

ZUSAMMENHANGSMAßE FÜR DIE VIERFELDERTAFEL

Martin-Luther-Universität Halle-Wittenberg

Institut für Soziologie

Übung Einführung in die deskriptive Statistik

Agenda

- Wiederholung
- Vorbemerkungen zu Zusammenhangsmaßen
 - Auswahlkriterien
 - gewünschte Eigenschaften
- Zusammenhangsmaße für die Vierfeldertafel I
 - Prozentsatzdifferenz
 - Odds-Ratio
 - Yules Q

Aufgabe 1: Geschlecht und Rauchen

Im ALLBUS 2014 wurden die Befragten gebeten anzugeben, ob sie Raucher sind oder nicht. Für die Verteilung nach Geschlechtern ergab sich folgende Kreuztabelle:

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	586	412	998
	NEIN	1174	1297	2471
Gesamtsumme		1760	1709	3469

- Berechnen Sie die Spalten-, Zeilen- und Gesamtprozent.
- Wie viel Prozent der Raucher waren männlich?
- Wie viel Prozent waren Raucher unter den Frauen?
- Wie viel Prozent der Befragten mit gültigen Werten bei beiden Variablen waren Männer und zugleich Raucher?

Aufgabe 1a: Lösung I

Spaltenprozente:

			GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
			MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	Anzahl % in GESCHLECHT, BEFRAGTE<R>	586 33,3%	412 24,1%	998 28,8%
	NEIN	Anzahl % in GESCHLECHT, BEFRAGTE<R>	1174 66,7%	1297 75,9%	2471 71,2%
Gesamtsumme		Anzahl % in GESCHLECHT, BEFRAGTE<R>	1760 100,0%	1709 100,0%	3469 100,0%

Quiz: Woran erkennen wir,
dass es sich hier um Spaltenprozente handelt?

Aufgabe 1a: Lösung I

Spaltenprozent:

			GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
			MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	Anzahl % in GESCHLECHT, BEFRAGTE<R>	586 33,3%	412 24,1%	998 28,8%
	NEIN	Anzahl % in GESCHLECHT, BEFRAGTE<R>	1174 66,7%	1297 75,9%	2471 71,2%
Gesamtsumme		Anzahl % in GESCHLECHT, BEFRAGTE<R>	1760 100,0%	1709 100,0%	3469 100,0%

Aufgabe 1a: Lösung II

Zeilenprozente:

			GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
			MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	Anzahl	586	412	998
		% in RAUCHEN SIE?	58,7%	41,3%	100,0%
	NEIN	Anzahl	1174	1297	2471
		% in RAUCHEN SIE?	47,5%	52,5%	100,0%
Gesamtsumme		Anzahl	1760	1709	3469
		% in RAUCHEN SIE?	50,7%	49,3%	100,0%

Quiz: Woran erkennen wir,
dass es sich hier um Zeilenprozente handelt?

Aufgabe 1a: Lösung II

Zeilenprozent:

			GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
			MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	Anzahl	586	412	998
		% in RAUCHEN SIE?	58,7%	41,3%	100,0%
	NEIN	Anzahl	1174	1297	2471
		% in RAUCHEN SIE?	47,5%	52,5%	100,0%
Gesamtsumme		Anzahl	1760	1709	3469
		% in RAUCHEN SIE?	50,7%	49,3%	100,0%

Aufgabe 1a: Lösung II

Gesamtprozente:

			GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
			MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	Anzahl	586	412	998
		% des Gesamtergebnisses	16,9%	11,9%	28,8%
	NEIN	Anzahl	1174	1297	2471
		% des Gesamtergebnisses	33,8%	37,4%	71,2%
Gesamtsumme		Anzahl	1760	1709	3469
		% des Gesamtergebnisses	50,7%	49,3%	100,0%

Quiz: Woran erkennen wir,
dass es sich hier um Gesamtprozente handelt?

Aufgabe 1a: Lösung I

Gesamtprozente:

			GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
			MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	Anzahl	586	412	998
		% des Gesamtergebnisses	16,9%	11,9%	28,8%
	NEIN	Anzahl	1174	1297	2471
		% des Gesamtergebnisses	33,8%	37,4%	71,2%
Gesamtsumme		Anzahl	1760	1709	3469
		% des Gesamtergebnisses	50,7%	49,3%	100,0%

Aufgabe 1b: Lösung II

Zeilenprozent:

			GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
			MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	Anzahl	586	412	998
		% in RAUCHEN SIE?	58,7%	41,3%	100,0%
	NEIN	Anzahl	1174	1297	2471
		% in RAUCHEN SIE?	47,5%	52,5%	100,0%
Gesamtsumme		Anzahl	1760	1709	3469
		% in RAUCHEN SIE?	50,7%	49,3%	100,0%

58,7 Prozent der Raucher waren männlich (586 von 998 Rauchern)

