


Statistische Tests für Kontingenztabellen

Häufig ist der Zusammenhang zwischen Variablen interessant. Um das geeignete Maß dafür zu finden, ist das Messniveau von Bedeutung. Bei metrischen Variablen können Korrelationen oder Streudiagramme als erstes Maß in Betracht gezogen werden. Hingegen stellt für diskrete oder klassierte Merkmale die Kontingenztafel (auch „Kreuztafel“ genannt) eine Darstellungsform für den Zusammenhang zweier simultan beobachteter Merkmale dar. Mit Hilfe von Kontingenztabellen können auch Tests auf den Zusammenhang zwischen den beiden Variablen durchgeführt werden. Im Folgenden sollen vor allem Abhängigkeiten zwischen zwei diskreten Merkmalen getestet werden.

Wie bei dem Kapitel der Linearen Regression werden Skripte in vier Statistikprogrammen bereit gestellt. Diese sind auf der rechten Seite zu finden.

Inhaltsverzeichnis

- [Kontingenztabellen](#)
- [Zusammenhangsmaße](#)
 - [Phi-Koeffizient](#)
 - [Kontingenz-Koeffizient](#)
 - [Cramer's V](#)
- [Tests auf Unabhängigkeit](#)
 - [Chi-Quadrat-Test](#)
 - [Exakter Fisher-Test](#)
- [McNemar-Test](#)

 fu:stat bietet regelmäßig Schulungen für [Hochschulen](#) sowie für [Unternehmen](#) und [weitere Institutionen](#) an. Die Inhalte reichen von Statistikgrundlagen (Deskriptive, Testen, Schätzen, lineare Regression) bis zu Methoden für Big Data. Es werden außerdem Kurse zu verschiedenen Software-Paketen gegeben. Auf Anfrage können wir auch gerne individuelle [Inhouse-Schulungen](#) bei Ihnen anbieten.

Kontingenztabellen

Diskrete und klassierte Merkmale können in Kontingenztabellen dargestellt werden. Dabei werden üblicherweise zwei Merkmale in Verbindung gebracht. Diese Merkmale haben jeweils bedingte und unbedingte Ausprägungen. Nehmen wir z.B. die beiden Merkmale Geschlecht und Erwerbstätigkeit. In der Stichprobe gibt es eine gewisse Anzahl an Männern und Frauen unabhängig von ihrem Erwerbsstatus. Ebenso gibt es eine gewisse Anzahl an Personen, die z.B. ganztags erwerbstätig oder eben nicht erwerbstätig sind. Nun könnte man aber auch Kombinationen betrachten, also z.B. die Männer, die ganztags erwerbstätig sind. All diese unterschiedlichen Ausprägungen der beiden Merkmale können mit Hilfe von Kontingenztabellen übersichtlich dargestellt werden. Dabei bezeichnen wir im Folgenden das eine Merkmal mit X und das zweite Merkmal mit Y:

X	Y						Σ
	y_1	y_2	...	y_I	...	y_L	
x_1	n_{11}	n_{12}	...	n_{1I}	...	n_{1L}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2I}	...	n_{2L}	$n_{2\bullet}$
...							
x_m	n_{m1}	n_{m2}	...	n_{mI}	...	n_{mL}	$n_{m\bullet}$
...							
x_M	n_{M1}	n_{M2}	...	n_{MI}	...	n_{ML}	$n_{M\bullet}$
Σ	$n_{\bullet 1}$	$n_{\bullet 2}$...	$n_{\bullet I}$...	$n_{\bullet L}$	n

