

How to use Table 1-1

The notation of the Fourfold Table

a	b
c	d

increasing sequence of cell frequencies	read under the sequence	sign change
a b c d	a b c d	no
a b d c	a b c d	no
a c b d	a c b d	no
a c d b	a c b d	no
a d b c	a b c d	no
a d c b	a c b d	no
b a c d	b a d c	yes
b a d c	b a d c	yes
b c a d	b a d c	yes
b c d a	b d a c	yes
b d a c	b d a c	yes
b d c a	b d a c	yes
c a b d	c a d b	yes
c a d b	c a d b	yes
c b a d	c a d b	yes
c b d a	c d a b	yes
c d a b	c d a b	yes
c d b a	c d a b	yes
d a b c	d b c a	no
d a c b	d c b a	no
d b a c	d b c a	no
d b c a	d b c a	no
d c a b	d c b a	no
d c b a	d c b a	no

Step A

Arrange the cell frequencies *in order of increasing size* (the sequence is arbitrary with equal frequencies)

Example:

Starting sequence numerically

a	b	c	d
13	9	3	12

increasing size sequence

c	b	d	a
---	---	---	---

Step B

Locate the increasing size sequence in the left column of the table to the left. Note the corresponding reading sequences from the middle column. Rearrange the observed fourfold table according to this reading sequence. Note from the right column whether the sign of ϕ must be changed.

In the example:

increasing size sequence

c	b	d	a
---	---	---	---

reading sequence

c	d	a	b
---	---	---	---

numerically

3	12	13	9
---	----	----	---

Change of sign of ϕ ? Yes.

Step C

Enter Table 1-1 with the reading sequence and locate ϕ there, the two-sided and the one-sided probability.

In the example, under the reading sequence

3	12	13	9
---	----	----	---

on page 219 one finds:

3	12	13	9	-0.387	.0409	.0204
---	----	----	---	--------	-------	-------

The ϕ of the observed table is $\phi = + 0.387$ (sign change!). The two-sided probability $P_2 = 0.0409$, the one-sided probability $P_1 = 0.0204$.

Die Anwendung der Tafel 1 – 1

Die Notation der Vierfeldertafel

a	b
c	d

Größenfolge der Zellenfrequenzen	lies ab unter der Folge	Vorz. wechs.
a b c d	a b c d	nein
a b d c	a b c d	nein
a c b d	a c b d	nein
a c d b	a c b d	nein
a d b c	a b c d	nein
a d c b	a c b d	nein
b a c d	b a d c	ja
b a d c	b a d c	ja
b c a d	b a d c	ja
b c d a	b d a c	ja
b d a c	b d a c	ja
b d c a	b d a c	ja
c a b d	c a d b	ja
c a d b	c a d b	ja
c b a d	c a d b	ja
c b d a	c d a b	ja
c d a b	c d a b	ja
c d b a	c d a b	ja
d a b c	d b c a	nein
d a c b	d c b a	nein
d b a c	d b c a	nein
d b c a	d b c a	nein
d c a b	d c b a	nein
d c b a	d c b a	nein

Schritt A

Ordne die Zellfrequenzen *in aufsteigender Größenfolge* (bei gleichen Frequenzen ist die Anordnung beliebig)

Beispiel:

Ausgangsfolge:	a	b	c	d
numerisch	13	9	3	12
Größenfolge:	c	b	d	a

Schritt B

Gehe mit der Größenfolge in die linke Spalte der nebenstehenden Tabelle. Lies dort in der mittleren Spalte die Ablesefolge ab. Ordne die beobachtete Vierfeldertafel nach dieser Ablesefolge um. Vermerke aus der rechten Spalte der Tabelle, ob das Vorzeichen von ϕ gewechselt werden muß.

Im Beispiel:

Größenfolge:	c	b	d	a
Ablesefolge:	c	d	a	b
numerisch	3	12	13	9
Vorzeichenwechsel von ϕ ?	ja			

Schritt C

Gehe mit der Ablesefolge in Tafel 1 – 1 ein und lies dort ϕ , die zweiseitige und die einseitige Überschreitungswahrscheinlichkeit ab.

Im Beispiel ist unter der Ablesefolge

	3	12	13	9
auf Seite 219 zu finden:				
	3	12	13	9
		-0.387	.0409	.0204

Das ϕ der beobachteten Tafel ist $\phi = +0.387$ (Vorzeichenwechsel!)

Die zweiseitige Überschreitungswahrscheinlichkeit $P_2 = 0.0409$, die einseitige Überschreitungswahrscheinlichkeit $P_1 = 0.0204$.

The Fourfold Table

Die Vierfeldertafel

Die Vierfeldertafel
Sozial- und Biowissenschaften

Von

P. Krause · W. Lehmacher · K.-D. Wall (†)



Statistical Tables in the Social and Biological Sciences

Statistische Tafeln für Sozial- und Biowissenschaftler



Gustav Fischer Verlag · Stuttgart · New York

1981

W

The Fourfold Table

up to $N = 80$

Die Vierfeldertafel

bis $N = 80$

Inhaltsverzeichnis

Contents

By/Von

H.-P. Krüger · W. Lehmacher · K.-D. Wall (†)

1. Vorhandene Tabellen für den exakten Test in 2×2 -Felder-Tabellen 3

2. Der exakte Test in 2×2 -Felder-Tabellen 5

2.1. Das Prinzip des exakten Tests 5

2.2. Die Alternativenypothesen 9

2.3. Die Prüfung auf Kontingenz 13

2.4. Interpretation eines Ergebnisses 12

2.5. Die Beschreibung der Tabellen 14

2.6. Formeln für die asymptotische Prüfung 16

3. Anwendungen 17

3.1. Die Prüfung der Homogenität bei zwei unabhängigen Stichproben 17

3.2. Die Prüfung auf Kontingenz 18

1. Available Tables for the Exact Test in 2×2 -Fold Tables 3

2. The Exact Test in 2×2 -Fold Tables 5

2.1. Principle of Exact Testing 5

2.2. Alternative Hypotheses 9

2.3. Test against Contingency 13

2.4. Interpretation of Results 12

2.5. Description of the Tables 14

2.6. Formulas for the Asymptotic Test 16

3. Applications 17

3.1. Test with Homogeneity in Two Independent Samples 17

3.2. Test against Contingency 18



Gustav Fischer Verlag · Stuttgart · New York

1981

Addresses of Authors:
Anschriften der Autoren:

Professor Dr. Hans-Peter Krüger
Institut für Psychologie der Universität, Lehrstuhl Psychologie I,
Domerschulstr. 13, 8700 Würzburg

Dr. Walter Lehmacher
Gesellschaft für Strahlen- und Umweltforschung mbH, Institut für Medizinische Informatik und Systemforschung
Ingolstädter Landstr. 1, 8042 Neuherberg

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Krüger, Hans-Peter:

The fourfold table : up to N = 80

= Die Vierfeldertafel / by H.-P. Krüger ;

W. Lehmacher ; K.-D. Wall. – Stuttgart ; New York : Fischer, 1981.

(Statistical tables in the social and biological sciences)

ISBN 3-437-30320-1

NE: Lehmacher, Walter;; Wall, Klaus-Dieter;; GT

© Gustav Fischer Verlag · Stuttgart · New York · 1981

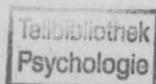
All rights reserved / Alle Rechte vorbehalten

Typesetting/Satz: Bauer und Bökeler Filmsatz KG, Denkendorf

Printing/Druck and Bookbinding/Einband: Graph. Großbetrieb Friedrich Pustet, Regensburg

Printed in Germany

ISBN 3-437-30320-1



Lehmacher LST III

Hinweis für eilige Anwender:

Bitte entnehmen Sie dem Buch die beiliegende Karte. Auf ihr sind alle Informationen zur Benutzung der Vertafelung zusammengestellt!

Note for hurried users:

Please refer to the card inserted in the book. It contains all information necessary for using the tables!

Inhaltsverzeichnis

Vorwort 1

1. Vorhandene Tafeln für den exakten Test in 2 × 2-Felder-Tafeln 3

2. Der exakte Test in 2 × 2-Felder-Tafeln 5

 2.1. Das Prinzip des exakten Testens 5

 2.2. Die Alternativhypothesen 8

 2.3. Die Prüfung auf Kontingenz 12

 2.4. Interpretation eines Ergebnisses 12

 2.5. Die Beschreibung der Tafeln 14

 2.6. Formeln für die asymptotische Prüfung . . . 16

3. Anwendungen 17

 3.1. Die Prüfung der Homogenität bei zwei unabhängigen Stichproben 17

 3.2. Die Prüfung auf Kontingenz 18

Literaturverzeichnis 20

Tafeln 21

Die Anwendung der Tafel 1-1 22

Tafel 1-1 von N = 3(1)80 für alle Möglichkeiten, wenn die einseitige Überschreitungswahrscheinlichkeit $P_1 \leq 0.10$ und die zweiseitige Überschreitungswahrscheinlichkeit $P_2 \geq 0.0001$ 25

Tafel 1-2 Tafel der Normalverteilung und der χ^2 -Verteilung mit 1 Freiheitsgrad 437

Contents

Preface 1

1. Available Tables for the Exact Test in 2 × 2-Fold Tables. 3

2. The Exact Test in 2 × 2-Fold Tables. 5

 2.1. Principle of Exact Testing. 5

 2.2. Alternative Hypotheses. 8

 2.3. Test against Contingency. 12

 2.4. Interpretation of Results. 12

 2.5. Description of the Tables. 14

 2.6. Formulas for the Asymptotic Test. 16

3. Applications 17

 3.1. Test for Homogeneity with Two Independent Samples. 17

 3.2. Test against Contingency. 18

References 20

Tables 21

How to use Table 1-1 23

Table 1-1 Beginning with N = 3(1)80 for all frequencies, when the one-tailed probability $P_1 \leq 0.1$ and the two-tailed probability $P_2 \geq 0.0001$ 25

Table 1-2 Table of normal distribution and χ^2 -distribution with one degree of freedom. . 437

Vorwort

Das Anwachsen quantitativer Forschung im sozial- und biowissenschaftlichen Bereich verschärft in der Datenanalyse drei Probleme:

- In größerem Umfang werden nur qualitativ zu fassende Variablen in die Untersuchungen einbezogen.
- Bei quantitativen Variablen wird häufig festgestellt, daß die Normalverteilungsannahme nicht oder nur unzureichend zutrifft. Das führt zu einer verstärkten Anwendung von verteilungsfreien Verfahren.
- Viele Untersuchungsgebiete zeichnen sich dadurch aus, daß nur kleine Stichproben zur Verfügung stehen, so daß auch nicht über den zentralen Grenzwertsatz die asymptotische Normalverteiltheit der Prüfgrößen angenommen werden kann.

Verteilungsfreie Verfahren zeichnen sich in ihrer exakten Version aber durch einen hohen Rechenaufwand aus. Ihre Approximationsgeschwindigkeit an stetige Prüfverteilungen ist oft relativ gering, so daß für kleine Stichprobenumfänge eigentlich grundsätzlich auf den exakten Test rekuriert werden muß. In dem vorliegenden Band ist die Vierfeldertafel, eines der praktisch wichtigsten Prüfinstrumente, bis zu einem Stichprobenumfang von $N = 80$ mit ein- und zweiseitigen Überschreitungswahrscheinlichkeiten vertafelt. Weitere wichtige Prüfverfahren sollen in den folgenden Bänden vertafelt werden.

Die Arbeit an den Tafeln ist überschattet von dem Tod unseres Kollegen und Freundes Dr. KLAUS-DIETER WALL, der am 11.3.1978 den Folgen eines Autounfalls erlag. Sein Anteil an diesen Tafeln ist nicht zu unterschätzen: sie tragen in Idee und Durchführung in weiten Teilen seine Handschrift.

Zu danken haben wir Herrn Prof. Dr. Dr. G. A. LIENERT, Nürnberg, der uns mit vielen Ratschlägen zur Seite stand. Ebenso danken wir Herrn Dr. J. DWYER von der State University of New York für seine Hinweise zur Übersetzung des Textes. Dem Verlag Gustav Fischer und dort Herrn Dr. VON LUCIUS sei für die Bereitschaft gedankt, eine solche Veröffentlichung zu unterstützen.

Nürnberg und München 1981

HANS-PETER KRÜGER
WALTER LEHMACHER

Preface

The growth of quantitative research in the social and biological sciences has made three problems of data analysis more acute:

- Variables which can only be measured qualitatively are increasingly involved in investigation.
- When using quantitative variables it is frequently found that the assumption of a normal distribution is either incorrect or not sufficiently accurate. This has led to increased use of distribution-free methods.
- Many areas of research are characterized by the availability of only small samples, and thus asymptotic normal distribution of a test statistic cannot be assumed on the basis of the central limit theorem.

Distribution-free methods have an exact version but require substantial computation. Continuous statistics like the chisquare converge relatively slowly to the exact probabilities as sample size increases. Thus, one must fall back on an exact test when using a small-sized sample. In the present volume the fourfold table, one of the most important testing instruments, is discussed for random sample sizes up to $N = 80$. Both one-tailed and two-tailed probabilities are provided. Significance tests for analysis of larger contingency tables will be presented in the subsequent volumes.

The work on these tables has been overshadowed by the loss of our colleague and friend Dr. KLAUS-DIETER WALL, who died of injuries from an auto accident on March 11, 1978. His contribution to this work cannot be underestimated. His hand can be seen in conceptual basis and implementation.

We are grateful to Prof. Dr. Dr. G. A. LIENERT, Nürnberg, for many helpful suggestions, also Dr. J. DWYER, State University of New York, for his help to translate the text. Our thanks go also to the publishers, Gustav Fischer Verlag, and their Dr. VON LUCIUS for the support of a publication of this type.

