

BEZIEHUNGEN ZWISCHEN NOMINALLEN VARIABLEN

Martin-Luther-Universität Halle-Wittenberg

Institut für Soziologie

Übung Einführung in die deskriptive Statistik

Was machen wir heute?

- Wiederholung
- Konzept der stochastischen Unabhängigkeit und entsprechende symmetrische nominale Zusammenhangsmaße
- Konzept der proportionalen Fehlerreduktion und entsprechende asymmetrische nominale Zusammenhangsmaße

Aufgabe 1: Gewerkschaftsmitgliedschaft

Eine Arbeitsmarktforscherin interessiert sich dafür, ob es einen Einfluss der Beschäftigung im öffentlichen Dienst auf die Gewerkschaftsmitgliedschaft gibt. Mithilfe von SPSS erhält sie folgende Kreuztabelle:

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	115	170	285
	NEIN	321	1091	1412
Gesamtsumme		436	1261	1697

- Welche Ihnen bekannte Maßzahl ist zur Untersuchung des Zusammenhangs geeignet? Begründen Sie kurz Ihre Wahl.
- Berechnen Sie das Zusammenhangsmaß.
- Interpretieren Sie Ihr Ergebnis statistisch und inhaltlich.

Aufgabe 1a: Lösung

asymmetrische
Fragestellung:
Beschäftigung im
öffentlichen Dienst (X) →
Gewerkschafts-
mitgliedschaft (Y)

Eine Arbeitsmarktforscherin interessiert sich dafür, ob es einen Einfluss der Beschäftigung im öffentlichen Dienst auf die Gewerkschaftsmitgliedschaft gibt.

Vierfeldertafel

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	115	170	285
	NEIN	321	1091	1412
Gesamtsumme		436	1261	1697

geeignetes Maß:
Prozentsatzdifferenz
 $d_{YX}\%$

Aufgabe 1b: Lösung

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	115	170	285
	NEIN	321	1091	1412
Gesamtsumme		436	1261	1697

- Berechnung:

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

- $d_{YX}\% = 100 * \left(\frac{115}{436} - \frac{170}{1261} \right) = 12,9$

Aufgabe 1c: Lösung

- Interpretation:
 - Die Prozentsatzdifferenz hinsichtlich der Gewerkschaftsmitgliedschaft zwischen denen, die im öffentlichen Dienst tätig sind, und denen, die nicht im öffentlichen Dienst tätig sind, liegt bei 12,9 Prozentpunkten.
 - Das positive Vorzeichen deutet darauf hin, dass die a-d-Diagnale überwiegt. Folglich sind Personen im öffentlichen Dienst tendenziell eher Gewerkschaftsmitglied als Personen, die nicht im öffentlichen Dienst beschäftigt sind.

Aufgabe 1c: Lösung II

% in IM OEFFENTLICHEN DIENST TAETIG?

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsumme
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	26,4%	13,5%	16,8%
	NEIN	73,6%	86,5%	83,2%
Gesamtsumme		100,0%	100,0%	100,0%

$$d_{YX}\% = 26,4\% - 13,5\% = 12,9 \text{ Prozentpunkte}$$

Quiz

% in IM OEFFENTLICHEN DIENST TAETIG?

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	26,4%	13,5%	16,8%
	NEIN	73,6%	86,5%	83,2%
Gesamtsumme		100,0%	100,0%	100,0%

Wie müsste die Tabelle aussehen, wenn die Tätigkeit im öffentlichen Dienst keinen Einfluss auf die Gewerkschaftsmitgliedschaft hätte?

Die Prozentsatzdifferenz $d_{YX}\% = 0$ sein. Folglich müssten die bedingten Verteilungen für die Gewerkschaftsmitgliedschaft ja und nein mit unbedingten Verteilung übereinstimmen.

Stochastischen Unabhängigkeit: Idee der Indifferenztabelle

% in IM OEFFENTLICHEN DIENST TAETIG?

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsumme
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	16,8%	16,8%	16,8%
	NEIN	83,2%	83,2%	83,2%
Gesamtsumme		100,0%	100,0%	100,0%

bedingte Spaltenprozentage stimmen überein mit der unbedingten Verteilung

Stochastische Unabhängigkeit: Idee der Indifferenztabelle II

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA			285
	NEIN			1412
Gesamtsumme		436	1261	1697

Zeilen-
summen
 $n_{i,+}$

Spalten-
summen
 $n_{+,j}$

Gesamt-
summe n

- Übertragung der Randverteilung:

- $e_{i,j} = n_{+,j} * \frac{n_{i,+}}{n} = \frac{n_{i,+} * n_{+,j}}{n}$ für jede Zelle

- Beispiel:

- $e_{1,1} = 436 * \frac{285}{1697} = 436 * 0,168 \dots = 73,2$

Aufgabe 2.1: Indifferenztabelle

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA			285
	NEIN			1412
Gesamtsumme		436	1261	1697

Berechnen Sie die fehlenden Zellen der Indifferenztabelle mit Hilfe der

$$\text{Formel } e_{i,j} = \frac{n_{i,+} * n_{+,j}}{n}$$

Aufgabe 2.1: Lösung

Erwartete Anzahl

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	73,2	211,8	285,0
	NEIN	362,8	1049,2	1412,0
Gesamtsumme		436,0	1261,0	1697,0

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	16,8%	16,8%	16,8%
	NEIN	83,2%	83,2%	83,2%
Gesamtsumme		100,0%	100,0%	100,0%

in Spalten-
prozenten

Stochastische Unabhängigkeit: Idee χ^2

- Idee:

- Inwieweit weichen unsere beobachteten absoluten Häufigkeiten (Kontingenztafel) von den theoretisch erwarteten Häufigkeiten bei stochastischer Unabhängigkeit (Indifferenztafel) ab?

- Berechnung:

- $$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{i,j} - e_{i,j})^2}{e_{i,j}}$$

- „Bilde dann für jede Zelle die Abweichung der beobachteten von den erwarteten Werten und quadriere dies, dann teile den Wert durch die erwarteten Fälle in der jeweiligen Zelle. Summiere das Ergebnis auf und du erhältst den Wert für χ^2 “

Quiz: Wann erhalten wir einen Chi-Quadratwert von 0?
Was darf nicht passieren?

