_i^X$ and $y_j = r_j^Y$ for all i, j , $R_S = R$.

References: Hays (1988, chapter 14); Kendall & Stuart (1973, Vol. II, section 31.19: The permutation distribution of R); Pearson (1920).

5.10 Tests for two columns = two groups

When $c = 2$, that is, the number of columns is two:

- Kruskal-Wallis' $H = \text{Mann-Whitney-Wilcoxon (MWW)}$, and is equivalent to R_S^2 , $(n - 1)H = z^2$, where z is the (asymptotically standard normal) z -score of MWW.
- $\eta^2 = R^2$, and
- R is equivalent to the t-test for two independent samples.

5.11 Which test statistic?

Table 5: Choosing a statistic based on the alternative hypothesis.

Alternative hypothesis	Statistic	Mnemonic	Comment
1. General, X and Y categorical	X^2	X2	Modelling with many degrees of freedom lowers the power of the test substantially.
2. X ordered; columns are groups; one-way anova; groups differ in means (of scores or rankscores)	H η^2	K-W ETA2	if values of X are midranks. if values of X are scores at interval level.
3. Both variables are at least ordered, and higher values on X tend to go together with higher values on Y (or with lower values of Y)	τ_b R_S R	TAU RS R	ordinal association. ordinal association. linear association: X and Y at interval level, and the alternative assumes linear dependence.

The choice of a test statistic depends on the type of alternative hypothesis one has in mind. Table 5 gives an overview of the relations between the statistics available in FISHER and the corresponding alternative hypotheses. Lehmann (1957, Ch. 7.13, p.306; 1986, Ch. 8.8, p. 480) stated

“It is both a strength and a weakness of the χ^2 test . . . that its asymptotic power depends only on the weighted sum of squared deviations, not on the signs of these deviations and their distribution over [the cells]. This is an advantage if no knowledge is available concerning the alternatives since the test then provides equal protection against all alternatives that are equally distant from [the null hypothesis]. However, frequently one does know the type of deviations to be expected if the hypothesis is not true, and in such cases the test can be modified so as to increase its asymptotic power against the alternatives of interest by concentrating it on these alternatives.”

A lapse made frequently by many users of statistics is to give insufficient thought to the proper choice of the alternative hypothesis. Strictly speaking, choosing an alternative hypothesis that is unnecessarily general, is not incorrect, but it has a high price: loss of power, and in many cases a very substantial loss!

In the social sciences there is a popular misunderstanding that for numerical variables one *must* use R and η^2 , and one *may not* use the nonparametric counterparts R_S , τ , and H . This really is a choice between a little more power if the distributional assumptions hold, and a lot more robustness if they fail. For R the distributional assumption is binormality of the distribution of X and Y , and for η^2 it is normality in each column plus the same variance in each column. As we are dealing with tables, these assumptions are violated anyway. Little is known about how much this affects the power.

Figure 3 gives a graphical display of the values X^2 , η^2 , τ_c , and R in the isomarginal family of a 2×3 table. Note that X^2 and η^2 are second degree polynomials in the observed values, while R is linear. In tables with 2 degrees of freedom also Kendall's τ is linear, but in larger tables τ is quadratic. We have not plotted H and R_S because they can be seen as special cases of η^2 and R , respectively.

5.12 Scores for numerical variables

Note that the correlation ratio η^2 requires X to be a numerical variable, and the correlation coefficient R requires both variables to be numerical. For these cases the values (= scores) of X and Y may be supplied by the user. These values are denoted by x_i for the i -th row and y_j for the j -th column. Their default values are $1, 2, 3, \dots$. We assume that the values are strictly increasing:

$$x_1 < x_2 < \dots < x_r, \text{ and } y_1 < y_2 < \dots < y_c$$

.
FISHER tests this assumption.

5.13 Midranks for ordinal variables

The nonparametric counterparts of η^2 and R are Kruskal-Wallis' H and Spearman's R_S , respectively. These are based on ranks rather than on scores. Because we are dealing with tables, many ranks are tied. To resolve the problem of ranking ties, we use midranks.

The **midrank** r_i^X is defined as the average of the ranks of the observations in that row. Similarly we define the midranks r_i^Y of Y . See the Table 6 on page 18.

Table 6: The calculation of midranks.

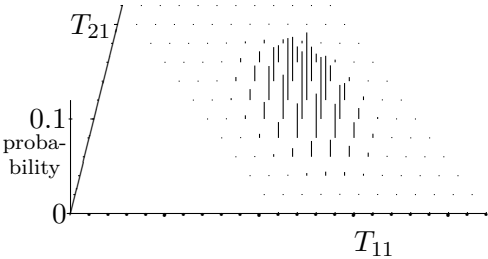
table	row marginal	row ranks	midranks of X and Y
$t_{11}, t_{12}, \dots, t_{1c}$	t_{1+}	$1, 2, \dots, t_{1+}$	$r_1^X = \frac{1}{2}(1 + t_{1+})$
$t_{21}, t_{22}, \dots, t_{2c}$	t_{2+}	$t_{1+} + 1, \dots, t_{1+} + t_{2+}$	$r_2^X = \frac{1}{2}t_{1+} + (1 + t_{2+})$
\vdots	\vdots	\vdots	\vdots
$t_{r1}, t_{r2}, \dots, t_{rc}$	t_{r+}	$n - t_{r+} + 1, \dots, n$	$r_r^X = n - \frac{1}{2}t_{r+} + \frac{1}{2}$
$t_{+1}, t_{+2}, \dots, t_{+c}$	n		$r_1^Y = \frac{1}{2}(1 + t_{+1})$
			\dots
			$r_c^Y = n - \frac{1}{2}t_{+c} + \frac{1}{2}$

An example

table	row marginal	row ranks	midranks of X and Y
2 3 1	6	1,2,3,4,5,6	$r_1^X = 3\frac{1}{2}$
1 4 0	5	7,8,9,10,11	$r_2^X = 9$
3 7 1	$n = 11$		$r_1^Y = 2, r_2^Y = 7, r_3^Y = 11$

Figure 3: 3D Plots of p , X^2 , η^2 , τ , and R in the isomarginal family with margins 22, 22 and 22, 11, 11. Note the very different scales on the vertical axes. Also note that τ and R are negative in the left half of their graph.

T_{11}	.	22
T_{21}	.	11
.	.	11
22	22	44



Note that Kruskal-Wallis coincides with the correlation ratio if $x_i = r_i^X$, and that Spearman coincides with the correlation coefficient if $x_i = r_i^X$ and $y_j = r_j^Y$ (cf. their definitions). Hence generally it seems useful to compare Kruskal-Wallis with the correlation ratio and to compare Kendall's τ_b , Spearman's rank correlation and the ordinary correlation coefficient.

Finally, recall that Kendall's τ_b only depends on the order of rows and columns.

6 Large sample tests

6.1 Limit distributions

If the expected values are not too small, asymptotic limit distributions can be used to approximate the exact distributions. What is 'small', is still a matter of investigation. Some papers with results for X^2 are Agresti & Wackerly (1977), Larntz (1978), Agresti, Wackerly, & Boyett (1979), Koehler & Larntz (1980), Kroonenberg & Verbeek (1980), Cressie & Read (1984), and Read & Cressie (1988). Klotz & Teng (1977) present some results for Kruskal & Wallis' H .

In 7 we present the asymptotic limit distributions under the null hypothesis of the statistics calculated in FISHER. Some of the statistics have to be transformed before the approximating distribution can be used.

Table 7: Asymptotic approximations to the null distributions of suitable transformations of the statistics.

Stat.	Mnem.	Trans- formation	Limit distr.	Degrees of freedom	References
X^2	X2		χ^2	$(r-1)(c-1)$	Mood et al. (1974).
$\frac{H}{\eta^2}$	K-W ETA2	$\frac{(n-1)H}{(n-1)\eta^2}$	$\frac{\chi^2}{\chi^2}$	$\frac{c-1}{c-1}$	Lehman (1975) pag. 396 ex 3.
τ_b	TAU	τ_b/σ_{τ_b}	standard normal	—	Kendall (1975) sect. 5.6 & 4.9.
R_S	RS	$R_S \sqrt{\frac{n-2}{1-R_S^2}}$	Student's t	$n-2$	Kendall & Stuart (1973) sect. 31.19:
R	R	$R \sqrt{\frac{n-2}{1-R^2}}$	Student's t	$n-2$	The permutation distribution of R.

6.2 Should most expected values be larger than about 1?

If one is testing against a general alternative (for example with X^2 , G^2 , F - T , or EPT), it seems rather natural to require that almost all expected values are larger than about 1. After all, cells with an expected value less than about 1 only give information with respect to the null hypothesis, if the observed value is very large. If most expected and observed values are around 1, there is hardly any information with respect to the null hypothesis, and the power of any test will be very small. In such a case, testing does not make much sense. Typically the mentioned test statistics are not equivalent; remember that they are only asymptotically equivalent. Here, they order the tables in different ways, see for example Table 2 on page 8. Thus an arbitrary or unjustified choice of test statistic does influence the significance. And we know of no way to justify any choice.

Another way to put this is the following. If an expected value is much smaller than 1, the observed value is zero with a high probability ($\Pr(\text{obs}=0) \approx \exp(-\text{expected value})$). Hence, we have hardly any (degree of) freedom in this cell. For such a case, Cochran proposed to approximate the distribution by a product of a multinormal distribution (for cells with sizeable expected values) and Poisson distributions (for cells with small expected values).

Of course, for tests against more specific alternatives, such restrictions are not needed. Implemented in FISHER are Anova's correlation ratio η^2 , Kruskal-Wallis' H , Kendall's τ_s , and Spearman's and Kendall's correlation coefficients R_S and R . These can also be used for truly continuous variables. (For η^2 and H the grouping variable, Y , should be discrete, of course.) Suppose both X and Y are truly continuous, and the observed values would be tabulated in a 'contingency table'. Then every marginal frequency would be 1 (almost surely), and every expected frequency is $1/n$, which is very small indeed. Nevertheless such a test can be meaningful and powerful.

7 Small sample tests

In this section we discuss three issues. Firstly, the calculation of the probability of the observed value itself, $\Pr(S = S_{\text{obs}})$, can pose numerical problems if the statistic S is real-valued rather than integer-valued. (Rescaling may solve this problem. For example, τ can be transformed linearly into an integer valued statistic.) The next two subsections are devoted to subtleties of one-sided and two-sided small sample tests based on τ_b , R_S , or R . The problem of conditioning or not is a small sample problem as well. It fades when the sample size becomes large. However, we will not discuss this problem here (see, however, Appendix B.2).

7.1 The calculation of the probability mass at the observed value

As a test statistic R_S is easily seen to be equivalent with $\sum_{ij} 4(R_i^X)(R_j^Y)T_{ij}$. Note that $2R_i^X$, $2R_j^Y$ and T_{ij} are integers, so the latter statistic is integer-valued. Similarly τ can easily be transformed linearly into an integer-valued statistic. For the other statistics however, a simple, monotonic transformation to an integer-valued statistic (which must stay clear from integer overflow), is not easily available. For a real-valued statistic S the problem to compute $\Pr(S = S_{\text{obs}})$ is **ill-conditioned**. Consider two tables T_1 and T_2 , for which S has the values S_1 and S_2 . Using approximate, floating point arithmetic it is very difficult to ascertain whether S_1 and S_2 are exactly equal. In floating point arithmetic S is only approximated. There are roundoff errors, and we may make type I errors as well as type II errors: We may think that S_1 and S_2 are different when the difference is only due to roundoff errors, or we may think they are mathematically equal while they are not.