Aufgabe 1c: Lösung I

Spaltenprozente:

			GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
			MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	Anzahl	586	412	998
		% in GESCHLECHT, BEFRAGTE<R>	33,3%	24,1%	28,8%
	NEIN	Anzahl	1174	1297	2471
		% in GESCHLECHT, BEFRAGTE<R>	66,7%	75,9%	71,2%
Gesamtsumme		Anzahl	1760	1709	3469
		% in GESCHLECHT, BEFRAGTE<R>	100,0%	100,0%	100,0%

24,1 Prozent der Frauen waren Raucher (412 von 1709 Frauen)

Aufgabe 1d: Lösung I

Gesamtprozente:

			GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
			MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	Anzahl	586	412	998
		% des Gesamtergebnisses	16,9%	11,9%	28,8%
	NEIN	Anzahl	1174	1297	2471
		% des Gesamtergebnisses	33,8%	37,4%	71,2%
Gesamtsumme		Anzahl	1760	1709	3469
		% des Gesamtergebnisses	50,7%	49,3%	100,0%

16,9 Prozent der Befragten mit gültigen Werten waren Raucher und Männer. (586 von insgesamt 3469 Befragten mit gültigen Werten)

Zusammenhangsmaße

- Versuch die Stärke und die ab ordinalem Skalenniveau auch die Richtung eines Zusammenhangs mithilfe einer einzigen statistischen Kennzahl erfassen
- Auswahlkriterien zur Bestimmung:
 - Vierfeldertafel oder Mehrfeldertafel
 - Skalenniveau der betrachteten Variablen
 - symmetrische oder asymmetrische Beziehung
 - ggfs. gewünschte sonstige Eigenschaften des Zusammenhangsmaßes

Vierfeldertafel vs. Mehrfeldertafel

- Vierfeldertafel:

		GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	586	412	998
	NEIN	1174	1297	2471
Gesamtsumme		1760	1709	3469

- Solche Tabellen entstehen immer dann, dann wenn die gemeinsamen Verteilungen von zwei dichotomen Variablen betrachtet werden. Namensgebend sind die vier Zellen der gemeinsamen Verteilungen.
- Als Spezialfall der Mehrfeldertafel besteht hier die Möglichkeit, die Berechnung bestimmter Zusammenhangsmaße einfacher vorzunehmen, darüber hinaus existieren Maßzahlen, die sich nur für die Vierfeldertafel sinnvoll bestimmen lassen.

Vierfeldertafel: spezielle Konvention

Kreuztabelle mit 2 Zeilen und 2 Spalten		Spaltenvariable X		Zeilensummen
		$X = 1$	$X = 2$	
Zeilenvariable Y	$Y = 1$	a	b	$a + b$
	$Y = 2$	c	d	$c + d$
Spaltensummen		$a + c$	$b + d$	$n =$ $a + b + c + d$

Vierfeldertafel vs. Mehrfeldertafel II

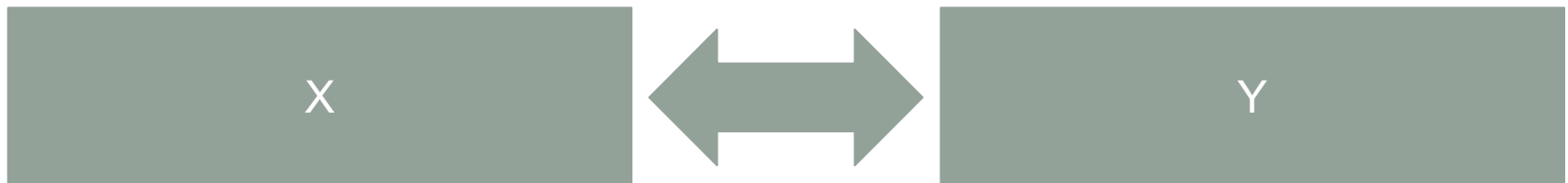
- Mehrfeldertafel:

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
BEFRAGTE<R> BERUFSTAETIG?	HAUPTBERUFL. GANZTAGS	1026	545	1571
	HAUPTBERUFL. HALBTAGS	41	309	350
	NEBENHER BERUFSTAE.	73	135	208
	NICHT ERWERBSTAETIG	619	720	1339
Gesamtsumme		1759	1709	3468

- Wenn man Variablen kreuztabelliert, bei denen mindestens eine der betrachteten Variablen polytom ist, entstehen Mehrfeldertafeln. Diese zeichnen sich dadurch aus, dass es mehr als vier Zellen bei den gemeinsamen Verteilungen gibt.
- Eine Anwendung von Maßzahlen für die Vierfeldertafel ist nicht mehr zulässig. Jedoch lassen sich alle Zusammenhangsmaße für Mehrfeldertafeln auch auf Vierfeldertafeln anwenden.