Nürnberg and München 1981

HANS-PETER KRÜGER
WALTER LEHMACHER

1. Vorhandene Tafeln für den exakten Test in 2×2 -Felder-Tafeln

Eine Fülle von statistischen Problemen ist direkt oder nach einfachen Umformungen mit der Vierfeldertafel anzugehen. Ihre Beliebtheit resultiert vor allem aus der einfachen Auswertung nach dem X^2 von PEARSON, das asymptotisch χ^2 -verteilt ist. Allerdings sollte dieser Test nur angewendet werden, wenn die Erwartungswerte hinreichend groß sind. COCHRAN (1954) und WISE (1963) nennen Werte zwischen 2 und 5 als untere Grenze. Diese Grenzen sind aber nur dann realistisch, wenn die Erwartungswerte etwa gleich groß sind (WALL, 1972). Sind diese (oftmals zu optimistischen) Bedingungen nicht gegeben, muß auf den exakten Test rekurriert werden, der unter dem Namen FISHER-YATES-Test bekannt wurde (FISHER, 1935, YATES, 1934). Dessen hoher Rechenaufwand führte schnell zu einer Vertafelung für kleinere Stichprobenumfänge. Den Anfang machte FINNEY (1948) mit einer Vertafelung bis $N = 30$ (die Tafeln sind zu finden in den BIOMETRIKA TABLES von PEARSON u. HARTLEY, 1969). Es folgten MAINLAND u. MURRAY (1952), LATSCHA (1953), ARMSSEN (1955). BROSS u. KASTEN (1957) approximierten die hypergeometrische Verteilung über eine Binomialverteilung und konnten so Werte für Randsummen < 50 angeben. Ebenfalls anwendbar sind die Tafeln von LIEBERMANN u. OWEN (1961) über die hypergeometrische Verteilung. BENNETT und HORST (1966) erweiterten die Tafeln von FINNEY, LATSCHA, BENNETT und HSU (1963) bis $N = 100$. HENZE (1976) vertafelte in einer neuen, noch unveröffentlichten Arbeit bis $N = 10000$ in dem Stichprobenumfang angepaßten Sprüngen von N . Abweichend von FISHERS Procedere vertafelten MCDONALD, DAVIS und MILLIKEN (1977) einen nicht-randomisierten Test bis $N = 30$.

Trotz dieses großen Angebots an Tafeln werden diese in der Praxis wenig genutzt. Dafür gibt es u. E. zwei Gründe:

- a. die meisten Tafeln beschränken sich auf zwei oder mehrere α -Niveaus, wobei meist auch nur die einseitigen Überschreitungswahrscheinlichkeiten vertafelt sind,
- b. die beobachtete Vierfeldertafel muß mehrfach umgeordnet werden, bevor in das Tafelwerk eingegangen werden kann.

Beide Gründe resultieren aus dem begrenzten Raum, der für eine Vertafelung zur Verfügung steht. Durch Umordnung kann der Platzbedarf bis auf rund ein Achtel reduziert werden. Noch drastischer ist die Ersparnis bei der Einführung von Niveaus statt der Vertafelung der exakten P-Werte. Auch das vorliegende Tafelwerk zollt dem Buchumfang seinen Tribut: der Anwender muß die beobachtete Vierfeldertafel in der Regel umordnen.

1. Available Tables for the Exact Test in 2×2 -Fold Tables

A wealth of statistical problems can be approached with the fourfold table. The popularity of this table can be attributed to the ease of analysis based PEARSON'S X^2 distribution. This test should only be used, however, when the expected cell frequencies are sufficiently high. COCHRAN (1954) and WISE (1963) give values between 2 and 5 as a lower bound. Such requirement is only realistic, however, when the expected values are about the same size (WALL, 1972). If these (often too optimistic) conditions are not satisfied, one must fall back on the exact test which has become known under the name FISHER-YATES test (FISHER, 1935; YATES, 1934). The large amount of computation involved in the test quickly led to the working out of tables for small sample sizes. FINNEY (1948) opened the way with tables up to $N = 20$ (the tables can be found in the Biometrika tables of PEARSON and HARTLEY, 1969). He was followed by MAINLAND and MURRAY (1952), LATSCHA (1953), and ARMSSEN (1955). BROSS and KASTEN (1957) approximated the hypergeometric distribution using a binomial distribution and were thus able to provide values for marginal sample sizes < 50 . The tables of LIEBERMANN and OWEN (1961) on the hypergeometric distribution are also useful. BENNETT and HORST (1966) expanded the tables of FINNEY, LATSCHA, BENNETT and HSU (1963) up to $N = 100$. In a new, unpublished paper HENZE (1976) worked out tables up to $N = 10000$ with steps of N corresponding to sample size. Deviating from FISHER'S procedure, MCDONALD, DAVIS and MILLIKEN (1977) presented tables for a nonrandomized test up to $N = 30$.

Despite the large supply of tables, they remain little used in practice. Two of the main reasons for this are that:

- a. Most tables are limited to two or more α -levels where only one-sided probabilities are usually given.
- b. The observed fourfold tables must be frequently rearranged before actual use can be made of them.

Both of the above reasons result from the limited space available for printing the tables. Rearrangement permits approximately an eightfold reduction in the amount of required printing space. The savings are even more pronounced when levels are used instead of tabling the exact P values. The tables in this volume could also not be presented without some consideration for limitations of space: the user will generally have to rearrange the observed fourfold table.

- Dafür bietet es im Gegensatz zu den anderen Tafeln
- die exakten Überschreitungswahrscheinlichkeiten unter Verzicht auf vorgegebene α -Niveaus
 - die Angabe von zweiseitigen Überschreitungswahrscheinlichkeiten im Sinne von FREEMAN u. HALTON (1951).

Aufgenommen in die Tabellierung sind alle Vierfeldertafeln, deren

- einseitige Überschreitungswahrscheinlichkeit nicht größer als 10 % ist und
- deren zweiseitige Überschreitungswahrscheinlichkeit gleich oder größer 0.0001 (die Stellengenauigkeit der Vertafelung) ist.

Durch diesen Genauigkeitsgrad ist die Vertafelung vor allem dann auch noch brauchbar, wenn (etwa bei mehreren simultanen Tests) mit der α -Adjustierung gemäß der BONFERRONI-Ungleichung gerechnet werden muß (siehe dazu etwa MILLER, 1966).

Zur Frage randomisiertes vs. nicht-randomisiertes Testen

Ist eine Prüfgröße T (wie etwa beim Vierfeldertest) diskret verteilt, so existiert nicht zu jedem Niveau α ein kritischer Wert $t_{(\alpha)}$, für den unter der Nullhypothese die Überschreitungswahrscheinlichkeit *genau* α ist. In der Regel wählt man in dieser Situation zu einem vorgegebenen α als kritischen Wert das kleinste $t_{(\alpha')}$, für das die Überschreitungswahrscheinlichkeit α' kleiner gleich α ist. Das Ergebnis ist in der Regel ein *konservativer* Test, da das faktische Niveau α' kleiner ist als das nominelle (vorgewählte) α .

Einen Ausweg bietet das Prinzip des *randomisierten Testens* (zu seiner Begründung siehe etwa WITTING, 1974, 173 ff.). Bezeichnet $t_{(\alpha'')}$ den größten Wert t mit $t_{(\alpha'')} < t_{(\alpha')}$ (d. h. den nächstkleineren Nachbarn von $t_{(\alpha')}$), dann wird in dem Fall, wo $t_{(\alpha'')}$ angenommen wird, die Entscheidung für oder gegen die Nullhypothese von einem zusätzlichen Zufallsexperiment abhängig gemacht (Randomisation). Hat $t_{(\alpha'')}$ unter der Nullhypothese die Punktwahrscheinlichkeit $p_0 (= \alpha'' - \alpha')$, so muß dabei die Wahrscheinlichkeit γ , mit der bei Vorliegen der Beobachtung von $t_{(\alpha'')}$ die Entscheidung gegen H_0 getroffen wird, so gewählt werden, daß gilt: $\gamma p_0 + \alpha' = \alpha$. Durch dieses Verfahren wird das gesamte Niveau α ausgeschöpft. Analoge Prozeduren sind beim zweiseitigen Test zu konstruieren. Wendet man dieses Prinzip auf den Test bei Vierfeldertafeln an (sog. randomisierte FISHER-YATES-Tests), so lassen sich gleichmäßig beste unverfälschte Tests konstruieren. Entsprechende Vertafelungen finden sich zu den Niveaus von 10 %, 5 % und 1 % bis $N = 50$ bei SIEGERSTETTER u. WELZL (1975).

Das Konzept des randomisierten Testens mag zwar theoretisch befriedigen, da es zu schärferen Tests führt, stößt beim Praktiker aber auf Schwierigkeiten, da die Testent-

- In contrast to these other tables, this book offers
- exact probabilities of deviations from the null hypothesis, thus foregoing the use of preset α -levels.
 - two-sided probabilities of deviations from the null hypothesis along the lines of FREEMAN and HALTON (1951).

The tabulation includes all fourfold tables whose

- one-sided probability of deviation from the null hypothesis does not exceed 10 %, and whose
- two-sided probability of deviation from the null hypothesis is equal or exceeds 0.0001 (the accuracy of the entries in the table).

This degree of precision makes the table even usable when (e. g., with several simultaneous tests) calculations must be made using the α -adjustment based on the BONFERRONI inequality (c. f., MILLER, 1966, for instance).

On the Question of Randomized vs. Nonrandomized Testing

If a test statistic T (e. g., in the fourfold test) is discretely distributed, then at every level α a critical value $t_{(\alpha)}$ does not exist for which the probability under the null hypothesis is *exactly* α . For a given α in this situation one generally selects as critical value the smallest $t_{(\alpha')}$ for which the probability α' is less than or equal to α . The result is generally a *conservative* test, as the actual level α' is less than the nominal α .

A way out is offered by the principle of *randomized testing* (for substantiation, c. f., WITTING, 1974, 173 ff.). If $t_{(\alpha'')}$ is the largest value of t where $t_{(\alpha'')} < t_{(\alpha')}$ (i. e., the smallest neighbor to $t_{(\alpha')}$), then in the case where $t_{(\alpha'')}$ is accepted, the decision for or against the null hypothesis is made dependent on an additional experiment of chance (randomization). If $t_{(\alpha'')}$ has the point probability $p_0 (= \alpha'' - \alpha')$ under the null hypothesis, then the probability γ with which the decision against H_0 is made when $t_{(\alpha'')}$ is observed must be chosen such that $\gamma p_0 + \alpha' = \alpha$. The entire α -level is exhausted with this procedure. Analogous procedures should be set up for the two-sided tests. Application of this principle to the test with fourfold tables (randomized FISHER-YATES tests) permits construction of uniformly most powerful unbiased tests. Such tables for the 10 %, 5 % and 1 % levels up to $N = 50$ can be found in SIEGERSTETTER and WELZL (1975).

The concept of randomized testing may be theoretically satisfying because it leads to more powerful tests. It causes problems in practice, however, because the test decision depends not only on the data, but also on the outcome of an additional experiment of chance. Or consider the way BERKSON (1977, 41) put it: «But what investigating scientist would decide which of two drugs is the more effective by tossing a coin?» This remark refers to FISHER's randomizing technique in which the marginal values are considered as

scheidung nicht nur von seinen Daten, sondern auch vom Ausgang eines zusätzlichen Zufallsexperiments abhängt. Oder wie BERKSON (1977, 41) es ausdrückt: «But what investigating scientist would decide which of two drugs is the more effective by tossing a coin?» Diese Bemerkung bezieht sich auf die Randomisierungstechnik FISHERS, wobei die Randsummen als «ancillary» betrachtet werden. Vor allem PEARSON (1947) hat dagegen Einwände erhoben. Andere Tests wurden vorgeschlagen (z. B. von PLACKETT, 1977, McDONALD, DAVIS und MILLIKEN, 1977). Letzten Endes verbirgt sich hinter der Diskussion die Frage, ob man FISHERS Haltung oder die von NEYMAN-PEARSON einnimmt, wonach α genau die Ablehnungswahrscheinlichkeit angibt.

Vor allem aus praktischen Gründen haben wir uns in dieser Vertafelung auf den nicht-randomisierten Test beschränkt. Er ist in aller Regel konservativer. Bezieht man jedoch die Überlegung ein, daß die Unterscheidung randomisiert-, nichtrandomisiert vor allem bei kleinen Stichprobenumfängen relevant wird, erscheint ein prinzipiell eher konservatives Vorgehen zu rechtfertigen.

2. Der exakte Test in 2×2 -Feldertafeln

Vor der Darstellung der Tafeln und ihrer Anwendung wird das Prinzip des exakten Testens exemplarisch am Beispiel der Prüfung zweier Stichproben auf Homogenität dargestellt.

2.1. Das Prinzip des exakten Testens

In einem Experiment soll geklärt werden, ob ein neues aktivierendes Medikament Ratten schneller ein Labyrinth lernen läßt. Dazu wird eine Gruppe von $N = 15$ Ratten zufällig auf eine Experimentalgruppe ($X+$) mit $N_1 = 7$ Ratten und eine Kontrollgruppe ($X-$) mit $N_2 = 8$ Ratten aufgeteilt. Lediglich die Experimentalgruppe erhält das Medikament. Tabelle 1 zeigt den Ausgang des Experiments.

«ancillary». PEARSON (1947) mainly had objections to this. Other tests have been proposed (e. g., from PLACKETT, 1977; McDONALD, DAVIS and MILLIKEN, 1977). In the final analysis the discussion comes down to the question of whether one supports FISHER's standpoint, or that of NEYMAN-PEARSON which holds that α represents the exact probability of rejecting the null hypothesis when it is true.

We limited the tables in this volume to the nonrandomized test primarily for practical reasons. It is more conservative as a rule. When one considers that differentiation between randomized and nonrandomized is primarily relevant in samples of small size, a principally more conservative procedure seems to be justified.