For exact testing this has the following consequence. When we carry out the algorithm sketched in Section 4, and we accumulate the hypergeometric probabilities of all tables with $S \geq S_{\text{obs}}$, we again may make both type of errors. We may include terms in the sum that do not belong there, and we may omit terms that do belong there. Note that we are not dealing with an error in the last few digits; *we are dealing with an error caused by summing over the wrong index set!* That is, a certain table from the isomarginal family *is* included in the critical region on one machine but not on another, only due to very small differences in rounding. Even with the same FORTRAN program, two different machines or two different compilers on the same machine, or even different options of the same compiler may occasionally classify tables differently with respect to the question: Is $S \geq S_{\text{obs}}$? **Thus on different machines the same program may occasionally produce rather widely different answers for the same significance!**

Hence it is very hard to determine numerically the significance $\Pr(S \geq S_{\text{obs}})$ and the size of the ‘probability mass’ $\Pr(S = S_{\text{obs}})$ precisely. Also it is not easy to estimate the possible error sharply. FISHER handles this problem as follows. It computes

$$\begin{aligned} \text{significance} &=_{\text{def}} \Pr(S \geq S_{\text{obs}} - \varepsilon), \text{ and} \\ \text{probability mass} &=_{\text{def}} \Pr(S_{\text{obs}} - \varepsilon \leq S < S_{\text{obs}} + \varepsilon). \end{aligned}$$

Here ‘ S ’ and ‘ S_{obs} ’ should be interpreted as their rounded machine representations. The FORTRAN text contains expressions like:

```
IF (S.GE.(SOBS-EPS2)) ....
```

Commonly $\varepsilon = .0005$, except for input mode ‘DATA M’ (input of margins and statistics). Then ε is determined by the number of digits specified in the entered observed value of each statistic. If one enters $X2=3.84$, $\varepsilon = .01$. If one enters $X2=3.8400$, $\varepsilon = .0001$. Now, **different results produced by different machines do not differ more than the probability mass**. Actually they will differ less, but it is hard to give an efficient algorithm for a better upper bound.

The implicit assumption is that two S values are likely to be equal if their absolute difference is smaller than ε . If in reality two values are different, then the significance indicated is too large (that is, too insignificant), but at most the size of the probability mass. This algorithm may be conservative. If a borderline case with a large probability mass occurs – and an extensive analysis is justified – one may print the entire distribution to check the members with an S close to S_{obs} . However, discriminating between ‘ S_{obs} ’ and, say, ‘ $S_{\text{obs}} - \varepsilon$ ’ almost invariably presupposes a precision not available in the data.

We feel, that if one is careful enough to compute the exact distribution rather than its asymptotic approximation, one should also pay tribute to the inherent discreteness of the exact distribution by glancing at the probability mass.

Other approaches to this problem are:

- Ignore the problem altogether.
- Don’t mention it to the user, but subtract some small $\varepsilon > 0$ from the observed value(s). This guarantees that the significance reported is not too small (that is, not exaggerated).
- Apply exact, rational arithmetic (working with pairs of integers without integer overflow) rather than real arithmetic.

Admittedly the first two approaches are more user friendly in the sense that they arouse less discomfort, and in many situations they are quite sufficient. However, this approach has the drawback, that **one has no information about the possible error in the computation of the significance**. Occasionally, this can be serious, because the computed significances can be wrong, in a machine dependent way, by much more than just the last few digits.

In theory the third approach seems applicable to many statistics, since many are rational, or can be transformed monotonically into a rational-valued statistic. We have not seriously tried this, because the efficiency of such calculations seems rather low, and the programming task rather involved. We do not think that today’s machinery is fast enough to make this successful.

7.2 One-sided tests

Tests of independence based on τ_b , R_S , or R can be against a two-sided alternative or against a one-sided alternative. This also holds for tests of independence in 2×2 tables. In general, we can model this situation as follows. Let S be a statistic

defined on $r \times c$ tables that can be positive and negative, and that is zero under exact independence. Assume that S is the sample equivalent of a parameter ϑ , and assume that independence implies $\vartheta = 0$. In general there will be nuisance parameters as well, and then $\vartheta = 0$ does not imply independence. For example, zero correlation includes more cases than independence, unless we restrict ourselves to a suitable one-dimensional family of distributions. Note, that our null hypothesis remains ‘independence’, and not $\vartheta = 0$. In general, a critical region of size 5% under independence has a much larger size under the set of distributions with $\vartheta = 0$.

For a one-sided alternative there is little disagreement about the definition of significance. For the right-sided alternative $\vartheta > 0$:

$$\text{significance} = \Pr(S \geq S_{\text{obs}}),$$

and for a left-sided alternative $\vartheta < 0$:

$$\text{significance} = \Pr(S \leq S_{\text{obs}}).$$

Note that these two probabilities add up to

$$1 + \Pr(S = S_{\text{obs}}). \tag{3}$$

As explained above in section 7.1, $\Pr(S = S_{\text{obs}})$ is hard to determine for a real valued statistic like R . In FISHER it is estimated by the ‘probability mass’, which is guaranteed to be equal $\Pr(S = S_{\text{obs}})$ or too large. Hence in FISHER the left and right p-values add up to:

$$1 + \text{probability mass}.$$

7.3 Two-sided tests

For two-sided alternatives there are various methods to distribute the critical region over both tails of the distribution of S . Table 2 on page 8 gives a typical example of a 2×2 table in which the critical regions of X^2 and G^2 differ.

Each of the following three, heuristic methods leads to a nested family of critical regions, but in general each method interlaces the left tail and right tail differently. The significance or p-value of an observation is defined as the size of the smallest critical region containing the observation. Now we are going to define a two-sided critical region of size α . Due to discreteness one can only do this exactly for some values of α .

- Tails of equal size: take a left-tail critical region and a right-tail critical region, each of size $\alpha/2$. Due to discreteness, this usually can only be done approximately. A ‘best’ approximation need not be defined uniquely.
- Take S^2 as test statistic: make critical regions of the form ‘ $S \leq -c$ or $S \geq c$ ’. These regions are characterized by being symmetric. A problem with this approach is that, for example, X^2 and G^2 are equivalent as one-sided test statistics in a 2×2 table, but not as two-sided test statistics; they may interlace both tails differently (as in Table 2).

- 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 unbiased, or at least locally unbiased. For discrete distributions this usually requires randomization, but it is rather obvious (and unique) how to define 'nearest' non-randomized tests.

As a consequence there is no unanimously accepted definition of the 'two-sided p-value'.

To avoid the ambiguity, FISHER only gives the left-sided and right-sided p-values of τ_b , R_S , and R . If one wants two-sided p-values, one can do the following.

Case (i) Just double the p-value. However, most likely there is no unrandomized critical region exactly of this size. One can also study the distribution of S , using the command **DISTR**.

Case (ii) Run the analysis twice, once with the observed table, and once with the opposite value of the statistic of interest. To this end, use **DATA M** to reenter the margins, and minus the observed value of the statistic.

Case (iii) If you know c_L and c_R , run the analysis twice, using **DATA M**, as for case (ii) above. Otherwise, use **DISTR** and sort the file output by FISHER with the values of S as sort key, to obtain the complete distribution of S . (For details see subsection 11.3 in Part III.)

8 Criterion robustness versus inference robustness

For a general discussion of criterion robustness and inference robustness see the original paper of Box & Tiao (1964) or Chapter 3 of their book Box & Tiao (1973).

The program FISHER arose out of a concern for the accuracy of using the χ^2 distribution for the null distribution of Pearson's X^2 and similar statistics, when testing for independence against an arbitrary alternative. FISHER solves the question of *inference robustness* by using the null distributions of the test statistics themselves, and avoiding the use of an approximation.

The question of *criterion robustness*, however, is still open. That is:

Is the statistic used the proper one?

Let us confine ourselves to the general alternative: no independence. Then the question is whether Pearson's X^2 is a proper test statistic when the samples are small. Almost all justifications for Pearson's X^2 known to us rely on large sample arguments. However, most alternatives of X^2 , like G^2 , Freeman-Tukey's statistic $F-T$, and the H_o probability of the observed table are asymptotically equivalent to X^2 as test statistics. Hence, all justifications for X^2 apply to them just as well, and they do not discriminate between them at all. A property that makes G^2 desirable in some cases, is its additivity, when three or more nested models are tested. This possibility to decompose an overall G^2 into components, just like a total sum of squares in Anova, helps its interpretation.

Note, furthermore, that authors proposing corrections to Pearson's X^2 , for example in the context of 2×2 tables seem to aim for inference robustness. However, sometimes they create a non-equivalent test statistic, that interlaces the right and left tail differently than X^2 . By modifying X^2 so that its distribution agrees as closely as possible to the χ^2 distribution, these authors implicitly assume that the χ^2 distribution is the proper criterion. Remarkably, this assumption is often implicit. We think that its only justification is an appeal to convenience.

9 More on Pearson's X^2

9.1 X^2 terms and residuals

In most cases significance testing should be considered as a “matter of hygiene”, performed without much ado to avoid the disgrace of discussing an effect that can be attributed entirely to random fluctuations. This refreshing analogy is borrowed from Ehrenberg (1975, p. 323). Once significance has been established, one proceeds to the heart of the matter: An investigation and discussion of the structure and the size of the effects (here: dependencies). Unfortunately many users of statistics stop after significance testing. If dependency is established, one should look next at the residuals. These (raw) residuals are defined as

$$\text{residuals} = \text{obs} - \text{exp}.$$

Examining the residuals is essential for building or refining any model. FISHER offers three options for the residuals:

- Raw residuals (see above).
- X^2 terms = (standardized residual)² = $\frac{\text{residual}^2}{\text{exp}} = \frac{(T_{ij} - t_{i+}t_{+j}/n)^2}{t_{i+}t_{+j}/n}$
- Adjusted residuals = $\frac{\text{residual}}{\text{est.st.dev.}} = \frac{T_{ij} - t_{i+}t_{+j}/n}{\sqrt{t_{i+}t_{+j}(1 - t_{i+}/n)(1 - t_{+j}/n)/n}}$

The X^2 -term is the contribution of a cell to Pearson's X^2 , and it is the square of the standardized residual.

Note that the absolute value of a standardized residual of a cell will always be smaller than its adjusted residual. The difference will be negligible when t_{+j}/n and t_{i+}/n are small. The adjusted residuals have the advantage that under H_0 they have asymptotically standard normal distributions.

For a visual inspection, (re)arrange rows and columns so that the most likely interaction would show a simple pattern. If you have no idea about the possible pattern of interaction, and both X and Y are unordered, rearrange X and Y so that the marginal frequencies are weakly increasing. Then the expected value increase more or less smoothly from the smallest value in the cell $(1, 1)$ to the largest value in cell (r, c) . When the original table is ordered this way, cells with exceptionally large residuals stand out more clearly, and when the residuals are ordered this way, they sometimes show a simple pattern, corresponding to Tukey's ‘One degree of freedom for non-additivity’.

References: Everitt (1977, section 3.4.3), Haberman (1973), Tukey(1949)

9.2 Collapsing and grouping

By *grouping* a continuous variable we mean, recoding it into a discrete variable. By *collapsing* a discrete variable we mean, recoding it into a variable with fewer categories. By collapsing an $r \times c$ table we mean, getting a smaller table by collapsing one or both variables. Hence ‘grouping’ is a kind of ‘collapsing continuous variables into discrete ones’. In Table 8 we illustrate the old rule, that one may not collapse a table *in a way that is based on the data*, and then test as if the collapsed table is the only table we have seen. A fortiori this also holds for grouping. However, the following result shows that this problem can be circumvented by comparing X^2 to the χ^2 distribution with as many df as the original table.

Lemma. Collapsing a table cannot increase X^2 . Collapsing two columns or rows leaves X^2 (nearly) invariant, if and only if the conditional distributions in these two columns or rows are (nearly) the same, that is, the frequencies are (nearly) proportional.

Proof. This only has to be verified for collapsing two columns. For rows, the proof is similar, and for arbitrary collapsing one simply repeats the argument.

First, we write $X^2 = -N + \sum \text{obs}^2 / \text{exp}$. Now suppose we collapse columns j and j' . For brevity write $o_1 = \text{obs}_{ij}, o_2 = \text{obs}_{ij'}, e_1 = \text{exp}_{ij}, e_2 = \text{exp}_{ij'}$. Consider what happens to the X^2 terms of cells ij and ij' . By collapsing their joint contribution changes by $-o_1^2/e_1 - o_2^2/e_2 + (o_1 + o_2)^2 / (e_1 + e_2)$, which equals $-(o_1/e_2 - o_2/e_1)^2 / (1/e_1 + 1/e_2)$. Hence this change is nonpositive. It is (nearly) zero if and only if o_1/e_1 is (nearly) equal to o_2/e_2 .

Corollary. If the X^2 of a collapsed table is significant when compared to the χ^2 distributed with the number of df of the *uncollapsed* table, the uncollapsed X^2 is significant as well.

