Inferenzstatistik für nominale Daten: Gliederung

1 Tests für Häufigkeitsunterschiede bei zwei unabhängigen Gruppen: Vierfelder χ^2- und Fisher-Yates Test

Exkurs: Gewichtung von Fällen in SPSS

2 Tests für k x m Kontingenztafeln

3 Weitere Tests für Häufigkeitsunterschiede

4 Simpsons Paradox

Einführende Literatur

📖 Bortz, J. & Schuster, Ch. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Auflage). Berlin: Springer. [Kap. 9]

📖 Diehl, J. M. & Arbinger, R. (2001). *Einführung in die Inferenzstatistik* (3. Auflage). Eschborn bei Frankfurt: Klotz Verlag. [Kap. 19]

📖 Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz. [Kap. 11.4 & 12.2]

Weiterführende Literatur

📖 Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). New York: Wiley.

📖 Bortz, J., Lienert, G. A. & Boehnke, K. (2010). *Verteilungsfreie Methoden in der Biostatistik*. (3. Auflage). Berlin: Springer.

- Liegen zwei nominalskalierte Variablen vor, so kann man den Zusammenhang der beiden Variablen in einer Kontingenztafel darstellen. Im einfachsten Fall sind die beiden Variablen nur dichotom (Vierfeldertafel). Besteht ein bivariater Zusammenhang, so ...
  - sind beide Variablen abhängig, andernfalls unabhängig.
  - sind die beiden relativen Anteile im Merkmal Y in den beiden durch X definierten Gruppen $p_1 = h_{11} / h_{1\bullet}$ und $p_2 = h_{21} / h_{2\bullet}$ unterschiedlich (bzw. umgekehrt).
  - weichen die bei Unabhängigkeit zu erwartenden Häufigkeiten $e_{ij} = h_{i\bullet} \cdot h_{\bullet j} / n$ von den beobachteten Häufigkeiten h_{ij} ab, aufsummiert in der χ^2-Statistik: $$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(h_{ij} - e_{ij})^2}{e_{ij}}$$
  - ist der die Enge des Zusammenhangs (bei natürlich dichotomen Variablen) quantifizierende φ-Koeffizient ungleich 0.

- Für φ gilt:

$$|\varphi| = \sqrt{\frac{\chi^2}{n}} = 0.87 = \sqrt{\frac{0.762}{100}}$$

$$\begin{array}{cc|c|c|c} & & \multicolumn{2}{c}{Y} & \\ & & y_1 & y_2 & \\ \hline X & x_1 & h_{11} & h_{12} & h_{1\bullet} \\ \hline & x_2 & h_{21} & h_{22} & h_{2\bullet} \\ \hline & & h_{\bullet 1} & h_{\bullet 2} & n \end{array}$$

	Nicht-raucher	Raucher		
♀	37	13	50	$p_1 = .74$
♂	33	17	50	$p_2 = .66$
	70	30	100	

Vierfelder χ^2-Test

- <u>Bezeichnung</u>: Vierfelder χ^2-Test.

- <u>Einsatzbereich</u>: Prüfung der Hypothese, dass sich in zwei unabhängigen Stichproben die relativen Auftretenshäufigkeiten (Anteile, Auftretenswahrscheinlichkeiten) eines Merkmals $p_1 = h_{11} / h_{1\bullet}$ und $p_2 = h_{21} / h_{2\bullet}$ unterscheiden. Dies entspricht der Prüfung auf Unabhängigkeit der beiden Variablen X und Y.

- Mögliche Konstellationen sind:
  - Aus den Populationen der Frauen und der Männer werden jeweils Zufallsstichproben mit $n = 50$ gezogen und erfasst, ob diese rauchen oder nicht.
  - Man weist 60 Personen einer Zufallsstichprobe mit einer klinischen Störung per Randomisation einer behandelten Experimentalgruppe (EG) oder einer Kontrollgruppe (KG) zu und erfasst die Häufigkeiten, mit der in beiden Gruppen ein Behandlungserfolg eingetreten ist oder nicht.
  - Es wird eine Zufallsstichprobe von Männern gezogen und erfasst, ob sie regelmäßig Sport treiben und ob sie regelmäßig Alkohol trinken.
  - Es wird eine Zufallsstichprobe von 200 Personen gezogen und die Zahl der durchschnittlich pro Tag gerauchten Zigaretten und das Alter erfasst und beide Variablen am Median dichotomisiert (nicht sinnvoll!).

	Nicht raucher	Raucher	
♀			50
♂			50

	Erfolg	kein Erfolg	
EG			30
KG			30

	Alkohol	kein Alkohol
Sport		
kein Sport		

	mehr	weniger	
älter			100
jünger			100
	100	100	200

Vierfelder χ^2-Test

- Hypothesen: zweiseitig: H_0: $\pi_1 = \pi_2$ und H_1: $\pi_1 \neq \pi_2$ (mit den π als Parameter für die Anteile p_1 und p_2) oder entsprechend einseitig. Alternativ kann man zweiseitig auch formulieren: H_0: $\pi_{ij} = \pi_{i\bullet} \cdot \pi_{\bullet j}$ für alle i, j (mit den $\pi_{i\bullet}$ und $\pi_{\bullet j}$ als Populations-Randwahrscheinlichkeiten).

- Voraussetzungen: Die Gruppen sind unabhängig und die erwartete Häufigkeiten in allen Zellen nicht „zu klein".

In vielen (älteren) Lehrbüchern wird gefordert, dass alle $e_{ij} \geq 5$ sein sollten. Auch eine Empfehlung von Cochran wird häufig zitiert, wonach nicht mehr als 20% der erwarteten Häufigkeiten zwischen 1 and 5 liegen sollten und keine davon kleiner als 1 sein sollte.

Computersimulationen zeigen, dass bei Stichprobenumfängen von $n \geq 8$ eine ausreichende Kontrolle des α-Fehlers erfolgt (Ruxton & Neuhäuser, 2010). Die Power ist dann allerdings, wie immer bei so kleinem n, sehr gering.

Stichprobe	Population
p_1, p_2	π_1, π_2
$h_{i\bullet}/n,\ h_{\bullet j}/n$	$\pi_{i\bullet},\ \pi_{\bullet j}$

		Y: y_1	y_2	
X	x_1	h_{11}	h_{12}	$h_{1\bullet}$
	x_2	h_{21}	h_{22}	$h_{2\bullet}$
		$h_{\bullet 1}$	$h_{\bullet 2}$	n

	Nicht-raucher	Raucher	
♀	37	13	50
♂	33	17	50
	70	30	100

$$p_1 = h_{11} / h_{1\bullet} = 37/50 = 0.74$$

$$p_2 = h_{21} / h_{2\bullet} = 33/50 = 0.66$$

Vierfelder χ^2-Test

- Vorgehen: Unter der Bedingung, dass alle erwarteten Häufigkeiten in den Zellen nicht zu klein sind, folgt die Prüfgröße

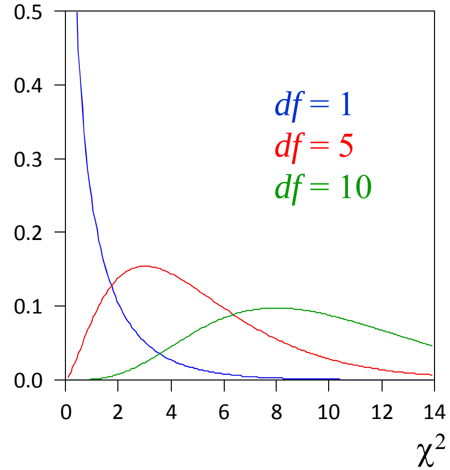
$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(h_{ij}-e_{ij})^2}{e_{ij}}$$

  bei Gültigkeit der H_0 einer χ^2-Verteilung mit $df = 1$ Freiheitsgrad.