X hat die Ausprägungen $\{x_1\}$ bis $\{x_M\}$ und Y die Ausprägungen $\{y_1\}$ bis $\{y_L\}$. Für $\{i=1 \dots M\}$ und $\{j=1 \dots L\}$ entspricht $\{n_{ij}\}$ der Häufigkeit $\{(X=x_i; Y=y_j)\}$. Die Zeilen enthalten L gemeinsame absolute Häufigkeiten oder anders gesagt absolute Häufigkeiten von Y bedingt durch Ausprägungen von $\{X=x_i\}$, z.B. die Elemente $\{n_{i1}\}$ bis $\{n_{iL}\}$ sind absolute Häufigkeiten von Ausprägungen von Y unter der Bedingung, dass $\{X=x_i\}$. Am Ende dieser Zeile steht die entsprechende Randhäufigkeit $\{n_{i\bullet}\}$ von X. Um bedingte relative Häufigkeiten von Y zu bekommen, dividiert man die Elemente $\{n_{i1}\}$ bis $\{n_{iL}\}$ durch die Randhäufigkeit $\{n_{i\bullet}\}$. Analog enthält jede Spalte M gemeinsame Häufigkeiten oder absolute Häufigkeiten von X bedingt durch Ausprägungen von $\{Y=y_j\}$.

Im Kontext des [Datenbeispiels](#) wäre interessant zu untersuchen, ob sich die Erwerbstätigkeit zwischen Frauen und Männern unterscheidet. Eine mögliche Betrachtung dieser Frage ermöglicht die folgende Kontingenztafel für die Variablen BERUFSTAETIG und GESCHL:

	MAENNlich	WEIBlich	$\{(\text{sum})\}$
HAUPTBERUFL.GANZTAGS	1026	545	1571
HAUPTBERUFL.HALBTAGS	41	309	350

Datei	Geändert
Datei Kontingenztafel len.R	21.04.2016 by Ann-Kristin Kreutzmann
Datei Kontingenztafel len.do	21.04.2016 by Ann-Kristin Kreutzmann
Datei Kontingenztafel len.sps	23.05.2016 by Alena Churakova
Datei Kontingenztafel len.sas	23.05.2016 by Alena Churakova

[Alle herunterladen](#)

NEBENHER BERUFSTAE.	73	135	208
NICHT ERWERBSTAETIG	619	720	1339
\sum	1759	1709	3468

Aus der Kontingenztabelle ist z.B. abzulesen:

- 41 Männer sind hauptberuflich halbtags erwerbstätig (BERUFSTAETIG = 2 ,GESCHL = 1)
- 545 Frauen sind hauptberuflich ganztags erwerbstätig (BERUFSTAETIG = 1, GESCHL = 2)

Die Werte in der letzten Spalte stellen die Randverteilung von BERUFSTAETIG dar und die letzte Zeile die Randverteilung von GESCHL. Die Frage zur Erwerbstätigkeit wurde von 1759 Männern und 1709 Frauen beantwortet.

Die obige Kontingenztabelle enthält absolute Häufigkeiten. Es lassen sich auch relative Häufigkeiten in einer Kontingenztabelle darstellen. Dafür werden alle absoluten Häufigkeiten durch die Gesamtanzahl an Beobachtungen geteilt.

	MAENNLICH	WEIBLICH	\sum
HAUPTBERUFL.GANZTAGS	0.296	0.157	0.453
HAUPTBERUFL.HALBTAGS	0.012	0.089	0.101
NEBENHER BERUFSTAE.	0.021	0.039	0.060
NICHT ERWERBSTAETIG	0.178	0.208	0.386
\sum	0.507	0.493	1.000

Auch Kontingenztabelle mit sog. bedingten relativen Häufigkeiten in den Zeilen oder Spalten können hilfreich sein. Bedingte relative Häufigkeiten lassen sich zeilen- oder spaltenweise ausrechnen.

Bezogen auf die Spalten werden die gemeinsamen Ausprägungen in Relation zur Gesamtanzahl von Frauen und Männern gesetzt.

	MAENNLICH	WEIBLICH
HAUPTBERUFL.GANZTAGS	0.583	0.318
HAUPTBERUFL.HALBTAGS	0.023	0.181
NEBENHER BERUFSTAE.	0.042	0.079
NICHT ERWERBSTAETIG	0.352	0.421
\sum	1.000	1.000

Innerhalb der Teilstichprobe von Männern (1759 Beobachtungen) sind $\frac{1026}{1759} \approx 0.583$ hauptberuflich ganztags erwerbstätig, während $\frac{41}{1759} \approx 0.023$ hauptberuflich halbtags erwerbstätig, $\frac{73}{1759} \approx 0.042$ nebenher erwerbstätig und $\frac{619}{1759} \approx 0.352$ nicht erwerbstätig sind.