Proof. By the lemma, the uncollapsed X^2 is at least as large as the collapsed X^2 .

It is well known that significance of Pearson's X^2 in a collapsed table does not imply significance in the larger table. The reason is that there are so many more df in the larger table. If one looks at an uncollapsed table, and then collapses over a few rows or columns where the conditional distribution is about the same, one loses little in X^2 and one kills a lot of df. Hence, the result gets more significant. But this is a kind of data mining like ‘looking for the largest contrast (or: the largest correlation coefficient in a correlation matrix; or: the most significant predictor in the next step of stepwise regression) and then testing it, *as if it were the only test*’. This activity is wrong, illegal, sinful, and abundant in statistics. If one bases the choice of the test on the data, looking for big effects, and tests as if it were the only test under consideration, one gets a p-value that is too small; often it is very much too small.

Hence grouping or collapsing based on the observed values can be dangerous. It is not dangerous if it is based on the interpretations of the categories, rather than on

Table 8: Example of collapsing. In each table $X^2 = 8$. Note how much p varies. Collapsing may not be based on the data, only on theory, i.e. on the formulation of the null hypothesis and the alternative hypothesis

	a	b	c	d	e	f	g	h		p_{exact}	= 1		
I	1	1	1	1	0	0	0	0	4	p_{asympt}	= .333	X^2	Pr
II	0	0	0	0	1	1	1	1	4	df	= 7	8	1
	1	1	1	1	1	1	1	1	8	isom.fam.	= 70		
	abcd	e	f	e	g					p_{exact}	= .029	X^2	Pr
I	4	0	0	0	0	4				p_{asympt}	= .092(!)	4	36/70
II	0	1	1	1	1	4				df	= 5	5	32/70
	4	1	1	1	1	8				isom.fam.	= 8	8	2/70
	ab	cd	ef	eg						p_{exact}	= .086	X^2	Pr
I	2	2	0	0	4					p_{asympt}	= .05	0	16/70
II	0	0	2	2	4					df	= 3	4	48/70
	2	2	2	2	8					isom.fam.	= 19	8	6/70
	abcd	efgh								p_{exact}	= .029	X^2	Pr
I	4	0	4							p_{asympt}	= .005	0	30/70
II	0	4	4							df	= 1	2	32/70
	4	4	8							isom.fam.	= 5	8	2/70

the observed frequencies. Of course, having fewer categories strongly increases the power of the test, and avoiding categories with small numbers increases the chances that Cochran's rule applies. The classic paper by Cochran (1954) deals with both aspects, and is still very much worthwhile reading.

Table 8 shows what can happen by collapsing. Each of the four tables has $X^2 = 8$; the df are 7, 5, 3, and 1, making $X^2 = 8$ range from (asymptotically) insignificant, to highly significant. Of course we also give the exact p-values, which happen to vary somewhat less. Note that in each table X^2 has its extreme value (= 8), so the probability masses are equal to the p-values. Also note that in 3 cases out of 4, the asymptotic p-value lies nicely between the exact p-value $\Pr(X^2 \geq 8)$ and $0 = \Pr(X^2 > 8)$. For curiosity sake, the sizes of the isomarginal family have been included as well.

In this case, collapsing changes the p-value very much. Which one is the right one? If we have no theory before we look at the data, we should look at the uncollapsed 2×8 table, from which there is little to be learned. If there is a theory saying that the smallest, i.e. 2×2 table, contains the information of interest, then we should only look at that table. Then we have much more power, and in this case we find a significant result. We should not collapse a table on the basis of the data, but on the basis of a subject matter theory and on the basis of the power available to

test that theory. Such a theory should specify the relevant null hypothesis and the relevant alternative. Considerations of power favour collapsing, if one can choose between various alternatives.

In the 8×2 table $p_{\text{exact}} = 1$ because all row marginals have the value 1. In that case, all tables in the isomarginal family (here: 70) have exactly the same X^2 , because they can be obtained from each other by permutations.

9.3 Quick methods

The 95-percentile of the χ^2 distributions. In many cases X^2 is so large, and so obviously significant that a blind horse can see it. Then we don't have to bring in fine tools like asymptotic or exact testing. If one needs a rough comparison with the 95-percentile of a χ^2 distribution, one can approximate the latter by

$$\chi_{\text{df}}^2[.95] \approx \text{df} + 3\sqrt{\text{df}}$$

This is an overestimate, $\Pr(\chi^2 > \text{df} + 3\sqrt{\text{df}})$ is a decreasing function of df, starting at .046 for df=1, and approaching .01 for $\text{df} \rightarrow \infty$.

Simplification. In many cases it is easy to establish significance when X^2 is really large. Accumulate a few large X^2 terms, perhaps after simplifying the table by collapsing columns and rows that are nearly proportional, or that do not contribute much to X^2 anyway. By the lemma above this gives an lower bound for X^2 . If this lower bound is larger than the 95-percentile of the χ^2 distribution with the number of df before collapsing, we have a significant X^2 , at least according to asymptotic methods. We guess, that if the *collapsed* table is a Cochran table, this conclusion is justified. More formally, this conjecture can be formulated as follows.

Let ν be the number of df before collapsing, X_0^2 and X_1^2 the value of Pearson's X^2 before and after collapsing, respectively. Suppose that after collapsing, the table satisfies Cochran's rule, and suppose $\Pr(\chi_\nu^2 > X_1^2) < .05$. Then in the uncollapsed table, Pearson's X_0^2 is significant at least at the .07 level that Cochran's rule promises.

If, after collapsing, the table is significant compared to the χ^2 distribution with the *new* df, this gives little information about the uncollapsed table (see section 9.2).

Rearranging the table. For a quick inspection of the table, it can be useful to reorder rows and columns so that marginal frequencies are nondecreasing. Then the smallest expected values are in the upper left corner, and the largest are in the lower right. Hence it is easier to pick out large X^2 . It is also easier to find the smallest expected value (now in cell (1,1)), and to count the number of cells with an expected value below 5. Finally, simple patterns in the residuals may show up after rearranging the table suitably. In any case it is worthwhile to give some thought to the 'natural' order of rows and columns in terms of the possible interaction effects.

Part II

Example of a FISHER job

10 Input

Below we give the input of a FISHER job in **teletype font**, interspersed with comments in *italics*. Thereafter the output of this job, also in **teletype font**, is discussed. Another example of a FISHER job can be executed by starting FISHER, and typing **in2 name=demo** at the first prompt.

FISHER *has online help available.*

help

Next we enter the frequencies of a 2×3 table.

data 2 3 4 10 10 12 6 6

The exact significance of Pearson's X^2 is requested as follows.

ana x2 exact

The row and column variable of the following 5×5 table are ordinal.

data 5 5 3 2 5 10 11 11 8 16 35 19 28 13 23 33 6 27 11 13 12 5 63 10 9 4 0

With such a large table, obviously with rather strong interactions, one should not embark automatically on an exact analysis. An approximate (=asymptotic) analysis is requested as follows.

ana all approx

*The keyword **all** means X^2 , η^2 , $K-W$, τ , R , and RS . Next inspect the expected values and the (adjusted) residuals.*

exres adj

*Finally, we demonstrate how the entire distribution (rather than a single significance) can be obtained. This distribution is written to a file. This option is activated, and the file name is specified by the following command. (**distr** applies to each subsequent **anal** ; it does not have to precede **data**.)*

distr name=d replace data 2 3 5 1 0 1 2 3 ana exa nostat

*With the **nostat** option, the analysis command writes the frequencies of each table in the isomarginal family to file. For each table the values of all statistics, and the hypergeometric probability are written to file by*

ana exa all

Rather than the entire distribution, one may write only a Monte Carlo sample to file. In this demonstration we choose a ridiculously small sample size.

mc 1 ana mc all distr off

This is how to stop writing distributions to file. Here, this is not really needed, because we will end our FISHER session now

end

*The keywords **quit**, **bye**, or **stop** will also work.*

Part III

Reference Manual

11 The user interface

11.1 General points

- + There is a **HELP** command.
- + There is full input checking. Diagnostic messages are preceded by ‘!!!’, potentially fatal errors by ‘***’.
- + There are internal consistency checks on exact significances.
- + All input is automatically echoed onto a log file. Thus it is easy to check what has been done, and to rerun a session, perhaps after the correction of some errors.
- + One may enter the table of observed frequencies or, alternatively, one may enter the marginals and statistics of a table.
- + There are two input files: **FISHER** starts reading from the keyboard (= primary input), but one can switch to read commands and data from a file (=secondary input). When **FISHER** reaches an end-of-file, it returns to primary input.
- + Output has two possible formats: **LONG** which is suitable for printing, and **SHORT** which usually gives all output of one analysis on a single screen.
- + There are five output channels; usually primary output is connected to the screen, and the secondary output can be used to keep a record of all output or a selection thereof. The output channels may have a **SHORT** or a **LONG** output format, independent of another. If requested, the entire distribution is written to the third output file. The fourth output file is the log file onto which all input is echoed. The fifth output file is the dump file, that is only used in case **FISHER** finds an internal inconsistency.
- + In batch mode the program prints the input table, its expected values and its residuals, after successful entry of a table. In interactive mode these are displayed upon request.
- + For residuals, one has a choice of raw residuals, Haberman adjusted residuals, or X^2 terms. In view of Cochran’s rule (see the Introduction) expected values less than 5 are labeled with a ‘#’. Haberman’s adjusted residuals are labeled similarly, if their absolute value exceeds 3.
- + Output is labeled by a time stamp on all output files, indicating the *start* of the run, or of the last **REINIT**. So all analyses from one session get the same time stamp. If you want to update the time stamp, start a new session with **REIN**.
- For very large tables, several formats may overflow in a rather untidy manner. **FISHER** has not been designed to handle such cases.
- There are no labels (for categories, tables, or data). Analyses are done so easily, that labelling is hardly worthwhile.
- There is no internal editor. For commands this is hardly necessary. Keywords are very short, and most keywords keep their values until redefined. Tables cannot be edited either. Fortunately, they are small too.
- **FISHER** can only handle frequency data in the form of an $r \times c$ table. Data in the form of a data matrix cannot be handled.

11.2 Files

FISHER can read from or write to the following seven files.

On PC, FISHER accepts input from the keyboard, and writes output to the screen. Of course, both files may be redirected, using the DOS mechanism for file redirection (> and <). For an example, see Appendix A.4.

A more convenient way to read from or write to a file, however, is provided by secondary input (the command `IN2`) and secondary output (the command `OU2`). On the secondary output file, one can set the format (`SHORT` or `LONG`) and the inclusion of messages (`MESSAGE` or `NOMESSAGES`) independently of what is displayed on screen (= primary output). When on secondary input an end-of-file is reached, input returns automatically to primary input (= keyboard).

When entire distributions are being written (the command `DISTRIB`), this is done to a special file, whose name is specified by the user. Don't forget to switch `DIS OFF` when you want to proceed with analyses that should not be written to file. FISHER automatically creates a log file or diary with all syntactically correct input of a session. On PC this file is called `LOG.FIS`. It can be renamed by the DOS command `rename`, but not within FISHER.

Finally, when FISHER detects an inconsistency (for example when the sum of all probabilities is not near enough equal to one) it generates a dump file, and says so on screen. If this should happen, please print this file, and send a copy to us. It contains all the necessary information about the current input table, all parameter settings, and the command that has caused the dump. On PC the dump file is called `DUMP.FIS`.

11.3 Writing the entire distribution to file

Rather than giving only the p-values, FISHER can also write the entire distribution to a file, or a Monte Carlo sample thereof. This option is activated with the command

```
DIS NAME = filename
```

and deactivated after one or more analyses with

```
DIS OFF
```

`DIS` is short for `DISTRIBUTION`. As with all parameters '`NAME=filename`' is optional. If omitted, FISHER takes the last file accessed by `DIS`, or if this is the first call, '`NAME=DISTR.FIS`' is assumed.

Writing the entire distribution to a file is much slower than simple enumeration, and hence it can be quite time consuming, and it can create a very big file. In general, it is recommended to do the analysis first without writing to a file, in order to learn what the size of the isomarginal family is. The file size in bytes is about 30–80 times the size of the isomarginal family.