2. The Exact Test in 2×2 -Fold Tables

Before presenting the tables and describing how they can be used, the principle of exact testing will be illustrated with an example involving a test for homogeneity of two samples.

2.1. Principle of Exact Testing

An experiment is used to clarify whether or not a new stimulatory drug enables rats to learn a maze more quickly. A group of $N = 15$ rats is randomly divided into an experimental group ($X+$) with $N_1 = 7$ rats and a control group ($X-$) with $N_2 = 8$ rats. The medication is administered only to the experimental group. Table 1 shows the outcome of the experiment.

Tabelle 1

		gelernt Y+	nicht gelernt Y-	Σ
Experimentalgruppe	X+	5 = a	2 = b	7 = N ₁
Kontrollgruppe	X-	1 = c	7 = d	8 = N ₂
	Σ	6 = M ₁	9 = M ₂	15 = N

Y+ bedeutet «in einer definierten Zeit durch das Labyrinth gekommen», Y- bedeutet «Mißerfolg». ¹ Die Nullhypothese H₀ lautet: Das Medikament erbringt keine Veränderung der Lernleistung, während die Alternativhypothese H₁ eine Veränderung der Lernleistung postuliert (die verschiedenen Formen von H₁ werden weiter unten besprochen).

Das Prinzip des exakten Testens besteht darin, alle möglichen Ausgänge eines solchen Experiments in die Entscheidung für oder gegen H₀ einzubeziehen.

P_{X+} bezeichne die Wahrscheinlichkeit, daß in der Experimentalgruppe X+ mit N₁ = 7 genau a = 5 Ratten das Labyrinth lernten.

Nach der Binomialverteilung berechnet sich

$$P_{X+} = \binom{N_1}{a} \pi_{X+}^a (1 - \pi_{X+})^{N_1-a} \quad (1)$$

π_{X+} bezeichnet die im allgemeinen unbekannte Wahrscheinlichkeit, daß eine Ratte der Experimentalgruppe X+ das Labyrinth lernt. Ebenso berechnet sich die Wahrscheinlichkeit P_{X-}, daß in der Kontrollgruppe X- mit N₂ = 8 genau c = 1 Ratte das Labyrinth lernt nach

$$P_{X-} = \binom{N_2}{c} \pi_{X-}^c (1 - \pi_{X-})^{N_2-c} \quad (2)$$

Die Wahrscheinlichkeit, genau das beobachtete Ergebnis in der Experimentalgruppe X+ und in der Kontrollgruppe X- zu erhalten, berechnet sich nach dem Multiplikationssatz der Wahrscheinlichkeit, da beide Gruppen voneinander unabhängig sind:

$$P_{X+P_{X-}} = \binom{N_1}{a} \pi_{X+}^a (1 - \pi_{X+})^{N_1-a} \cdot \binom{N_2}{c} \pi_{X-}^c (1 - \pi_{X-})^{N_2-c} \quad (3)$$

¹ Üblicherweise wird die Vierfeldertafel verschieden notiert, je nachdem, ob auf Homogenität oder Zusammenhang geprüft wird (siehe z. B. LIENERT, 1973). Da unsere Tabellen für beide Fragestellungen anwendbar sind, haben wir uns für eine gemeinsame Notation mit den bekannten Namen für die Zellfrequenzen a, b, c und d entschieden.

Table 1

		learned Y+	not learned Y-	Σ
experimental group	X+	5 = a	2 = b	7 = N ₁
control group	X-	1 = c	7 = d	8 = N ₂
	Σ	6 = M ₁	9 = M ₂	15 = N

Y+ means «passed through the maze within a set period of time»; Y- means «failure». ¹ The null hypothesis H₀ says that the medication does not yield an alteration in learning performances, whereas the alternative hypothesis H₁ postulates a change in learning performance (the various forms of H₁ will be discussed later on).

The principle of exact testing consists of including all possible outcomes of such an experiment in the decision for or against H₀.

P_{X+} stands for the probability that exactly a = 5 rats in the experimental group X+ with N₁ = 7 learn the maze. According to binomial distribution

$$P_{X+} = \binom{N_1}{a} \pi_{X+}^a (1 - \pi_{X+})^{N_1-a} \quad (1)$$

π_{X+} represents the generally unknown probability that a rat from the experimental group X+ learns the maze. In the same manner the probability that exactly c = 1 rat in the control group X- with N₂ = 8 rats learns the maze is calculated by

$$P_{X-} = \binom{N_2}{c} \pi_{X-}^c (1 - \pi_{X-})^{N_2-c} \quad (2)$$

The probability of finding exactly the outcome observed in the experimental group X+ and in the control group X- is the product of both of the above probabilities because both groups are independent of each other:

$$P_{X+P_{X-}} = \binom{N_1}{a} \pi_{X+}^a (1 - \pi_{X+})^{N_1-a} \cdot \binom{N_2}{c} \pi_{X-}^c (1 - \pi_{X-})^{N_2-c} \quad (3)$$

¹ Generally the notation of the fourfold table varies depending on whether homogeneity or contingency is being tested (c.f., LIENERT, 1973). As our tables can be used for both questions, we have decided on a joint notation with the known cell frequencies labelled a, b, c and d.

H_0 postuliert, daß die beiden Wahrscheinlichkeiten π_{X+} und π_{X-} sich nicht unterscheiden: unter H_0 hat das Medikament keinen Effekt. Berücksichtigt man diese Annahme und benützt die Beziehung $N_1 - a = b$, $N_2 - c = d$, $a + c = M_1$ und $b + d = M_2$, verkürzt sich Formel (3) zu Formel (4)

$$P_{X+P_{X-}} = \binom{N_1}{a} \binom{N_2}{c} \pi^{M_1} (1-\pi)^{M_2} \quad (4)$$

Vereinigt man die beiden Stichproben, haben $M_1 = 6$ Ratten das Labyrinth gelernt, $M_2 = 9$ scheiterten. Die Wahrscheinlichkeit P_T ($T = \text{Total}$), genau dieses Ergebnis zu erhalten, ist mit Formel (5) wieder durch die Binomialverteilung gegeben:

$$P_T = \binom{N}{M_1} \pi^{M_1} (1-\pi)^{M_2} \quad (5)$$

π_T bezeichnet die unbekannte Wahrscheinlichkeit, daß eine Ratte (gleichgültig, ob aus $X+$ oder $X-$) das Labyrinth lernt. Diese Wahrscheinlichkeit ist bei Gültigkeit von H_0 genau dieselbe wie in den beiden Stichproben: $\pi_{X+} = \pi_{X-} = \pi_T$. Mit p° bezeichnen wir jetzt die Wahrscheinlichkeit, daß M_1 erfolgreiche und M_2 erfolglose Ratten *zufällig* sich so in zwei Stichproben des Umfangs N_1 und N_2 aufteilen, daß in der einen Stichprobe $X+$ insgesamt a erfolgreiche, in der zweiten Stichprobe $M_1 - a = c$ erfolgreiche Ratten sich befinden. Die a bzw. c erfolgreichen Fälle werden dabei *unter der Bedingung* betrachtet, daß es insgesamt M_1 erfolgreiche Fälle gibt und die Stichproben gerade die Größen N_1 und N_2 haben. Dieses p° berechnet sich als bedingte Wahrscheinlichkeit aus dem Quotienten

$$p^\circ = \frac{P_{X+P_{X-}}}{P_T} = \frac{\binom{N_1}{a} \binom{N_2}{c} \pi^{M_1} (1-\pi)^{M_2}}{\binom{N}{M_1} \pi^{M_1} (1-\pi)^{M_2}} = \frac{\binom{N_1}{a} \binom{N_2}{c}}{\binom{N}{M_1}} \quad (6)$$

Wie Formel (6) zeigt, kürzt sich die unbekannte Wahrscheinlichkeit π heraus: die Wahrscheinlichkeit p° (also genau das beobachtete Ergebnis zu erhalten), ist keine Funktion der Wahrscheinlichkeit in der Population mehr, sondern kann unter Gültigkeit von H_0 aus den Beobachtungswerten berechnet werden.

Berücksichtigt man die Definition

$$\binom{N}{x} = \frac{N!}{x!(N-x)!}$$

wird aus Formel (6) Formel (7)

$$p^\circ = \frac{N_1! N_2! M_1! M_2!}{N! a! b! c! d!} \quad (7)$$

H_0 postulates that both probabilities π_{X+} and π_{X-} do not differ: according to H_0 the medication has no effect. If one makes this assumption and uses the relation $N_1 - a = b$, $N_2 - c = d$, $a + c = M_1$ and $b + d = M_2$, equation (3) can be shortened to (4):

$$P_{X+P_{X-}} = \binom{N_1}{a} \binom{N_2}{c} \pi^{M_1} (1-\pi)^{M_2} \quad (4)$$

If one combines both samples, $M_1 = 6$ rats have learned the maze, $M_2 = 9$ have failed. The probability P_T ($T = \text{total}$) of obtaining exactly this outcome is provided in equation (5), again from the binomial distribution:

$$P_T = \binom{N}{M_1} \pi^{M_1} (1-\pi)^{M_2} \quad (5)$$

π_T is the unknown probability that a rat (regardless of whether from $X+$ or $X-$) learns the maze. When H_0 is true this probability is exactly the same in both samples: $\pi_{X+} = \pi_{X-} = \pi_T$. We now let p° represent the probability that M_1 successful and M_2 nonsuccessful rats are *accidentally* distributed in two samples of size N_1 and N_2 such that one sample $X+$ contains a total of a successful rats and the second sample contains a total of $M_1 - a = c$ successful rats. Let us consider the a and c successful cases *under the condition* that there is a total of M_1 successful cases and the samples have exactly the sizes N_1 and N_2 . The above mentioned p° can be calculated as the conditional probability from the following quotient:

$$p^\circ = \frac{P_{X+P_{X-}}}{P_T} = \frac{\binom{N_1}{a} \binom{N_2}{c} \pi^{M_1} (1-\pi)^{M_2}}{\binom{N}{M_1} \pi^{M_1} (1-\pi)^{M_2}} = \frac{\binom{N_1}{a} \binom{N_2}{c}}{\binom{N}{M_1}} \quad (6)$$

As equation (6) shows, the unknown probability π cancels out. The probability p° (of obtaining precisely the observed outcome) is no longer a function of the probability in the population. Instead, it can be calculated from the observed values provided H_0 is valid.

If we apply the definition

$$\binom{N}{x} = \frac{N!}{x!(N-x)!}$$

then we obtain equation (7) from equation (6)

$$p^\circ = \frac{N_1! N_2! M_1! M_2!}{N! a! b! c! d!} \quad (7)$$

p° wird als «Punktwahrscheinlichkeit» bezeichnet und berechnet sich für das Beispiel in Tabelle 1 zu

$$p^\circ = \frac{7!8!6!9!}{15!5!2!1!7!} = 0.03357$$

Jeder anderen Kombination von Zellenbesetzungen, die den Bedingungen der Randsummen N_1, N_2, M_1 und M_2 gehorcht, läßt sich ebenfalls eine Punktwahrscheinlichkeit p° zuschreiben. Sie gehen im einzelnen aus Tabelle 2 hervor.

Tabelle 2

a	b	c	d	p°	P_1	P_2
0	7	6	2	.00559	.00559	.00699
1	6	5	3	.07832	.08391	.11888
2	5	4	4	.29371	.37762	.60839
3	4	3	5	.39161	.62238	1.00000
4	3	2	6	.19580	.23077	.31468
5	2	1	7	.03357	.03497	.04056
6	1	0	8	.00140	.00140	.00140

(P_1 und P_2 werden im nächsten Abschnitt behandelt).

p° is referred to as the «point probability» and amounts to the following for the example in Table 1.

$$p^\circ = \frac{7!8!6!9!}{15!5!2!1!7!} = 0.03357$$

Every other combination of cell frequencies which satisfies the conditions for the marginal values N_1, N_2, M_1 and M_2 can also be given a point probability p° . The individual cases are presented in Table 2.

Table 2

a	b	c	d	p°	P_1	P_2
0	7	6	2	.00559	.00559	.00699
1	6	5	3	.07832	.08391	.11888
2	5	4	4	.29371	.37762	.60839
3	4	3	5	.39161	.62238	1.00000
4	3	2	6	.19580	.23077	.31468
5	2	1	7	.03357	.03497	.04056
6	1	0	8	.00140	.00140	.00140

(P_1 and P_2 will be treated in the next section).

2.2. Die Alternativhypothesen

Lediglich zwei Punktwahrscheinlichkeiten (für $a = 0$ und $a = 6$) sind kleiner als das $p^\circ = .03357$ für $a = 5$ in unserem Beispiel. Das Problem der Signifikanzbestimmung eines beobachteten Ergebnisses muß gemäß den Hypothesen stattfinden, die der Experimentator aufgestellt hat. Die Beschreibung der verschiedenen Hypothesen wird deutlicher, wenn man von den relativen statt von den absoluten Häufigkeiten in der Vierfeldertafel ausgeht, die ja nicht direkt vergleichbar sind. Das Grundmuster einer Vierfeldertafel im Fall des Vergleichs zweier Stichproben zeigt Tabelle 3, in der die Wahrscheinlichkeiten und die entsprechenden relativen Häufigkeiten eingetragen sind.

Tabelle 3

	Erfolg	kein Erfolg	
Stichprobe 1	$\pi_E = (a/N_1)$	$1 - \pi_E = (1 - a/N_1)$	1
Stichprobe 2	$\pi_C = (c/N_2)$	$1 - \pi_C = (1 - c/N_2)$	1

2.2. Alternative Hypotheses

Only two point probabilities (for $a = 0$ and $a = 6$) are lower than $p^\circ = 0.03357$ for $a = 5$ in our example. The problem of determining the significance of an observed outcome must take into account the hypotheses set up by the experimenter. The description of the various hypotheses becomes clearer if one starts from the relative, instead of the absolute, frequencies in the fourfold table, which of course are not comparable. The basic make up of a fourfold table in a comparison of two samples is presented in Table 3 which contains the probabilities and the corresponding relative frequencies.