Bezogen auf die Zeilen werden die gemeinsamen Ausprägungen in Relation zur Gesamtzahl in den unterschiedlichen Formen von Erwerbstätigkeit gesetzt.

	MAENNLICH	WEIBLICH	\sum

HAUPTBERUFL.GANZTAGS	0.653	0.347	1.000
HAUPTBERUFL.HALBTAGS	0.117	0.883	1.000
NEBENHER BERUFSTAE.	0.351	0.649	1.000
NICHT ERWERBSTAETIG	0.462	0.538	1.000

Insgesamt sind 1339 Personen in der Stichprobe nicht erwerbstätig, nämlich 619 Männer und 720 Frauen. Die relativen Häufigkeiten unter der Bedingung von keiner Erwerbstätigkeit sind $\frac{619}{1339} \approx 0.462$ und $\frac{720}{1339} \approx 0.538$, d.h. 46.23% der nicht erwerbstätigen Befragten sind männlich.

Zusammenhangsmaße

Wenn man an dem Zusammenhang zwischen zwei metrischen Variablen interessiert ist, wird häufig als erstes die Korrelation zwischen den beiden Variablen berechnet. Bei nominalen Variablen kann zwar keine Korrelation berechnet werden, aber dennoch gibt es Maße, die etwas über die Stärke des Zusammenhangs der Variablen aussagen, ohne dass ein Test angewendet werden muss. Im Folgenden werden drei Maße kurz vorgestellt: der Phi-Koeffizient, der Kontingenz-Koeffizient und Cramer's V. Alle drei Maße basieren auf dem χ^2 -Wert. Dies ist die Prüfgröße eines Tests auf Unabhängigkeit der beiden Merkmale.

Phi-Koeffizient

Der Phi-Koeffizient wird für 2x2 Kontingenztabellen angewendet. Der Koeffizient liegt zwischen 0, d.h. es gibt keinen Zusammenhang, und 1, d.h. es gibt einen perfekten Zusammenhang. $\phi = \sqrt{\frac{\chi^2}{n}}$

Kontingenz-Koeffizient

Der Kontingenz-Koeffizient (KK) kann nur bei symmetrischen Kontingenztabellen angewendet werden, d.h. wenn alle Variablen die gleiche Anzahl an Kategorien haben. Grundsätzlich liegt der KK zwischen 0 und 1, die entsprechend für keinen oder einen sehr starken Zusammenhang stehen. Allerdings wird bei kleinen Kontingenztabellen die 1 nicht erreicht, selbst wenn der Zusammenhang zwischen den beiden Variablen perfekt ist. Daher kann der KK nicht genutzt werden, um den Zusammenhang zwischen Variablen aus unterschiedlich großen Tabellen zu vergleichen. $KK = \sqrt{\frac{\chi^2}{\chi^2 + n}}$

Cramer's V

Cramer's V ist besonders gut, um die Zusammenhangsmaße von Variablen unterschiedlicher Kontingenztabellen miteinander zu vergleichen. Dabei gilt weiterhin, dass ein Wert von 0 angibt, dass kein Zusammenhang besteht und steigende Werte einen stärkeren Zusammenhang bedeuten mit einem Maximum von 1. $V = \sqrt{\frac{\chi^2}{n(k-1)}}$ Dabei gibt k die kleinere Anzahl an Zeilen oder Spalten an. Wenn es z.B. vier Zeilen und zwei Spalten gibt, ist k=2.