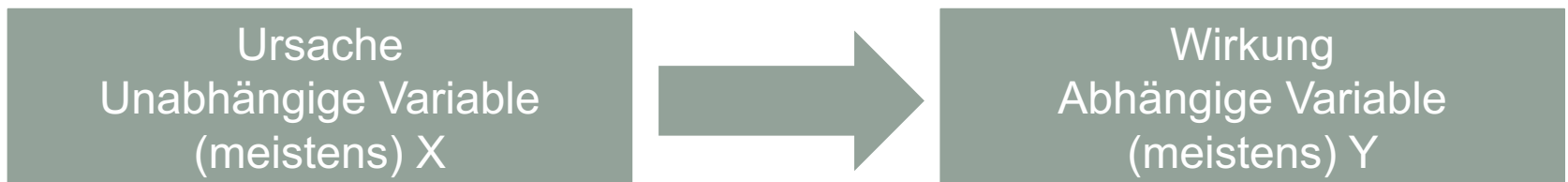
Symmetrische vs asymmetrische Beziehungen

- symmetrische Beziehung:



Schlüsselworte: hängen zusammen, gehen miteinander einher... (keine Richtung)

- asymmetrische Beziehung:



Schlüsselworte: hängt ab von..., beeinflusst..., übt Einfluss aus..., bewirkt...

Sonstiges: Gütekriterien nach Galtung I

- Statistische Kennwerte sollten allgemein:
 - so viele Informationen wie möglich berücksichtigen,
 - robust gegen kleinere Veränderungen sein,
 - einfach berechenbar sein und Vergleiche ermöglichen,
 - unabhängig von der Fallzahl sein,
 - unabhängig von der Anzahl der Ausprägungen der Variablen sein,
 - eine klare (inhaltliche) Interpretation aufweisen und
 - invariant gegenüber zulässigen Transformationen der Variablenwerte sein.

Galtung (1970: 208ff.)

Sonstiges: Gütekriterien nach Galtung II

- Zusammenhangsmaße im Speziellen sollten:
 - (nur) bei bestehender Unabhängigkeit den Wert null annehmen,
 - (nur) ihr Maximum erreichen, wenn ein perfekter Zusammenhang besteht,
 - ab ordinalem Skalenniveau die Richtung des Zusammenhangs angeben,
 - einen normierten Wertebereich aufweisen,
 - asymmetrisch sein, wenn dies notwendig ist, und
 - empfindlich für den Unterschied zwischen linearer und kurvilinearere Abhängigkeit sein
- (Galtung 1970: 211 ff.).

Zusammenhangsmaße für die Vierfeldertafel

asymmetrische Zusammenhangsmaße

- Vierfeldertafel:
 - Prozentsatzdifferenz d_{YX} %
 - Prozentsatzdifferenz d_{XY} %
 - Lambda λ_{YX}
 - Lambda λ_{XY}

symmetrische Zusammenhangsmaße

- Vierfeldertafel
 - Odds-Ratio
 - Yules Q
 - Phikoeffizient ϕ
 - Cramers V

Prozentsatzdifferenz

Prozentsatzdifferenz $d_{YX}\%$

- Verwendung:
 - mithilfe der X-Variablen (Spaltenvariablen) soll der Unterschied in der Y-Variablen (Zeilenvariablen) erklärt werden
- Formel:
 - $d_{YX}\% = 100 \left(\frac{a}{a+c} - \frac{b}{b+d} \right)$

Prozentsatzdifferenz $d_{XY}\%$

- Verwendung:
 - mithilfe der Y-Variablen (Zeilenvariablen) soll der Unterschied in der X-Variablen (Spaltenvariablen) erklärt werden
- Formel:
 - $d_{XY}\% = 100 \left(\frac{a}{a+b} - \frac{c}{c+d} \right)$

Prozentsatzdifferenz $d_{YX}\%$: Interpretation

- Interpretation:
 - Wert zwischen 0 (kein Einfluss) und ± 100 gebunden (maximaler Einfluss)
 - Unterschied in den Spaltenprozenten in Prozentpunkten

		X: Zufriedenheit Wirtschaftslage		Summe
		1 unzufrieden	2 zufrieden	
Y: Demos- trations- teilnahme	1 nein	40 %	40 %	40 %
	2 ja	60 %	60 %	40 %
Summe		100 %	100 %	100%

- $d_{YX}\% = 40 - 40 = 0$
- kein Einfluss der Zufriedenheit (X) auf Demonstrationsteilnahme (Y)

		X: Zufriedenheit Wirtschaftslage		Summe
		1 unzufrieden	2 zufrieden	
Y: Demos- trations- teilnahme	1 nein	0 %	100 %	50 %
	2 ja	100 %	0 %	50 %
Summe		100 %	100 %	100%

- $d_{YX}\% = 0 - 100 = -100$
- maximaler Einfluss der Zufriedenheit (X) auf Demonstrationsteilnahme (Y)