Table 3

	success	no success	
sample 1	$\pi_E = (a/N_1)$	$1 - \pi_E = (1 - a/N_1)$	1
sample 2	$\pi_C = (c/N_2)$	$1 - \pi_C = (1 - c/N_2)$	1

Drei verschiedene Alternativhypothesen sind möglich:

a. **Einseitig:** die Experimentalgruppe zeigt eine gerichtete Änderung.

Zum Beispiel: das Medikament erhöht die Leistung

$$\pi_E > \pi_C \quad \text{oder} \quad \pi_E < \pi_C$$

b. **Zweiseitig:** die Experimentalgruppe zeigt eine Wirkung, die positiv *oder* negativ sein kann

$$\pi_E \neq \pi_C$$

Die Entscheidung über die Signifikanz eines Ergebnisses ist für jeden der drei Fälle anders.

Die einseitige Alternative

In diesem Fall bezieht sich die Alternativhypothese auf ein a priori vorhandenes Wissen und spezifiziert ein bestimmtes Ergebnis, entweder daß π_E größer ist als π_C oder daß π_E kleiner ist als π_C . In unserem Beispiel wäre es plausibel anzunehmen, daß das aktivierende Medikament die Wahrscheinlichkeit eines positiven Ausgangs erhöht. Das heißt, die Experimentalgruppe sollte höhere Lernleistungen zeigen als die Kontrollgruppe: $\pi_E > \pi_C$. Bei gegebenen Randsummen ist nur ein Ergebnis extremer als das beobachtete (siehe Tabelle 2): nämlich dasjenige, bei dem alle $M_1 = 6$ Ratten mit Erfolg aus der Experimentalgruppe gekommen wären ($a = 6$). Dieses Ereignis hat die Punktwahrscheinlichkeit $p^\circ = 0.00140$ (Tabelle 2, Spalte p°). Die Wahrscheinlichkeit eines Ergebnisses, das gleich oder extremer ist als das beobachtete, ist unter der einseitigen Alternativhypothese dann gegeben durch

$$P_1 = 0.03357 + 0.00140 = 0.03497$$

(Tabelle 2, Spalte P_1) (das Subskript «1» bedeutet «einseitig»). P_1 wird manchmal «obere einseitige Überschreitungswahrscheinlichkeit» genannt. Im Fall einer Hypothese $\pi_E < \pi_C$ gibt es folglich auch eine «untere einseitige Überschreitungswahrscheinlichkeit». Wie das berechnete P_1 zeigt, könnten wir auf einem vorher vereinbarten Signifikanzniveau von $\alpha = 5\%$ davon ausgehen, daß das Medikament die Lernleistung von Ratten steigert.

Um die einseitige Überschreitungswahrscheinlichkeit zu bestimmen, werden also alle Punktwahrscheinlichkeiten p° derjenigen Felderkombinationen aufsummiert, die gleich oder extremer sind als die beobachtete in Richtung auf die vorher festgelegte Alternativhypothese. Das Ergebnis wird in unseren Tafeln als «einseitige Überschreitungswahrscheinlichkeit» P_1 bezeichnet.

Die zweiseitige Alternative

Gibt es keine a priori-Information über die Richtung, hat sich der Untersucher für eine zweiseitige Alternativhypothese zu entscheiden. Im parametrischen Fall definiert ein vorgegebenes Signifikanzniveau α an jedem Ende der Prüfverteilung einen Ablehnungsbereich der Größe $\alpha/2$. Beim exakten Testen ist die Prüfverteilung aber diskret und in aller Re-

Three different alternative hypotheses are possible:

a. **One-sided:** the experimental group reveals a directional change.

For example: the drug increases performance.

$$\pi_E > \pi_C \quad \text{or} \quad \pi_E < \pi_C$$

b. **Two-sided:** The experimental group reveals an effect which can either be positive *or* negative.

$$\pi_E \neq \pi_C$$

The decision regarding the significance of an outcome is different in each of the three cases.

The one-sided alternative

In this case the alternative hypothesis is based on a priori existing information and specifies a certain result, either that π_E is higher than π_C , or that π_E is lower than π_C . In our example it would be plausible to assume that the stimulating drug increases the probability of a positive outcome. In other words, the experimental group should reveal better learning performances than the control group: $\pi_E > \pi_C$. With given marginal values only one outcome is more extreme than the observed outcome (see Table 2): namely that outcome in which all $M_1 = 6$ rats from the experimental group are successful ($a = 6$). Such an outcome has a point probability $p^\circ = 0.00140$ (Table 2, column p°). The probability of an outcome which is equal to or more extreme than the observed result under the one-sided alternative hypothesis is given by

$P_1 = 0.03357 + 0.00140 = 0.03497$ (Table 2, column P_1)

(the subscript «1» stands for «one-sided»). P_1 is often referred to as «upper one-sided probability». In the case of a $\pi_E < \pi_C$ hypothesis there is thus also a «lower one-sided probability». As shown by the calculated value of P_1 , we could assume, based on a previously established level of significance of $\alpha = 5\%$, that the drug increased the learning performance of rats.

The one-sided probability is thus determined by adding up all point probabilities p° of those field combinations which are equal to or more extreme than the observed combination in terms of the previously set down hypothesis. The result is termed «one-sided probability» in our tables.

The two-sided alternative

If there is no a priori information on the direction, the investigator must decide on a two-sided alternative hypothesis. In a parametric case a previously prescribed significance level α defines a rejection region of the size $\alpha/2$ at each end of the distribution of the test statistic. In an exact test, however, this distribution is discrete and as a rule asymmetrical.

gel asymmetrisch. Wie KRÜGER (1979) gezeigt hat, stehen für diesen Fall drei Möglichkeiten zur Bildung von zweiseitigen Überschreitungswahrscheinlichkeiten zur Verfügung:

a. das Verfahren nach FREEMAN und HALTON (1951). Die Autoren empfehlen, alle Punktwahrscheinlichkeiten p° aufzusummieren, die kleiner oder gleich der Punktwahrscheinlichkeit der beobachteten Tafel sind. Bei der Prüfung auf Homogenität werden damit in der Regel alle Tafeln in die Ablehnungsregion von H_0 einbezogen, deren Anteilsdifferenzen $a/N_1 - c/N_2 = \pi_E - \pi_C$ größer oder gleich der beobachteten Anteilsdifferenz sind. Im Fall der Kontingenzprüfung werden in der Regel alle Tafeln einbezogen, deren Kontingenzkoeffizient ϕ größer oder gleich dem ϕ der beobachteten Tafel ist.

b. die «Niveau-Verfahren», die bei stetigen Prüfverteilungen die Verfahren der Wahl sind. Dabei sind zwei Alternativen zu sehen:

b₁: das $2P_1$ -Verfahren. Man bestimmt die einseitige Überschreitungswahrscheinlichkeit einer Tafel und prüft diese gegen ein Niveau von $\alpha/2$. Rechentechisch äquivalent ist eine Verdoppelung von P_1 und eine Prüfung gegen α . Dieses Verfahren wird häufig in der Literatur angegeben (z. B. bei PEARSON u. HARTLEY, 1969, 73 und bei OWEN, 1962, 480).

b₂: das $\alpha/2$ -Verfahren. An beiden Enden der Prüfverteilung werden von den Extremen her die Punktwahrscheinlichkeiten solange aufsummiert, bis jeweils das $\alpha/2$ eines vorgegebenen Signifikanzniveaus möglichst nah erreicht, nicht aber überschritten wird. Die unterhalb dieser Grenzen liegenden Tafeln bilden die Ablehnungsregion. Überschreitet die kleinste Punktwahrscheinlichkeit auf einer Seite der Verteilung bereits dieses $\alpha/2$, muß H_0 beibehalten werden. In aller Regel sind die beiden Verfahren b₁ und b₂ identisch. Sie unterscheiden sich lediglich in dem Fall, daß die kleinste Punktwahrscheinlichkeit bereits größer als $\alpha/2$ ist. Dann ist nach b₂ kein zweiseitiger Test möglich. Als Konsequenz muß H_0 beibehalten werden.

Diese drei Verfahren der Bestimmung von zweiseitigen Überschreitungswahrscheinlichkeiten können für dieselben Vierfeldertafeln einer gegebenen Randverteilung zu verschiedenen Entscheidungen führen. Das geht aus dem Beispiel der Tabelle 1 hervor, dessen Punktwahrscheinlichkeit in Tabelle 2, Spalte p° verzeichnet sind. In Spalte P_2 stehen die Überschreitungswahrscheinlichkeiten nach FREEMAN-HALTON (oben Verfahren a.). Zu der Punktwahrscheinlichkeit einer Tafel wurden jeweils alle Punktwahrscheinlichkeiten hinzuaddiert, die kleiner oder gleich waren. Beispiel: die Überschreitungswahrscheinlichkeit P_2 für $a = 1$ ergibt sich als Summe der Punktwahrscheinlichkeiten $p_{a=0} + p_{a=1} + p_{a=5} + p_{a=6} = .00559 + .07832 + .03357 + .00140 = .11888$. Zwei Beispiele mögen die Verschiedenheit der Bildung von zweiseitigen Überschreitungswahrscheinlichkeiten verdeutlichen:

As shown by KRÜGER (1979), there are three possibilities for generating two-sided probabilities.

a. The method according to FREEMAN and HALTON (1951). The authors recommend adding up all point probabilities p° which are smaller than or equal to the point probability of the observed table. The test for homogeneity will thereby generally include all tables in the rejection region of H_0 whose proportional differences $a/N_1 - c/N_2 = \pi_E - \pi_C$ are greater than or equal to the observed proportional difference. When testing against contingency all tables generally are included for which the coefficient of contingency ϕ is greater than or equal to the ϕ of the observed table.

b. The «level methods» which are the preferred approach for continuous statistics: Two types are recognizable:

b₁: The $2P_1$ method. One determines the one-sided probability in a table and checks this against a level of $\alpha/2$. Doubling P_1 and checking against α is statistically equivalent. This method is cited often in the literature (e. g., in PEARSON and HARTLEY, 1969, 73; and in OWEN, 1962, 480).

b₂: The $\alpha/2$ method. The point probabilities are added together starting from the extremes at both ends of the sampling distribution, until in both cases $\alpha/2$ approaches a given significance level as closely as possible but does not exceed this level. The tables below this border constitute the rejection region. If the lowest point probability on one side of the distribution already exceeds $\alpha/2$, then H_0 must be accepted. Generally methods b₁ and b₂ provide identical results. They differ merely in the case that the smallest point probability is already higher than $\alpha/2$. In such a case a two-sided test with b₂ is impossible. The consequence is that H_0 must be accepted.

These three methods of determining two-sided probabilities can lead to different decisions with a given marginal distribution when using the same fourfold tables. This can be drawn from the example of Table 1, the point probabilities of which are shown in column p° of Table 2. The P_2 column shows the probabilities according to FREEMAN-HALTON (method a. above). All point probabilities which were less than or equal to the point probabilities of a table were added on to the table probability. An example: The probability P_2 for $a = 1$ can be taken as the sum of the point probabilities for $p_{a=0} + p_{a=1} + p_{a=5} + p_{a=6} = .00559 + .07832 + .03357 + .00140 = .11888$. Two examples should illustrate generation of two-sided probabilities:

1. Hätten wir eine zweiseitige Alternativhypothese mit einem $\alpha = 1\% = 0.01$ gebildet, führten die drei Verfahren zu folgenden Entscheidungen, wenn ein $a = 0$ beobachtet worden wäre:

- a. Das FREEMAN-HALTON-Verfahren: $P_2 = .00699 < \alpha$, daher Zurückweisung von H_0
- b. das $2P_1$ -Verfahren: dazu wird die einseitige Überschreitungswahrscheinlichkeit P_1 gegen ein Niveau von $\alpha/2 = .005$ getestet. Da $P_1 = .00559 > \alpha/2$, muß H_0 beibehalten werden.
- c. das $\alpha/2$ -Verfahren: auf jeder Seite der Verteilung wird eine Ablehnungsregion der Größe $\alpha/2$ definiert. Da die Punktwahrscheinlichkeit für $a=0$ bereits größer als $\alpha/2$ ist, ist kein zweiseitiger Test möglich. Als Konsequenz muß H_0 beibehalten werden.

2. Im zweiten Beispiel soll ebenfalls eine zweiseitige Alternativhypothese zugrundeliegen, wobei das Signifikanzniveau mit $\alpha = 5\%$ festgelegt wird. Beobachtet sei eine Tafel mit $a = 5$, also die Tabelle 1. Entschieden werden muß

- a. nach FREEMAN-HALTON: P_2 ist nach Spalte P_2 der Tabelle 2 für $a = 5$ $P_2 = .04056 < \alpha$, folglich wird H_0 verworfen.
- b. nach dem $2P_1$ -Verfahren: hierzu wird P_1 nach $\alpha/2$ beurteilt. Da $P_1 = .03497 > .025$, wird H_0 beibehalten.
- c. nach dem $\alpha/2$ -Verfahren: in die «untere Ablehnungsregion» der Größe $\alpha/2 = .025$ fällt die Tafel mit $a = 0$ ($P_1 = .00559$), in die «obere Ablehnungsregion» der gleichen Größe fällt nur die Tafel mit $a = 6$ ($P_1 = .00140$). Die beobachtete Tafel mit $a = 5$ gehört also nicht zur Ablehnungsregion. Folglich muß H_0 beibehalten werden.