Tests auf Unabhängigkeit

Ganz allgemein kann man mit Hilfe von bedingten Häufigkeiten feststellen, ob man von einem fehlenden Zusammenhang zwischen zwei kategoriellen Merkmalen sprechen kann, d.h. von Unabhängigkeit der Merkmale. Intuitiv sind Merkmale dann unabhängig, wenn Ausprägungen von X keinen Einfluss auf Ausprägungen von Y haben und umgekehrt. Daraus folgt, dass die absoluten bedingten Häufigkeiten von Y unter der Bedingung $(X=x_i)$ nicht davon abhängen, welche Ausprägung $(x_1), (x_2) \dots (x_M)$ gewählt wird. Dann sollten die relativen bedingten Häufigkeiten $\frac{n_{m1}}{n_{m\bullet}} = \frac{n_{m2}}{n_{m\bullet}} = \dots = \frac{n_{mL}}{n_{m\bullet}}$

identisch sein.

Einer der bekanntesten Tests unter den Hypothesentests ist der [Chi-Quadrat-Test](#). Genauer genommen ist der Chi-Quadrat-Test eine Gruppe von Tests, deren Prüfgröße χ^2 verteilt ist. Der Chi-Quadrat-Test kann als Verteilungstest, Homogenitätstest und Unabhängigkeitstest fungieren. Hier soll jedoch im Folgenden nur der Unabhängigkeitstest beschrieben werden. Ebenso wird der [Exakte Fisher-Test](#) vorgestellt, der im Vergleich zum Chi-Quadrat-Test jedoch keine Anforderungen an den Stichprobenumfang stellt.

Chi-Quadrat-Test



Beispiele

Für die Zusammenhangsmaße gibt es keine Beispiele in den Statistikprogrammen. Aber diese [Zusammenfassung](#) erklärt jedes der drei Maße mit einfachen Rechenbeispielen.

Datei	Geändert
Datei Kontingenztable _Tests.R	04.05.2016 by Ann-Kristin Kreutzmann
Datei Kontingenztable _Tests.do	04.05.2016 by Ann-Kristin Kreutzmann
Datei Kontingenztable n_Tests.sps	23.05.2016 by Alena Churakova

Der Chi-Quadrat-Test als Unabhängigkeitstests prüft stochastische Unabhängigkeit zwischen zwei diskreten Zufallsvariablen X und Y. Dies führt zu folgender Nullhypothese und Alternative:

- H_0 : X und Y sind stochastisch unabhängig
- H_1 : X und Y sind stochastisch abhängig

Die Prüfgröße für den Chi-Quadrat-Unabhängigkeits-Test beruht auf dem Vergleich von erwarteten Häufigkeiten mit tatsächlich beobachteten Häufigkeiten. Wenn die erwarteten Häufigkeiten nicht zu sehr von den beobachteten Häufigkeiten abweichen, spricht man von Unabhängigkeit. Im Folgenden sollen zunächst die erwarteten Häufigkeiten für das Datenbeispiel berechnet werden: $\tilde{n}_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$, d.h. zum Beispiel $\frac{1571 \cdot 1759}{3468} = 796.825$ für die Zelle männlich und hauptberuflich ganztags erwerbstätig.

	MAENNLICH	WEIBLICH	$\backslash(\text{sum})$
HAUPTBERUFL.GANZTAGS	796.83	774.18	1571
HAUPTBERUFL.HALBTAGS	177.52	172.48	350
NEBENHER BERUFSTAE.	105.50	102.50	208
NICHT ERWERBSTAETIG	679.15	659.85	1339
$\backslash(\text{sum})$	1759	1709	3468

Um erwartete und beobachtete Häufigkeiten zu vergleichen, können Differenzen zwischen den beiden Häufigkeiten in Betracht gezogen werden. Diese können jedoch sowohl positiv als auch negativ sein. Daher werden quadrierte Differenzen für die Prüfgröße des Unabhängigkeitstests genutzt. Es werden alle Terme $\frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}}$ summiert. In einer Kontingenztafel mit i Zeilen kann man zuerst diese Terme in jeder Zeile summieren und dann die Summe von Zeilensummen bilden. $\chi^2 = \sum_{i=1}^M \sum_{j=1}^L \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}}$