Prozentsatzdifferenz $d_{XY}\%$: Interpretation

- Interpretation:
 - Wert zwischen 0 (kein Einfluss) und ± 100 gebunden (maximaler Einfluss)
 - Unterschied in den Zeilenprozenten in Prozentpunkten

		X: Zufriedenheit Wirtschaftslage		Summe
		1 unzufrieden	2 zufrieden	
Y: Demos- trations- teilnahme	1 nein	40%	60%	100%
	2 ja	40%	60%	100%
Summe		40%	60%	100%

		X: Zufriedenheit Wirtschaftslage		Summe
		1 unzufrieden	2 zufrieden	
Y: Demos- trations- teilnahme	1 nein	0%	100%	100%
	2 ja	100%	0%	100%
Summe		50%	50%	100%

- $d_{YX}\% = 40 - 40 = 0$
- kein Einfluss der Demonstrationsteilnahme (Y) auf Zufriedenheit (X)

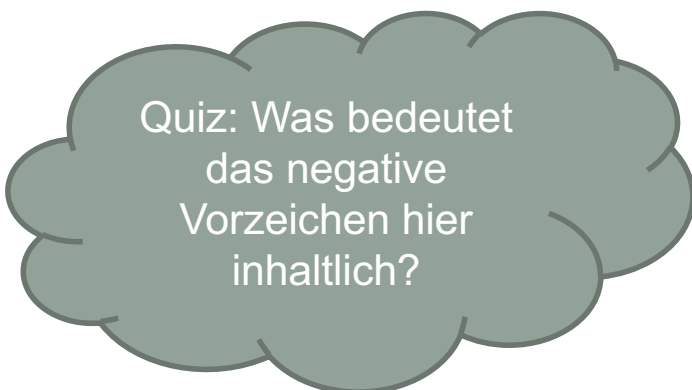
- $d_{YX}\% = 0 - 100 = -100$
- maximaler Einfluss der Demonstrationsteilnahme (Y) auf Zufriedenheit (X)

Prozentsatzdifferenz: Interpretation Vorzeichen

- Vorzeichen gibt „Richtung“ an
 - + deutet auf Dominanz der a-d-Diagonale hin
 - - deutet auf Dominanz der b-c-Diagonale hin

		X: Zufriedenheit Wirtschaftslage		Summe
		1 unzufrieden	2 zufrieden	
Y: Demos- trations- teilnahme	1 nein	0%	100%	10%
	2 ja	100%	0%	100%
Summe		50%	50%	100%

- $d_{YX}\% = 0 - 100 = -100$



Quiz: Was bedeutet das negative Vorzeichen hier inhaltlich?

Aufgabe 2: Rauchen II

Im ALLBUS 2014 wurden die Befragten gebeten anzugeben, ob Sie Raucher oder nicht. Für Männer und Frauen ergab sich hierbei folgende Kreuztabelle:

		GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	586	412	998
	NEIN	1174	1297	2471
Gesamtsumme		1760	1709	3469

Berechnen Sie die Prozentsatzdifferenz, wenn sie sich für den Einfluss des Geschlechts auf das Rauchverhalten interessieren

Aufgabe 2: Lösung

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	586	412	998
	NEIN	1174	1297	2471
Gesamtsumme		1760	1709	3469

- Berechnung:

- $d_{YX}\% = 100 * \left(\frac{a}{a+c} - \frac{b}{b+d} \right)$

- $d_{YX}\% = 100 * \left(\frac{586}{1760} - \frac{412}{1709} \right)$

- $d_{YX}\% = 9,2$

- Interpretation:

- Es besteht ein Einfluss des Geschlechts auf das Rauchverhalten.
 - Unter den Männern ist der Anteil der Raucher um 9,2 Prozentpunkte höher als unter den Frauen.

Aufgabe 2: Lösung II

		GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	586	412	998
	NEIN	1174	1297	2471
Gesamtsumme		1760	1709	3469

Spaltenprozenzte

		GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	33,3%	24,1%	28,8%
	NEIN	66,7%	75,9%	71,2%
Gesamtsumme		100,0%	100,0%	100,0%

$$d_{YX}\% = 33,3\% - 24,1\% = 9,2 \text{ Prozentpunkte}$$

Prozentsatzdifferenz: Eigenschaften

- **Vorteil:**
 - einfache Berechnung
 - einfache und klare Interpretation
- **Nachteil:**
 - anfällig für schiefe Verteilungen der abhängigen Variablen
 - bei Gleichverteilung der unabhängigen Variablen, aber schiefer Verteilung der abhängigen Variablen kann die Prozentsatzdifferenz nur kleine Werte annehmen (Problem additive Modellierung)

Odds-Ratio

- basiert auf dem Konzept der Odds
- Anwendbarkeit:
 - bei zwei dichotomen Variablen (Vierfeldertafeln)
 - vermuteter symmetrischer Zusammenhang
- Relevanz:
 - beliebt in der medizinischen Forschung
 - steigende Relevanz in Forschungsliteratur
 - logistische Regression