- Entscheidung: Zurückweisung der H_0, falls $\chi^2 > \chi^2_{crit}$ mit $\chi^2_{crit} = \chi^2_{1,1-\alpha}$ bei zweiseitiger Prüfung und $\chi^2_{crit} = \chi^2_{1,1-2\cdot\alpha}$ bei einseitiger Prüfung (und korrekter Richtung). Nähere Erläuterung auf folgender Folie.

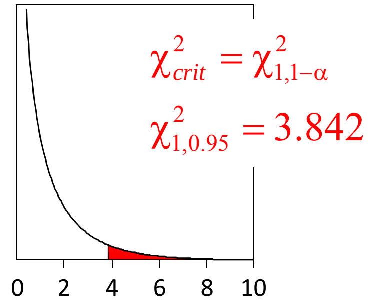
- Hinweis: Der χ^2-Test kann auch zur inferenzstatistischen Prüfung der Abweichung der φ-Korrelation von 0 verwendet werden (oder alternativ ein z-Test; vgl. Diehl & Arbinger, 1993, S. 391 & 496f) .

- Wiederholung (vgl. rechts): Die χ^2-Verteilung ist asymmetrisch, „beginnt“ bei $\chi^2 = 0$ und weist nur „rechts“ eine Zurückweisungsregion auf.



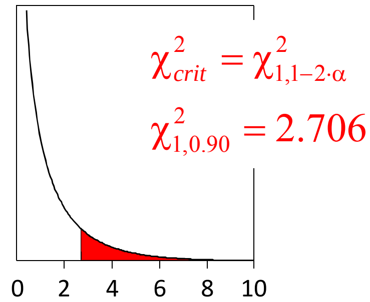
- Durch die Quadrierung der Abweichungen in der Prüfgröße vergrößern Abweichungen in beide Richtungen ($p_1 > p_2$ und $p_1 < p_2$) den χ^2-Wert.

- Bei zweiseitiger Prüfung wird der kritische Wert entsprechend so festgelegt, dass er den Anteil $1-\alpha$ der Verteilung abschneidet.



	Nicht-raucher	Raucher	
♀	37	13	$p_1 = .74$
♂	33	17	$p_2 = .66$

- Bei einseitiger Prüfung wird der kritische Wert entsprechend so festgelegt, dass er den Anteil $1-2\cdot\alpha$ der Verteilung abschneidet, wobei immer dann, wenn die Richtung von $p_1 - p_2$ hypothesenkonträr ist, die H_0 sowieso beibehalten wird.



♀	33	17	$p_1 = .66$
♂	37	13	$p_2 = .74$

In beiden Fällen ist $\chi^2 = 0.762$

Vierfelder χ^2-Test

- **Beispiel**: An einer Stichprobe von 100 Personen wurden die beiden dichotomen Variablen Geschlecht (X, codiert $0 = ♀$, $1 = ♂$) und Rauchverhalten erhoben (Y, codiert 0 = Nichtraucher, 1 = Raucher). Frage: Unterscheidet sich der Anteil der Nichtraucher bei Männern und Frauen? Die beiden Anteile betragen

$$p_1 = p_{♀} = 37/50 = 0.74;\ \ p_2 = p_{♂} = 33/50 = 0.66$$

Vp	X	Y
1	1	0
2	0	0
3	1	1
…		
100	1	1

- Die Frage soll zweiseitig bei $\alpha = 0.05$ getestet werden. Es resultiert:

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(h_{ij}-e_{ij})^2}{e_{ij}} = \frac{(37-35)^2}{35} + \ldots + \frac{(17-15)^2}{15} = 0.762$$

h_{ij}	0 = Nichtraucher	1 = Raucher	
0 = ♀	37	13	50
1 = ♂	33	17	50
	70	30	100

Der χ^2-Verteilung entnehmen wir für $\alpha = 0.05$ und $df = 1$ sowie zweiseitiger Testung den Wert $\chi^2_{crit} = \chi^2_{1,0.95} = 3.842$.

z.B.

$$e_{11} = \frac{h_{1\bullet}\cdot h_{\bullet 1}}{n} = \frac{50\cdot 70}{100} = 35$$

e_{ij}	0	1
0	35	15
1	35	15

Da $\chi^2 = 0.762 < \chi^2_{crit} = 3.842$, schließen wir also, dass der Anteil der Nichtraucher sich bei Männern und Frauen nicht statistisch signifikant unterscheidet.

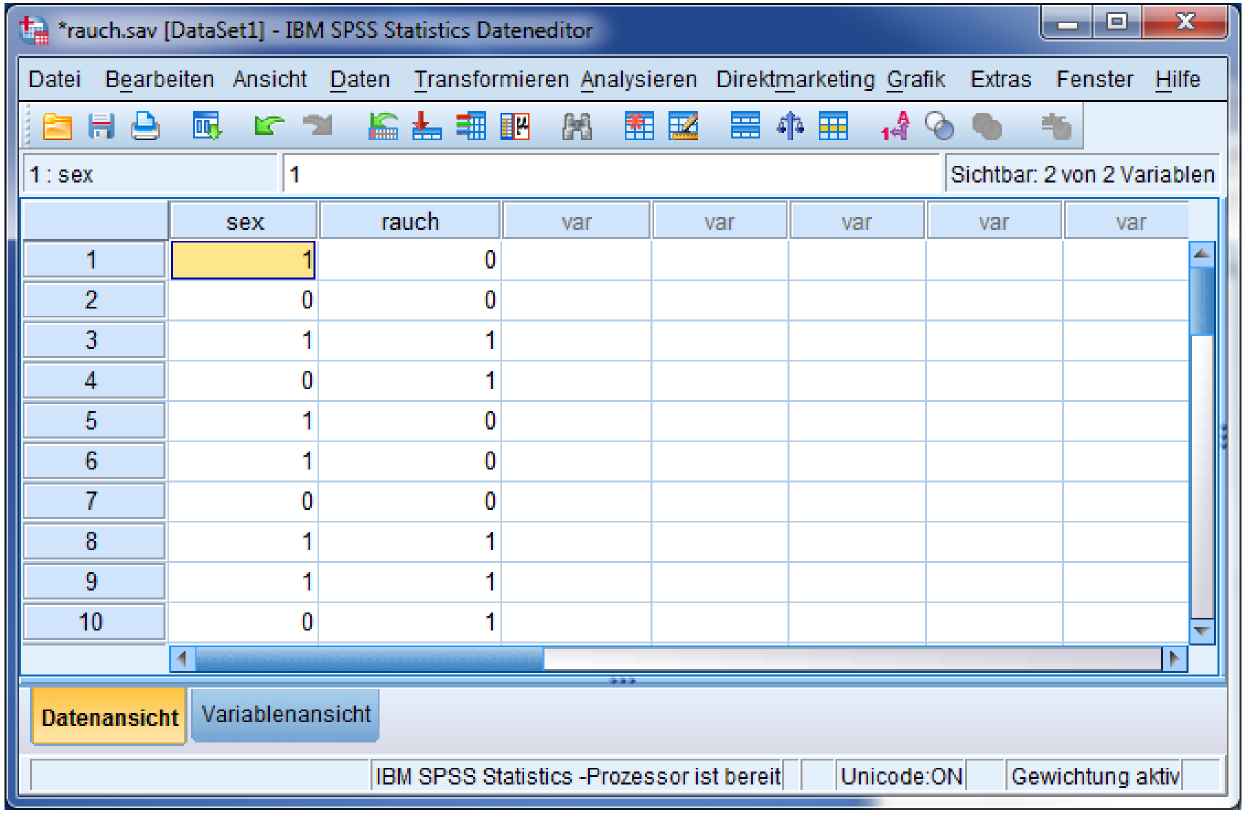
Vierfelder χ^2- und Fisher-Yates Test

- Alternative I: Bei sehr kleinen Stichproben (und auch sonst) kann alternativ Fishers exakter Test (=Fisher-Yates-Test) angewendet werden. In dem Test werden für alle möglichen Kontingenztafeln, die unter den gegebenen Randverteilungen möglich sind, und die so stark wie die empirisch gegebene Verteilung oder stärker von der unter der H_0 resultierenden Verteilung abweichen, deren Wahrscheinlichkeiten (die einer hypergeometrischen Verteilung folgen) kumuliert und geprüft, ob das Resultat kleiner als α ist, was zur Zurückweisung der H_0 führen würde. (Der Test ist also konditional zu beiden als fest angenommenen Randverteilungen.)