Dieser χ^2 -Koeffizient ist die Prüfgröße für den Chi-Quadrat-Unabhängigkeits-Test und ist unter der Nullhypothese asymptotisch χ^2 -verteilt mit $(M-1) \cdot (L-1)$ Freiheitsgraden. Je größer der χ^2 -Koeffizient ist, desto eher trifft die alternative Hypothese zu. Der Ablehnungsbereich des Tests zum Niveau α ist rechts von einem Schwellenwert $\chi^2_{(M-1)(L-1), 1-\alpha}$. Im Datenbeispiel wird die Prüfgröße wie folgt berechnet:

$$\chi^2 = \frac{(1026-796.83)^2}{796.83} + \frac{(41-177.52)^2}{177.52} + \dots + \frac{(720-659.85)^2}{659.85} = 377.94$$

Der Wert der Prüfgröße, 377.94, wird mit den Schwellenwerten $\chi^2_{(3, 0.95)} = 7.81$ auf dem 5% Signifikanzniveau und $\chi^2_{(3, 0.99)} = 11.34$ auf dem 1% Signifikanzniveau verglichen. Sowohl für α von 5%, als auch von 1% wird die Nullhypothese abgelehnt. Daraus kann geschlossen werden: "Zwischen Erwerbstätigkeit und Geschlecht der befragten Personen besteht ein Zusammenhang".

Damit die Prüfgröße als annähernd χ^2 -verteilt angenommen werden kann, müssen Voraussetzungen an die erwarteten Häufigkeiten gegeben sein. Alle gemeinsamen Häufigkeiten \tilde{n}_{ij} müssen größer als Null sein und für nicht mehr als 20% der Werte darf gelten $\tilde{n}_{ij} < 5$. Sollten diese Voraussetzungen nicht erfüllt sein, kann der Exakte Fisher-Test angewendet werden.

Exakter Fisher-Test

Der Exakte Fisher-Test überprüft wie der Chi-Quadrat-Test die Unabhängigkeit zweier diskreter Merkmale. Dabei stellt der Exakte Fisher-Test jedoch keine Voraussetzungen an die Stichprobengröße. Der ursprüngliche Test ist auf 2x2 Kontingenztabellen ausgelegt und händisch berechenbar. Für Erweiterungen sind Statistikprogramme zur Berechnung erforderlich (siehe Infobox rechts). Der Exakte Fisher-Test unterscheidet sich von vielen Signifikanztests dahingehend, dass die Signifikanz einer Abweichung von der Nullhypothese exakt berechnet werden kann.

Der Exakte Fisher-Test wird anhand einer neuen Fragestellung beschrieben: "Hängt das Rauchverhalten der Befragten vom Geschlecht ab?" bzw. "Unterscheidet sich das Rauchverhalten der ältesten Befragten zwischen den Geschlechtern?" Die folgenden zwei Tabellen zeigen nämlich, dass, wenn nur die Befragten betrachtet werden, die vor 1933 geboren wurden, in zwei Zellen die erwarteten Häufigkeiten $\tilde{n}_{ij} < 5$ sind und somit der Chi-Quadrat-Test nicht mehr angewendet werden sollte.

Alle Befragten:

Beobachtung/erwartete Häufigkeit	MAENNLICH	WEIBLICH	$\backslash(\text{sum})$
RAUCHER: JA	586 / 506.34	412 / 491.66	998
RAUCHER: NEIN	1174 / 1253.66	1297 / 1217.34	2471
$\backslash(\text{sum})$	1760	1709	3469

Erweiterungen vom Exakten Fisher-Test

Der Exakte Fisher-Test kann auch für größere Kontingenztabellen durchgeführt werden. Unter diesem [Link](#) können Verweise auf Rechner gefunden werden, die die Teststatistik für größere Kontingenztabellen berechnen können. In **R** hat die Funktion `fisher.test` ein Argument `hybrid`, über das auch größere Kontingenztabellen genutzt werden können (siehe [Details](#)).