Odds I (Chance)

- stammt ursprünglich aus dem Bereich der Sportwetten
- Angabe als Wahrscheinlichkeit für das Auftreten eines Ereignisses im Vergleich zum Nichtauftreten
- „Gewinnchance“
- gibt an, wie das Verhältnis dafür ist, dass ein Ereignis eintritt, im Verhältnis dazu, dass das Ereignis nicht eintritt
- Allgemeine Berechnung:

$$\bullet \textit{ Odd} = \frac{\textit{Wahrscheinlichkeit,dass Ereignis eintritt}}{\textit{Wahrscheinlichkeit,dass Ereignis nicht eintritt}} = \frac{p_1}{1-p_1}$$

Quiz

In einer Stichprobe von 10 Personen befinden sich 2 Raucher. Wir ziehen zufällig eine Person.



Wie hoch ist die Wahrscheinlichkeit, dass wir einen Raucher erwischen?

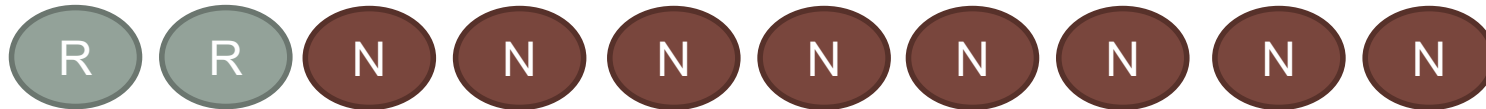
$$\text{Antwort: } p = \frac{2}{10} = 0,2 \rightarrow 20\%$$

Wie hoch ist die Wahrscheinlichkeit, dass wir einen Raucher erwischen?

$$\text{Antwort: } 1 - p = \frac{8}{10} = 0,8 \rightarrow 80\%$$

Quiz

In einer Stichprobe von 10 Personen befinden sich 2 Raucher.
Wir ziehen zufällig eine Person.



Wie hoch ist die Chance dafür, dass wir einen Raucher erwischen?

$$Odd = \frac{p_{Raucher}}{1 - p_{Raucher}}$$

Antwort:

$$Odd = \frac{\frac{2}{10}}{\frac{8}{10}} = \frac{2}{8} = \frac{1}{4} = 0,25$$



Die Chance einen Raucher zu erwischen liegt bei 0,25 gegenüber keinen Raucher zu erwischen. Es gibt 0,25mal so viele Raucher wie Nichtraucher.

Odds II

- Übertragung des Konzepts auf Vierfeldertafel
- Berechnung konditioneller Odds:
 - $Odds_{Y=1|X=1} = \frac{a}{c}$ (unter der Bedingung das x1 vorliegt)
 - $Odds_{Y=1|X=2} = \frac{b}{d}$ (unter der Bedingung das x2 vorliegt)
- Wertebereich:
 - $[0; \infty]$
 - 1: gleiche Häufigkeit von Zähler und Nenner
 - <1 : Häufigkeit im Zähler kleiner als im Nenner
 - >1 : Häufigkeiten im Zähler größer als im Nenner

Odds III

- Interpretation:
 - Faktor um den die im Zähler aufgeführte Ausprägung einer Variablen häufiger oder seltener vorkommt als die Ausprägung der Variablen im Nenner
 - Prozentwert: wenn man vom Ergebnis 1 abzieht und anschließend mit 100 multipliziert...

Odds IV: Beispiel

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	586	412	998
	NEIN	1174	1297	2471
Gesamtsumme		1760	1709	3469

- Berechnung:

- $Odds_{ja|männlich} = \frac{a}{c}$ (Bedingung: männlich)

- $Odds_{ja|männlich} = \frac{586}{1174}$

- $Odds_{ja|männlich} = 0,499$

- Interpretation:

- Unter den Männer gibt es 0,499 mal so viele Raucher wie Nichtraucher, also fast halb so viele Raucher wie Nichtraucher.

- Prozentwert: $(0,499 - 1) * 100 = -50,1$. Unter den Männern gibt folglich 50,1% weniger Raucher als Nicht-Raucher.

Odds V: Beispiel

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	586	412	998
	NEIN	1174	1297	2471
Gesamtsumme		1760	1709	3469

- Berechnung:

- $Odds_{ja|weiblich} = \frac{b}{d}$ (Bedingung: weiblich)

- $Odds_{ja|weiblich} = \frac{412}{1297}$

- $Odds_{ja|weiblich} = 0,318$

- Interpretation:

- Unter den Frauen gibt es 0,318 mal so viele Raucher wie Nichtraucher.

- Prozentwert: $(0,318 - 1) * 100 = -68,2$.

- Unter den Frauen gibt folglich 68,2% weniger Raucher als Nicht-Raucher.