Stochastische Unabhängigkeit: Kontingenztabelle und Indifferenztabelle

- Kontingenztabelle:
 - tatsächlich beobachtete absolute Häufigkeiten $n_{i,j}$:

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	115	170	285
	NEIN	321	1091	1412
Gesamtsumme		436	1261	1697

- Indifferenztabelle:
 - erwartete absolute Häufigkeiten $e_{i,j}$:

Erwartete Anzahl

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	73,2	211,8	285,0
	NEIN	362,8	1049,2	1412,0
Gesamtsumme		436,0	1261,0	1697,0

Stochastische Unabhängigkeit: Idee χ^2 II

Kreuztabelle MITGLIED IN EINER GEWERKSCHAFT? *IM OEFFENTLICHEN DIENST TAETIG?

			IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
			JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	Anzahl	115	170	285
		Erwartete Anzahl	73,2	211,8	285,0
	NEIN	Anzahl	321	1091	1412
		Erwartete Anzahl	362,8	1049,2	1412,0
Gesamtsumme		Anzahl	436	1261	1697
		Erwartete Anzahl	436,0	1261,0	1697,0

Anzahl: $n_{i,j}$

erwartete Anzahl:
 $e_{i,j}$

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{i,j} - e_{i,j})^2}{e_{i,j}}$$

Aufgabe 2.2: Berechnung χ^2

Zeile i	Spalte j	$n_{i,j}$	$e_{i,j}$	$n_{i,j} - e_{i,j}$	$(n_{i,j} - e_{i,j})^2$	$\frac{(n_{i,j} - e_{i,j})^2}{e_{i,j}}$
1	1	115	73,2			
1	2	170	211,8			
2	1	321	362,8			
2	2	1091	1049,2			
		1697	1697			$\chi^2 =$

Bestimmen Sie mithilfe der Tabelle den Wert von $\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{i,j} - e_{i,j})^2}{e_{i,j}}$!

Stochastische Unabhängigkeit: Hilfstabelle Berechnung χ^2 II

Zeile i	Spalte j	$n_{i,j}$	$e_{i,j}$	$n_{i,j} - e_{i,j}$	$(n_{i,j} - e_{i,j})^2$	$\frac{(n_{i,j} - e_{i,j})^2}{e_{i,j}}$
1	1	115	73,2	41,8	1747,24	23,87
1				-41,8	1747,24	8,25
2				-41,8	1747,24	4,82
2	2	1091	1049,2	41,8	1747,24	1,67
		1697	1697			$\chi^2 =$ 38,61

Besonderheit in der Vierfeldertafel

Was sagt dieser Wert?

Stochastische Unabhängigkeit:

χ^2 : Interpretation

- Wertebereich:
 - $[0; \infty]$
- Interpretation:
 - wenn $\chi^2 > 0$, dann besteht ein Zusammenhang zwischen den beiden untersuchten Variablen, sie sind stochastisch abhängig

Stochastische Unabhängigkeit: Eigenschaften von Chiquadrat

- Eigenschaften:
 - Wert von χ^2 ist nach oben offen \rightarrow kein direktes Maß der Stärke
 - Größenordnung hängt nicht nur von Stärke des Zusammenhangs, sondern auch von Fallzahl und Zeilen- und Spaltenzahl der Tabelle ab
- Lösungsansatz: Normierung
 - korrigierter Kontingenzkoeffizient C^*
 - Cramér's V (Mehrfeldertafel)
 - Phikoeffizient (Vierfeldertafel)

Exkurs: Kontingenzkoeffizient C

- Berechnung:

- $C = \sqrt{\frac{\chi^2}{n + \chi^2}}$

- Eigenschaften:

- nicht mehr abhängig von Fallzahl
 - Zeilen- und Spaltenzahl beeinflussen Größenordnung
 - produziert tendenziell recht hohe Zusammenhangswerte

- Lösungsansatz:

- Normierung am Maximalwert

- $C_{Max} = \sqrt{\frac{\min(I-1, J-1)}{\min(I, J)}}$

I: Anzahl Zeilen
J: Anzahl Spalten

Exkurs: Korrigierter Kontingenzkoeffizient C^*

- Berechnung:

- $C^* = \frac{C}{C_{Max}}$

- $C^* = \sqrt{\frac{\chi^2}{n+\chi^2}} / \sqrt{\frac{\min(I-1, J-1)}{\min(I, J)}}$

- Eigenschaften:

- Koeffizient ist zwischen 0 (perfekte Unabhängigkeit) und 1 (perfekte Abhängigkeit) gebunden
 - Höhe des Zusammenhangs wird tendenziell stärker eingeschätzt als bei anderen nominalen Zusammenhangsmaßen
 - heute kaum mehr gebräuchlich, da direkte Normierung an Maximum von χ^2 möglich und eingängiger

Cramér's V

- Zusammenhangsmaß für symmetrische Zusammenhänge zwischen nominalen Variablen

- Berechnung:

$$V = \sqrt{\frac{\chi^2}{\chi^2_{max}}} = \sqrt{\frac{\chi^2}{n * \min(I-1, J-1)}}$$

I: Anzahl der Zeilen
J: Anzahl der Spalten

Minimum der beiden
Werte (Zeilenanzahl-1
oder Spaltenanzahl-1)

- Eigenschaften:

- Wertebereich von [0;1]
- 0 kein Zusammenhang
- 1 perfekter Zusammenhang

Aufgabe 3.1: Berechnung Cramér's V

<i>Zeile i</i>	<i>Spalte j</i>	$n_{i,j}$	$e_{i,j}$	$n_{i,j} - e_{i,j}$	$(n_{i,j} - e_{i,j})^2$	$\frac{(n_{i,j} - e_{i,j})^2}{e_{i,j}}$
1	1	115	73,2	41,8	1747,24	23,87
1	2	170	211,8	-41,8	1747,24	8,25
2	1	321	362,8	-41,8	1747,24	4,82
2	2	1091	1049,2	41,8	1747,24	1,67
		1697	1697		$\chi^2 =$	38,61

Berechnen Sie Cramér's V!