- Alternative II: Manchmal wird bei der Berechnung der χ^2-Statistik (vor allem bei kleinem n) noch die auf Yates zurückgehende Stetigkeitskorrektur der χ^2-Prüfgröße angewendet, bei der im Zähler von der Differenz der Häufigkeiten jeweils noch der Wert 0.5 subtrahiert wird $(\left|h_{ij} - e_{ij}\right| - 0.5)^2$ (was zu konservativeren Entscheidungen führt).

- Von den meisten Statistikern wird diese Yates-Korrektur aber nicht mehr empfohlen. Die Frage, ob die χ^2-Statistik oder Fishers exakter Test besser geeignet ist, wird kontrovers diskutiert (hinsichtlich der Kontrolle des α-Fehlers, der Power, der Annahme fixer Randverteilungen etc., vgl. Ruxton & Neuhäuser, 2010). In den meisten Fällen führen aber beide Tests zu identischen statistischen Entscheidungen.

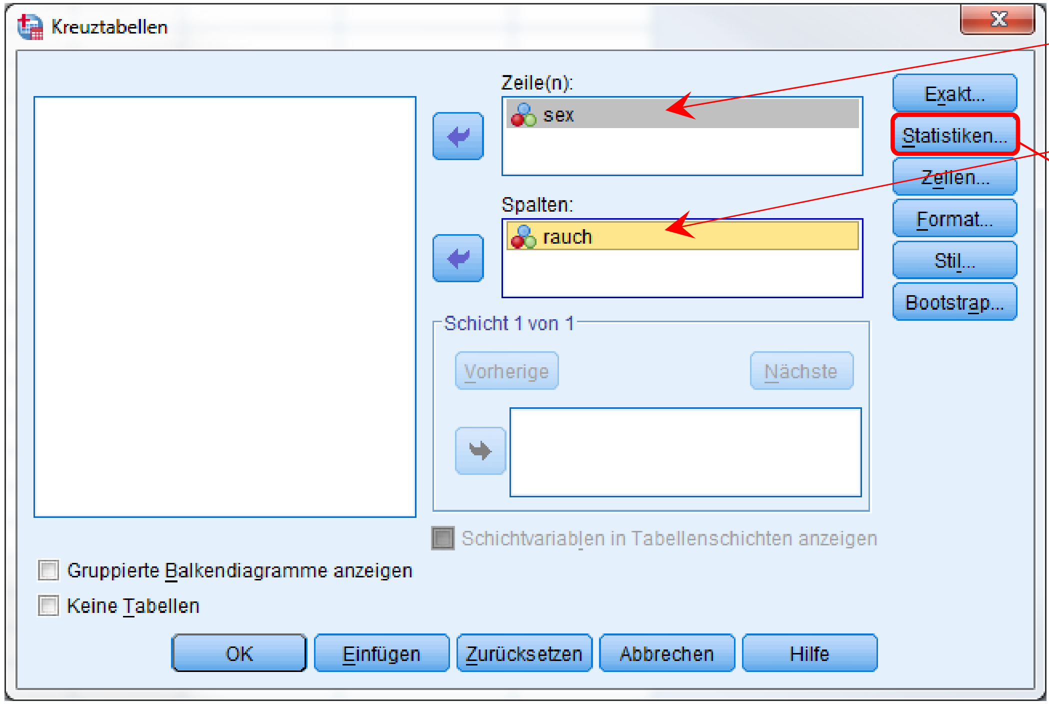
- Den χ^2- und den Fisher-Yates Test für unabhängige Stichproben (für den allgemeinen Fall einer $k \times m$ Kontingenztafel) erhält man unter: `Analysieren/Deskriptive_Statistiken/ Kreuztabellen…` .



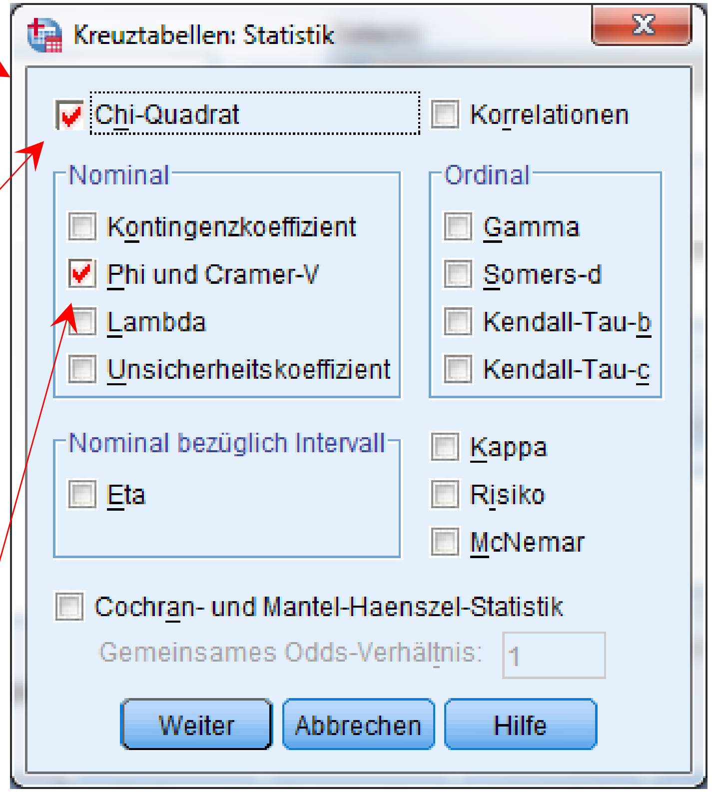
- Die Daten sind wie üblich angeordnet (im Beispiel die $n = 100$ Zeilen = Fälle in den beiden dichotomen 0-1 codierten Variablen `sex` und `rauch`).