Before the tables from the isomarginal family of an analysis are written to file, first the margins, time stamp (of the beginning of the session), and the format used are written in a header.

Not all tables are written to file, but only those tables for which the output values of the probability are positive, that is larger than $.5E - 8$. All analyses (X2, X2+, ALL, NOSTAT) write the same set of tables in exactly the same order to file.

To obtain the frequencies of each table, rather than the values of certain statistics, one must use ANAL NOSTAT, after DIS is activated.

The file written by FISHER after the DIS command is not a nicely polished product intended for visual inspection. Because the file contains so much information that can be used in so many different ways, we have provided a format that allows for easy handling by other programs. In order to obtain the cumulative distribution of a statistic this file must be edited (the header must be stripped off), sorted (for example by the DOS command SORT), and probabilities of tables with the same value of the statistic must be added (for example with a statistical package).

11.4 Numbers

Most numbers are input or output in the form of integers (like 12, 112, or 3013) or fixed point reals (like 3.84 or .0031). On input one may also use the E notation, and as FISHER is not case-sensitive 'e' may be used instead of 'E'. For example, $1000000 = 1E6 = 1e6$, and $1e-6 = 0.000001$. This can be convenient for the commands ISOM and MC. Some output may also use this E notation, especially to indicate the size of a huge isomarginal family.

Moreover, frequencies are entered in free format (that is, separated by one or more spaces), and on input a number may not take more than 10 positions.

12 FISHER command syntax

12.1 General rules

- Each command should be on one line of at most 80 characters.
- All keywords may be abbreviated to the first 3 characters. The remainder of any keyword is ignored.
- FISHER is not case-sensitive.
- After '!' any text on that line is treated as comment. That is, it is ignored. Empty lines, and lines with only comments are allowed.
- One or more spaces and/or commas and/or '=' serve as a single separator.
- All parameters are optional.
- All parameters are initialized.
- Almost all parameters retain their values until they are redefined. Hence, if you want to resubmit a command with only one parameter changed, it suffices to re-enter the command with only this single parameter. This property will be referred to as '**sticky parameters**'.

The only exceptions are the file parameters *<new?>*, *<off?>*, *<format>*, and *<message>*. In this case, you do not really want parameters that retain their old values, when you specify a (new) file name. Hence, with a filename these parameters have defaults and do not retain their previous values.

If a command contains a syntax error, its parameter values are not retained either.

- All parameters are order independent, with the obvious exception of those in **DATA** *<nr>* *<nc>*, **XVAL** *<x-values>*, and **YVAL** *<y-values>*.
- This information, the syntax of all commands, and the file settings can be inspected interactively with the **HELP** command.
- The values of all parameters can be inspected with the command **PARAM**.
- Entry of numbers is free format.
- Numbers may not be longer than 10 positions.

12.2 Command descriptions

In the command descriptions below

- *<item>* stands for a value that has to be supplied by the user, without ‘<’ and ‘>’.
- {*item*}... means that *item* may be repeated, without the { and }.
- There are two types of input lines:
 - data lines (after **DATA**), and
 - command lines.

Each command lines has the following format.

command {*parameter*}...

A command line starts with a command. Optionally, and depending on the command, it may be followed by zero, one or more parameters. Each parameter has one of the following forms:

keyword

value

or

keyword = *value*

Remember: **each parameter is optional** and each parameter that is not (re)specified retains its previous value. Moreover, all parameters are initialized. The input table, however, is not a parameter and is not initialized.

12.3 Data entry commands

DATA *<nr>* *<nc>* *<input mode>* *<residuals>*

<nr> is the number of rows. It must be the first parameter.

<nc> is the number of columns. It must be the second parameter.

<input mode> can be T, TM, or M.

<residuals> can be NORES, RES, ADJRES, or X2TERMS.

This command sets the stage for data entry. It should be followed by information of the table (frequencies, and/or marginal frequencies and/or statistics), in a format determined by *<input mode>*.

<input mode>:

T	table frequencies only
TM	table frequencies plus marginals
M	marginals plus one or more <i><statistics></i>
<i><residuals></i> :	
NORES	no residuals
RES	raw residuals
ADJRES	Haberman's adjusted residuals
X2TERMS	Terms of Pearson's X^2

For the definitions of the various residuals see section 9.1 'X²-terms and residuals'.

Initialization

```
DATA 2 2 T RES
```

Data entry

If *<input mode>*=T, the DATA command should be followed by *nr* lines, each with *nc* observed frequencies:

```
<freq11> ... <freq1,nc>
<freq21> ... <freq2,nc>
:
<freqnr,1> ... <freqnr,nc>
```

If *<input mode>*=TM, the DATA command should be followed by *nr* lines, each with *nc* observed frequencies and one row marginal, followed by a line with the column marginals and the total number of observations:

```
<freq11> ... <freq1,nc> <row marg1>
<freq21> ... <freq2,nc> <row marg2>
:
<freqnr,1> ... <freqnr,nc> <row margnr>
<col marg1> ... <col margnc> <n>
```

If *<input mode>*=M, the DATA command should be followed by two lines with observed marginal frequencies, and a third line with zero or more statistics:

```
<row marg1> ... <row margnr>
<col marg1> ... <col margnc>
{<statistic> = <value>} ...
```

<statistic> can be X2, K-W, ETA2, TAU, RS, R, LR, F-T, or p.

One may specify all, some or no statistics in any order. Even if no statistics are specified, there has to be an (empty) statistics line. For omitted statistics no output will be given, but this does not reduce the execution time. After an empty statistics line, the ANAL command is only legal with *<statistics>* = NOSTAT.

For the definitions of the various statistics see section 5 'Test statistics'.

In batch mode, successful entry of a table is automatically followed by a print-out of the table, its expected values and its residuals. For input mode M the cell

frequencies and residuals are not known. In this case, the statistics will be printed instead.

In interactive mode, the entry of frequencies can be aborted by ‘*’ in the first position of a line.

Examples

```
DATA 2 3 TM
```

```
2 3 4 9
```

```
1 0 0 1
```

```
3 3 4 10
```

!Note both margins.

```
REIN
```

```
DAT
```

!After startup or REIN, ‘DAT’ is equivalent to ‘DATA 2 2 T RES’

```
1 2
```

```
3 4
```

```
dat 7 8 m
```

```
1 2 3 4 5 6 7
```

```
1 2 3 0 4 5 6 7
```

```
ETA2=.5, K-W=.8 r=-.4
```

!Note: There is no case-sensitivity.

! One or more spaces and/or commas serve as separator.

XVALUES $\langle value_1 \rangle \langle value_2 \rangle \dots \langle value_{nc} \rangle$

YVALUES $\langle value_1 \rangle \langle value_2 \rangle \dots \langle value_{nr} \rangle$

Reassign ‘values’ or ‘scores’ to the categories of the X-variable and the Y-variable, respectively. A value may not consist of more than 10 characters. Values must be strictly increasing, and only the characters 0-9, =, -, E, e are valid.

These values remain into effect until another **XVAL**, **REIN**, or **DVAL** command, or until a table with different dimensions is given. In the latter case both X-values and Y-values are re-initialized to 1, 2, 3,

DVALUES

Reset the values of the categories of the X-variable and the Y-variable to the defaults, that is to 1, 2, 3,

XVALUES

Show the current values of the categories of the X-variable.

YVALUES

Show the current values of the categories of the Y-variable.

12.4 Analysis commands

ANALYSIS *<output mode>* *<statistics>* *<threshold>*
<output mode> can be APPROX, EXACT, MC or EXMC
<statistics> can be NOSTAT, X2, X2+, X2PLUS, or ALL
<threshold> can be NOTHRESH, THRESH, or THRESH=*<value>*

This command computes statistics and approximate or exact significances, or Monte Carlo estimates of the latter.

<output mode>:

- | | |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| APPROX | Approximate significances. The output includes <ul style="list-style-type: none"> - the statistics of the current table; - the approximate significances of these statistics, obtained by asymptotic approximations; - an estimate of the size of the isomarginal family; - the hypergeometric probability of the table (if <i><input mode></i> = T or TM). |
| EXACT | Exact significances. The output includes, in addition to the above items, the exact significances obtained by enumerating all members of the isomarginal family. However, if the estimated size of the isomarginal family exceeds a preset limit (see command ISOM), the exact calculations will not be carried out. |
| MC | Monte Carlo estimates. In addition to the output produced by APPROX, the output includes Monte Carlo estimates of the significances. |
| EXMC | Like EXA if the isomarginal family is sufficiently small, otherwise like MC |

<statistics>:

- | | |
|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| NOSTAT | No statistics. This options provides a means to assess the exact size of the isomarginal family much faster than with statistics. This result is accurate, but much slower than the calculation of the Gail-Mantel estimate obtained with APPROX. |
| X2 | Pearson's X^2 . |
| X2+ or X2PLUS | Four χ^2 statistics: Pearson's X^2 , the $-2 \log$ likelihood ratio G^2 , the Freeman-Tukey statistic $F-T$, and the hypergeometric probability itself. |
| ALL | Pearson's X^2 , η^2 , Kruskal Wallis' H , Kendall's τ , and Pearson's and Spearman's correlation coefficient. |

For the definitions of the statistics see section 5.

<threshold>:

When, during the enumeration of the isomarginal family, the significance of X^2 exceeds a given threshold, one may decide that the table does not deviate much

from independence, and one might wish to terminate the enumeration in order to save time. The threshold value is set by this parameter. It is effective only in case `<statistics>=X2`. A threshold less than .000001 is replaced by 2., and thus made ineffective.

THRESH is equivalent to THRESH=.1, which is also the initial value.

Initialization

```
ANA EXMC X2 THRES=.1
```

Examples

```
ANAL X2 NOTHRESH EXACT
```

```
ana x2 not exa           !(equivalent to the line above)
```

```
ANAL ALL THRESH
```

```
           !(THRESH will be ineffective <statistics> is not X2)
```

```
ANAL APP NOSTAT
```

TABLE

Print the current table of observed values. This command has no parameters.

EXRES <i><residuals></i>

Print the tables of expected values and residuals. *<residuals>* can have the same values as with the DATA command.

In batch mode (see BATCH and INTER), TABL and EXRE are generated automatically after successful table entry.

12.5 Control commands

STOP
BYE
QUIT
END

Any of these command terminates the execution of FISHER.

HELP <i><help option></i>
<i><help option></i> can be empty, ALL, TABLE, ANALYSE, VALUES, CONTROL, FILES, or F0

This command provides a short explanation of all commands

<help option>:

empty	gives a survey of the syntax and of all help options
ALL	gives all help screens
TAB	for DATA TABL EXRE
ANA	for ANAL
VAL	for XVAL YVAL DVAL
CON	for QUIT STOP END REIN INTER BATCH ISOM MC SEED PARAM
FIL	for OUT OU2 IN2 DISTR
F0	for the initial file settings

Initialization

HELP

Examples

HEL

HELP ALL

help tab

INTERACTIVE
BATCH

These two commands switch between interactive mode and batch mode. In batch mode all input is echoed to output, some output is generated automatically (namely that of TAB and EXRE after DAT), and fatal errors terminate the execution. In interactive mode all input is echoed to the log file, and all errors are dealt with interactively.

Initialization is interactive mode:

INT

ISOMARGINAL <i><limit></i>

If the estimated size of the isomarginal family exceeds $\langle limit \rangle$, no exact analysis will be performed. This applies to $\langle output\ mode \rangle = \text{EXA}$, or EXMC of ANAL.

If $\langle limit \rangle$ is omitted, the current value is shown.

Initialization

ISOM 100000

Examples

ISO

ISOM 1000000

ISOM 1E6

The last two lines are equivalent.

MC $\langle n \rangle$

This command sets the size of the Monte Carlo sample to n . This applies to $\langle output\ mode \rangle = \text{MC}$, or EXMC of ANAL.

If $\langle n \rangle$ is omitted, the current value is shown.

Initialization

MC 2000

Examples

MC

MC 10000

SEED $\langle n \rangle$

This command sets the seed of the random number generator for a Monte Carlo analysis to n , which can be any integer. This applies to $\langle output\ mode \rangle = \text{MC}$ and EXMC of ANAL.

If $\langle n \rangle$ is omitted, the current value is shown.