Berechnung von Cramér's V: Beispiel II

Zeile i	Spalte j	$n_{i,j}$	$e_{i,j}$	$n_{i,j} - e_{i,j}$	$(n_{i,j} - e_{i,j})^2$	$\frac{(n_{i,j} - e_{i,j})^2}{e_{i,j}}$
1	1	115	73,2	41,8	1747,24	23,87
1	2	170	211,8	-41,8	1747,24	8,25
2	1	321	362,8	-41,8	1747,24	4,82
2	2	1091	1049,2	41,8	1747,24	1,67
		1697	1697		$\chi^2 =$	38,61

$$\bullet V = \sqrt{\frac{\chi^2}{n \cdot \min(I-1, J-1)}}$$

$$\bullet V = \sqrt{\frac{38,61}{1697 \cdot \min(2-1, 2-1)}} = \sqrt{\frac{38,61}{1697}} = 0,151$$

Cramér's V: Interpretation

Wert von Cramér's V	Stärke des Zusammenhangs
0	kein Zusammenhang
$0 < V < 0,05$	vernachlässigbarer Zusammenhang
$0,05 < V < 0,25$	geringer Zusammenhang
$0,25 < V < 0,50$	mittlerer Zusammenhang
$0,50 < V < 1$	starker Zusammenhang
1	perfekter Zusammenhang

Wie lässt sich Cramér's V für den vorliegenden Fall interpretieren?

Cramér's V weist einen Wert von 0,151. Zwischen einer Tätigkeit im öffentlichen Dienst und einer Gewerkschaftsmitgliedschaft besteht folglich ein schwacher Zusammenhang. Schaut man in die Tabelle so wird ersichtlich, dass Angehörige des öffentlichen Dienstes eher Gewerkschaftsmitglieder sind.

Phikoeffizient ϕ

- Spezialfall von Cramér's V für die Vierfeldertafel
 - Betrag entspricht dem von Cramér's V
 - je nach verwendeter Formel und Ausgangsdaten lässt sich auch das Vorzeichen sinnvoll interpretieren

- Berechnung:

- $\phi = \sqrt{\frac{\chi^2}{n}}$

Maximum von χ^2 in der Vierfeldertafel ist die Fallzahl n (vgl. Formel Cramér's V)

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

Diese Formel ist wesentlich schneller und behält das Vorzeichen als Hinweis auf die „Richtung“!

Phikoeffizient ϕ : Berechnung

		IM OEFFENTLICHEN DIENST TAETIG?		Gesamtsum me
		JA	NEIN	
MITGLIED IN EINER GEWERKSCHAFT?	JA	115	170	285
	NEIN	321	1091	1412
Gesamtsumme		436	1261	1697

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

$$\phi = \frac{(115 * 1091) - (170 * 321)}{\sqrt{285 * 1412 * 436 * 1261}} = 0,151$$

Fallbeispiel: Geschlecht und Beruf 2014

Eine Ungleichheitsforscherin interessiert sich dafür, ob das Geschlecht einen Einfluss auf die Berufstätigkeit ausübt. Mithilfe des ALLBUS 2014 erhält sie folgende Kreuztabelle:

		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

Quiz: Was für ein Skalenniveau liegt bei beiden Variablen vor?
Handelt es sich um eine symmetrische oder asymmetrische Fragestellung?

Geschlecht und Beruf II: Analyse

asymmetrische
Fragestellung:
Einfluss Geschlecht(X)
auf Berufstätigkeit(Y)

Mehrfeldertafel mit einer
nominalen Variablen (X)
und einer ordinalen
Variablen (Y)
→ größtes
gemeinsames Niveau:
nominal

Eine Ungleichheitsforscherin interessiert sich dafür, ob das Geschlecht einen Einfluss auf die Berufstätigkeit ausübt.

		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

Quiz

		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

Quiz: Ist die Prozentsatzdifferenz ein geeignetes Zusammenhangsmaß?
Warum (nicht)?

Logik der proportionalen Fehlerreduktion

- Ausgangsfrage:
 - Um wie viel Prozent reduziert sich der Vorhersagefehler für die abhängige Variable durch Kenntnis der unabhängigen Variablen?

- allgemeine Formel:

- $PRE = \frac{E_0 - E_1}{E_0}$

E_0 : Fehler **ohne** Kenntnis der unabhängigen Variablen

E_1 : Fehler **mit** Kenntnis der unabhängigen Variablen

Goodmans und Kruskals λ

- einfaches PRE-Maß für nominale Variablen
- Formel:
 - $PRE = \frac{E_0 - E_1}{E_0}$
- zwei asymmetrische Varianten:
 - λ_{YX}
 - λ_{XY}
- zusätzlich eine symmetrische Variante:
 - $\lambda_{symmetrisch}$
- Wertebereich: 0 bis 1
(entspricht Fehlerreduktion von 0% bis 100%)

Goodmans und Kruskals λ_{YX}

- Verwendung:
 - asymmetrisches Zusammenhangsmaß für nominale Variablen
 - Einfluss der Spaltenvariablen X auf die Zeilenvariable Y soll untersucht werden

- Bestimmung des E0-Fehlers:

- $E_0 = n - \max_i(n_{i,+})$

n : Gesamtsumme
 $\max_i(n_{i,+})$: Modus von Y

- Bestimmung des E1-Fehlers:

- $E_1 = \sum_{j=1}^J (n_{+,j} - \max_i(n_{i,j}))$

$n_{+,j}$: Summe der Spalte j
 $\max_i(n_{i,j})$: Modus in Spalte j

Goodmans und Kruskals λ_{XY}

- Verwendung:
 - asymmetrisches Zusammenhangsmaß für nominale Variablen
 - Einfluss der Zeilenvariablen Y auf die Spaltenvariablen X soll untersucht werden

- Bestimmung des E0-Fehlers:

$$E_0 = n - \max_j(n_{+,j})$$

n : Gesamtsumme
 $\max_j(n_{+,j})$: Modus von X

- Bestimmung des E1-Fehlers:

$$E_1 = \sum_{i=1}^I \left(n_{i,+} - \max_j(n_{i,j}) \right)$$

$n_{i,+}$: Summe der Zeile i
 $\max_j(n_{i,j})$: Modus in Zeile i

Exkurs: Goodmans und Kruskals λ_S

- Verwendung:
 - Zusammenhangsmaß für nominale Daten, wenn keine klare Wirkungsrichtung angegeben werden kann