h_{ij}	0 = Nichtraucher	1 = Raucher	
0 = ♀	37	13	50
1 = ♂	33	17	50
	70	30	100

Vierfelder χ^2-Test in SPSS



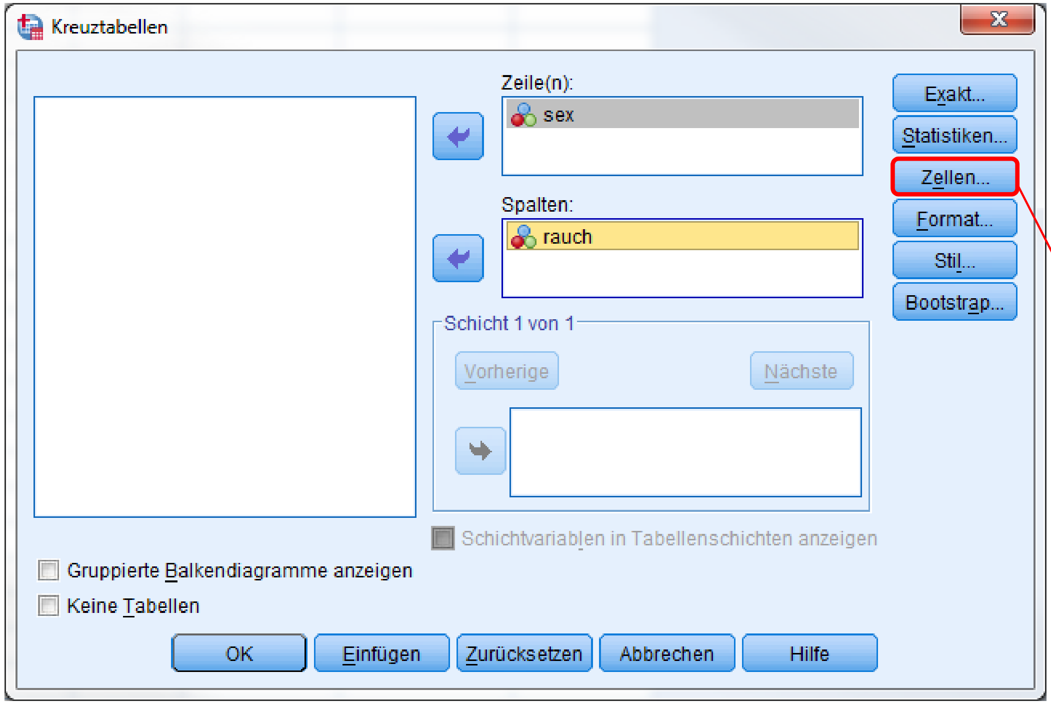
Von beiden nominalen Variablen sind die eine unter "Zeilen" und die andere unter "Spalten" einzutragen.



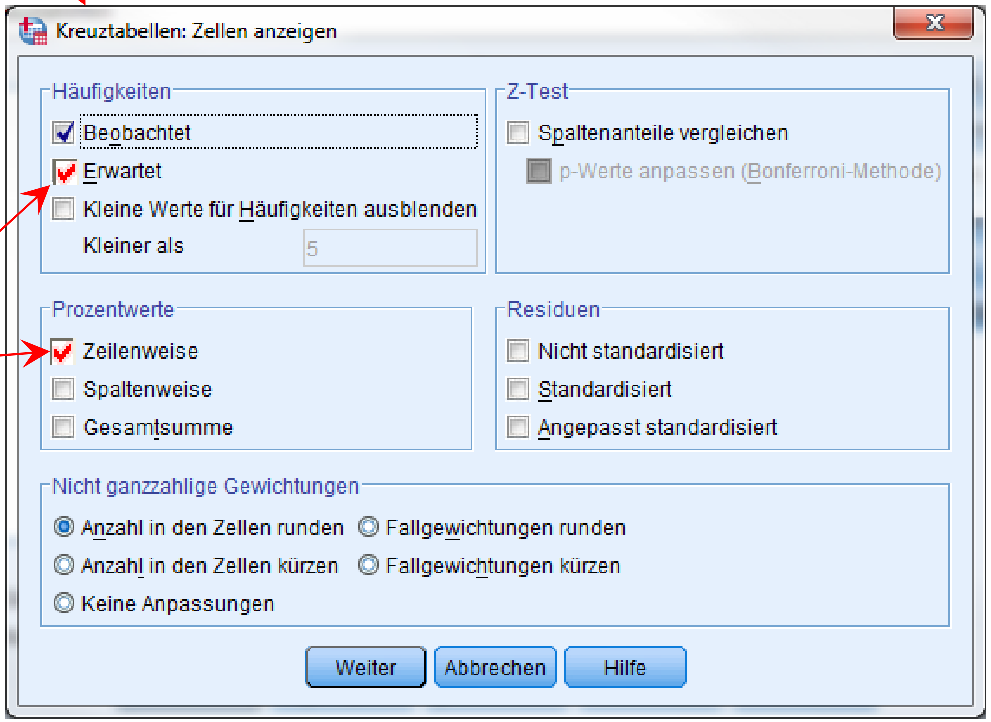
Wird unter (`Statistiken...`) die Option "Chi-Quadrat" aktiviert, so werden (unter anderem) die χ^2-Prüfgröße (incl. Kontinuitätskorrektur bei 2 x 2-Tafeln) sowie Fishers exakter Tests (standardmäßig nur im Falle einer 2 x 2-Tafel) ausgegeben.

Hier kann der φ-Koeffizient angefordert werden.

Vierfelder χ^2-Test in SPSS



Unter (`Zellen…`) kann die zusätzliche Ausgabe der erwarteten Häufigkeiten angefordert werden sowie eine Tabelle mit den zeilen- bzw. spaltenweise oder auf die Gesamtzahl prozentualisierten Zellhäufigkeiten.



Kreuztabellen

sex * rauch Kreuztabelle

Anzahl

		rauch		
		0 Nichtraucher(in)	1 Raucher(in)	Gesamt
sex	0 weiblich	37	13	50
	1 männlich	33	17	50
Gesamt		70	30	100

Standardmäßig wird die Kontingenztafel mit den absoluten Häufigkeiten ausgegeben.

Kreuztabellen

sex * rauch Kreuztabelle

			rauch		
			0 Nichtraucher(in)	1 Raucher(in)	Gesamt
sex	0 weiblich	Anzahl	37	13	50
		Erwartete Anzahl	35,0	15,0	50,0
		% innerhalb von sex	74,0%	26,0%	100,0%
	1 männlich	Anzahl	33	17	50
		Erwartete Anzahl	35,0	15,0	50,0
		% innerhalb von sex	66,0%	34,0%	100,0%
Gesamt		Anzahl	70	30	100
		Erwartete Anzahl	70,0	30,0	100,0
		% innerhalb von sex	70,0%	30,0%	100,0%

Diese Ausgabe erfolgt, wenn man zusätzlich die Ausgabe der erwarteten Häufigkeiten und der zeilenweisen Prozentwerte anfordert.

Die bei Unabhängigkeit zu erwartende Häufigkeit für Raucherinnen beträgt $e_{12} = 15$.

66% der Männer rauchen nicht.

30% aller Personen rauchen.

Vierfelder χ^2-Test in SPSS

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	,762[a]	1	,383		
Kontinuitätskorrektur[b]	,429	1	,513		
Likelihood-Quotient	,764	1	,382		
Exakter Test nach Fisher				,513	,257
Zusammenhang linear-mit-linear	,754	1	,385		
Anzahl der gültigen Fälle	100				

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 15,00.

b. Wird nur für eine 2x2-Tabelle berechnet

Im χ^2-Test zeigt die Prüfgröße $\chi^2(df=1) = 0.762$ keinen statistisch signifikanten Geschlechtsunterschied bzgl. des Rauchens an ($0.383 > \alpha = 0.05$). D.h.: Es besteht kein statistisch signifikanter Zusammenhang zwischen dem Geschlecht und dem Rauchen. Oder anders formuliert: Der Anteil der Nichtraucher unter den Frauen unterscheidet sich nicht statistisch signifikant von dem bei den Männern.

Vierfelder χ^2-Test in SPSS

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	,762[a]	1	,383		
Kontinuitätskorrektur[b]	,429	1	,513		
Likelihood-Quotient	,764	1	,382		
Exakter Test nach Fisher				,513	,257
Zusammenhang linear-mit-linear	,754	1	,385		
Anzahl der gültigen Fälle	100				

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 15,00.

b. Wird nur für eine 2x2-Tabelle berechnet

Auch der Fisher-Yates Test zeigt mit $p = 0.513 > \alpha = 0.05$ keinen statistisch signifikanten Zusammenhang an.