REINITIALIZE

Restart FISHER: discard all data, re-initialize all parameters, except INTER/BATCH and except IN2 if secondary input is active, that is ON.

PARAMETERS

Show the current values of all parameters and all file names.

12.6 File commands

The following file commands may reset certain output parameters and (de)activate alternative input and output files. They share a number of parameters, and these will be explained after the commands.

IN2 NAME=<filename> <new?> <off?>

This refers to secondary input. It can be used to enter a file of commands into FISHER. After successfully finishing these commands, input returns to primary input, that is the keyboard.

OUTPUT <format>

This command can be used to change the output format of the primary output channel (=screen) from SHORT to LONG and back.

OU2 NAME=<filename> <new?> <off?> <format> <message>

This refers to secondary output. It can be used to write output to a file, parallel to screen output (=primary output). Primary and secondary output formats need not be the same.

DISTR NAME=<filename> <new?> <off?>

This refers to a file for writing the entire isomarginal family to, that is the entire distribution of all statistics requested by ANAL. ANAL produces this output automatically, if and only if DISTR is activated. Make sure that the isomarginal family is small, otherwise ANAL will take very much time and produce a very big file. The file size is about 30-80 times the size of the isomarginal family. See also sections C (p. 60ff.) and 11.3 (p. 38).

Below is a description of the parameters common to all file commands.

<filename> is a DOS filename, optionally with path and/or drive <new?> can be REPLACE, or NEW <off?> can be ON, or OFF <format> can be SHORT, or LONG <message> can be MESSAGE, or NOMESSAGES

With NAME=<filename> each unspecified parameter resumes its DEFAULT value (rather than the previously assigned value). Moreover, if the file exists and REPLACE has been specified, the file will be overwritten, and the old information is lost.

Without NAME=<filename> the file must exist, and the file pointer is not moved. For output, this means that new information is appended to the end of the file. (For input, it has no meaning, because FISHER cannot stop reading in the middle of a (secondary input) file.) Without NAME=<filename>, <new?> may not be specified. The reason is that NEW is meaningless, while REPLACE would be dangerous, if you do not specify the file name.

<new?>:

NEW	The file must not exist, and will be created. If it does exist, this command is aborted with an error message. This is the default.
REPLACE	The file may exist. If it exists, it will be overwritten. If it doesn't, it will be created.
<off?>:	
OFF	This deactivates the file, but does not close it. Use the same command with ON, but without NAME=<filename> to continue at the same point. A file is closed by using the same command with a different NAME=<filename>
ON	Re-activates the file is in use, but inactive. ON is the default.
<format>:	
SHORT	Results from one analysis usually fit on a single screen.
LONG	More nicely formatted output, suitable for printing, but too verbose for the screen
<message>:	
MES	Echo input and display error messages on this file.
NOM	Do not echo input and do not display error messages on this file. This is default.

Defaults: NEW ON NOMESS. In interactive mode the default format is SHORT, in batch mode it is LONG.

Initialization

```

IN2 NAME=IN2.FIS
OUTPUT SHORT                !If INTERACTIVE
OUTPUT LONG                 !If BATCH
OU2 NAME=OU2.FIS SHORT NEW MESS !If INTERACTIVE
OU2 NAME=OU2.FIS LONG NEW MESS  !If BATCH
DIS NAME=DISTR.FIS OFF NEW

```

Examples

```

IN2 NAM=A:\DEMO
OU2 NAM=C:\TMP\FISHER.LOG REPLACE

```

Part IV

Appendices

A Installing and running FISHER on IBM PCs and compatibles

A.1 System requirements

FISHER 3.0 required a mathematical coprocessor in your XT or AT. The new version of FISHER, version 3.10, uses a coprocessor if it is present on your system. If not, the program will run as well, but it will be much, much slower.

A.2 The distribution disk

The distribution disk contains the following files:

filename	ext	size (kB)	purpose
FISHER	EXE	312	the program
DEMO2		1	demonstration input for batch usage
DEMO1		1	demonstration input for interactive usage
README		1	last minute information
DISPLAY	EXE	21	utility for displaying
INSTALL	BAT	1	installation file

Displaying and scrolling the output of FISHER (thus also that of the DEMO), and the README file may be done by entering: `DISPLAY <filename>`, where *filename* is the file to be displayed.

A.3 Installation

Make sure to make a back-up copy of the distribution disk (e.g. with the DOS command `diskcopy`), and put this copy in a safe place.

Suppose you wish to install the software in a new directory `C:\FISHER`. To do so you first create this directory by typing `md C:\FISHER`. Then move to this directory: `cd C:\FISHER` - for details, please refer to the DOS manual. Now put the original distribution disk in drive A: (or B:), and move to that drive by typing `A:` (or `B:`). By typing `install c` and pressing the enter key, the installation of FISHER will be carried out.

A.4 Running FISHER

The program FISHER consists of a single file FISHER.EXE. To run FISHER *interactively* just type

```
FISHER
```

at the DOS prompt. After loading, which might take a little while, you should see the header, ending with the FISHER prompt `>>`. You can exit the program with `BYE`, `END`, `QUIT`, or `STOP`.

To run FISHER interactively *with input from the file IN*, type

```
IN2 NAME=IN
```

to FISHER's prompt ('>>'). Here, the command IN2 activates secondary input, from the file IN, see the sections 11.2 and 12.6 on files. When FISHER (successfully) reaches end of file on IN, it returns to primary input, that is input from the keyboard.

To run FISHER in 'batch mode' *with input from the file IN*, use DOS redirection:

```
FISHER <IN
```

When FISHER reaches end of file on IN, execution terminates. In this case, the file IN should start with the command BATCH.

A.5 Running the demo

For interactive usage, prepare running the demonstration input as follows.

- Move to the FISHER directory (by the DOS command CD).
- Start FISHER
- Type at the FISHER prompt: `in2 name=demo1`.

For batch usage, prepare running the demonstration input as follows.

- Move to the FISHER directory (by the DOS command CD).
- Type `FISHER <demo2`

You can inspect the demonstration input by typing

```
DISPLAY DEMO1
```

or

```
DISPLAY DEMO2
```

You can inspect the demonstration output by typing

```
DISPLAY DEMO1.OU1
```

or

```
DISPLAY DEMO2.OU2
```

B Additional technical information

B.1 Marginal zeroes

Rows and/or columns that consist entirely of zeroes, may be entered, and are handled smoothly, as long as there are at least two non-zero rows, and at least two non-zero columns. Of course, the number of degrees of freedom (used for the asymptotic approximations) is based on the number of non-zero rows and columns. Zero rows and/or columns matter for the calculation of r , and zero rows matter for η^2 , whenever the scores are the default ones, that is 1, 2, 3, In that case deletion of a zero row or column by hand changes the scores, hence affects the values and distribution of the statistic. For example, when the second row is deleted by hand, the scores of the other rows change from 1, 3, 4, . . . to 1, 2, 3, Note, that only r and η^2 depend on scores, and that for the other statistics addition or deletion of zero rows or columns makes no difference.

B.2 Conditioning

The question of conditioning on the marginal totals is rooted in controversial, philosophical foundations of statistical inference. Half a century of lively debate has failed to produce one dominating view. Moreover, the debate is rather technical, much more technical than the main body of this text. To report the debate, we need such notions as unbiased tests, conditional tests, and randomized tests, which are for instance well explained in Lehmann (1959, pp. 11, 62, 130, 125).

A test is called *unbiased* if its power exceeds the significance level for all alternative hypotheses. This is a special case of unbiasedness of decision rules ('at any incorrect decision the risk exceeds the risk at any correct decision'), of which unbiasedness of statistics is another special case.

A test at level α is called *conditional* (Lehmann says: the test has a Neymann structure) if the conditional level is $\leq \alpha$ for every value of the conditioning statistic.

A test is called *randomized* if the decision to reject depends not only on the data, but also on an independent random variable ('flipping a coin').

Let us consider how an exact, *unconditional* test based on the statistic S in a 2×2 table proceeds. Suppose we have 20 observations and the null distribution is multinomial. More precisely, the null distribution is multinomial with $n = 20$ and 4 cells, and the four cell probabilities (p_1, p_2, p_3, p_4) satisfy 'homogeneity', that is $p_1/p_2 = p_3/p_4$. A multinomial with 20 observations and 4 cells has a sample space of $\binom{23}{3} = 1,771$ points. Given $p = (p_1, p_2, p_3, p_4)$, to each point in the sample space corresponds a multinomial probability and a value of S . Hence we can derive the null distribution of S , and from this the p-value of S_{obs} , which depends on the vector p . What we want, is the supremum over all vectors p satisfying $p_1/p_2 = p_3/p_4$.

Note how much conditioning simplifies the technical calculation of the p-value. Firstly, the sample space is reduced considerably. Secondly, under conditioning the null hypothesis is a single point, and we do not have to take a supremum to find the

p-value. We get a different test, however, and this appendix is about the theoretical advantages and disadvantages of both tests, and not only about ease of computation.

Let us return to our problem of conditioning. We first confine the problem as follows. We are given the observed table and a specific statistic, and we are asked to report the p-value. Now there are two problems. Firstly:

Should we report the conditional or the unconditional p-value?

Secondly, even the very definition of the p-value for a discrete distribution is not uncontroversial. This second problem is not discussed in this Manual. At the moment we will only discuss the first problem, and define p-values as the probability (under the null hypothesis) of obtaining a result at least as extreme as the observed result ‘in a similar situation’. The crux is: What is ‘a similar situation’? Is it a situation where the margins are the same, or should the margins be taken as random, keeping only the grand total the same.

Fisher (1935) introduced conditioning in 2×2 tables with the following argument.

Let us blot out the contents of the table, leaving only the marginal frequencies. If it be admitted that these marginal frequencies by themselves supply no information on the point of issue, namely, as to the proportionality of the frequencies in the body of the table, we may recognize the information they supply as wholly ancillary; and therefore recognize that we are concerned only with the relative probabilities of occurrence of the different ways in which the table can be filled in, subject to these marginal frequencies.

This has an intuitive appeal, but it is also somewhat cryptic. Since 1935 Cox (1958), Plackett (1977, 1981), Barndorff-Nielsen (1978) and others have tried to fill in the details, while for example Barnard (1945a, 1945b, 1947), Boschloo (1970), Berkson (1978a, 1978b), McDonald, Davis, & Milliken (1977) have provided counterarguments. Summaries of more recent contributions to this discussion can be found in Yates (1984) and Agresti (1990, section 3.5.5). First we discuss two problems with Fisher’s argument.

Firstly, it is quite difficult to give a precise definition for ‘no information’. Technically it is called ‘ancillarity’. Barndorff-Nielsen (1978) discusses four variations. Not only does each variation have its own disadvantages or paradoxes, but more importantly, it turns out that according to three of the four definitions the margins of 2×2 tables contain some information about independence. Furthermore for example Berkson (1978b) explicitly argues that these margins contain information about independence. Hence the conclusion is that there is no single, widely accepted definition of ‘ancillarity’, and that according to some definitions the margins *contain information with respect to independence*.

Secondly, we want such a definition of ‘ancillarity’ that one can show that conditioning on such a statistic is in some sense good. If this were true for some definition of ancillarity, one should condition on all ancillary statistics. Unfortunately, Basu (1964) discusses a situation, where there is no maximal ancillary statistic. Hence

there can be no rule that one should *always* condition on *every* ancillary statistic. In the example of Basu we are faced with the peculiar choice on which ancillary statistic we ‘ought to’ condition. Why on one and not on the other?

Cox & Hinkley (1974, 2.3.iii on p.38) discuss a hypothetical situation where the choice is clear. There, conditioning on ancillary information makes a big difference, and it definitely seems the right thing to do. Hence, such situations do exist.

A mathematical argument in favour of conditioning is the following. For the sampling schemes and the null hypothesis we discuss, any unbiased test is conditional (Lehmann, 1959, Thm 4.3.2). This is a strong argument, but it is not conclusive. Let us mention two counterarguments.

- Approximate unbiasedness is very attractive, but *exact* unbiasedness is not a necessity. For example X^2 and G^2 are only asymptotically unbiased, even in most 2×2 tables for conditional, randomized tests. Moreover, shouldn’t one trade a little bias for a large gain in power, or in convenience, if this is possible?
- In our situation unbiased, conditional tests ‘almost certainly’ require randomization. This is not so much because only finitely many levels can be obtained, but more seriously, because different values of the marginals correspond to different sets of attainable levels.