- Berechnung:

- $\lambda_S = \frac{E_0 - E_1}{E_0}$

- $E_0 = E_0(\lambda_{YX}) + E_0(\lambda_{XY})$

- $E_1 = E_1(\lambda_{YX}) + E_1(\lambda_{XY})$

Addition der beiden E0- und E1-Fehler
für λ_{YX} und λ_{XY}

Geschlecht und Beruf III: Bestimmung des E_0 -Fehlers für λ_{YX}

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
BEFRAGTE<R> BERUFSTAETIG?	HAUPTBERUFL. GANZTAGS			1571
	HAUPTBERUFL. HALBTAGS			350
	NEBENHER BERUFSTAE.			208
	NICHT ERWERBSTAETIG			1339
Gesamtsumme				3468

Quiz: Wir kennen lediglich die Verteilung der Berufstätigkeit für alle Befragten. Was wäre hier der beste Tipp. Wenn wir raten müssten, welche Berufstätigkeit für einen einzelnen Befragten vorliegt?

Geschlecht und Beruf IV:

Bestimmung des E_0 -Fehlers für λ_{YX}

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MAENNLICH	WEIBLICH	
BEFRAGTE<R> BERUFSTAETIG?	HAUPTBERU GANZTAGS	<div style="border: 1px solid black; padding: 5px; text-align: center;"> bester Tipp: hauptberuflich ganztags (1571 mal richtig) </div>		1571
	HAUPTBERU HALBTAGS			350
	NEBENHER BERUFSTAE	<div style="border: 1px solid black; padding: 5px; text-align: center;"> maximal könnten wir mit unserem Tipp 3468 mal daneben liegen </div>		208
	NICHT ERWERBSTA			1339
Gesamtsumme				3468

Wenn wir unter ausschließlicher Nutzung der Verteilung von Y darauf tippen würden, dass der Befragte ganztags tätig ist, liegen wir also $E_0 = 3468 - 1571 = 1897$ mit diesem Tipp daneben.

Geschlecht und Beruf V:

Bestimmung des E_1 -Fehlers für λ_{YX}

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

Quiz: Welchen Tipp würden wir abgeben, wenn wir wissen, das unser Befragter ein Mann ist? Wie oft liegen wir damit daneben? Wie sieht es bei den Frauen aus?

Geschlecht und Beruf VI:

Bestimmung des E_1 -Fehlers für λ_{YX}

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

Lösung: idealer Tipp für Männer ist hauptberuflich tätig (733 mal daneben) und für Frauen ist der ideale Tipp nicht erwerbstätig (989 mal daneben). Insgesamt würden wir bei der Prognose unter Kenntnis des Geschlechts $733+989=1722$ mal daneben liegen. Das entspricht unserem E_1 -Fehler.

Geschlecht und Beruf VII:

Bestimmung λ_{YX}

		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

- Berechnung:

- $E_0 = 3468 - 1571 = 1897$

- $E_1 = (1759 - 1026) + (1709 - 720) = 1722$

- $\lambda_{YX} = \frac{(E_0 - E_1)}{E_0}$

Berechnen Sie λ_{YX} . Um wie viel Prozent verbessert sich die Vorhersage von Y unter Kenntnis von X?

Geschlecht und Beruf VIII:

Bestimmung λ_{YX}

- Berechnung:

- $E_0 = 3468 - 1571 = 1897$

- $E_1 = (1759 - 1026) + (1709 - 720) = 1722$

- $\lambda_{YX} = \frac{(E_0 - E_1)}{E_0}$

- $\lambda_{YX} = \frac{(1897 - 1722)}{1897} = 0,092$

- Interpretation:

- Durch Kenntnis des Geschlechts (X) verbessert sich die Vorhersage der Berufstätigkeit (Y) um 9,2% (im Vergleich zur Vorhersage ohne Kenntnis des Geschlechts).

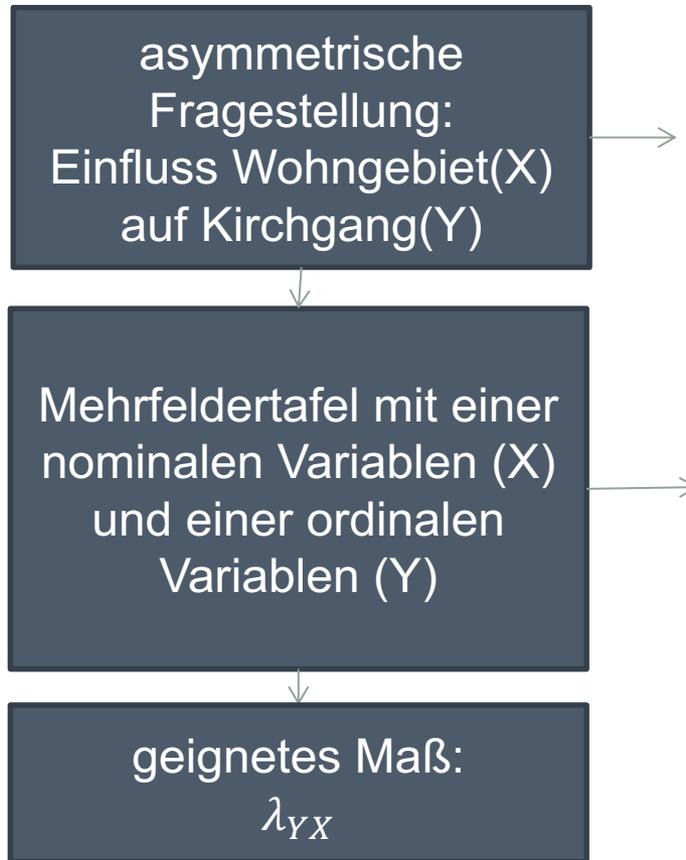
Aufgabe 4: Kirchgangshäufigkeit 2014

Eine Religionssoziologin interessiert sich dafür, ob sich durch Kenntnis des Wohngebietes des Befragten die Prognose der Kirchgangshäufigkeit verbessern lässt. Mithilfe des ALLBUS 2014 erhält sie folgende Kreuztabelle:

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
KIRCHGANGSHÄUFIGKEIT	UEBER 1X DIE WOCHE	45	10	55
	1X PRO WOCHE	113	23	136
	1-3X PRO MONAT	225	37	262
	MEHRMALS IM JAHR	458	132	590
	SELTENER	807	343	1150
	NIE	610	555	1165
Gesamtsumme		2258	1100	3358

- Welches Zusammenhangsmaß ist hier geeignet?
- Berechnen Sie ein geeignetes Zusammenhangsmaß.
- Interpretieren Sie dieses!