Die Diskrepanzen zwischen den verschiedenen Verfahren treten völlig unregelmäßig auf. Dabei ist die Entscheidung nach dem $\alpha/2$ -Verfahren (b_2) in aller Regel extrem konservativ. Insbesondere sorgt es dafür, daß der Test echt *zweiseitig* bleibt. Das kann das FREEMAN-HALTON-, wie das $2P_1$ -Prinzip nicht garantieren: so sind für $a = 6$ die einseitige wie die FREEMAN-HALTON-Überschreitungswahrscheinlichkeit gleich. Entsprechend den Empfehlungen von KRÜGER (1979) wird in der Vertafelung in diesem Band auf das $\alpha/2$ -Verfahren verzichtet. Dem Anwender stehen aber sowohl die einseitige Überschreitungswahrscheinlichkeit (mit der Möglichkeit zum $2P_1$ -Verfahren im zweiseitigen Fall) und die FREEMAN-HALTON-Überschreitungswahrscheinlichkeit zur Verfügung. Dabei gelten für die praktische Arbeit zwei Empfehlungen:

- a. wird bei der zweiseitigen Prüfung Wert auf die Richtung des Unterschieds gelegt, sollte das $2P_1$ -Verfahren angewendet werden. Dabei werden de facto zwei einseitige simultane Tests zum Niveau $\alpha/2$ durchgeführt: die Richtung des Unterschieds kann dann ebenfalls interpretiert werden. →

1. If we had taken a two-sided alternative hypothesis with $\alpha = 1\% = 0.01$, the three methods would lead to the following decisions when $a = 0$ is observed:

- a. FREEMAN-HALTON method: $P_2 = .00699 < \alpha$; thus rejecting of H_0 .
- b. $2P_1$ method: the one-sided probability is tested against a level of $\alpha/2 = .005$. Since $P_1 = .00559 > \alpha/2$, H_0 must be accepted.
- c. $\alpha/2$ method: a rejection region of size $\alpha/2$ is defined on both sides of the distribution. Since the point probability for $a = 0$ is already higher than $\alpha/2$, a two-sided test is impossible. As a result H_0 can not be rejected.

2. The second example is also based on a two-sided alternative hypothesis, whereby the level of significance is set at $\alpha = 5\%$. A table with $a = 5$ is observed, i. e., Table 1. We must decide according to

- a. FREEMAN-HALTON: P_2 for $a = 5$ is $.04056 < \alpha$ from column P_2 of Table 2. Therefore H_0 is rejected.
- b. The $2P_1$ method: P_1 is evaluated on the basis of $\alpha/2$. Since $P_1 = .03497 > .025$, H_0 is not rejected.
- c. The $\alpha/2$ method: only the table with $a = 0$ ($P_1 = .00559$) is in the «lower rejection region» with size $\alpha/2 = .025$. Only the table with $a = 6$ ($P_1 = .00140$) is in the «upper rejection region» of the same parameter. The observed table with $a = 5$ is thus not in the rejection region. Therefore H_0 can not be rejected.

The discrepancies between the various methods occur irregularly. A decision based on the $\alpha/2$ method (b_2) is generally extremely conservative. It especially ensures that the test truly remains *two-sided*. The same is not guaranteed by the FREEMAN-HALTON or the $2P_1$ approaches. Thus for $a = 6$ the one-sided and the FREEMAN-HALTON probability is the same. In agreement with the recommendation of KRÜGER (1979), the $\alpha/2$ method is not used in the present volume. Both the one-sided probability (with the possibility of using the $2P_1$ method in a two-sided case) and the FREEMAN-HALTON probability are available to the user, however. In this regard we would like to give two suggestions for using the tables:

- a. If the direction of the difference in the two-sided method is important, the $2P_1$ method should be used. In effect, two one-sided, simultaneous tests at the $\alpha/2$ level are hereby employed. The direction of the difference can also be interpreted.
 - b. If the two-sided test is used merely to check for a difference in the portions which are postulated as equal in the null hypothesis, then the FREEMAN-HALTON method should be used.
-
- b. geht es bei der zweiseitigen Prüfung lediglich um die Verschiedenheit der in der Nullhypothese postulierten Gleichheit der Anteile, ist das FREEMAN-HALTON-Verfahren indiziert.

2.3. Die Prüfung auf Kontingenz

Der zweite große Anwendungsbereich der Vierfeldertafel ist neben der Homogenitätsprüfung das Prüfen auf Kontingenz, des Zusammenhangs zwischen zwei alternativen Merkmalen. Als Maß für diesen Zusammenhang ist PEARSONS ϕ -Koeffizient eingeführt

$$\phi = \frac{ad - bc}{(a + b)(c + d)(a + c)(b + d)}$$

Es läßt sich zeigen, daß unter der Nullhypothese der Unabhängigkeit der beiden Merkmale die gleiche bedingte Verteilung entsteht wie bei der Homogenitätsprüfung (siehe etwa LEHMANN, 1975). Die Tabellen sind deshalb für diese Fragestellung in gleicher Weise anzuwenden. Dabei müssen die Überschreitungswahrscheinlichkeiten wie folgt interpretiert werden:

- einseitig (P_1): ist ϕ verschieden von 0 in Richtung des Vorzeichens von ϕ ? (Bsp.: ist ein $\phi = +.600$ ein signifikant positiver Zusammenhang?)
- zweiseitig (P_2): ist ϕ signifikant verschieden von 0? (Prüfung ohne Berücksichtigung des Vorzeichens von ϕ)

2.4. Interpretation eines Ergebnisses

Obwohl der exakte Test für Homogenitäts- und Unabhängigkeitsfragen derselbe ist, muß die Interpretation eines Ergebnisses das Datenmodell miteinbeziehen. Hier gibt es vier Möglichkeiten, die am Beispiel der Frage aufgezeigt werden sollen, ob Kaffeetrinken das Einschlafen behindert (die Nomenklatur folgt LEHMANN, 1975):

Modell 1: Eine *gegebene* Gruppe von N Versuchspersonen wird zufällig vom Experimentator in zwei Untergruppen der Größe N_1 und N_2 aufgeteilt. Die eine erhält zwei Stunden vor dem Zubettgehen eine definierte Menge Kaffee, die andere Gruppe dient zur Kontrolle. Abhängige Variable ist die Zeit, die die Versuchspersonen benötigen, um einzuschlafen. Diese Prüfungsmethode soll genannt werden: «Zufallsmodell für den Vergleich zweier Behandlungen».

Modell 2: Aus einer Population, die groß genug ist, um von der Unabhängigkeit ihrer Mitglieder untereinander ausgehen zu können, wird eine Zufallsstichprobe des Umfangs N gezogen. Diese Stichprobe wird nach Zufall in zwei Gruppen des Umfangs N_1 und N_2 aufgeteilt, wobei eine behandelt wird, die andere nicht. Die «gegebene Gruppe» des Modells 1 wird hier ersetzt durch eine «Zufallsstichprobe». Deshalb wird das Modell genannt: «Populationsmodell für den Vergleich zweier Stichproben».

2.3. Test against Contingency

The second main area of application of the fourfold table, i. e., in addition to the test for homogeneity, is the test against contingency, the correlation between two alternative categorizations. We would like to introduce PEARSON'S ϕ -coefficient as a measure of this correlation.

$$\phi = \frac{ad - bc}{(a + b)(c + d)(a + c)(b + d)}$$

It can be shown that under the null hypothesis of independence between the two dichotomous variables the same conditional distribution occurs as in the test for homogeneity (c. f., LEHMANN, 1975). The tables can therefore be used in the same manner for this question. In doing so the probabilities must be interpreted as follows:

- one-sided (P_1): is ϕ different from 0 in the direction of the sign of ϕ ? (e. g., is $\phi = +.600$ a significantly positive correlation?)
- two-sided (P_2): is ϕ significantly different from 0? (Test without consideration for sign of ϕ .)

2.4. Interpretation of Results

Although the exact test for questions of homogeneity and independence is the same, the interpretation of a result must take the model into consideration. There are four possibilities for doing so which will be illustrated using the example of the question as to whether or not coffee drinking inhibits falling asleep (the nomenclature corresponds with LEHMANN, 1975).

Model 1: A *given* group of N experimental persons is randomly divided by the experimenter into two subgroups of size N_1 and N_2 . One subgroup receives a given amount of coffee two hours before going to bed; the other subgroup functions as a control. The dependent variable is the time which the persons in the experiment require in order to fall asleep. We will call this testing method: «randomization model for the comparison of two treatments».

Model 2: A random sample of size N is taken from a population which is large enough to be able to assume independence of its members. The sample is randomly divided into two groups of size N_1 and N_2 , whereby one group is treated; the other is not. The «given group» in model 1 is replaced in this model by a «random sample». Therefore this model will be called: «population model for the comparison of two treatments».

Modell 3: Man könnte argumentieren, daß die beiden Merkmale «Kaffeetrinken am Abend» und «Schwierigkeiten beim Einschlafen» die gesamte Population bereits in zwei Subpopulationen teilt. Wie in Modell 2 werden diese beiden Subpopulationen als groß genug angenommen, um die Unabhängigkeit ihrer Mitglieder zu garantieren. Von diesen Subpopulationen werden je eine Zufallsstichprobe des Umfangs N_1 und N_2 gezogen. Der Anteil der Leute «die abends Kaffee trinken und Schwierigkeiten beim Einschlafen haben» muß nun verglichen werden mit dem Anteil der Leute, «die abends Kaffee trinken, aber keine Schwierigkeiten beim Einschlafen haben». Diese Prüfmethode wird genannt: «Vergleich zweier Merkmale oder Subpopulationen durch eine Stichprobe aus jeder Subpopulation».

Modell 4: Es dürfte in manchen Fällen sehr schwierig sein, die in Frage stehenden Subpopulationen zu finden. Man kann aus der Gesamtpopulation *eine* Zufallsstichprobe ziehen und in dieser Stichprobe die Anteile «Kaffeetrinken mit Einschlafschwierigkeit» und «Kaffeetrinken ohne Einschlafschwierigkeit» miteinander vergleichen. Die Anteile der Kaffeetrinker und der Personen mit Einschlafschwierigkeiten sind bei diesem Vorgehen Zufallsvariablen. Das Vorgehen wird genannt: «Vergleich zweier Merkmale oder Subpopulationen durch eine Stichprobe aus der Gesamtpopulation».

Es ist zu bemerken, daß diese vier Modelle *nicht* verschiedene Wahrscheinlichkeitsmodelle einer Vierfeldertafel darstellen. Betrachten wir nun einige Eigenschaften der Modelle. Modell 1 und 2 sind dadurch charakterisiert, daß die Behandlungen nach Zufall zugeteilt werden, während Modell 3 und 4 sich die bereits in der Population vorhandenen Variation der Merkmale zunutze machen. Das bedingt aber einen wichtigen Unterschied in der Interpretation. Modell 1 und 2 konstituieren eine kausale Beziehung zwischen Behandlung und Wirkung (unabhängiger und abhängiger Variablen), während es bei Modell 3 und 4 lediglich erlaubt ist, von einem Zusammenhang zwischen zwei Merkmalen zu sprechen. Die Zufallszuteilung in den ersten beiden Modellen garantiert, daß alle Variablen «balanciert» sind außer der Behandlungsvariablen. Es gibt keinen plausiblen Grund, daß eine «versteckte» Variable lediglich in *einer* Zufallsgruppe auftreten sollte, vorausgesetzt die Zufallszuteilung ist korrekt verlaufen. Das ist der Grund, weshalb in diesen beiden Modellen von einer kausalen Beziehung gesprochen werden darf.

Dagegen sind die Modelle 3 und 4 nicht gegen den Einfluß solcher versteckter Variablen geschützt. Vielleicht sind alle Kaffeetrinker geistig Arbeitende und Denken verhindert Einschlafen! Kaffeetrinken wäre dann nur ein Oberflächenphänomen und überhaupt nicht kausal für Schwierigkeiten beim Einschlafen. Deshalb erlaubt das Fehlen der Zufallszuteilung von Behandlungen lediglich, von Zusammenhängen zu sprechen, die selbst wieder nichts anderes zu sein brauchen als Pseudo-Zusammenhänge!

Model 3: One could argue that both variables «coffee drinking in the evening» and «difficulties in falling asleep» already divide the entire population into two subpopulations. As in model 2, both of these subpopulations are assumed to be large enough to ensure independence of their members. One random sample of size N_1 and N_2 is drawn from each of the subpopulations. The proportion of persons «who drink coffee in the evening and have difficulties falling asleep» now has to be compared with the proportion of persons «who drink coffee in the evening but do not have difficulties falling asleep». This testing method will be referred to as «comparison of two attributes or subpopulations through a sample from each».

Model 4: In some cases it may be very difficult to locate the subpopulation which comes into question. One can draw *one* random sample from the entire population and compare the proportions «coffee drinking and falling asleep difficulties» and «coffee drinking without falling asleep difficulties» with each other. The proportions of the coffee drinkers and the persons with falling asleep difficulties are random variables in this procedure. The method will be called «comparison of two attributes or subpopulations through a sample from the total population».

It should be noted that these four models do *not* represent different models of probability for a fourfold table.

Let us now consider some of the properties of the models. Models 1 and 2 are distinguished by the fact that treatments are delivered by chance, whereas models 3 and 4 utilize the variation of the variables which is already present in the population. This causes an important difference in interpretation, however. Models 1 and 2 constitute a causal relationship between treatment and effect (independent and dependent variables), whereas it is only permissible to speak of a relationship between two variables in models 3 and 4. The randomization in the first two models guarantees that all variables are «balanced», aside from the treatment variable. There is no plausible reason why a «hidden» variable should only appear in *one* random group, provided the randomization has been generated correctly. This is the reason why one may speak of a causal relationship in both of these models.