Hence one arrives at the question of appropriateness of randomized tests. Certainly there are situations where such tests are appropriate. As an example we mention recurrent decision problems where the loss in one week can be compensated by an expected gain in the next week. This means that one may optimize the gains and losses over a long run of such actual repetitions. In other words, one may optimize the expected outcome of the decision, and one knows precisely what expectation to consider. To us this seems opposite to the situation where one wants to say what we can learn from a single set of data. In the case of recurrent, compensating decision problems the question of randomization in conditional tests does not really arise. In such a situations the optimization problem is specified completely, and in principle it is easy to solve.

In a single, stand-alone experiment however, randomization does not seem appropriate to us. A client who has great interests in the correctness of our decision will not be impressed if we first look at the data and then decide to flip a coin to reach a decision. In a single, stand alone analysis, where we want ‘to say what we can learn from the data’, randomization does not seem appropriate. If we condition, we are left with an unrandomized conditional test.

The major disadvantage of unrandomized, conditional tests is the following. Suppose we are testing at a preset level α . Due to the discreteness of the attainable levels, we have to be satisfied with a test at a lower level. Hence one sacrifices power, and often a substantial amount, at no gain whatsoever. By considering unconditional tests, one may choose a critical region that has a different level for different values of the margins, as long as the average level is α or less. Unrandomized conditional tests should be of level α or less for each possible set of values of the margins. Clearly this is much more restrictive, and unconditional tests have (often much) more power (unconditionally), virtually ‘free of charge’. The question remains, whether this fact is relevant to the particular situation at hand:

- Is it allowed to compensate a higher conditional level of the test at one set of margins with a lower level at another set?
- Is the unconditional power relevant?
- Are we testing at a preset level, or are we reporting p-values?

Our personal opinion can be summarized as follows. There are situations in which conditioning definitely seems appropriate, and others in which it definitely seems inappropriate. In single, stand-alone analyses where we want ‘to say what we can learn from the data’, we go along with Cox (1958), and we would condition on the margins. In our consultations this covers the majority of analyses. But the border lines are not at all clear, and there remain many situations where the experts strongly disagree.

The next question is: Is there a need for a program that computes the unconditional p-values? We think that the answer is negative, and that one can rely on asymptotic approximations in the vast majority of applications. After all, asymptotic approximations tend to be much better for unconditional tests than for conditional ones. The reason is, that unconditional tests are averages of conditional ones. Asymptotic methods are already good for conditional tests for many remarkably small samples. They fail only occasionally, in a rather small set of tables. Moreover, this set seems to be rather ‘random’ with respect to the sets that unconditional tests average over. Hence, averaging strongly reduces the errors. FISHER certainly has a place, but not a big one. A program for unconditional p-values would have a very much smaller place.

Moreover, such a program has to find the supremum of the p-values over all points corresponding to independence, which makes the program computationally much more complex than a program like FISHER for conditional tests.

Let us conclude with an example of a situation where we do not condition, at least not strictly. On the basis of extensive calculations we have reached the conclusion that Cochran’s rule applies (in a relevant population of tables) in more than 99.9% of the cases. This argument is only reassuring if one is not a strict adherent of conditioning.

B.3 Monte Carlo estimation

The random number generator built in in FISHER is from Rubinstein (1981). It uses the following simple, linear congruential method.

```
randominteger = 16807 * randominteger MODULO 231 - 1
```

```
randomnumber = randominteger / 231 - 1
```

The seed is taken from system date and time, producing a different seed at each run. If the user provides a seed through the **SEED** command, a big number is added, to avoid that the seed (and one or two successive random numbers) are exceedingly small. Moreover, care is taken that the randominteger does not become 0 modulo $2^{31} - 1$, which would be disastrous.

For generating random tables (= random drawings from the hypergeometric distribution) we use the algorithm given in Patefield (1981).

The asymptotic standard error of the Monte Carlo estimate of the exact significance is *defined* as follows. In principle one can obtain exact, 95% confidence intervals for the exact significance. Unfortunately this interval is in general not symmetric. So to report them, we would have to report two numbers. Much rather we report a single number. To this end, we define the asymptotic standard error as (an approximation to) the longest half of this interval divided by 1.96. Hence, the estimated $p\text{-value} \pm 1.96 \times \text{the asymptotic standard errors}$ span a confidence interval with at least 95% confidence.

Let us denote the exact significance by π (because here it is a parameter to be estimated). Let N be the size of the Monte Carlo sample, and let p be the proportion of tables from the sample that lie in the critical region. So p is our Monte Carlo estimate of the exact significance, and Np is binomially distributed with parameters (N, π) . Hence p is approximately normally distributed with mean π and variance $\pi(1 - \pi)/n$.

Now we want to define the asymptotic standard error ε of p in such a way, that $p \pm 1.96\varepsilon$ defines a confidence interval for π with at least 95% confidence. First we will construct a confidence interval (π_1, π_2) for π . In general this interval will not be symmetric around p . If 1.96 is replaced by c , the endpoints $\pi_{1,2}$ are the roots of the quadratic equation

$$(p - \pi)^2 = c^2 \pi(1 - \pi)/N.$$

That is,

$$\pi_{1,2} = \frac{2pN + c^2 \pm \sqrt{4p(1-p)Nc^2 + c^4}}{2(n + c^2)}.$$

The asymptotic standard error can now be defined as

$$\begin{aligned} \varepsilon &=_{\text{def}} \max_{i=1,2} \frac{|\pi_i - p|}{c} \\ &= \frac{|1 - 2p| + \sqrt{1 + 4p(1-p)N/c^2}}{2c + 2n/c}. \end{aligned}$$

Taking $c = 2$ we get

$$\varepsilon = \frac{|1 - 2p| + \sqrt{1 + Np(1-p)}}{n + 4}.$$

This is the standard error provided by the program.

For $\frac{5}{N} < p < .1$ and $N > 1000$ this reduces to

$$\varepsilon \approx \sqrt{p/N}.$$

Almost just as well, we might have estimated the standard error simply as $\sqrt{p(1-p)/N}$. However, for example for $N = 1000, p = .01$ one has

$$\begin{aligned}\varepsilon &= .0043, \\ \sqrt{p/N} &= .0032, \\ \sqrt{p(1-p)/N} &= .0031,\end{aligned}$$

while the 95% confidence interval based on the Poisson distribution is $(.0053 < \pi < .0176)$, resulting in a ‘standard error’ of $.76/1.96 = .39$.

C Annotated output

```
X X2  FISHER EXACT SIGNIFICANCE TESTS
X      (C) (1988, 1991) AV/PMK/NSK/WPB
X X    VERSION 3.10 (c) ProGAMMA 87-91
```

```
DATE: 19 DEC '92    TIME: 10:27
```

The time and date at the beginning of the run are used to label all output. So later on, times refer to this time rather than the then current time

```
>> help
```

>> is the FISHER prompt. So help is entered by the user. Its output, given below, should be self-explanatory. For help about data entry, type: help table, etc.

The following help options are available:

```
HELP ALL
HELP TABLE      for: DATA  TABLE  EXRE
HELP ANALYSIS    for: ANAL
HELP VALUES     for: XVAL   YVAL   DVAL
HELP CONTROL     for: QUIT   STOP   END   INTER  BATCH
                  ISOM   MC   SEED   REIN   PARAM
HELP FILES       for: OUT   OU2   IN2   DISTR
HELP FO          for initial file settings
```

General information on the syntax.

- Each command should be on one line of ≤ 80 chars.
- Key words may be abbreviated to the first 3 chars.
- Any text after '!' is treated as comment.
- One or more spaces and/or commas and/or '=' serve as separator.
- All parameters are optional.
- All parameters are order independent, with the obvious exception of those in DATA <nr> <nc>, XVAL and YVAL.
- All parameters are initialized.
- All parameters retain their values until redefined (see PARAM), except file command parameters.

More information can be found in the manual.

Next we are going to do some real work. First we enter the frequencies of a 2×3 table. Then we request the exact p-value of Pearson's X^2 .

```
>> data 2 3
ENTER FREQUENCIES, ONE ROW PER LINE
4 10 10
12 6 6

>> ana x2 exact

=> ANAL X2      EXACT  THRESHOLD =  0.10000
```


The `analyse` command echoes the options specified, and shows which options are selected by default. Recall that all options have an initial value, and almost all options are 'sticky', i.e., they retain their last value, unless a new value is explicitly specified.

The option `threshold` is initialised at the value 0.10. It works as follows. If the p -value of only X^2 is requested, and FISHER discovers that its p -value exceeds the threshold, it stops the calculation of the exact p -value. When the p -value is larger than say 0.10, it is generally not of interest to know its exact value. This option can be deactivated with `nothresh`, or by assigning a value larger than 1 to `thresh`.

Below is the output of the analysis. It comes in two parts. The first half contains all calculations that can be done almost instantaneously. The second half contains the results that are obtained after enumerating the isomarginal family of all tables with the same margins, or after the generation of a Monte Carlo sample thereof.

```
.....
TABLE NUMBER      1      ANALYSIS NUMBER      1      10:27  19 DEC '92

NR =   2      NC =   3      2 DEGREES OF FREEDOM      N =      48
  ROW MARGINALS:      24      24
COLUMN MARGINALS:      16      16      16
ESTIMATED SIZE OF THE ISOMARGINAL FAMILY . . . . .      230
HYPERGEOMETRIC PROBABILITY OF THE TABLE. . . . .      0.003619

STATISTIC          X2
OBS VAL            6.000
APPROXIMATE
SIGNIFICANCE       0.0498
RUNNING
EXACT
SIGNIFICANCE       0.0774
PROB. MASS         0.0217

PROB. MASS IS THE PROBABILITY THAT THE STATISTIC IS IN THE INTERVAL
                                OBSERVED VALUE +-  0.000050
NUMBER OF TABLES ENUMERATED. . . . .      132
TIME REQUIRED FOR EXACT ANALYSIS . . . . .      0.17 SECONDS
```

Note that the Gail-Mantel estimate of the number of tables with the same margins may deviate by a factor 2 or more from the number of tables enumerated. There are two reasons. Most importantly, the ingenious approximation of Gail & Mantel (1977) is not always very accurate. Secondly, FISHER may skip the enumeration of certain tables as it is certain that their accumulated contribution to the p -value is less than 10^{-6} .

Next, we will reanalyse Table 10.3 in Maxwell (1977). The population consists of 387 patients of a psychiatrist. Each patient is rated on two ordinal 5-point scales: Tension (down), and Worry (across). This has produced the following 5×5 table.

```
>> data 5 5
```

```
ENTER FREQUENCIES, ONE ROW PER LINE
3 2 5 10 11
```

```

11 8 16 35 19
28 13 23 33 6
27 11 23 12 5
63 10 9 4 0

```

With such a large table, obviously with rather strong interaction, one should not embark automatically on an exact analysis. An approximate (= asymptotic) analysis is the obvious starting choice.

```
>> ana all approx
```

```
=> ANAL ALL      APP      THRESHOLD =  0.10000
```

```

.....
TABLE NUMBER      2      ANALYSIS NUMBER      2                      20:37  19 DEC '92

NR =   5      NC =   5      16 DEGREES OF FREEDOM                      N =      387
  ROW MARGINALS:      31      89      103      78      86
COLUMN MARGINALS:     132      44      76      94      41
X-VALUES:      1.000      2.000      3.000      4.000      5.000
Y-VALUES:      1.000      2.000      3.000      4.000      5.000
ESTIMATED SIZE OF THE ISOMARGINAL FAMILY . . . . . 0.127E+23
HYPERGEOMETRIC PROBABILITY OF THE TABLE. . . . . 0.590E-41

```

Note that the isomarginal family is far too big for an exact analysis

```

STATISTIC      X2      K-W      ETA2      TAU      RS      R
OBS VAL      136.157      0.2807      0.2798      -0.4451      -0.5286      -0.5265
APPROXIMATE
SIGNIFICANCE      0.0000      0.0000      0.0000      1.0000      1.0000      1.0000
TIME REQUIRED FOR APPROXIMATE ANALYSIS . . . . . 0.16 SECONDS

```

For a table with so much structure (two ordinal variables, strong deviation from independence) it is natural to inspect the residuals. Here we have chosen for Haberman's adjusted residuals, because under the null hypothesis they have asymptotically a standard normal distribution.