Aufgabe 4a: Analyse



Eine Religionssoziologin interessiert sich dafür, ob sich durch Kenntnis des Wohngebietes des Befragten die Prognose der Kirchgangshäufigkeit verbessern lässt.

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
KIRCHGANGSHÄUFIGKEIT	UEBER 1X DIE WOCHE	45	10	55
	1X PRO WOCHE	113	23	136
	1-3X PRO MONAT	225	37	262
	MEHRMALS IM JAHR	458	132	590
	SELTENER	807	343	1150
	NIE	610	555	1165
Gesamtsumme		2258	1100	3358

Aufgabe 4b: Lösung

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
KIRCHGANGSHAUEFIGKEIT	UEBER 1X DIE WOCHE	45	10	55
	1X PRO WOCHE	113	23	136
	1-3X PRO MONAT	225	37	262
	MEHRMALS IM JAHR	458	132	590
	SELTENER	807	343	1150
	NIE	610	555	1165
Gesamtsumme		2258	1100	3358

- Berechnung:

- $E_0 = 3358 - 1165$

- $E_0 = 2193$

Aufgabe 4b: Lösung II

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
KIRCHGANGSHAUFGEHET	UEBER 1X DIE WOCHE	45	10	55
	1X PRO WOCHE	113	23	136
	1-3X PRO MONAT	225	37	262
	MEHRMALS IM JAHR	458	132	590
	SELTENER	807	343	1150
	NIE	610	555	1165
Gesamtsumme		2258	1100	3358

- Berechnung:

- $E_1 = (2258 - 807) + (1100 - 555)$

- $E_1 = 1996$

Aufgabe 4b: Lösung III

- Berechnung:

- $E_0 = 2193$

- $E_1 = 1996$

- $\lambda_{YX} = PRE = \frac{E_0 - E_1}{E_0}$

- $\lambda_{YX} = \frac{2193 - 1996}{2193} = 0,09$

- Interpretation:

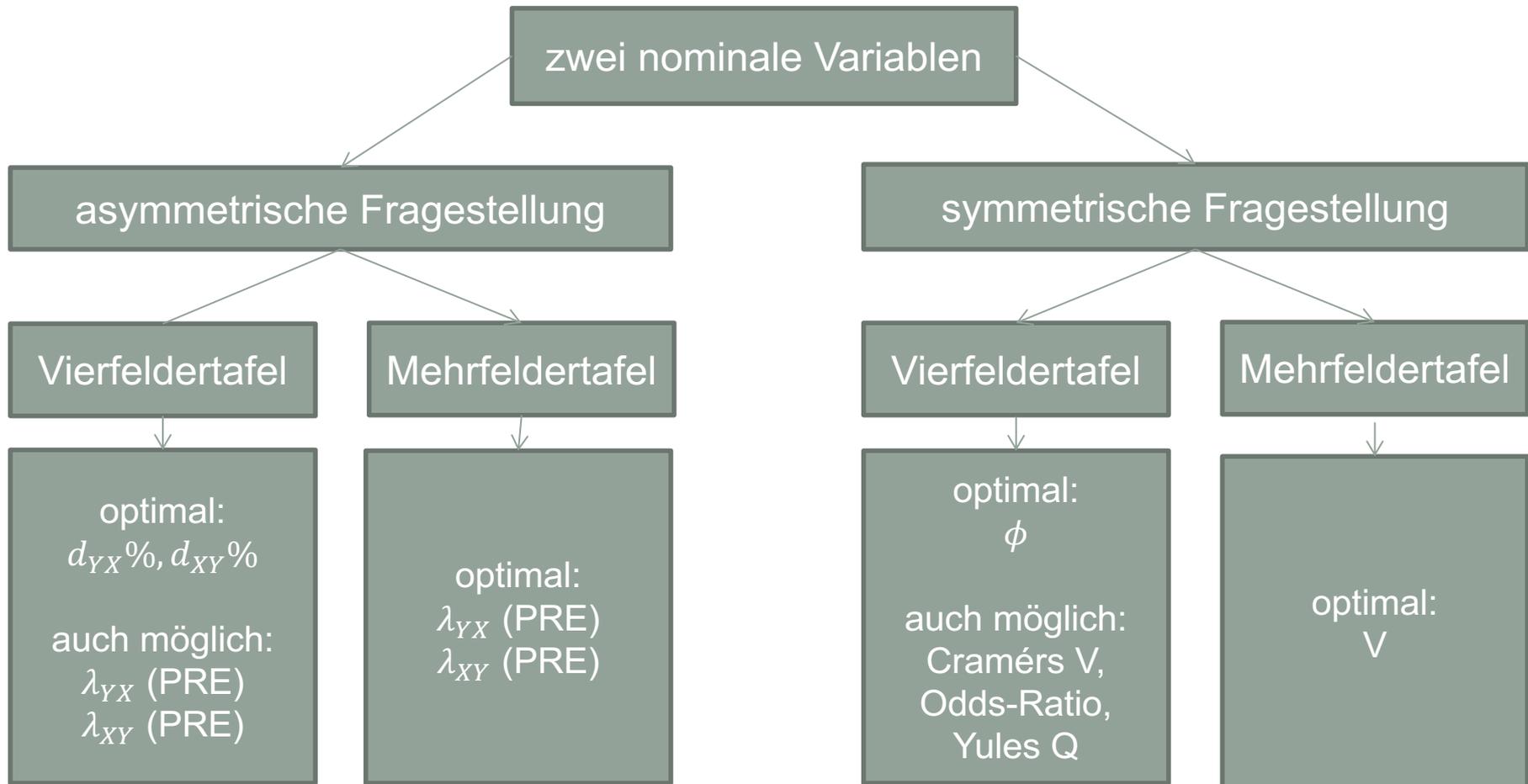
- Durch Kenntnis des Wohngebietes eines Befragten lässt sich die Prognose der Kirchgangshäufigkeit um 9% verbessern.