In contrast, models 3 and 4 are not protected from the influence of such hidden variables. Perhaps all coffee drinkers are people who work mentally, and thinking inhibits falling asleep! Coffee drinking would then be a superficial phenomenon and absolutely not causal for difficulties in falling asleep. Therefore the lack of randomization of treatment merely permits conclusion of relationships which themselves may be no more than pseudorelationships!

Der Unterschied zwischen Modell 1 und Modell 2 ist in der Reichweite des empirischen Schlusses zu sehen. Die begrenzte Zahl der Versuchspersonen in Modell 1 erlaubt lediglich, Folgerungen für die untersuchte Gruppe zu ziehen, nicht für die gesamte Population. Dagegen kann auf Grund der Tatsache, daß die Versuchspersonen in Modell 2 eine Zufallsstichprobe aus der Population darstellen, das Ergebnis solcher Experimente auf die Population verallgemeinert werden. Die Begründung dafür verläuft analog: Modell 1 ist nicht gegen eine versteckte Variable geschützt, die genau zu der Gruppe geführt hat, die untersucht wird (das wird besonders deutlich, wenn man z. B. an die «Gruppe meiner Kollegen» denkt).

Die Modelle 1 bis 4 führen alle bei Alternativmerkmalen zu den gleichen Vierfeldertafeln und zum gleichen Rationale, wenn es um die Prüfung der Signifikanz geht. Aber obwohl die Prüfmethode exakt die gleichen Berechnungsschritte geht, muß die Interpretation der Ergebnisse verschieden sein, je nachdem welches Modell zugrunde liegt. Das heißt: zwar können die verschiedenen Datenmodelle zur gleichen Prüfverteilung führen, aber nie zum gleichen Schluß!

2.5. Die Beschreibung der Tafeln

Im Interesse der Benutzerfreundlichkeit wurde ein neues Vertafelungsprinzip eingeführt. Ausgehend von der Notation der Vierfeldertafel in Tabelle 1 mit den Zellenfrequenzen, a, b, c und d wird zur Charakterisierung einer Vierfeldertafel das Quadrupel der vier Zellenfrequenzen herangezogen.

Wegen der Symmetriebeziehungen in einer Vierfeldertafel läßt sich das Gesamt der möglichen Tafeln in der Darstellung wesentlich verkürzen, wenn folgende, die Wahrscheinlichkeitsberechnung wie den Betrag von φ nicht verändernden Umordnungen benutzt werden:

a. Vertauschung der beiden Spalten

$$\begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{cc} b & a \\ d & c \end{array} \text{ bedingt Vorzeichenwechsel von } \varphi$$

b. Vertauschung der beiden Zeilen

$$\begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{cc} c & d \\ a & b \end{array} \text{ bedingt Vorzeichenwechsel von } \varphi$$

c. Vertauschung der beiden Diagonalfelder

entweder Hauptdiagonale

$$\begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{cc} d & b \\ c & a \end{array} \text{ kein Vorzeichenwechsel von } \varphi$$

oder Nebendiagonale

$$\begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{cc} a & c \\ b & d \end{array} \text{ kein Vorzeichenwechsel von } \varphi$$

oder Haupt- und Nebendiagonale

$$\begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{cc} d & c \\ b & a \end{array} \text{ kein Vorzeichenwechsel von } \varphi$$

The difference between model 1 and model 2 can be seen in the difference in the range of the respective empirical conclusion. The limited number of experimental persons in model 1 merely permits drawing conclusions for the investigated group, not for the entire population. In contrast, the fact that the experimental persons in model 2 represent a random sample of the population enables one to generalize the result of such experiments as valid for the population. The justification for this is analogous: model 1 is not protected against a hidden variable which has lead precisely to the group which is being investigated (this becomes especially clear when one thinks, e. g., of «my group of colleagues»).

With alternative interpretations, models 1 to 4 all lead to the same fourfold tables and the same procedure when testing for significance. While the testing method contains exactly the same computational steps, the interpretation of the results must be different depending on the model on which they are based. This means that the various data models may indeed lead to the same sampling distribution, but never to the same conclusion!

2.5. Description of the Tables

In the interest of user ease a new tabling principle has been introduced. Starting with the notation of the fourfold table in Table 1 with the cell frequencies a, b, c and d, the quadruple of the four cell frequencies is used to characterize a fourfold table.

Due to the symmetrical relationships in a fourfold table the whole of all possible tables can be presented in substantially reduced form when the following rearrangements are used. These rearrangements *change* neither the calculated probability *nor* the amount of φ .

a. Exchanging both columns

$$\begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{cc} b & a \\ d & c \end{array} \text{ causes change of sign of } \varphi$$

b. Exchanging both rows

$$\begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{cc} c & d \\ a & b \end{array} \text{ causes change of sign of } \varphi$$

c. Exchanging both diagonal fields

either main diagonals

$$\begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{cc} d & b \\ c & a \end{array} \text{ no change of sign of } \varphi$$

or secondary diagonals

$$\begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{cc} a & c \\ b & d \end{array} \text{ no change of sign of } \varphi$$

or main and secondary diagonals

$$\begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{cc} d & c \\ b & a \end{array} \text{ no change of sign of } \varphi$$

Es werden aufgrund dieser Symmetriebeziehungen folgende Restriktionen eingeführt:

$$\begin{aligned} a &\leq b \\ a &\leq d \\ b &\leq c \end{aligned}$$

Diese drei Restriktionen lassen sich immer durch Umordnung der Tafel nach den drei Prinzipien erreichen. Wer sich nicht der Mühe dieser Umordnung unterziehen will, kann sich der Tabelle 5 bedienen, die auch als Karte diesem Band beiliegt. Dazu müssen die Zellenfrequenzen in eine *nach ihrer Größe ansteigenden Ordnung* gebracht werden.

Tabelle 5: Umordnung der Zellenfrequenzen

Größenfolge der Zellenfrequenzen	lies ab unter der Folge	Vorz. wechs.
a b c d	a b c d	nein
a b d c	a b c d	nein
a c b d	a c b d	nein
a c d b	a c b d	nein
a d b c	a b c d	nein
a d c b	a c b d	nein
b a c d	b a d c	ja
b a d c	b a d c	ja
b c a d	b a d c	ja
b c d a	b d a c	ja
b d a c	b d a c	ja
b d c a	b d a c	ja
c a b d	c a d b	ja
c a d b	c a d b	ja
c b a d	c a d b	ja
c b d a	c d a b	ja
c d a b	c d a b	ja
c d b a	c d a b	ja
d a b c	d b c a	nein
d a c b	d c b a	nein
d b a c	d b c a	nein
d b c a	d b c a	nein
d c a b	d c b a	nein
d c b a	d c b a	nein

Im Beispiel

4 1
2 3

heißt die Ordnung

b	c	d	a
1	2	3	4

Due to the symmetrical relationships the following restrictions are introduced:

$$\begin{aligned} a &\leq b \\ a &\leq d \\ b &\leq c \end{aligned}$$

These three restrictions can always be fulfilled by rearrangement of the table according to the above 3 principles.

Whoever prefers not to make the effort of determining the rearrangement can refer to Table 5 which is also inserted in the present volume as a card. In doing so one should arrange the cell frequencies *in order of increasing size*.

Table 5: Rearrangement of Cell Frequencies

size sequence of cell frequencies	read in the sequence	sign change
a b c d	a b c d	no
a b d c	a b c d	no
a c b d	a c b d	no
a c d b	a c b d	no
a d b c	a b c d	no
a d c b	a c b d	no
b a c d	b a d c	yes
b a d c	b a d c	yes
b c a d	b a d c	yes
b c d a	b d a c	yes
b d a c	b d a c	yes
b d c a	b d a c	yes
c a b d	c a d b	yes
c a d b	c a d b	yes
c b a d	c a d b	yes
c b d a	c d a b	yes
c d a b	c d a b	yes
c d b a	c d a b	yes
d a b c	d b c a	no
d a c b	d c b a	no
d b a c	d b c a	no
d b c a	d b c a	no
d c a b	d c b a	no
d c b a	d c b a	no

An example:

4 1
2 3

has the sequence

b	c	d	a
1	2	3	4

Dieses Quadrupel b-c-d-a wird in Tabelle 5 aufgesucht. Als Ablesefolge ist dort

b	d	a	c
1	3	4	2

zu finden. Mit dem Quadrupel 1-3-4-2 ist dann in Tafel 1-1 einzugehen. Das «ja» hinter dem Quadrupel in Tabelle 5 bedeutet, daß das Vorzeichen von ϕ gewechselt werden muß. Sind Zellenfrequenzen gleich, ist die Anordnung beliebig.

Die zweite Restriktion betrifft die einseitige Überschreitungswahrscheinlichkeit P_1 . Ist P_1 (und mit ihr dann auch die zweiseitige Überschreitungswahrscheinlichkeit) größer als $0.10 (= 10\%)$, wurde die Vierfeldertafel nicht in Tafel 1-1 aufgenommen.

Als dritte Restriktion wurde eingeführt, daß die zweiseitige Überschreitungswahrscheinlichkeit nach dem Prinzip von FREEMAN-HALTON (und damit auch die einseitige Überschreitungswahrscheinlichkeit) nicht kleiner sein darf als die Stellengenauigkeit der Tabelle. Bei vier Stellen bedeutet das, daß P_2 nicht kleiner werden darf als 0.0001 , also 10^{-4} . Ist eine beobachtete Vierfeldertafel trotz hohem X^2 nicht zu finden, ist $P_2 < 10^{-4}$. Das X^2 dieser Tafeln ist dann in der Regel größer als $X^2 = 14$. Diese letzte Restriktion erscheint vor allem deshalb vertretbar, weil für diese extremsten Tafeln der exakte Test auch relativ einfach von Hand zu rechnen ist.

Ist eine beobachtete Vierfeldertafel also nicht in Tafel 1-1 zu finden, kann das drei Gründe haben:

1. Die Zellfrequenzen sind nicht entsprechend der Anweisung umgeordnet.
2. Bei geringem X^2 : die einseitige Überschreitungswahrscheinlichkeit ist größer als 10% .
3. Bei einem $X^2 > 14$: die Omnibus-Überschreitungswahrscheinlichkeit nach FREEMAN-HALTON ist kleiner als 10^{-4} .

2.6. Formeln für die asymptotische Prüfung

Soll eine Vierfeldertafel asymptotisch geprüft werden, sind folgende Prüfgrößen einschlägig (die Notation ist dabei wie in Tabelle 1):

1. X^2 mit Kontinuitätskorrektur (YATES, 1934).

$$X^2 = \frac{N(|ad-bc| - 0.5N)^2}{(a+b)(c+d)(a+c)(b+d)}$$

2. X^2 ohne Kontinuitätskorrektur:

$$X^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

This quadruple is located in Table 5. The sequence for reading is given as

b	d	a	c
1	3	4	2

One should then enter Table 1-1 with the quadruple 1-3-4-2. The «yes» after the quadruple in Table 5 means that the sign of ϕ must be changed. (If the cell frequencies found in Table 5 had been equal the sequence would have been arbitrary).

The second restriction involves the one-sided probability P_1 . If P_1 (and with it then the two-sided probability as well) is greater than $0.10 (= 10\%)$ the table was not included in Table 1-1.

The third restriction which has been incorporated is that the two-sided probability according to the principle of FREEMAN-HALTON (and thereby also the one-sided probability) is not allowed to be less than the accuracy of the table. With four elements this means that P_2 is not allowed to be less than 0.0001 , i. e., 10^{-4} . If an observed fourfold table cannot be found despite high X^2 , then $P_2 < 10^{-4}$. The X^2 of this table is then greater than $X^2 = 14$ as a rule. This last restriction seems to be justifiable above all because the exact test for most extreme tables can be calculated by hand relatively easily.

If an observed fourfold table cannot be found in Table 1-1, three reasons are possible:

1. The cell frequencies are not arranged in the prescribed sequence.
2. With low X^2 : the one-sided probability P_1 is greater than 10% .
3. With $X^2 > 14$: the two-sided probability P_2 according to FREEMAN-HALTON is less than 10^{-4} .

2.6. Formulas for the Asymptotic Testing

If a fourfold table is to be tested asymptotically, the following test statistics are of relevance (notation as in Table 1):

1. X^2 with correction for continuity (YATES, 1934)

$$X^2 = \frac{N(|ad-bc| - 0.5N)^2}{(a+b)(c+d)(a+c)(b+d)}$$

2. X^2 without correction for continuity:

$$X^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

3. PEARSONS ϕ :

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

4. $z = \phi\sqrt{N}$ und $X^2 = \phi^2 N$

Unter Gültigkeit der Nullhypothese ist X^2 asymptotisch χ^2 -verteilt mit 1 Freiheitsgrad, z ist asymptotisch standardnormalverteilt.

Die zweiseitigen Überschreitungswahrscheinlichkeiten für die z - bzw. X^2 -Werte aus diesen Formeln sind in Tafel 1-2 zu finden.

3. Anwendungen

3.1. Die Prüfung der Homogenität bei zwei unabhängigen Stichproben

Problemstellung: Gegeben sind zwei unabhängige Stichproben mit den Meßwerten $x_{11}, x_{12}, \dots, x_{1N_1}$ und $x_{21}, x_{22}, \dots, x_{2N_2}$, wobei X_1 und X_2 alternativ verteilte Zufallsvariablen mit den Mittelwerten μ_1 und μ_2 sind.

Hypothesen: Es soll untersucht werden, ob die Mittelwerte der Zufallsvariablen voneinander verschieden sind ($H_0: \mu_1 = \mu_2$). Dabei sind zwei Alternativhypothesen möglich:

- a. einseitig: $\mu_1 < \mu_1$ (bzw. $\mu_2 < \mu_1$)
- b. zweiseitig: $\mu_1 \neq \mu_2$

Testverfahren: Zur Festlegung der Verteilung unter H_0 ($\mu_1 = \mu_2$) werden 4 Parameter benötigt. Für die vorliegende Vertafelung wurden die 4 Zellfrequenzen a, b, c , und d der Vierfeldertafel ausgewählt.