Note that all significances in FISHER are probabilities of exceeding the observed value. The significances of 1.000 for τ , R , R_S indicate that, given the marginals, virtually all tables have values of the statistic larger than the negative value indicated. Said differently, there are no tables with a smaller R than -0.5265 which have a nonnegligible probability. So when we use a two-sided test, or test against negative association, as in this table, significances of 1.0000 for τ , R , R_S are highly significant.

*For X^2 , $K-W$, η^2 , and other χ^2 statistics, however, only **small values** indicate significance.*

```
>> exres adj
```

```

EXPECTED VALUES AND MARGINAL FREQUENCIES;
EXPECTED VALUES < 5.00 , IF ANY, ARE FLAGGED BY #.

```

```

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

Note that only two expected values are below 5. As they are above 1, Cochran's rule allows the use of the χ^2 distribution for Pearson's X^2 .

HABERMAN'S ADJUSTED RESIDUALS

(OBS.FREQ.-EXP.FREQ.) / ESTIMATED ST.DEV.

ABSOLUTE VALUES LARGER THAN 3.00 ARE FLAGGED BY #.

-3.0	-0.9	-0.5	1.1	4.7#
-4.9#	-0.8	-0.4	3.8#	3.8#
-1.7	0.5	0.8	2.1	-1.8
0.1	0.9	2.5	-2.1	-1.3
8.7#	0.1	-2.4	-4.8#	-3.6#

Note how much the residuals tell you about the association in the table. They are negative in the North-West and South East corners, and positive along the main diagonal of the table. Formal significance testing is not really needed with this example size and such strong association.

*Next, we will demonstrate how the entire distribution (rather than a single significance) can be obtained. This distribution is written to a file. The command **distr** activates this option, and a file name should be specified. The option applies to each subsequent **anal** until de-activated by **distr off**. It does not have to precede the command **data** as it does here.*

To save space, we have chosen for a table with a very small sample size and a very small isomarginal family. It is the same table that is discussed in Table Table 3 on page 10, and several subsequent tables.

```
>> distr name=d replace
```

```
>> data 2 3
```

```
ENTER FREQUENCIES, ONE ROW PER LINE
```

```
5 1 0
```

```
1 2 3
```

```
>> ana exa nostat
```

*With the **nostat** option, the analysis command writes the frequencies of each table in the isomarginal family to file. With **distr** active, each call of **anal** will produce the same subset of the isomarginal family in the same order, but skipping all tables with a probability less than .5E-8.*

```
=> ANAL NOSTAT EXACT THRESHOLD = 0.10000
```

```
.....
TABLE NUMBER    3    ANALYSIS NUMBER    3                10:27  19 DEC '92
```

```

NR = 2    NC = 3    2 DEGREES OF FREEDOM    N =    12
  ROW MARGINALS:      6      6
COLUMN MARGINALS:    6      3      3
ESTIMATED SIZE OF THE ISOMARGINAL FAMILY . . . . . 17
RUNNING
NUMBER OF TABLES ENUMERATED. . . . . 16
THE DISTRIBUTION HAS BEEN WRITTEN ONTO THE FILE. . . . . d
TABLES WITH VERY SMALL PROBABILITIES MAY HAVE BEEN IGNORED.
TIME REQUIRED FOR EXACT ANALYSIS . . . . . 0.16 SECONDS

```

Note the slowness of writing each table to file. Below is the file d.

```

-----
EXACT DISTRIBUTION OF ANALYSIS # 3 ON TABLE # 3 IN RUN 10:27 19 DEC '92
  ROW MARGINALS:      6      6
COLUMN MARGINALS:    6      3      3
FORMAT (I5,F11.8,I4,25(I4))
  P      -LOG P F11 F12 F13 F21 F22 F23
 1 0.00108225 2 0 3 3 6 0 0
 2 0.01948052 1 1 2 3 5 1 0
 3 0.01948052 1 1 3 2 5 0 1
 4 0.04870130 1 2 1 3 4 2 0
 5 0.14610390 0 2 2 2 4 1 1
 6 0.04870130 1 2 3 1 4 0 2
 7 0.02164502 1 3 0 3 3 3 0
 8 0.19480519 0 3 1 2 3 2 1
 9 0.19480519 0 3 2 1 3 1 2
10 0.02164502 1 3 3 0 3 0 3
11 0.04870130 1 4 0 2 2 3 1
12 0.14610390 0 4 1 1 2 2 2
13 0.04870130 1 4 2 0 2 1 3
14 0.01948052 1 5 0 1 1 3 2
15 0.01948052 1 5 1 0 1 2 3
16 0.00108225 2 6 0 0 0 3 3

```

The value of 'each' statistic for each table with the same margins can be obtained on file d by the following command.

```

>> ana exa all
=> ANAL ALL    EXACT  THRESHOLD = 0.10000

```

```

.....
TABLE NUMBER    3    ANALYSIS NUMBER    4    10:27 19 DEC '92

NR = 2    NC = 3    2 DEGREES OF FREEDOM    N =    12
  ROW MARGINALS:      6      6
COLUMN MARGINALS:    6      3      3
X-VALUES:      1.000    2.000
Y-VALUES:      1.000    2.000    3.000
ESTIMATED SIZE OF THE ISOMARGINAL FAMILY . . . . . 17
HYPERGEOMETRIC PROBABILITY OF THE TABLE. . . . . 0.019481

STATISTIC      X2      K-W      ETA2      TAU      RS      R

```

OBS VAL	6.000	0.5000	0.5000	0.6708	0.7071	0.7035
APPROXIMATE						
SIGNIFICANCE	0.0498	0.0639	0.0639	0.0095	0.0051	0.0053
RUNNING						
EXACT						
SIGNIFICANCE	0.1234	0.1234	0.1234	0.0206	0.0206	0.0206
PROB. MASS	0.1212	0.1212	0.1212	0.0195	0.0195	0.0195

PROB. MASS IS THE PROBABILITY THAT THE STATISTIC IS IN THE INTERVAL
 OBSERVED VALUE +- 0.000050
 FOR TAU AND RS, HOWEVER, PROB. MASS = P (OBSERVED STATISTIC).
 NUMBER OF TABLES ENUMERATED. 16
 THE DISTRIBUTION HAS BEEN WRITTEN ONTO THE FILE. d
 TABLES WITH VERY SMALL PROBABILITIES MAY HAVE BEEN IGNORED.
 TIME REQUIRED FOR EXACT ANALYSIS 0.22 SECONDS

Below is the second part of file d.

```

-----
EXACT DISTRIBUTION OF ANALYSIS # 4 ON TABLE # 3 IN RUN 10:27 19 DEC '92
  ROW MARGINALS:      6      6
  COLUMN MARGINALS:   6      3      3
  FORMAT (I5,F11.8,I4,F10.4,5F10.6)
    P      -LOG P      X2      ETA2      K-W      TAU      RS      R
  1 0.00108225  2 12.0000  1.000000  1.000000 -0.894427 -0.942809 -0.904534
  2 0.01948052  1  6.0000  0.500000  0.500000 -0.670820 -0.707107 -0.703526
  3 0.01948052  1  6.0000  0.500000  0.500000 -0.521749 -0.549972 -0.502519
  4 0.04870130  1  4.0000  0.333333  0.333333 -0.447214 -0.471405 -0.502519
  5 0.14610390  0  1.3333  0.111111  0.111111 -0.298142 -0.314270 -0.301511
  6 0.04870130  1  4.0000  0.333333  0.333333 -0.149071 -0.157135 -0.100504
  7 0.02164502  1  6.0000  0.500000  0.500000 -0.223607 -0.235702 -0.301511
  8 0.19480519  0  0.6667  0.055556  0.055556 -0.074536 -0.078567 -0.100504
  9 0.19480519  0  0.6667  0.055556  0.055556  0.074536  0.078567  0.100504
 10 0.02164502  1  6.0000  0.500000  0.500000  0.223607  0.235702  0.301511
 11 0.04870130  1  4.0000  0.333333  0.333333  0.149071  0.157135  0.100504
 12 0.14610390  0  1.3333  0.111111  0.111111  0.298142  0.314270  0.301511
 13 0.04870130  1  4.0000  0.333333  0.333333  0.447214  0.471405  0.502519
 14 0.01948052  1  6.0000  0.500000  0.500000  0.521749  0.549972  0.502519
 15 0.01948052  1  6.0000  0.500000  0.500000  0.670820  0.707107  0.703526
 16 0.00108225  2 12.0000  1.000000  1.000000  0.894427  0.942809  0.904534

```

Rather than the entire distribution, one may write only a Monte Carlo sample to file. In this demonstration we chose a ridiculously small sample size.

```
>> mc 1
```

THE SAMPLE SIZE OF A MONTE CARLO ANALYSIS IS SET TO 10 TABLES.

FISHER has a minimum sample size of 10 for Monte Carlo sampling. Requests to set the sample size to a smaller number results in this minimum size.

```
>> anal mc all
=> ANAL ALL MC THRESHOLD = 0.10000
```

```

.....
TABLE NUMBER      3      ANALYSIS NUMBER      5      10:27  19 DEC '92

NR =  2      NC =  3      2 DEGREES OF FREEDOM      N =      12
  ROW MARGINALS:      6      6
COLUMN MARGINALS:      6      3      3
X-VALUES:      1.000      2.000
Y-VALUES:      1.000      2.000      3.000
ESTIMATED SIZE OF THE ISOMARGINAL FAMILY . . . . .      17
HYPERGEOMETRIC PROBABILITY OF THE TABLE. . . . .      0.019481

STATISTIC      X2      K-W      ETA2      TAU      RS      R
OBS VAL      6.000      0.5000      0.5000      0.6708      0.7071      0.7035
APPROXIMATE
SIGNIFICANCE      0.0498      0.0639      0.0639      0.0095      0.0051      0.0053
RUNNING
MONTE CARLO
SIGNIFICANCE      0.1000      0.1000      0.1000      0.0000      0.0000      0.0000
M.C.ST.ERROR +-0.1556 +-0.1556 +-0.1556 +-0.1429 +-0.1429 +-0.1429
PROB. MASS      0.1000      0.1000      0.1000      0.0000      0.0000      0.0000
M.C.ST.ERROR +-0.1556 +-0.1556 +-0.1556 +-0.1429 +-0.1429 +-0.1429

```

Note that the exact p-values (on the previous page) are well inside the MC confidence intervals. We have chosen to report the standard error as \pm for emphasis. Here, with an estimate of 0, this may appear a little odd.

```

PROB. MASS IS THE PROBABILITY THAT THE STATISTIC IS IN THE INTERVAL
                                OBSERVED VALUE +-  0.000050
FOR TAU AND RS, HOWEVER, PROB. MASS = P (OBSERVED STATISTIC).
SEED WAS . . . . .      9999984
M.C. SAMPLE SIZE . . . . .      10
THE DISTRIBUTION HAS BEEN WRITTEN ONTO THE FILE. . . . .      d
TABLES WITH VERY SMALL PROBABILITIES MAY HAVE BEEN IGNORED.
TIME REQUIRED FOR MONTE CARLO ANALYSIS . . . . .      0.33 SECONDS

```

Finally, the last part written to file d. If we want to see which tables have been generated, we should reset the seed to the value indicated above, and request `ana nostat mc`.

```

-----
MC DISTRIBUTION OF ANALYSIS # 5 ON TABLE # 3 IN RUN 10:27  19 DEC '92
  ROW MARGINALS:      6      6
COLUMN MARGINALS:      6      3      3
FORMAT (I5,T21,F10.4,5F10.6)

      X2      ETA2      K-W      TAU      RS      R
1      0.6667  0.055556  0.055556 -0.074536 -0.078567 -0.100504
2      6.0000  0.500000  0.500000 -0.223607 -0.235702 -0.301511
3      1.3333  0.111111  0.111111  0.298142  0.314270  0.301511
4      0.6667  0.055556  0.055556  0.074536  0.078567  0.100504
5      0.6667  0.055556  0.055556  0.074536  0.078567  0.100504
6      1.3333  0.111111  0.111111 -0.298142 -0.314270 -0.301511
7      0.6667  0.055556  0.055556 -0.074536 -0.078567 -0.100504
8      1.3333  0.111111  0.111111 -0.298142 -0.314270 -0.301511
9      0.6667  0.055556  0.055556 -0.074536 -0.078567 -0.100504
10     0.6667  0.055556  0.055556  0.074536  0.078567  0.100504

```

Writing to d is discontinued by the following request. It is not really needed here, because we end the FISHER session. It would be needed if we were to proceed with other analyses for which we did not want to write out the entire distribution.