- Alternativ: Durch Kenntnis des Wohngebietes der Befragten lässt sich der Vorhersagefehler um 9% reduzieren im Vergleich zur ausschließlichen Kenntnis der Verteilung der Kirchgangshäufigkeit.

Lambda: Hinweis

- λ_{YX} und λ_{XY} geben als Maße der proportionalen Fehlerreduktion an, ob sich durch Kenntnis der unabhängigen Variablen die Prognose der abhängigen Variablen verbessert. Ein Wert von null deutet darauf hin, dass keine Verbesserung erzielt werden kann, bedeutet jedoch **nicht**, dass kein Zusammenhang besteht.
- Wenn für alle die gleiche Ausprägung vorhergesagt würde, wie mithilfe der Randverteilung, dann wird der Wert null, da die Information nichts bringt, um die Prognose verbessern.

Entscheidungsbaum nominale Zusammenhangsmaße



Literaturhinweise

- nominale Maße für die Vierfeldertafel:
 - Kerstin Völkl / Christoph Korb (2018): Deskriptive Statistik. Eine Einführung für Politikwissenschaftlerinnen und Politikwissenschaftler. S. 153-160.
 - Steffen-M. Kühnel / Dagmar Krebs (2012): Statistik für die Sozialwissenschaften. Kapitel 9.1, 9.2.1, 9.2.3, 9.3 (ohne standardisierte Residuen und Yules Q, Schätzen und Testen)
- nominale Maße für die Mehrfeldertafel:
 - Kerstin Völkl / Christoph Korb (2018): Deskriptive Statistik. Eine Einführung für Politikwissenschaftlerinnen und Politikwissenschaftler. S. 161-180.
 - Steffen-M. Kühnel / Dagmar Krebs (2012): Statistik für die Sozialwissenschaften. Kapitel 10.1, 10.1.1 (nur Berechnung Chiquadrat), 10.3

Literaturhinweise II

- Zu beidem:
 - Hans Benninghaus (2007): Deskriptive Statistik. Eine Einführung für Sozialwissenschaftler. Kapitel 4.2, 4.3 und 5.1 sowie 5.2 (Kapitel 4.2 und 4.3 auch als Ergänzung zu Kühnel / Krebs lesenswert)

Übungsaufgabe 1: Lebenspartner und Heirat

Ein Familiensoziologe geht davon aus, dass es einen Zusammenhang zwischen dem Befinden in einer Partnerschaft und der Antwort auf die Frage, ob man bei dauerndem Zusammenleben heiraten sollte, gibt. Mithilfe des ALLBUS 2014 erhält er folgende Tabelle:

		HABEN SIE EINEN FESTEN LEBENSPARTNER?		Gesamtsum me
		JA	NEIN	
HEIRAT BEI	JA	175	286	461
DAUERNDEN	NEIN	398	410	808
ZUSAMMENLEBEN	UNENTSCHIEDEN	92	148	240
Gesamtsumme		665	844	1509

- Berechnen Sie ein geeignetes Zusammenhangsmaß, wenn Sie von einem symmetrischen Zusammenhang ausgehen.
- Interpretieren Sie Ihr Ergebnis!

Übungsaufgabe 1: Analyse

symmetrische
Fragestellung:
keine Richtung

Ein Familiensoziologe geht davon aus, dass es einen Zusammenhang zwischen dem Befinden in einer Partnerschaft und der Antwort auf die Frage, ob man bei dauerndem Zusammenleben heiraten sollte, gibt.

Mehrfeldertafel mit
nominalen Variablen

		HABEN SIE EINEN FESTEN LEBENSPARTNER?		Gesamtsumme
		JA	NEIN	
HEIRAT BEI DAUERNDEN ZUSAMMENLEBEN	JA	175	286	461
	NEIN	398	410	808
	UNENTSCHEIDEN	92	148	240
Gesamtsumme		665	844	1509

geeignetes Maß:
Cramers V

Übungsaufgabe 1: Lösung

- Indifferenztabelle:

Kreuztabelle HEIRAT BEI DAUERNDEN ZUSAMMENLEBEN*HABEN SIE EINEN FESTEN LEBENSPARTNER?

Erwartete Anzahl

		HABEN SIE EINEN FESTEN LEBENSPARTNER?		Gesamtsumme
		JA	NEIN	
HEIRAT BEI DAUERNDEN ZUSAMMENLEBEN	JA	203,2	257,8	461,0
	NEIN	356,1	451,9	808,0
	UNENTSCHIEDEN	105,8	134,2	240,0
Gesamtsumme		665,0	844,0	1509,0

Übungsaufgabe 1: Lösung II

- Hilfstabelle Chiquadrat:

<i>Zeile i</i>	<i>Spalte j</i>	n_{ij}	e_{ij}	$n_{ij} - e_{ij}$	$(n_{ij} - e_{ij})^2$	$\frac{(n_{ij} - e_{ij})^2}{e_{ij}}$
1	1	175	203,2	-28,2	795,2	3,91
1	2	286	257,8	28,2	795,2	3,08
2	1	398	356,1	41,9	1756	4,93
2	2	410	451,9	-41,9	1756	3,88
3	1	92	105,8	-13,8	190,4	1,80
3	2	148	134,2	13,8	190,4	1,42
		1509	1509		$\chi^2 =$	19,03

Übungsaufgabe 1: Lösung III

- Berechnung Cramers V:

$$\bullet V = \sqrt{\frac{\chi^2}{n \cdot \min(I-1, J-1)}}$$

$$\bullet V = \sqrt{\frac{19,03}{1509 \cdot \min(3-1, 2-1)}} = \sqrt{\frac{19,03}{1509}} = \pm 0,112$$

- Interpretation:

- Es besteht ein schwacher Zusammenhang zwischen dem Befinden in einer Partnerschaft und der Einstellung zu der Frage, ob Personen, die dauerhaft zusammenleben heiraten sollten.