Man sieht, dass SPSS in der zweiten Zeile auch noch das nicht zu empfehlende stetigkeits- bzw. kontinuitätskorrigierte χ^2 anzeigt und in der Fußnote die Information ausgibt, wieviel Prozent der Zellen eine erwartete Häufigkeit kleiner als 5 aufweisen und was die minimale erwartete Häufigkeit ist.

Vierfelder χ^2-Test in SPSS

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	,762[a]	1	,383		
Kontinuitätskorrektur[b]	,429	1	,513		
Likelihood-Quotient	,764	1	,382		
Exakter Test nach Fisher				,513	,257
Zusammenhang linear-mit-linear	,754	1	,385		
Anzahl der gültigen Fälle	100				

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 15,00.

b. Wird nur für eine 2x2-Tabelle berechnet

Symmetrische Maße

		Wert	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,087	,383
	Cramer-V	,087	,383
Anzahl der gültigen Fälle		100	

Der zusätzlich angeforderte Zusammenhang zwischen beiden Variablen beträgt $\varphi = 0.087$ und ist (da mittels des χ^2-Tests geprüft) ebenfalls nicht statistisch signifikant.

1 Tests für Häufigkeitsunterschiede bei zwei unabhängigen Gruppen: Vierfelder χ^2- und Fisher-Yates Test

Exkurs: Gewichtung von Fällen in SPSS

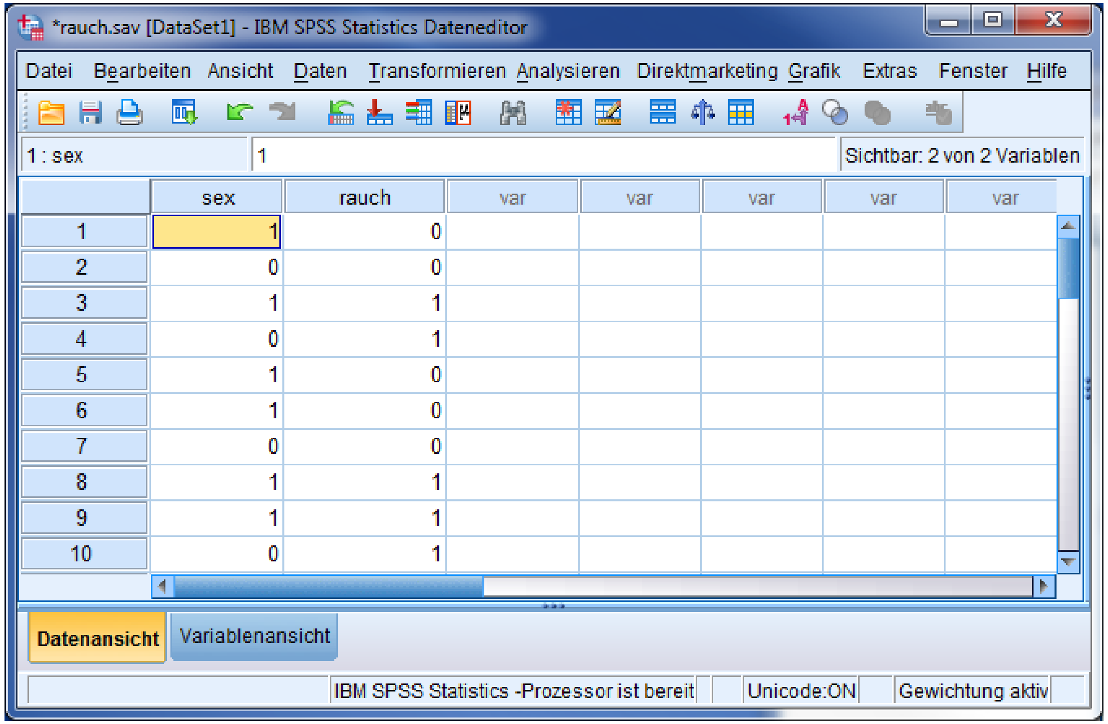
2 Tests für k x m Kontingenztafeln

3 Weitere Tests für Häufigkeitsunterschiede

4 Simpsons Paradox

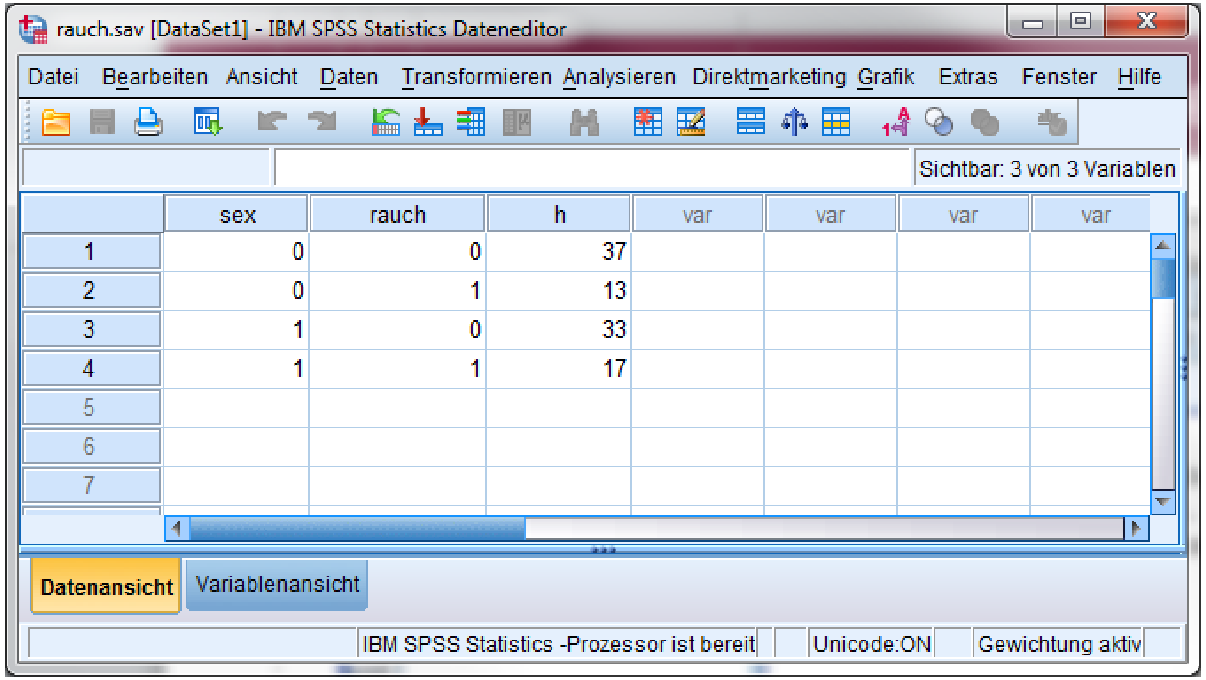
- Exkurs: Gewichtete Fälle in SPSS. Haben Personen in allen eine Analyse betreffenden Variablen die gleichen Werte, so können diese auch nur einmal eingegeben werden und in einer weiteren Variable (im Beispiel: h) die Häufigkeit angegeben werden, mit der dieses Datenprofil auftritt.
- SPSS kann dann mitgeteilt werden, dass alle nachfolgenden Analysen mit dieser Variable gewichtet werden sollen, d.h. für die Analyse jeder Fall so häufig „dupliziert" werden muss, wie in der Variable steht.
- Diese Option ermöglicht in diesem Fall (und auch in anderen Fällen) eine vereinfachte Dateneingabe, v.a. für die Sekundäranalyse von Daten.

h_{ij}	0 = Nicht-raucher	1 = Raucher
0 = ♀	37	13
1 = ♂	33	17



- Liegt beispielsweise in der Literatur die nebenstehende Kontingenztafel vor, so kann man die Daten leicht wie folgt unter Nutzung einer Gewichtungsvariable h eingeben:

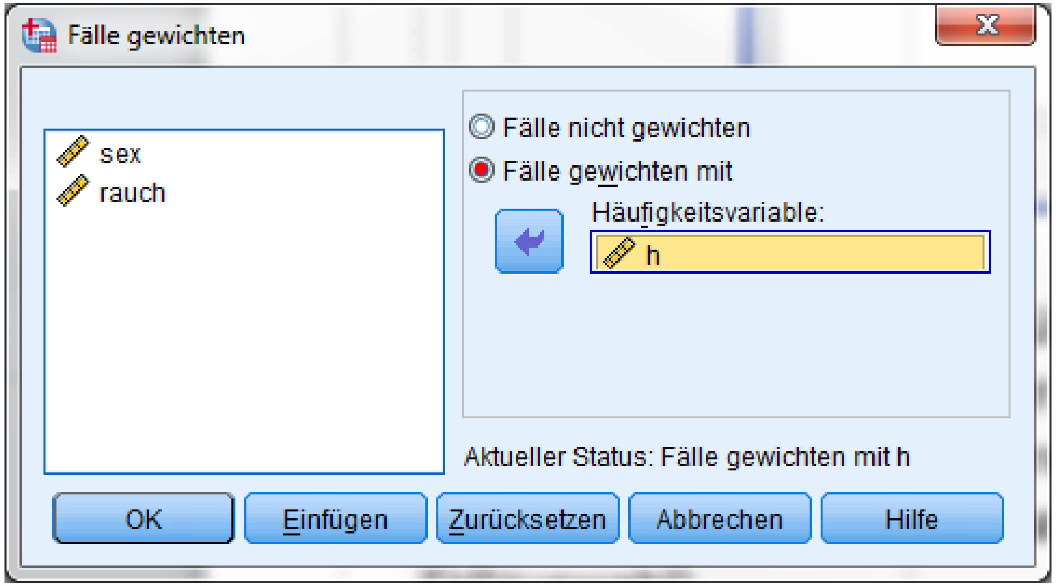
h_{ij}	0 = Nicht-raucher	1 = Raucher
0 = ♀	37	13
1 = ♂	33	17



- Die Gewichtung ist eine mächtige Funktion, die auch in anderen Kontexten eingesetzt werden kann.

  Z.B. wenn man eine geschichtete Zufallsauswahl mit disproportionaler Schichtung gezogen hat und die Disproportionalität wieder korrigieren will.

- Bevor nun die Analyse durchgeführt wird, muss SPSS unter `Daten/Fälle_gewichten…` mitgeteilt werden, dass alle nachfolgenden Analysen gewichtet werden sollen und wie die Gewichtungsvariable heißt:



- Alle im Folgenden durchgeführten Analysen werden nun mit der Variable h gewichtet. Die Gewichtung wird erst wieder aufgehoben, wenn (1) oben die Option „Fälle nicht gewichten“ aktiviert wird, (2) eine andere Häufigkeitsvariable unter „Fälle gewichten mit“ angegeben wird oder (3) eine andere Datendatei aktiviert wird bzw. SPSS verlassen wird (analog der Fallauswahl).

1 Tests für Häufigkeitsunterschiede bei zwei unabhängigen Gruppen: Vierfelder χ^2- und Fisher-Yates Test

Exkurs: Gewichtung von Fällen in SPSS

2 Tests für k x m Kontingenztafeln

3 Weitere Tests für Häufigkeitsunterschiede

4 Simpsons Paradox

Signifikanztest für $k \times m$ Kontingenztafeln

- Liegt statt einer 2×2 eine größere, zweidimensionale $k \times m$ Kontingenztafel vor, so kann man prüfen, ob sich mindestens eine der k Gruppen hinsichtlich ihrer Anteile von den übrigen Gruppen unterscheidet bzw. ob beide Variablen unabhängig sind.

- Unter den gleichen Voraussetzungen wie oben lässt sich dafür wieder die χ^2-Prüfgröße heranziehen:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(h_{ij} - e_{ij})^2}{e_{ij}} \quad \text{mit} \quad df = (k-1) \cdot (m-1)$$

- Besonderheit I: Wird hier die Nullhypothese gleicher Anteile in allen Gruppen (Unabhängigkeit) zurückgewiesen, so wissen wir noch nicht, zwischen welchen Gruppen Unterschiede in welchen Anteilen (Abhängigkeiten) bestehen. Dies muss im Anschluss an diesen Omnibustest ggf. mit Anschlusstests untersucht werden (vgl. z.B. Diehl & Arbinger, 2001, Kap. 19.4).

- Besonderheit II: Als exakter Test kommt ein verallgemeinertes Fisher-Yates Verfahren nach Freeman & Halton zum Einsatz. Bei größerem n (oder anderen ungünstigen Bedingungen) kann es passieren, dass Berechnungen sehr lange dauern oder auch gar nicht mehr durchgeführt werden können („insufficient memory“). In diesem Fall können alternativ Monte Carlo Methoden zum Einsatz kommen.

Signifikanztest für k x m Kontingenztafeln

- Fiktives Beispiel: Insgesamt 95 Studierende aus den Studienfächern Psychologie und Jura (X, $k = 2$, codiert 1=„Psychologie“ und 2=„Jura“) wurden danach gefragt, was ihnen bei der Wahl des künftigen Arbeitsplatzes am wichtigsten ist (Arbeitswerte Y, $m = 3$, mit Optionen 1=„gute Bezahlung“, 2=„nette Kollegen“ und 3=„interessante Arbeit“). Es interessiert die Frage, ob beide Variablen voneinander unabhängig sind oder ein Zusammenhang besteht.

- Es resultiert folgende Prüfgröße (Berechnung vgl. Statistik I):