Aufgabe 3: öffentlicher Dienst

Eine Arbeitsmarktforscherin interessiert sich dafür, ob hinsichtlich der Arbeit im öffentlichen Dienst Unterschiede zwischen Männern und Frauen bestehen.

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
IM OEFFENTLICHEN DIENST TAETIG?	JA	190	248	438
	NEIN	732	530	1262
Gesamtsumme		922	778	1700

- Wie hoch ist die Chance dafür, dass ein zufällig ausgewählter Mann im öffentlichen Dienst arbeitet im Vergleich dazu, dass er es nicht macht? Drücken Sie das Ergebnis auch in Prozent aus!
- Wie hoch ist die Chance dafür, dass eine zufällig ausgewählte Frau im öffentlichen Dienst arbeitet im Vergleich dazu, dass sie es nicht macht? Drücken Sie das Ergebnis auch in Prozent aus!

Aufgabe 3a: Lösung

		GESCHLECHT, BEFRAGTE <R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
IM OEFFENTLICHEN DIENST TAETIG?	JA	190	248	438
	NEIN	732	530	1262
Gesamtsumme		922	778	1700

- Berechnung:

- $Odds_{Y=1|X=1} = \frac{a}{c}$

- $Odds_{ja|Männer} = \frac{190}{732}$

- $Odds_{ja|Männer} = 0,2596$

- Interpretation:

- Es gibt also 0,259 mal so viele Männer die im öffentlichen Dienst arbeiten als solche, die außerhalb des öffentlichen Dienstes arbeiten.

- $(0,2596-1)*100=-74,0$.

- Es gibt also 74% weniger Männer im öffentlichen Dienst als außerhalb des öffentlichen Dienstes.

Aufgabe 3b: Lösung

		GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
		MAENNLICH	WEIBLICH	
IM OEFFENTLICHEN DIENST TAETIG?	JA	190	248	438
	NEIN	732	530	1262
Gesamtsumme		922	778	1700

- Berechnung:

- $Odds_{Y=1|X=2} = \frac{b}{d}$

- $Odds_{ja|Frauen} = \frac{248}{530}$

- $Odds_{ja|Frauen} = 0,468$

- Interpretation:

- Es gibt 0,468 mal so viele Frauen, die im öffentlichen Dienst arbeiten, als außerhalb.

- $(0,468-1)*100=-53,2$. Es gibt also 53,2% weniger Frauen, die im öffentlichen Dienst tätig sind als außerhalb.

Odds-Ratio (Chancenverhältnis)

- Maß für die Stärke des multiplikativen Zusammenhangs
- Quotient der beiden konditionalen Odds
- Berechnung:

$$\bullet \quad OR_{y|x_1, x_2} = \alpha = \frac{Odds_{y_1|x_1}}{Odds_{y_1|x_2}} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a*d}{b*c}$$

- Interpretation:
 - Wertebereich: $[0; \infty]$
 - 1: beide Odds gleich groß, kein Zusammenhang
 - <1 : Größenverhältnis unter Bedingung x_1 kleiner als unter Bedingung x_2
 - >1 : Größenverhältnis unter Bedingung x_1 größer als unter Bedingung x_2 (Zuwachs)

Odds-Ratio II

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	586	412	998
	NEIN	1174	1297	2471
Gesamtsumme		1760	1709	3469

- Berechnung:

- $$OR_{ja|Männlich,weiblich} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a*d}{b*c}$$

- $$OR_{ja|Männlich,weiblich} = \frac{\frac{586}{1174}}{\frac{412}{1297}} = \frac{586*1297}{412*1174}$$

- $$OR_{ja|Männlich,weiblich} = 1,571$$

- Interpretation:

- Zusammenhang zwischen Geschlecht und Rauchverhalten liegt vor, da der Wert von 1 abweicht.
- Die Odds unter den Männern sind 1,571 mal so hoch wie unter den Frauen.
- Die Chance unter Männern einen Raucher zu finden ist 1,571 mal so hoch wie unter Frauen.

Aufgabe 4: öffentlicher Dienst II

Eine Arbeitsmarktforscherin interessiert sich dafür, ob hinsichtlich der Arbeit im öffentlichen Dienst Unterschiede zwischen Männern und Frauen bestehen.

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
IM OEFFENTLICHEN DIENST TAETIG?	JA	190	248	438
	NEIN	732	530	1262
Gesamtsumme		922	778	1700

Berechnen Sie das entsprechende Odds-Ratio und interpretieren Sie Ihr Ergebnis!

Aufgabe 4: Lösung

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
IM OEFFENTLICHEN DIENST TAETIG?	JA	190	248	438
	NEIN	732	530	1262
Gesamtsumme		922	778	1700

- Berechnung:

- $OR_{ja|Männlich,weiblich} = \frac{a*d}{b*c}$
- $OR_{ja|Männlich,weiblich} = \frac{190*530}{248*732}$
- $OR_{ja|Männlich,weiblich} = 0,5547$

- Interpretation:

- Der Odds-Ratio-Wert von 0,5547 besagt, dass die Chance für Männer im öffentlichen Dienst beschäftigt zu sein 0,5547 mal so groß war wie für die Frauen.