Übungsaufgabe 2: Umweltbewusstsein 2010

Ein Forscher geht davon aus, dass es einen Zusammenhang zwischen der Teilnahme an einer Umweltdemonstration in den letzten fünf Jahren und dem gespendeten Geld für den selben Zweck im gleichen Zeitraum gibt. Mithilfe des ISSP 2010 erhält er folgende Tabelle:

		LETZTE 5 JAHRE: DEMONSTRATION F. UMWELT		Gesamtsumme
		JA	NEIN	
LETZTE 5 JAHRE: GELD GESPENDET F. UMWELT	JA	34	171	205
	NEIN	22	1130	1152
Gesamtsumme		56	1301	1357

- Berechnen Sie ein geeignetes chiquadrat-basiertes Zusammenhangsmaß.
- Interpretieren Sie Ihr Ergebnis.
- Welches andere symmetrische Zusammenhangsmaß hätten Sie berechnen können?

Übungsaufgabe 2: Analyse

symmetrische Fragestellung:
Zusammenhang Demo und Spende

Vierfeldertafel mit zwei dichotomen, nominalen Variablen

geeignete Maße:
 χ^2 -basierte Maße,
insbesondere: ϕ , auch
möglich: Cramers V

Ein Forscher geht davon aus, dass es einen Zusammenhang zwischen der Teilnahme an einer Umweltdemonstration in den letzten fünf Jahren und dem gespendeten Geld für den selben Zweck im gleichen Zeitraum gibt.

Kreuztabelle LETZTE 5 JAHRE: GELD GESPENDET F.UMWELT*LETZTE 5 JAHRE: DEMONSTRATION F. UMWELT

Anzahl

		LETZTE 5 JAHRE: DEMONSTRATION F. UMWELT		Gesamtsumme
		JA	NEIN	
LETZTE 5 JAHRE: GELD GESPENDET F.UMWELT	JA	34	171	205
	NEIN	22	1130	1152
Gesamtsumme		56	1301	1357

Übungsaufgabe 2: Lösung

**Kreuztabelle LETZTE 5 JAHRE: GELD GESPENDET F.UMWELT*LETZTE 5 JAHRE:
DEMONSTRATION F. UMWELT**

Anzahl

		LETZTE 5 JAHRE: DEMONSTRATION F. UMWELT		Gesamtsum me
		JA	NEIN	
LETZTE 5 JAHRE: GELD GESPENDET F.UMWELT	JA	34	171	205
	NEIN	22	1130	1152
Gesamtsumme		56	1301	1357

- **Berechnung:**

- $$\phi = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$
- $$\phi = \frac{(34*1130)-(171*22)}{\sqrt{(34+171)(22+1130)(34+22)(171+1130)}}$$
- $$\phi = +0,264$$

Übungsaufgabe 2: Lösung II

- Interpretation:
 - Phi beträgt 0,264. Es besteht somit ein mittlerer symmetrischer Zusammenhang zwischen der Demoteilnahme und der Spende für Umweltzwecke in den letzten fünf Jahren.
 - Da wir hier Phi mithilfe einer Formel berechnet haben, die ein Vorzeichen ergibt, lässt sich sagen, dass die AD-Diagonale dominiert. Menschen, die für den Umweltschutz gespendet haben, haben somit auch eher an Demonstrationen teilgenommen und umgekehrt.

Übungsaufgabe 2: alternative Lösung

1. Indifferenztabelle erstellen

<i>Indifferenz – tabelle</i>	Demoteilnahme	Keine Demoteilnahme	Gesamt
Spende	$= \frac{205 * 56}{1357}$ = 8,5	$= \frac{205 * 1301}{1357}$ = 196,5	205
Keine Spende	$= \frac{1152 * 56}{1357}$ = 47,5	$= \frac{1152 * 1301}{1357}$ = 1104,5	1152
Gesamt	56	1301	1357

2. Schritt: Hilfstabelle erstellen und Chi² ermitteln

Übungsaufgabe 2: alternative Lösung II

<i>Zeile i</i>	<i>Spalte j</i>	n_{ij}	e_{ij}	$n_{ij} - e_{ij}$	$(n_{ij} - e_{ij})^2$	$\frac{(n_{ij} - e_{ij})^2}{e_{ij}}$
1	1	34	8,5	25,5	650,25	76,50
1	2	171	196,5	-25,5	650,25	3,31
2	1	22	47,5	-25,5	650,25	13,69
2	2	1130	1104,5	25,5	650,25	0,59
		1357	1357	0	$\chi^2 =$	94,09

Übungsaufgabe 2: alternative Lösung III

3. Schritt: Cramérs V bestimmen:

$$\bullet V = \sqrt{\frac{\chi^2}{\chi^2_{max}}} = \sqrt{\frac{\chi^2}{N * \min(I-1, J-1)}} = \sqrt{\frac{94,09}{1357 * 1}} = \pm 0,264$$

4. Schritt: Interpretation:

- Cramers V beträgt 0,264. Es besteht somit ein mittlerer Zusammenhang zwischen der Demoteilnahme und der Spende für Umweltzwecke in den letzten fünf Jahren.

Bei Vierfeldertafeln ist der Weg über den Phikoeffizienten deutlich schneller. Wie erhalten hier das gleiche inhaltliche Ergebnis.

Übungsaufgabe 3: Politische Diskussion (USA)

Der General Social Survey ist das US-Pendant zum deutschen ALLBUS. Eine Politikwissenschaftlerin interessiert sich dafür, ob das Geschlecht der Befragten einen Einfluss auf die Häufigkeit der Teilnahme an politischen Diskussionen hat. Für das Jahr 2014 erhält sie folgende Tabelle:

Anzahl

		RESPONDENTS SEX		Gesamtsumme
		MALE	FEMALE	
HOW OFTEN R DISCUSS POLITICS	OFTEN	62	55	117
	SOMETIMES	205	197	402
	RARELY	185	234	419
	NEVER	101	216	317
Gesamtsumme		553	702	1255

- Welches Zusammenhangsmaß ist hier geeignet?
- Berechnen Sie dieses Maß und interpretieren Sie es!