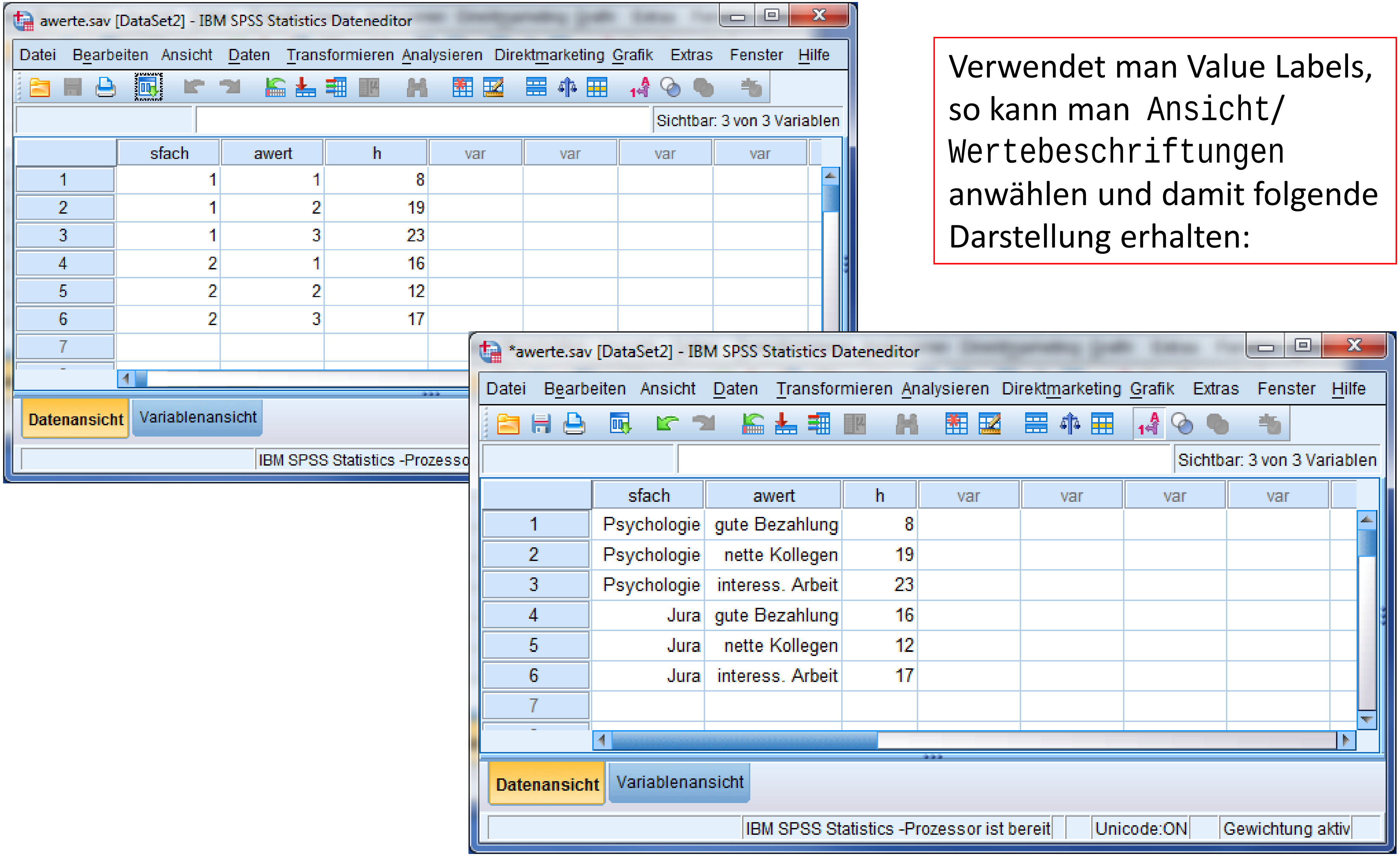
$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(h_{ij} - e_{ij})^2}{e_{ij}} = 4.898$$

h_{ij}	gute Bezahlung	nette Kollegen	interessante Arbeit	
Psychologie	8	19	23	50
Jura	16	12	17	45
	24	31	40	95

Der χ^2-Verteilung entnehmen wir für $\alpha = 0.05$ und $df = (k-1)\cdot(m-1) = (2-1)\cdot(3-1) = 2$ sowie zweiseitiger Testung den kritischen Wert $\chi^2_{crit} = \chi^2_{2,0.95} = 5.992$. (Eine einseitige Testung macht bei Kontingenztafeln größer 2 x 2 keinen Sinn.)

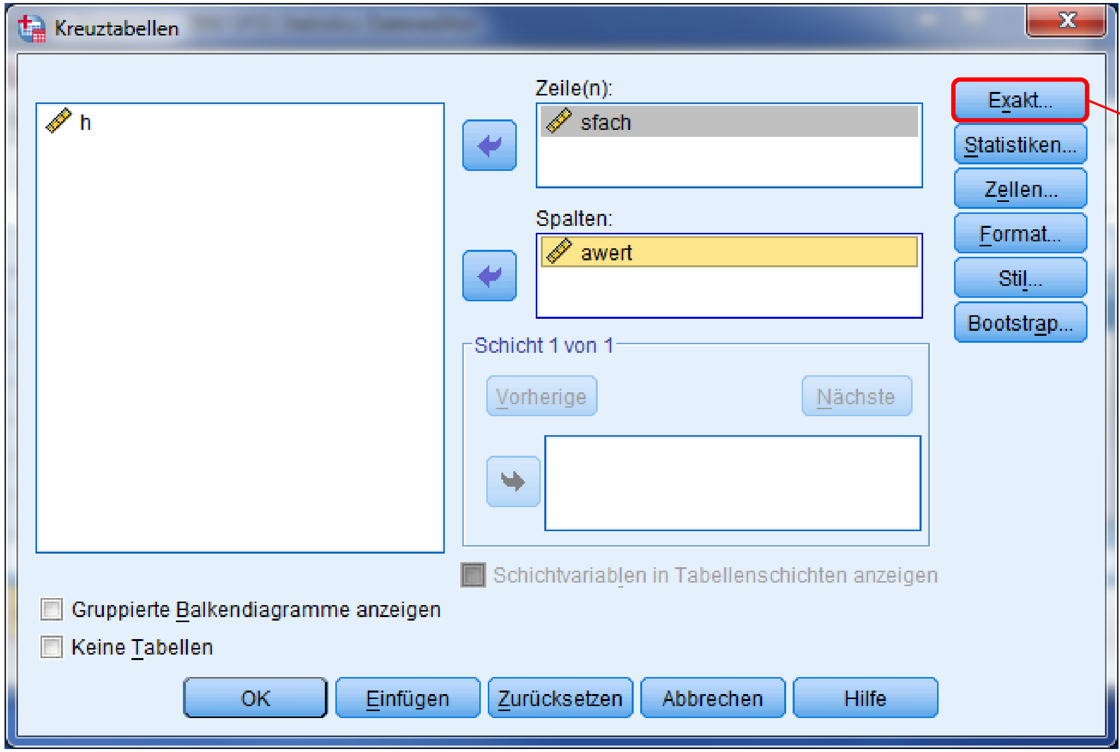
Da $4.898 < 5.992$ behalten wir die Nullhypothese bei und schließen, dass kein statistisch signifikanter Zusammenhang zwischen den beiden Studienfächern und der Präferenzen für Arbeitswerte besteht.

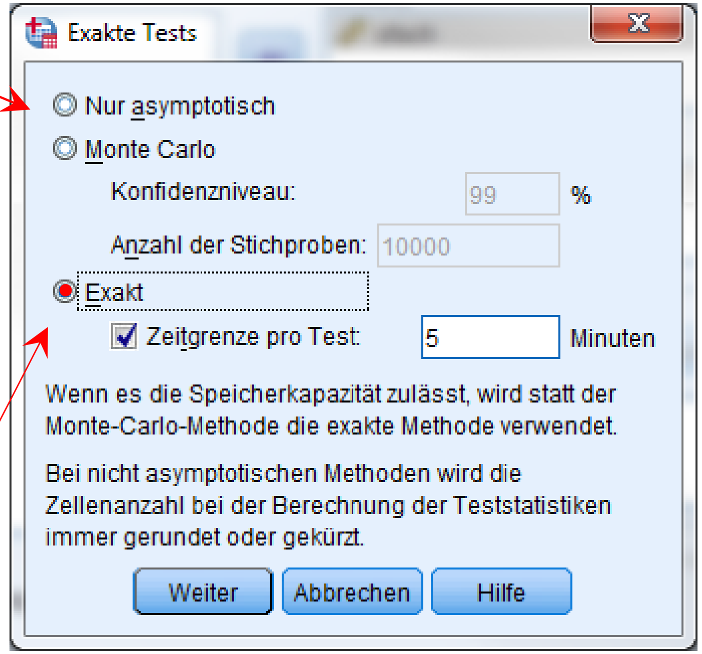
Verwendet man Value Labels, so kann man `Ansicht/ Wertebeschriftungen` anwählen und damit folgende Darstellung erhalten:



Signifikanztest für $k\ x\ m$ Kontingenztafeln in SPSS

- Den χ^2-Test für den allgemeinen Fall einer $k\ x\ m$ Kontingenztafel erhält man völlig analog zum $2\ x\ 2$-Fall durch die Festlegung der Variablen und die Aktivierung der Option „Chi-Quadrat) unter (`Statistiken`).