Alle Befragten, die vor 1933 geboren wurden:

	MAENNLICH	WEIBLICH	$\backslash(\text{sum})$
RAUCHER: JA	4 / 2.37	1 / 2.63	5
RAUCHER: NEIN	43 / 44.63	51 / 49.37	94
$\backslash(\text{sum})$	47	52	99

Für beide Tabellen gilt folgender Zusammenhang zwischen den Zelleinträgen:

	MAENNLICH	WEIBLICH	$\backslash(\text{sum})$
RAUCHER: JA	a	b	a + b
RAUCHER: NEIN	c	d	c + d
$\backslash(\text{sum})$	a + c	b + d	a + b + c + d (=n)

Die Wahrscheinlichkeit für eine Kombination der Einträge a-d folgt nach Fisher einer hypergeometrischen Verteilung und kann wie folgt berechnet werden $p = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{a! b! c! d! n!}$

Im Beispiel führt dies für die oben angegebene Kombination zu: $p = \frac{(4+1)! (43+51)! (4+43)!}{(1+51)! 4! 1! 43! 51! 99!} = 0.129678$

Beim Exakten Fisher-Test werden die Randhäufigkeiten immer als gegeben angesehen. In diesem Beispiel bedeutet dies, dass 5 der 99 älteren Befragten rauchen. Was ist die Wahrscheinlichkeit, unter der Annahme, dass Frauen und Männer mit gleicher Wahrscheinlichkeit rauchen, dass nur eine der 52 Frauen oder sogar keine raucht, aber dafür 4 oder 5 der 47 Männer? Die Wahrscheinlichkeit für den ersten Fall wurde oben berechnet. Laut Fisher müssen zur Berechnung des Signifikanzniveaus nun alle weiteren extremen Fälle berechnet werden, bei gleichbleibender Randhäufigkeit.

Dies bedeutet im Beispiel: $p = \frac{(5+0)! (42+52)! (5+42)! (0+52)!}{5! 0! 42! 52! 99!} = 0.02144675$

Die Addition der beiden Wahrscheinlichkeiten gibt den p-Wert=0.1511248 des einseitigen Exakten Fisher-Tests mit der Alternative, dass Männer höheren Alters eher rauchen als Frauen. Die Nullhypothese, dass es keinen Unterschied im Rauchverhalten zwischen älteren Männern und Frauen könnte auf dem 15,11% Niveau abgelehnt werden. Je kleiner der p-Wert, desto eher kann die Nullhypothese abgelehnt werden. Für einen zweiseitigen Test müssten auch die Extremfälle in die andere Richtung berücksichtigt werden. Dafür werden alle Kombinationen berücksichtigt, deren Wahrscheinlichkeit gleich oder geringer ist als die Wahrscheinlichkeit der beobachteten Tabelle.

McNemar-Test

Ein McNemar-Test wird angewendet, wenn das gleiche Merkmal vor und nach einem Treatment abgefragt wird. Dies kommt z.B. vor, wenn Kunden vor und nach einer Kampagne nach ihrer Meinung zu einem Produkt gefragt werden oder auch beim Testen eines Medikamentes. Dadurch, dass die gleichen Personen wieder befragt oder untersucht werden, sind die Stichproben nicht unabhängig, sondern miteinander verbunden.

Der McNemar-Test wird eingesetzt, um Unterschiede zwischen dem Merkmal vor und nach Durchführung eines beliebigen Treatments herauszufinden. Dabei kann jedoch nur eine Veränderung festgestellt werden, nicht jedoch in welche Richtung eine Veränderung vorliegt.

Die Kontingenztabelle verändert sich dahingehend, dass nun in den Zeilen und Spalten das gleiche Merkmal aus unterschiedlichen Stichproben steht:

	Stichprobe 1		\sum
Stichprobe 2	0	1	
0	a	b	a + b
1	c	d	c + d
\sum	a + c	b + d	n

Veränderungen werden vor allem in den Zellen b und c deutlich. Daher werden diese beiden Zellen in den Teststatistiken genutzt. Auch beim McNemar-Test gilt, dass es einen exakten Test gibt für geringe Beobachtungsumfänge, d.h. wenn $b + c < 25$, und eine χ^2 -Teststatistik für größere Beobachtungsumfänge.