Beispiel: Zur Prüfung der Wirkungsweise eines neuen Medikaments wird eine Gruppe von $N = 20$ Versuchspersonen (V_{pn}) nach Zufall in zwei Gruppen der Größe $N_1 = N_2 = 10$ V_{pn} aufgeteilt, wobei die eine Gruppe das neue Medikament, die andere ein bereits bekanntes Standardmedikament erhält. Die Wirkung wird durch ein Expertenteam auf den Stufen «deutlich gebessert» und «nicht oder nicht deutlich gebessert» beurteilt.

$H_0: \mu_1 = \mu_2$ (der Anteil der «deutlich gebessert» ist unter beiden Bedingungen gleich)

$H_1: \mu_1 \neq \mu_2$ (die Medikamente unterscheiden sich, zweiseitige Fragestellung).

Angenommenes Signifikanzniveau: $\alpha = 5\%$

3. PEARSONS ϕ :

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

4. $z = \phi\sqrt{N}$ and $X^2 = \phi^2 N$

When the null hypothesis holds, X^2 is asymptotically χ^2 distributed with one degree of freedom, z is asymptotically normally standard distributed.

The two-sided probability for the z and X^2 values from these formulas can be found in Table 1-2.

3. Applications

3.1. Test for Homogeneity with Two Independent Samples

Problem: Two independent samples with the values $x_{11}, x_{12}, \dots, x_{1N_1}$ and $x_{21}, x_{22}, \dots, x_{2N_2}$, are given, whereby X_1 and X_2 are alternatively distributed random variables with means μ_1 and μ_2 .

Hypotheses: We are to investigate whether or not the means of the random variables differ from one another ($H_0: \mu_1 = \mu_2$). Two alternative hypotheses are possible:

- a. one-sided: $\mu_1 < \mu_1$ (or $\mu_2 < \mu_1$)
- b. two-sided: $\mu_1 \neq \mu_2$

Testing Procedure: 4 parameters are required to determine the distribution under H_0 ($\mu_1 = \mu_2$). In view of the available tables the four cell frequencies a, b, c and d are selected.

Example: To test the mode of action of a new drug a group of $N = 20$ experimental subjects (S_s) is randomly divided into two groups of size $N_1 = N_2 = 10$ S_s . One group receives the new drug, the other receives an already known standard drug. The effect is evaluated as «clearly improved» or «not or not clearly improved» by a team of experts.

$H_0: \mu_1 = \mu_2$ (the proportion «clearly improved» is the same under both conditions).

$H_1: \mu_1 \neq \mu_2$ (the drugs differ, two-sided problem).

Assumed level of significance: $\alpha = 5\%$.

Die Untersuchung erbrachte als Ergebnis die Tabelle 6.
Tabelle 6

	deutlich gebessert	nicht oder nicht deutlich gebessert	Σ
neues Medikament	a = 8	b = 2	10 = N ₁
Standardmedikament	c = 4	d = 6	10 = N ₂
Σ	12	8	20 = N

Nach den Restriktionen des Kapitels 2.5. (siehe auch in diesen Band eingelegte Karte) werden die Zellfrequenzen zuerst in aufsteigender Reihenfolge geordnet:

2 4 6 8
b c d a

Mit diesem Quadrupel b-c-d-a wird in Tabelle 5 eingegangen. Dort ist als Ablesefolge das Quadrupel

b d a c
2 6 8 4

zu finden, ebenso der Hinweis, daß diese Vertauschung das Vorzeichen von ϕ nicht ändert.

Mit dem Quadrupel 2-6-8-4 wird in Tafel 1-1 eingegangen. Dort ist folgende Zeile zu finden:

2 6 8 4 -0.408 .1698 .0849
 ϕ P₂ P₁

Die zweiseitige Überschreitungswahrscheinlichkeit P₂ beträgt damit P₂ = .1698. Da P₂ > α , ist H₀ nicht zu verwerfen, wonach zwischen den beiden Medikamenten kein Unterschied besteht.

Hinweis: Das gleiche Procedere gilt für mehrklassige diskrete wie für stetige Zufallsvariablen, wenn diese vorher durch geeignete Transformationen (Dichotomierung) in Alternativvariablen übergeführt werden. Eine solche Technik liegt z. B. dem Median-Test zugrunde. Ähnlich auch der WESTENBERG-Test (1948), der zur Bestimmung der Dispersionsgleichheit in 2 Stichproben die Zahl der Fälle innerhalb und außerhalb der Quartilgrenzen Q₁ und Q₃ verwendet.

3.2. Die Prüfung auf Kontingenz

Problemstellung: Gegeben sind 2 paarweise verbundene Stichproben (x_i, y_i) für i = 1, 2, ..., N einer zweidimensionalen Verteilung von Alternativdaten.

Hypothesen: Es wird geprüft, ob der Kontingenzkoeffizient ϕ von Null verschieden ist (H₀: $\phi = 0$). Dazu wählt man als Alternativhypothese entweder

- a. einseitig: $\phi > 0$ (bzw. $\phi < 0$) oder
- b. zweiseitig: $\phi \neq 0$

The results of the investigation are shown in Table 6.
Table 6

	clearly improved	not or not clearly improved	Σ
new drug	a = 8	b = 2	10 = N ₁
standard drug	c = 4	d = 6	10 = N ₂
Σ	12	8	20 = N

In compliance with the restrictions of section 2.5. (see also the card inserted in this volume) the cell frequencies are first arranged in order of increasing size:

2 4 6 8
b c d a

Table 5 is entered with this quadruple b-c-d-a. The reading sequence of the quadruple

b d a c
2 6 8 4

can be found there, along with the note that this rearrangement does not change the sign of ϕ .

Table 1-1 is entered with the quadruple 2-6-8-4.

There we find:

2 6 8 4 -0.408 .1698 .0849
 ϕ P₂ P₁

The two-sided probability P₂ thus amounts to P₂ = .1698. Since P₂ > α , H₀ cannot be rejected. Thus there is no difference between the two drugs.

Note: The same procedures apply to the analysis of discrete and continuous random variables when they have been transformed to dichotomous variables by suitable means. The median test, for example, is based on such a technique. A similar case is the WESTENBERG test (1948) which uses the number of cases within and outside of the quartile borders Q₁ and Q₃ to determine the uniformity of dispersion in 2 samples.

3.2. Test against Contingency

Problem: 2 dichotomous variables X and Y are observed for a sample of N persons.

Hypotheses: We are to test whether the coefficient of contingency ϕ differs from zero (H₀: $\phi = 0$). As an alternative hypothesis one selects either

- a. one-sided: $\phi > 0$ (or $\phi < 0$) or
- b. two-sided: $\phi \neq 0$

Testverfahren: Zur Festlegung der Verteilung unter H_0 werden 4 Parameter benötigt. Für die vorliegende Vertafelung wurden die vier Zellfrequenzen a, b, c und d der Vierfelder-tafel ausgewählt.

Beispiel: Es soll geprüft werden, ob der Beruf des Vaters auch die Berufswahl des Sohnes beeinflusst. Dazu werden auf dem Gelände der Universität zufällig $N = 39$ Studenten ausgewählt und sowohl nach dem Beruf des Vaters wie nach der eigenen Studienrichtung befragt. Die Berufe werden den zwei Kategorien «geisteswissenschaftlich» und «naturwissenschaftlich» orientiert zugeordnet.

$H_0: \phi = 0$ (es besteht kein Zusammenhang zwischen dem Beruf des Vaters und der Studienrichtung)

$H_1: \phi > 0$ (die Studienrichtung hängt mit dem Beruf des Vaters zusammen – einseitige Fragestellung)

Angenommenes Signifikanzniveau: $\alpha = 5\%$

Tabelle 7 zeigt das Ergebnis der Befragung.

Tabelle 7

		SOHN		
		geistes- wissenschaftlich	natur- wissenschaftlich	
V	geisteswissenschaftlich	a = 17	b = 7	24 = N_1
A	naturwissenschaftlich	c = 5	d = 10	15 = N_2
T				
E	Σ	$M_1 = 22$	$M_2 = 17$	39 = N

Nach den Restriktionen der Tabelle 2.5. (siehe auch in diesen Band eingelegte Karte) werden die Zellfrequenzen zuerst in aufsteigender Reihenfolge geordnet:

5 7 10 17
c b d a

Mit diesem Quadrupel c-b-d-a wird in Tabelle 5 eingegangen. Dort ist als Ablesefolge das Quadrupel

c d a b
5 10 17 7

zu finden, ebenso der Hinweis, daß das Vorzeichen von ϕ in der Tafel nach dieser Transformation getauscht werden muß. Mit dem Quadrupel 5-10-17-7 wird in Tafel 1-1 eingegangen. Dort ist folgende Zeile zu finden:

5 10 17 7 -0.368 .0448 .0244
 ϕ P_2 P_1

Der Kontingenzkoeffizient von $\phi = + 0.368$ (Vorzeichen muß getauscht werden), deutet auf einen positiven Zusammenhang im Sinne der Alternativhypothese. Die einseitige Überschreitungswahrscheinlichkeit $P_1 = .0244$ ist kleiner als α , so daß zum vorgegebenen Signifikanzniveau H_0 zurückgewiesen werden kann.

Hinweis: Das gleiche Procedere gilt für mehrklassige diskrete und stetige Zufallsvariablen, wenn diese vorher durch geeignete Transformationen (Dichotomierung) in Alternativvariablen übergeführt werden.

Testing Procedure: 4 parameters are required to determine the distribution under H_0 . In view of the available tables, the four cell frequencies a, b, c, and d of the fourfold table are selected.

Example: We are to investigate whether or not a father's vocation also influences the choice of vocation of his son. For this purpose $N = 39$ students on the university grounds are chosen at random and asked to state their father's vocation and their own concentration of studies. The vocations are broken down into two categories: «liberal arts» and «natural sciences».

$H_0: \phi = 0$ (there is no correlation between the father's vocation and the concentration).

$H_1: \phi > 0$ (the concentration correlates with the father's vocation – one-sided problem).

Assumed level of significance: $\alpha = 5\%$

Table 7 shows the results of the survey.

Table 7

		SON		
		liberal arts	natural science	
F				
A	liberal arts	a = 17	b = 7	24 = N_1
T	natural sciences	c = 5	d = 10	15 = N_2
H				
E	Σ	$M_1 = 22$	$M_2 = 17$	39 = N

In compliance with the restrictions in section 2.5. (see also the card inserted in this volume) the cell frequencies are first arranged in order of increasing size:

5 7 10 17
c b d a

Table 5 is entered with this quadruple c-b-d-a. The reading sequence found there is the quadruple

c d a b
5 10 17 7

which is accompanied by the note that the sign of ϕ in the table must be changed after this transformation. Table 1-1 is entered with quadruple 5-10-17-7 and the following line is located:

5 10 17 7 -0.368 .0448 .0244
 ϕ P_2 P_1

The coefficient of contingency of $\phi = + 0.368$ (sign must be changed) suggests a positive correlation along the lines of the alternative hypothesis. The one-sided probability $P_1 = .0244$ is less than α and thus H_0 can be rejected at the preset level of significance.

Note: The same procedure applies to general discrete and continuous random variables when they are previously transformed to dichotomous variables by suitable means.

Literaturverzeichnis/References

- ARMSEN, P.: Tables for significance tests of 2×2 contingency tables. *Biometrika*, 42, 1955, 494–511.
- BENNETT, B.M. & HORST, C.: Supplement to tables for testing significance in a 2×2 contingency table. Cambridge: University Press, 1966.
- BERKSON, J.: In dispraise of the exact test. Do the marginal totals of the 2×2 table contain relevant information respecting the table proportions? *Journal of Statistical Planning and Inference*, 2, 1978, 27–42.
- BROSS, I.D. & KASTEN, E.L.: Rapid analysis of 2×2 tables. *J.Americ.Stat.Ass.*, 52, 1957, 18.
- COCHRAN, W.G.: Some methods for strengthening the common X^2 test. *Biometrics*, 10, 1954, 417–451.
- MCDONALD, L.L., DAVIS, B.M. & MILLIKEN, G.A.: A nonrandomized unconditional test for comparing two proportions in 2×2 contingency tables. *Technometrics*, 19, 1977, 145–157.
- FINNEY, D.J.: The FISHER-YATES-test of significance in 2×2 contingency tables. *Biometrika*, 35, 1948, 145–156.
- FINNEY, D.J., LATSCHA, R., BENNETT, B.M. & HSU, P.: Tables for testing significance in a 2×2 contingency table. Cambridge: University Press, 1963.
- FISHER, R.A.: The design of experiments. Edinburgh: Oliver & Boyd, 1935.
- FREEMAN, G.H. & HALTON, J. H.: Note on an exact treatment of contingency, goodness of fit, and other problems of significance. *Biometrika*, 38, 1951, 141–149.
- HENZE, F.H.: Ein- und zweiseitige Überschreitungswahrscheinlichkeiten des exakten FISHER-Tests. Vortrag gehalten auf dem Biometrischen Kolloquium 1976 in Bad Nauheim.
- KRÜGER, H.-P.: Zum Problem der Bestimmung zweiseitiger Überschreitungswahrscheinlichkeiten beim exakten Vierfeldertest. *EDV in Biologie und Medizin*, 10, 1979, 19–21.
- LATSCHA, R.: Tests of significance in a 2×2 contingency table: Extension of FINNEY's tables. *Biometrika*, 40, 1953, 74–86.
- LEHMANN, E.L.: Nonparametrics: Statistical methods based on ranks. New York: McGraw-Hill, 1975.
- LIEBERMANN, G.J. & OWEN, D.B.: Tables of the hypergeometric probability distribution. Stanford California: Stanford University Press, 1961.
- LIENERT, G. A. Verteilungsfreie Methoden in der Biostatistik. Band 1, 2. Aufl. Meisenheim: Hain, 1973.
- MAINLAND, D. & MURRAY, I. M.: Tables for use in fourfold contingency tables. *Science*, 116, 1952, 591–594.
- MILLER, R. G.: Simultaneous statistical inference. New York: McGraw-Hill, 1966.
- OWEN, D. B.: Handbook of statistical tables. Reading, Mass.: Addison-Wesley, 1962.
- PEARSON, E. S.: The choice of statistical tests illustrated in the interpretation of data classed in a 2×2 table. *Biometrika*, 34, 1947, 139–167.
- PEARSON, E. S. & HARTLEY, H. O.: *Biometrika tables for statisticians*. 3rd ed. Cambridge: University Press, 1969.
- PLACKETT, R. D.: The marginal totals of a 2×2 table. *Biometrika*, 64, 1977, 37–42.
- SIEGERSTETTER, J. & WELZL, G.: Tafeln zur Durchführung des zweiseitigen exakten Tests nach Fisher für Vierfeldertafeln. GSF-Bericht MD 145 (Gesellschaft für Strahlen- und Umweltforschung, München), 1975.
- WALL, K.-D.: Kombinatorische Analyse von Kontingenztafeln. Inaug.Diss. Berlin, 1972.
- WESTENBERG, J.: Significance test for median and interquartile range in samples from continuous populations of any form. *Proc. Kon. Ned. Ak. Wet (A)*51, 1948, 252–261.
- WISE, M. E.: Multinomial probabilities and the χ^2 and X^2 distribution. *Biometrika*, 50, 1963, 145–154.
- WITTING, H.: *Mathematische Statistik*. Stuttgart: Teubner, 1974.
- YATES, F.: Contingency tables involving small numbers and the X^2 -test. *Suppl. J. R. Stat. Soc. (B)*1, 1934, 217–235.