```
>> distr off
```

```
>> end
```

```
END OF RUN  10:27  19 DEC '92  
DID    5 ANALYSES ON    3 TABLES.  
THIS RUN REQUIRED  2 SEC.
```


References

- Agresti, A.** (1990). *Categorical data analysis*. New York: Wiley.
- Agresti, A., & D. Wackerly** (1977). Some exact conditional tests of independence for $R \times C$ cross-classification tables. *Psychometrika*, 42, 111-125.
- Agresti, A., D. Wackerly, & J.M. Boyett** (1979). Exact conditional tests for cross-classifications: approximation of attained significance levels. *Psychometrika*, 44, 75-83.
- Barnard, G.A.** (1945a, b). A new test for 2×2 -tables. *Nature*, 156, 177, and 783-784.
- Barnard, G.A.** (1947). Significance tests for 2×2 -tables. *Biometrika*, 34, 123-138.
- Barndorff-Nielsen** (1978). *Information and exponential families*. New York: Wiley.
- Basu, D.** (1964). Recovery of ancillary information. *Sankhyā, A*, 26, 3-16.
- Berkson, J.** (1978a). In dispraise of the exact test. *Journal of Statistical Planning and Inference*, 2, 27-42.
- Berkson, J.** (1978b). Do marginal totals of the 2×2 -table contain relevant information respecting the table proportions. *Journal of Statistical Planning and Inference*, 2, 43-44.
- BMDP** See Dixon & Brown.
- Borel, E.** (1924). *Traité du calcul des probabilités. Tome I, Fase I* (p. 35,36). Paris.
- Boschloo, R.D.** (1970). Raised conditional levels of significance for the 2×2 -table when testing the equality of two probabilities. *Statistica Neerlandica*, 24, 1-35.
- Box, G.E.P., & G.C. Tiao** (1964). A note on criterion robustness and inference robustness. *Biometrika*, 51, 169-173.
- Box, G.E.P., & G.C. Tiao** (1973). *Bayesian inference in statistical analysis*. Reading, MA: Addison-Wesley.
- Cochran, W.G.** (1952). The χ^2 test of goodness of fit. *Annals of Mathematical Statistics*, 23, 315-345.
- Cochran, W.G.** (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, 10, 417-451.
- Cox, D.R.** (1958). Some problems connected with statistical inference. *Annals of Mathematical Statistics* 29, 357-372.
- Cox, D.R., & D.V. Hinkley** (1974). *Theoretical statistics* London: Chapman and Hall.
- Cressie, N., & T.R.C. Read** (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society B*, 46, 440-464
- Dixon, W.J., & M.B. Brown** (1977). *BMDP-77*. (Especially program BMDP1F and appendix A5.) Berkeley: University of California Press.
- Ehrenberg, A.S.C.** (1975). *Data reduction*. New York: Wiley.
- Everitt, B.S.** (1977). *The analysis of contingency tables*. London: Chapman and Hall.
- Fienberg, S.E.** (1980). *The analysis of cross-classified categorical data*. Cambridge, MA: The MIT Press.

- Freeman, G.H., & J.H. Halton** (1951). Note on the exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika*, 38, 141-149.
- Fisher, R.A.** (1925, 1970) *Statistical methods for research workers*. (14th ed.) New York: Hafner Press.
- Fisher, R.A.** (1935). The logic of inductive inference. *Journal of the Royal Statistical Society*, 98, 39-54.
- Freeman, M.F., & J.W. Tukey** (1950). Transformations related to the angular and the square root. *Annals of Mathematical Statistics*, 21, 607-611.
- Gail, M., & N. Mantel** (1977). Counting the number of $r \times c$ contingency tables with fixed margins. *Journal American Statistical Association*, 72, 859-862.
- Goodman, L.A., & W.H. Kruskal** (1979). *Measures of association for cross classifications*. Springer Series in Statistics. Berlin: Springer.
- Haberman, S.J.** (1973). The analysis of residuals in cross-classified tables. *Biometrics*, 29, 205-220.
- Hays, W.L.** (1988) *Statistics* (4th ed.). London: Holt, Rinehart and Winston.
- Kendall, M.** (1975). *Rank correlation methods* (4th ed.). London: Griffin.
- Kendall, M.G., & A. Stuart** (1967). *The advanced theory of statistics, Vol. 2, 2nd ed* (Especially pp. 567-582). New York: Hafner, 1967.
- Klotz, J.N., & J. Teng** (1977). One-way layout for counts and the exact enumeration of the Kruskal-Wallis H distribution with ties. *Journal American Statistical Association*, 72, 165-169.
- Koehler, K.J., & K. Larntz** (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials. *Journal American Statistical Association*, 75, 336-344.
- Kroonenberg, P.M., & A. Verbeek** (1980). *Exact distributions of χ^2 -tests in contingency tables with small numbers*. (WEP-Reeks, WR 80-18-EX.) The Netherlands: Vakgroep Wijsgerige en Empirische Pedagogiek, Rijksuniversiteit Leiden.
- Kruskal, W.H., & W.A. Wallis** (1952a). A nonparametric test for the several sample problem. *Annals of Mathematical Statistics*, 23, 525-540.
- Kruskal, W.H., & W.A. Wallis** (1952b, 1953). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47, 583-612, and 48, 907-911
- Larntz, K.** (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *Journal American Statistical Association*, 73, 253-263.
- Lehmann, E.L.** (1959, 1986²). *Testing Statistical Hypotheses*. New York: Wiley.
- Maxwell, A.E.** (1977). *Multivariate analysis in behavioural research*. London: Chapman and Hall.
- McDonald, L.L., B.M. Davis, & G.A. Milliken** (1977). A nonrandomized unconditional test for comparing two proportions in 2×2 contingency tables. *Technometrics*, 19, 145-157.
- Mehta, C.R., & N.R. Patel** (1983). A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *Journal American Statistical Association*, 78, 427-434.

- Mehta, C.R., & N.R. Patel** (1986). A Fortran subroutine for Fisher's exact test on unordered $r \times c$ contingency tables. *ACM Transaction on Mathematical Software*, 12, 154-161.
- Mood, A.M., F.A. Graybill, & D.C. Boes** (1974). *Introduction to the Theory of Statistics* (3rd ed.). Tokyo: McGraw Hill Kogakusha Ltd.
- Norušis, M.J.** (1979). *SPSS, Statistical Algorithms*. Chicago, IL: SPSS Inc.
- Patefield, W.M.** (1981). Algorithm AS 159. An efficient method of generating random $R \times C$ tables with given row and column totals. *Applied Statistics*, 30, 91-97.
- Pearson, K.** (1900). On a criterion that a system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen in random sampling. *Philosophical Magazine, Series 5*, 50, 157-175
- Pearson, K.** (1920). Notes on the history of correlation. *Biometrika*, 13, 25-45.
- Plackett, R.L.** (1977, 1981²). *The analysis of categorical data*. Griffin's Statistical Monographs and Courses, No. 35. London: Griffin.
- Read, T.R.C., & N.A.C. Cressie** (1988). *Goodness-of-fit statistics for discrete multivariate data*. Springer Series in Statistics. Berlin: Springer.
- Rubinstein, R.Y.** (1981). *Simulation and the Monte Carlo method*. New York: Wiley.
- Siegel, S.** (1956). *Nonparametric statistics for the behavioral sciences*. Tokyo: McGraw-Hill Kogakusha.
- Spearman, C.** (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- Tukey, J.W.** (1949). One degree of freedom for non-additivity. *Biometrics*, 5, 232-242.
- Verbeek, A., & P.M. Kroonenberg** (1985). A survey of algorithms for exact distributions of test statistics in $r \times c$ contingency tables with fixed margins. *Computational Statistics & Data Analysis*, 3, 159-185.
- Wilks, S.S.** (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 26, 64-68.
- Yates, F.** (1984). Tests of significance for 2×2 contingency tables (with discussion). *Journal Royal Statistical Society, Series A*, 147, 426-463.

FISHER COMMAND SUMMARY

General

- For most commands, parameters keep their values until redefined.
Exception: most file command parameters; commands with syntax errors.
- All parameters are optional.
- Almost all parameters are order independent.
Exceptions: XVAL, YVAL, and DATA <nr> <nc>.
- Keywords may be abbreviated to the first 3 characters.
- Consecutive blanks and commas form one separator.
- After '!' any text on that line is treated as comment.
- REINITIALIZE does not re-initialize BATCH or INTERACTIVE, and also not IN2 if this is active.

Limitations

- Each command and each row of observed frequencies should be on one line of at most 80 characters.
- Maximum size of numerical input: 10 positions.
- The size of the work space is installation dependent.
- Maximum frequency: 100 000. (But most likely the work space will be exceeded for much lower values.)

Initializations

(See PARAM after begin or after REIN).

INTERACTIVE			
DATA	2 2 T	RES	
EXRES	RES		
XVAL	1 2 3	etc.	
YVAL	1 2 3	etc.	
HELP			
ISOM	100000		
MC	2000		
ANAL	X2 EXMC	THRESH=.1	
SEED	different	for each run	
primary input	CON	connected	active
secondary input	IN2.FIS	non-existent/not-connected	
primary output	CON	connected	active SHORT MESSAG
secondary output	OU2.FIS	non-existent/not-connected	
distribution file	DISTR.FIS	non-existent/not-connected	
log file	LOG.FIS	connected	active

Data entry commands

DATA *<nr>* *<nc>* *<input mode>* *<residuals>*

<nr> is the number of rows. It must be the first parameter.

<nc> is the number of columns. It must be the second parameter.

<input mode> is T, TM, or M.

<residuals> is NORES, RES, ADJRES, or X2TERMS.

With input mode T or TM, DATA must be followed by the observed frequency table without or with marginals respectively. With input mode M, DATA must be followed by two lines with marginal frequencies, and by a third line with zero or more *<statistics>* = *<value>*, where

<statistic> is X2, K-W, ETA2, TAU, RS, R, LR, F-T, or p.

XVALUES *<value₁>* ... *<value_{nc}>* Change or show category scores,

YVALUES *<value₁>* ... *<value_{nr}>* to be used for η^2 and R .

DVALUES Reset scores to 1 2 3 ...

Analysis commands

ANALYSIS *<output mode>* *<statistics>* *<threshold>*

<output mode> is APPROX, EXACT, MC, or EXMC

<statistics> is NOSTAT, X2, or ALL

<threshold> is NOTHRESH, THRESH, or THRESH=*<value>*

TABLE Show the table of observed frequencies

EXRES *<residuals>* Show the expected values and residuals

Control commands

STOP, BYE, QUIT, or END Terminate execution

HELP *<help option>* Show command syntax

<help option> is empty, ALL, TABLE, ANAL, VAL, CONTROL, FILES, or FO

INTERACTIVE Change to interactive mode

BATCH Change to batch mode

ISOMARGINAL *<limit>* Set or show limit to isomarginal family

MC *<n>* Set or show MC sample size

SEED *<n>* Set or show seed for MC analysis

REINITIALIZE Reset all parameters, except IN2 and INT/BAT

PARAMETERS Show all current values

File handling commands

IN2 NAME=*<file>* *<new?>* *<off?>* Secondary input

OUTPUT *<fmt>* Primary output format

OU2 NAME=*<file>* *<new?>* *<off?>* *<fmt>* *<msg>* Secondary output

DISTR NAME=*<file>* *<new?>* *<off?>* Write isomarginal family or MC sample to file

<file> is a DOS filename, optionally with path and/or drive

<new?> is REPLACE or NEW Default is NEW

<off?> is ON or OFF Default is ON

<fmt> is SHORT or LONG Default depends on BAT/INT

<msg> is MESSAGE or NOMESS Default is NOMESS