Exakter Test: $b + c < 25$

Für den exakten Test werden die Beobachtungen b und c als Zufallszahlen angesehen, sodass die Hypothesen mit der Wahrscheinlichkeit p für entweder b oder c folgen:

$$H_0: p = 0.5$$

$$H_1: p \neq 0.5$$

Die Anzahl der Beobachtungen b, bzw. c sind dann binomial verteilt mit $(b \sim \text{Bin}(b+c, 0.5), c \sim \text{Bin}(b+c, 0.5))$.

χ^2 -Teststatistik

Die Hypothese bleibt weiterhin, dass es keine Veränderung gibt, gegenüber der Alternativen, dass es eine Veränderung gibt, d.h. $a + b = a + c$ bzw. $c + d = b + d$ und übersetzt in Wahrscheinlichkeiten, dass

$$H_0: p_a + p_c = p_a + p_b$$

$$H_1: p_a + p_c \neq p_a + p_b$$

Die Teststatistik $\chi^2 = \frac{(b-c)^2}{(b+c)}$ ist approximativ χ^2 -verteilt mit einem Freiheitsgrad. Da die χ^2 -Verteilung stetig ist, aber die Beobachtungen diskret, gibt es einen Approximationsfehler, der durch die Yates-Korrektur $\chi^2 = \frac{(|b-c| - 0.5)^2}{(b+c)}$, bzw. mit der Edwards-Korrektur $\chi^2 = \frac{(|b-c| - 1)^2}{(b+c)}$ verringert werden kann. R und SPSS verwenden die Edwards-Korrektur. Ist die errechnete Teststatistik größer als der entsprechende Wert der χ^2 -Verteilung, kann die Nullhypothese abgelehnt werden und es liegt eine Veränderung vor. Wie schon erwähnt wird keine Aussage darüber getroffen, in welche Richtung eine Veränderung vorliegt.

Anhand des Beispiels, ob Kunden ihre Meinung über ein Produkt nach einer Werbekampagne ändern, wird der McNemar-Test illustriert. Für dieses Beispiel kann der Beispieldatensatz nicht genutzt werden, da die ALLBUS-Umfrage die Befragten nur zu einem Zeitpunkt befragt und kein Treatment zwischen zwei Befragungszeitpunkten vornimmt.

Wir nehmen folgende Beobachtungen fiktiv an für eine Entscheidung darüber, ob sich die Kaufentscheidung nach der Kampagne ändert.

	Kaufentscheidung nach der Kampagne		\sum
Kaufentscheidung vor der Kampagne	Kaufen	Nicht kaufen	
Kaufen	18	15	33
Nicht kaufen	26	21	47
\sum	44	36	80

Die Hypothese ist in diesem Fall, dass die Wahrscheinlichkeit, dass es einen Wechsel von der Entscheidung Kaufen zu Nicht kaufen gibt, gleich der Wahrscheinlichkeit, dass es einen Wechsel von Nicht kaufen zu Kaufen gibt ist. Die Alternative ist, dass sich die Wahrscheinlichkeiten unterscheiden. Sollte die Nullhypothese also abgelehnt werden, liegt eine Veränderung vor. Jedoch besagt der McNemar-Test nicht, in welche Richtung.

Die Teststatistik berechnet sich unter Berücksichtigung der McNemar-Korrektur nach Edwards, gegeben, dass $b = 15$ und $c = 26$, wie folgt:

$$\chi^2 = \frac{(|15-26| - 1)^2}{(15+26)} = 2.439024$$

Der kritische Wert der χ^2 - Verteilung entspricht auf einem 5%-Signifikanzniveau und einem Freiheitsgrad $\chi^2_{(1,0.95)} = 3.84$. Da die Teststatistik kleiner ist als der kritische Wert, kann die Nullhypothese nicht abgelehnt werden. Dies ist ein Indiz dafür, dass die Kampagne nicht für Veränderungen im Kaufverhalten sorgt hat.