Tafeln

Zeichenerklärung

- a, b, c, d Zellfrequenzen
- ϕ phi-Koeffizient
- P_1 einseitige Überschreitungswahrscheinlichkeit
- P_2 zweiseitige Überschreitungswahrscheinlichkeit

Tables

Key to the symbols used

- a, b, c, d cell frequencies
- ϕ phi-coefficient
- P_1 one-tailed transgression probability
- P_2 two-tailed transgression probability

Die Anwendung der Tafel 1 – 1

Die Notation der Vierfeldertafel

a	b
c	d

Größenfolge der Zellenfrequenzen	lies ab unter der Folge	Vorz. wechs.
a b c d	a b c d	nein
a b d c	a b c d	nein
a c b d	a c b d	nein
a c d b	a c b d	nein
a d b c	a b c d	nein
a d c b	a c b d	nein
b a c d	b a d c	ja
b a d c	b a d c	ja
b c a d	b a d c	ja
b c d a	b d a c	ja
b d a c	b d a c	ja
b d c a	b d a c	ja
c a b d	c a d b	ja
c a d b	c a d b	ja
c b a d	c a d b	ja
c b d a	c d a b	ja
c d a b	c d a b	ja
c d b a	c d a b	ja
d a b c	d b c a	nein
d a c b	d c b a	nein
d b a c	d b c a	nein
d b c a	d b c a	nein
d c a b	d c b a	nein
d c b a	d c b a	nein

Schritt A

Ordne die Zellfrequenzen *in aufsteigender Größenfolge* (bei gleichen Frequenzen ist die Anordnung beliebig)

Beispiel:

Ausgangsfolge: a b c d
numerisch 13 9 3 12

Größenfolge: c b d a

Schritt B

Gehe mit der Größenfolge in die linke Spalte der nebenstehenden Tabelle. Lies dort in der mittleren Spalte die Ablesefolge ab. Ordne die beobachtete Vierfeldertafel nach dieser Ablesefolge um. Vermerke aus der rechten Spalte der Tabelle, ob das Vorzeichen von ϕ gewechselt werden muß.

Im Beispiel:

Größenfolge: c b d a
Ablesefolge: c d a b
numerisch 3 12 13 9

Vorzeichenwechsel von ϕ ? ja

Schritt C

Gehe mit der Ablesefolge in Tafel 1 – 1 ein und lies dort ϕ , die zweiseitige und die einseitige Überschreitungswahrscheinlichkeit ab.

Im Beispiel ist unter der Ablesefolge

3 12 13 9

auf Seite 219 zu finden:

3 12 13 9 -0.387 .0409 .0204

Das ϕ der beobachteten Tafel ist $\phi = +0.387$ (Vorzeichenwechsel!)

Die zweiseitige Überschreitungswahrscheinlichkeit $P_2 = 0.0409$, die einseitige Überschreitungswahrscheinlichkeit $P_1 = 0.0204$.

How to use Table 1-1

The notation of the Fourfold Table

a	b
c	d

increasing sequence of cell frequencies	read under the sequence	sign change
a b c d	a b c d	no
a b d c	a b c d	no
a c b d	a c b d	no
a c d b	a c b d	no
a d b c	a b c d	no
a d c b	a c b d	no
b a c d	b a d c	yes
b a d c	b a d c	yes
b c a d	b a d c	yes
b c d a	b d a c	yes
b d a c	b d a c	yes
b d c a	b d a c	yes
c a b d	c a d b	yes
c a d b	c a d b	yes
c b a d	c d a b	yes
c b d a	c d a b	yes
c d a b	c d a b	yes
c d b a	c d a b	yes
d a b c	d b c a	no
d a c b	d c b a	no
d b a c	d b c a	no
d b c a	d b c a	no
d c a b	d c b a	no
d c b a	d c b a	no

Step A

Arrange the cell frequencies *in order of increasing size* (the sequence is arbitrary with equal frequencies)

Example:
 Starting sequence numerically: a b c d
 13 9 3 12
 increasing size sequence: c b d a

Step B

Locate the increasing size sequence in the left column of the table to the left. Note the corresponding reading sequences from the middle column. Rearrange the observed fourfold table according to this reading sequence. Note from the right column whether the sign of ϕ must be changed.

In the example:
 increasing size sequence: c b d a
 reading sequence: c d a b
 numerically: 3 12 13 9
 Change of sign of ϕ ? Yes.

Step C

Enter Table 1-1 with the reading sequence and locate ϕ there, the two-sided and the one-sided probability.

In the example, under the reading sequence
 3 12 13 9
 on page 219 one finds:

3 12 13 9 -0.387 .0409 .0204

The ϕ of the observed table is $\phi = +0.387$ (sign change!). The two-sided probability $P_2 = 0.0409$, the one-sided probability $P_1 = 0.0204$.

Wird eine beobachtete Vierfeldertafel nicht in Tafel 1-1 gefunden, kann das drei Gründe haben:

Grund a: die Zellfrequenzen sind falsch angeordnet

Grund b: die einseitige Überschreitungswahrscheinlichkeit P_1 ist größer als 0.10 (da $P_1 \leq P_2$, ist P_2 in diesem Fall ebenfalls ≥ 0.10)

Grund c: die zweiseitige Überschreitungswahrscheinlichkeit P_2 ist kleiner als 0.0001 (da $P_2 \geq P_1$, ist P_1 in diesem Fall ebenfalls ≤ 0.0001)

Beispiel für Grund a

Die Vierfeldertafel mit den Frequenzen $a=12$, $b=3$, $c=5$, $d=14$ ist in dieser Reihenfolge nicht in Tafel 1-1 zu finden. Kontrolle der richtigen Anordnung entsprechend den Schritten A-C auf der Rückseite:

Ausgangsfolge a b c d
numerisch 12 3 5 14

Größenfolge b c a d
Ablesefolge b a d c Vorzeichenwechsel: ja
numerisch 3 12 14 5

Auf Seite 219 ist zu finden:

3 12 14 5 -0.533 .0026 .0024

Die beobachtete Tafel hat damit ein $\phi = +0.533$ (Vorzeichenwechsel!), eine zweiseitige Überschreitungswahrscheinlichkeit $P_2 = .0026$ und eine einseitige Überschreitungswahrscheinlichkeit $P_1 = .0024$

Beispiel für Grund b

Ausgangsfolge a b c d
numerisch 2 12 15 26

Größenfolge a b c d
Ablesefolge a b c d Vorzeichenwechsel: nein
numerisch 2 12 15 26

Diese Tafel ist nicht zu finden. Der letzte vertafelte Wert auf S. 174 ist 2 12 15 25 mit einer einseitigen Überschreitungswahrscheinlichkeit $P_1 = 0.0982$. Da die beobachtete Tafel weniger extrem ist, ist ihr $P_1 > 0.10$.

Beispiel für Grund c

Ausgangsfolge a b c d
numerisch 3 14 15 1

Größenfolge d a b c
Ablesefolge d b c a Vorzeichenwechsel: nein
numerisch 1 14 15 3

Diese Tafel ist nicht zu finden. Der letzte vertafelte Wert auf S. 131 ist 1 14 15 6 mit einer zweiseitigen Überschreitungswahrscheinlichkeit $P_2 = 0.0001$. Da die beobachtete Tafel noch extremer ist, ist $P_2 < 0.0001$. Das ϕ dieser Tafel ist $\phi = -0.7638$. Nach $X^2 = \phi^2 N = (-0.7638)^2 \cdot 33 = 19.2540$ ist $X^2 > 14$.

If an observed fourfold table is not found in Table 1-1 there can be three reasons:

Reason a: The sequence of the cell frequencies is incorrect.

Reason b: The one-sided probability P_1 is greater than 0.10 (since $P_1 \leq P_2$, P_2 is also ≥ 0.10).

Reason c: The two-sided probability is less than 0.0001 (since $P_2 \geq P_1$, P_1 in this case is also ≤ 0.0001).

Example for reason a

The fourfold table with the frequencies $a=12$, $b=3$, $c=5$, $d=14$ cannot be found in this sequence in Table 1-1. Check for correct sequence in accordance with steps A-C on the reverse side:

initial sequence a b c d
numerically 12 3 5 14

increasing size b c a d
sequence

reading sequence b a d c sign change: yes
numerically 3 12 14 5

Taken from page 219:

3 12 14 5 -0.533 .0026 .0024

The observed table thereby has a $\phi = +0.533$ (sign change!), a two-sided probability $P_2 = .0026$ and a one-sided probability $P_1 = .0024$.

Example for reason b

Initial sequence a b c d
numerically 2 12 15 26

increasing size a b c d
sequence

reading sequence a b c d sign change: no
numerically 2 12 15 26

This table cannot be found. The last tabled value on page 174 is 2 12 15 25 with a one-sided probability $P_1 = 0.0982$. Since the observed table is less extreme, its $P_1 > 0.10$.

Example for reason c

Initial sequence a b c d
numerically 3 14 15 1

increasing size d a b c
sequence

reading sequence d b c a sign change: no
numerically 1 14 15 3

This table cannot be located. The last table value on page 131 is 1 14 15 6 with a two-sided probability $P_2 = 0.0001$. Since the observed table is even more extreme, $P_2 < 0.0001$. The ϕ of this table is $\phi = -0.7638$. Since $X^2 = \phi^2 N = (-0.7638)^2 \cdot 33 = 19.2540$, $X^2 > 14$.

Tafel 1-2

Werte der Normalverteilung und der χ^2 -Verteilung mit 1 Freiheitsgrad

In Tafel 1-2 sind die zweiseitigen Überschreitungswahrscheinlichkeiten für

z-Werte von 0 (0.1) 5.0

und die entsprechenden χ^2 -Werte angegeben. Da bei 1 Freiheitsgrad

$$z = \sqrt{\chi^2}$$

gilt, entsprechen die χ^2 dem Quadrat der z-Werte. Für einseitige Fragestellungen muß die angegebene Wahrscheinlichkeit lediglich halbiert werden.

Table 1-2

Values of the normal distribution and the χ^2 -distribution with 1 degree of freedom

Table 1-2 contains the two-sided probability for z-values of 0 (0.1) 5.0

and the corresponding χ^2 values. Since

$$z = \sqrt{\chi^2}$$

applies with one degree of freedom, the χ^2 s correspond to the squares of the z-values. For one-sided problems the given probability needs merely be halved.

z	χ^2	P	z	χ^2	P
0.1	0.0	0.920344	2.6	6.8	0.009322
0.2	0.0	0.841480	2.7	7.3	0.006934
0.3	0.1	0.764177	2.8	7.8	0.005110
0.4	0.2	0.689156	2.9	8.4	0.003732
0.5	0.3	0.617075	3.0	9.0	0.002700
0.6	0.4	0.548506	3.1	9.6	0.001935
0.7	0.5	0.483927	3.2	10.2	0.001374
0.8	0.6	0.423711	3.3	10.9	0.000967
0.9	0.8	0.368120	3.4	11.6	0.000674
1.0	1.0	0.317310	3.5	12.3	0.000465
1.1	1.2	0.271332	3.6	13.0	0.000318
1.2	1.4	0.230139	3.7	13.7	0.000216
1.3	1.7	0.193601	3.8	14.4	0.000145
1.4	2.0	0.161513	3.9	15.2	0.000096
1.5	2.3	0.133614	4.0	16.0	0.000063
1.6	2.6	0.109599	4.1	16.8	0.000041
1.7	2.9	0.089131	4.2	17.6	0.000027
1.8	3.2	0.071861	4.3	18.5	0.000017
1.9	3.6	0.057433	4.4	19.4	0.000011
2.0	4.0	0.045500	4.5	20.3	0.000007
2.1	4.4	0.035729	4.6	21.2	0.000004
2.2	4.8	0.027807	4.7	22.1	0.000003
2.3	5.3	0.021448	4.8	23.0	0.000002
2.4	5.8	0.016395	4.9	24.0	0.000001
2.5	6.3	0.012419	5.0	25.0	0.000001