Odds-Ratio II

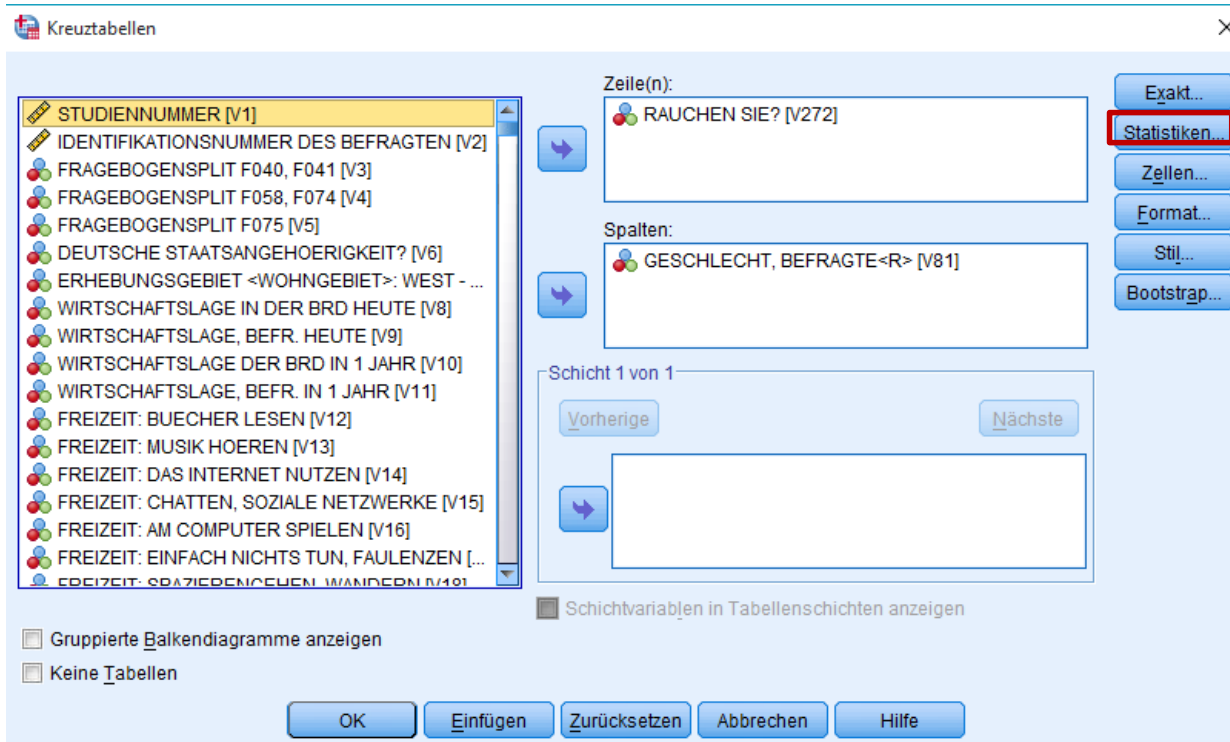
- Vorteil:
 - Unempfindlichkeit gegen Veränderungen der Randverteilungen
- Nachteil:
 - wenig intuitive Interpretation
 - Wertebereich zwischen 0 und unendlich



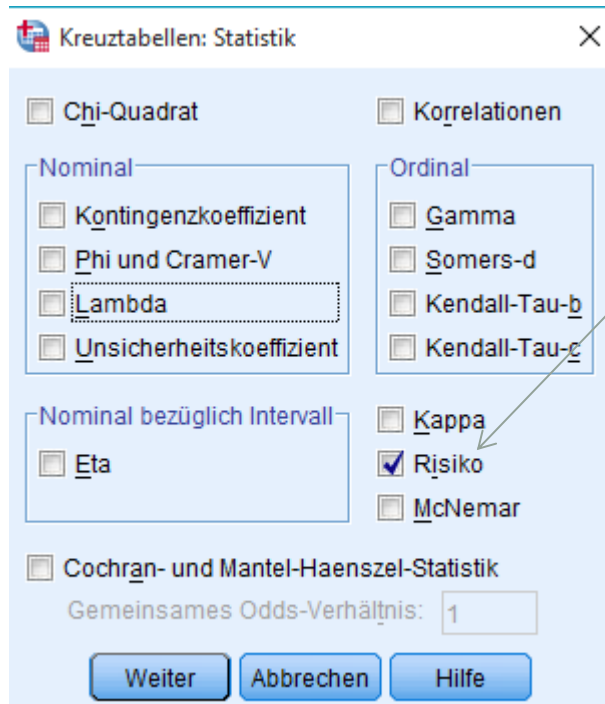
suboptimal für ein
Zusammenhangsmaß

SPSS: Odds-Ratio bestimmen

- Analysieren → Deskriptive Statistiken... → Kreuztabellen.. → Statistiken...