Der einzige Unterschied besteht darin, dass der exakte Test nicht automatisch ausgegeben wird. Dieser muss unter (`Exakt...`) durch die Anwahl der Option „Exakt" (immer bei gleichzeitig aktivierter Option „Chi-Quadrat") angefordert werden.

Signifikanztest für k x m Kontingenztafeln in SPSS

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)
Chi-Quadrat nach Pearson	4,898[a]	2	,086
Likelihood-Quotient	4,953	2	,084
Zusammenhang linear-mit-linear	2,802	1	,094
Anzahl der gültigen Fälle	95		

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 11,37.

Symmetrische Maße

		Wert	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,227	,086
	Cramer-V	,227	,086
Anzahl der gültigen Fälle		95	

Es besteht kein statistisch signifikanter Zusammenhang zwischen den beiden Studienfächern und der Präferenzen für Arbeitswerte, $V = 0.227$, $\chi^2(df{=}2) = 4.898$, $p = 0.086$.

Signifikanztest für k x m Kontingenztafeln in SPSS

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)	Punkt-Wahrschein lichkeit
Chi-Quadrat nach Pearson	4,898[a]	2	,086	,082		
Likelihood-Quotient	4,953	2	,084	,085		
Exakter Test nach Fisher	4,817			,085		
Zusammenhang linear-mit-linear	2,802[b]	1	,094	,099	,061	,025
Anzahl der gültigen Fälle	95					

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 11,37.

b. Die standardisierte Statistik ist -1,674.

Hier wird zusätzlich der exakte Test angefordert. Es wird das Ergebnis des verallgemeinerten Fisher-Yates Tests (auch als Freeman-Halton Test bezeichnet) ausgegeben, der ebenfalls anzeigt, dass kein statistisch signifikanter Zusammenhang besteht $p = 0.085 > \alpha = 0.05$.

1 Tests für Häufigkeitsunterschiede bei zwei unabhängigen Gruppen: Vierfelder χ^2- und Fisher-Yates Test

Exkurs: Gewichtung von Fällen in SPSS

2 Tests für k x m Kontingenztafeln

3 Weitere Tests für Häufigkeitsunterschiede

4 Simpsons Paradox

Weitere Signifikanztests für nominale Variablen

➢ Inferenzstatistische Tests für nominale (kategoriale) Variablen existieren auch für folgende Konstellationen, die wir nicht näher behandeln (vgl. Bortz, Lienert & Boehnke, 2010):

- Es liegt nur eine nominalskalierte Variable vor (eindimensional) und man will wissen, ob die Häufigkeiten in einer oder mehreren Kategorien von einer Gleichverteilung oder anderen Verteilung abweichen ($m = 2$ ☞ Binomialtest, $m > 2$ ☞ χ^2-Test).

  Beispiel: Erkranken Menschen im Winter häufiger an Depression als im Sommer? Ist der Anteil der Kinder aus Akademikerfamilien in der Psychologie höher als sonst, wo er 60% beträgt?

- Es liegen zwei abhängige Gruppen vor (z.B. eine Stichprobe zu zwei Messzeitpunkten) und es soll untersucht werden, ob Unterschiede in der Häufigkeitsverteilung eines dichotomen Merkmals (☞ McNemar-Test) oder mehrkategoriellen Merkmals (☞ Bowker-Test) bestehen .

  Beispiel: Ist eine Aufklärungskampagne bezüglich des Rauchens wirksam? Eine Stichprobe von 100 Personen wird vor und nach der Kampagne befragt, ob sie rauchen.

- Für den Fall, dass man untersuchen will, ob der Zusammenhang in einer Kontingenz-tafel von einer dritten kategorialen Variablen abhängt (Logik ähnlich der Partial-korrelation; ☞ Mantel-Haenszel Statistik). Beispiele folgen.

Simpsons Paradox

- Studie 1: In den Jahren 1972 bis 1974 wurde in einer englischen Studie in einer Befragung von Frauen auch danach gefragt, ob die Frauen rauchen oder nicht. 20 Jahre später wurde überprüft, welche der Frauen verstorben waren (Appleton, French & Vanderpump, 1996). Es zeigten sich die folgenden Ergebnisse:

		Verstorben			
		ja	nein		
Raucherin	ja	139	443	582	23%
	nein	230	502	732	31%
		369	945	1314	

- Ergebnis: Raucherinnen starben statistisch signifikant seltener, $\chi^2(df=1) = 9.12, p < 0.01$.

- Fiktive Studie 2: Eine Universität prüft, ob sie Frauen bei der Zulassung zu den Studiengängen diskriminiert und kommt zu folgendem Ergebnis:

	Zugelassen			
	ja	nein		
Frauen	320	680	1000	32%
Männer	600	400	1000	60%
	920	1080	2000	

- Ergebnis: Männer wurden statistisch signifikant häufiger zugelassen als Frauen, $\chi^2(df=1) = 157.81, p < 0.01$.

- Studie 1: Schlüsselt man die Ergebnisse getrennt für verschiedene Altersgruppen auf (schichtet also für Alter), so zeigen sich in jeder Gruppe die entgegen gesetzten Ergebnisse:

...

45-54 Jahre		ja	nein		
Raucherin	ja	27	103	130	21%
	nein	12	66	78	18%
		39	169	208	

55-64 Jahre		ja	nein		
Raucherin	ja	51	64	115	44%
	nein	40	81	121	33%
		91	145	236	

		Verstorben			
65 – 74 Jahre		ja	nein		
Raucherin	ja	29	7	36	81%
	nein	101	28	129	78%
		130	35	165	

...

- Der Befund, dass ein Zusammenhang verschwindet oder sich sogar ins Gegenteil verkehrt, wenn man ihn für Subgruppen aufschlüsselt, bezeichnet man auch als Simpsons Paradox.
- Wie ist dieses Ergebnis zu erklären? Je älter die Frauen, desto eher sterben sie. In der Gruppe der älteren Frauen sind aber überproportional Nichtraucher-innen (78% vs. 51% vs. 37%; aufgrund eines Selektionseffektes und auch, weil es in dieser Gruppe noch nicht so populär war zu rauchen), die dann in die Summe eine große Häufigkeit verstorbener Nichtraucherinnen einbringen.

Simpsons Paradox

- **Studie 2**: Schlüsselt man die Ergebnisse getrennt für Studienfächer auf (wir nehmen an, es gäbe nur zwei), so können sich ebenfalls in jeder Gruppe ganz andere Ergebnisse zeigen als in der aggregierten Tabelle.

Zugelassen

Alle	ja	nein		
Frauen	320	680	1000	32%
Männer	600	400	1000	60%
	920	1080	2000	

Psycho-logie	ja	nein		
Frauen	160	640	800	20%
Männer	40	160	200	20%
	200	800	1000	

Maschi-nenbau	ja	nein		
Frauen	160	40	200	80%
Männer	560	240	800	70%
	720	280	1000	

- Wie ist dieses Ergebnis zu erklären? Es hängt hier damit zusammen, dass sich die Frauen in einer großen Zahl in dem Fach bewerben, in dem die Zulassungsquote generell geringer ist. Auf diese Weise geht in die aggreggierte Kontingenztabelle eine große Zahl zurückgewiesener Frauen aus dem Fach Psychologie ein (bzw. eine große Zahl angenommener Männer im Maschinenbau) und verzerrt so die Gesamtstatistik.

Zitierte Quellen

- Appleton, D. R., French, J. M. & Vanderpump, M. P. J. (1996). Ignoring a covariate: An example of Simpson‘s paradox. *The American Statistician*, *50*, 340-341.
- Ruxton, G. D. & Neuhäuser, M. (2010). Good practice in testing for an association in contingency tables. *Behavioral Ecology and Sociobiology*, *64*, 1505-1513.