SPSS: Odds-Ratio bestimmen II



„Risiko“ auswählen

SPSS: Odds-Ratio bestimmen III

Risikoschätzung

	Wert	95%-Konfidenzintervall	
		Unterer	Oberer
Odds-Verhältnis für RAUCHEN SIE? (JA / NEIN)	1,571	1,354	1,823
Für Kohort GESCHLECHT, BEFRAGTE<R> = MAENNLICH	1,236	1,156	1,321
Für Kohort GESCHLECHT, BEFRAGTE<R> = WEIBLICH	,787	,724	,855
Anzahl der gültigen Fälle	3469		

Odds-Ratio

Yules Q

- symmetrisches Maß auf Basis der Odds-Ratios
- Vorteil:
 - Normierung auf einen Wertebereich von $[-1;1]$
- Berechnung:
 - $$Q = \frac{OR-1}{OR+1} = \frac{a*d-b*c}{a*d+b*c}$$
- Interpretation:
 - 0: kein Zusammenhang
 - -1/+1: perfekter Zusammenhang
 - Vorzeichen analog zur Prozentsatzdifferenz interpretiert
 - Betrag des Zusammenhangsmaßes per Daumenregel interpretierbar

Interpretation Stärke „normaler“ Zusammenhangsmaße

- Daumenregel nach Kühnel / Krebs (2012: 332):

Wert	Stärke
0	kein Zusammenhang
$0 < ZM < 0,05$	vernachlässigbarer Zusammenhang
$0,05 < ZM < 0,25$	geringer Zusammenhang
$0,25 < ZM < 0,50$	mittlerer Zusammenhang
$0,50 < ZM < 1$	starker Zusammenhang
1	perfekter Zusammenhang

Yules Q: Beispiel

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	586	412	998
	NEIN	1174	1297	2471
Gesamtsumme		1760	1709	3469

- Berechnung:

- $Q = \frac{a*d - b*c}{a*d + b*c}$

- $Q = \frac{586*1297 - 412*1174}{586*1297 + 412*1174}$

- $Q = 0,22$

- Interpretation:

- Es besteht ein schwacher Zusammenhang zwischen dem Geschlecht und dem Rauchverhalten. Das positive Vorzeichen deutet darauf hin, dass die A-D-Diagonale überwiegt. Dies bedeutet, dass Männer tendenziell eher Raucher sind als Frauen und umgekehrt.

Yules Q: Alternative Berechnung

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
RAUCHEN SIE?	JA	586	412	998
	NEIN	1174	1297	2471
Gesamtsumme		1760	1709	3469

- Berechnung:

- $Q = \frac{OR-1}{OR+1}$

- $Q = \frac{1,571-1}{1,571+1}$

- $Q = 0,22$

Problem: Yules Q erreicht in der Praxis relativ hohe Werte, schwer die Stärke des Zusammenhangs abzuschätzen.

Aufgabe 5: öffentlicher Dienst III

Ein Sozialforscherin vermutet, dass es einen Zusammenhang zwischen dem Geschlecht und der Arbeit im öffentlichen Dienst gibt. Sie möchte ein symmetrisches Zusammenhangsmaß bestimmen.

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
IM OEFFENTLICHEN DIENST TAETIG?	JA	190	248	438
	NEIN	732	530	1262
Gesamtsumme		922	778	1700

Berechnen Sie ein Ihnen bekanntes Zusammenhangsmaß und interpretieren Sie es!

Aufgabe 5: Lösung

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNlich	WEIBlich	
IM OEFFENTLICHEN DIENST TAETIG?	JA	190	248	438
	NEIN	732	530	1262
Gesamtsumme		922	778	1700

- Berechnung:

- $Q = \frac{a*d - b*c}{a*d + b*c}$

- $Q = \frac{190*530 - 248*732}{190*530 + 248*732} = -0,288$

- Interpretation:

- Es besteht ein mittlerer Zusammenhang zwischen dem Geschlecht und der Tätigkeit im öffentlichen Dienst.
 - Das negative Vorzeichen deutet daraufhin, dass die B-C-Diagonale dominiert. Männer sind tendenziell seltener im öffentlichen Dienst beschäftigt als Frauen und umgekehrt.

Literaturhinweise

- Kerstin Völkl / Christoph Korb (2018): Deskriptive Statistik. Eine Einführung für Politikwissenschaftlerinnen und Politikwissenschaftler. S. 149-160.
- Steffen-M. Kühnel / Dagmar Krebs (2012): Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. S. 317-343.
- Hans Benninghaus: Deskriptive Statistik. Eine Einführung für Sozialwissenschaftler. S. 98-136.