Übungsaufgabe 3a: Analyse

asymmetrische
Fragestellung:
Einfluss Geschlecht(X)
auf pol. Diskussion (Y)

Eine Politikwissenschaftlerin interessiert sich dafür, ob das Geschlecht der Befragten einen Einfluss auf die Häufigkeit der Teilnahme an politischen Diskussionen hat.

Mehrfeldertafel
X: nominal, Y: ordinal
höchstes gemeinsames
Skalenniveau: nominal

Anzahl

		RESPONDENTS SEX		Gesamtsumme
		MALE	FEMALE	
HOW OFTEN R DISCUSS	OFTEN	62	55	117
POLITICS	SOMETIMES	205	197	402
	RARELY	185	234	419
	NEVER	101	216	317
Gesamtsumme		553	702	1255

geeignetes Maß:
 λ_{YX}

Übungsaufgabe 3b: Lösung

Anzahl

		RESPONDENTS SEX		Gesamtsumme
		MALE	FEMALE	
HOW OFTEN R DISCUSS POLITICS	OFTEN	62	55	117
	SOMETIMES	205	197	402
	RARELY	185	234	419
	NEVER	101	216	317
Gesamtsumme		553	702	1255

- Berechnung:

- $\lambda_{YX} = PRE = \frac{E_0 - E_1}{E_0}$

- $E_0 = 1255 - 419 = 836$

- $E_1 = (553 - 205) + (702 - 234) = 816$

- $\lambda_{YX} = PRE = \frac{836 - 816}{836} = 0,024$

- Interpretation:

- Durch Kenntnis des Geschlechts der Befragten lässt sich die Prognose der Teilnahme an politischen Diskussionen um lediglich 2,4% verbessern.

Übungsaufgabe 4: Zeitungskonsum(USA)

Der General Social Survey ist das US-Pendant zum deutschen ALLBUS. Eine Politikwissenschaftlerin interessiert sich dafür, ob das Geschlecht der Befragten auf die Häufigkeit des Zeitungskonsums hat. Für das Jahr 2014 erhält sie folgende Tabelle:

Anzahl

		HOW OFTEN DOES R READ NEWSPAPER					Gesamtsumme
		EVERYDAY	FEW TIMES A WEEK	ONCE A WEEK	LESS THAN ONCE WK	NEVER	
RESPONDENTS SEX	MALE	222	120	98	123	191	754
	FEMALE	195	140	148	148	290	921
Gesamtsumme		417	260	246	271	481	1675

- Welches Zusammenhangsmaß ist hier geeignet?
- Berechnen Sie dieses Maß und interpretieren Sie es!

Übungsaufgabe 4a: Analyse

asymmetrische
Fragestellung:
Geschlecht(Y) auf
Zeitungskonsum (X)

Eine Politikwissenschaftlerin interessiert sich dafür, ob das Geschlecht der Befragten einen Einfluss auf die Häufigkeit der Teilnahme an politischen Diskussionen hat.

Mehrfeldertafel
Y: nominal, X: ordinal
höchstes gemeinsames
Skalenniveau: nominal

Anzahl

		HOW OFTEN DOES R READ NEWSPAPER					Gesamtsumme
		EVERYDAY	FEW TIMES A WEEK	ONCE A WEEK	LESS THAN ONCE WK	NEVER	
RESPONDENTS SEX	MALE	222	120	98	123	191	754
	FEMALE	195	140	148	148	290	921
Gesamtsumme		417	260	246	271	481	1675

geeignetes Maß:
 λ_{XY}

Übungsaufgabe 4b: Lösung

Anzahl

		HOW OFTEN DOES R READ NEWSPAPER					Gesamtsumme
		EVERYDAY	FEW TIMES A WEEK	ONCE A WEEK	LESS THAN ONCE WK	NEVER	
RESPONDENTS SEX	MALE	222	120	98	123	191	754
	FEMALE	195	140	148	148	290	921
Gesamtsumme		417	260	246	271	481	1675

- Berechnung:

- $\lambda_{XY} = PRE = \frac{E_0 - E_1}{E_0}$

- $E_0 = 1675 - 481 = 1194$

- $E_1 = (754 - 222) + (921 - 290) = 1163$

- $\lambda_{YX} = PRE = \frac{1194 - 1163}{1194} = 0,026$

- Interpretation:

- Durch Kenntnis des Geschlechts der Befragten lässt sich die Prognose der Häufigkeit des Zeitungskonsums um lediglich 2,6% verbessern.

Übungsaufgabe 5: Unterschied Ost-West Mehrheit uninteressiert an Mitmenschen?

Ein Soziologe interessiert sich dafür, ob es einen Unterschied zwischen Ost- und Westdeutschen hinsichtlich der Frage gibt, ob die Mehrheit der Menschen an ihren Mitmenschen uninteressiert ist. Mithilfe des ALLBUS 2012 erhält er folgende Kreuztabelle:

Anzahl

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
MEHRHEIT UNINTERESSIERT AN MITMENSCHEN	BIN DERS.MEINUNG	1604	833	2437
	BIN ANDERER MEINUNG	693	266	959
Gesamtsumme		2297	1099	3396

- Welches Zusammenhangsmaß ist hier geeignet?
- Berechnen Sie dieses Maß und interpretieren Sie es!

Übungsaufgabe 5: Analyse

asymmetrische
Fragestellung:
Einfluss Wohngebiet (X)
auf Einstellung (Y)

Ein Soziologe interessiert sich dafür, ob es einen Unterschied zwischen Ost- und Westdeutschen hinsichtlich der Frage gibt, ob die Mehrheit der Menschen an ihren Mitmenschen uninteressiert ist.

Vierfeldertafel mit zwei
dichotomen, nominalen
Variablen

Anzahl

		ERHEBUNGSGEBIET <WOHNGBEBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLAE NDER	NEUE BUNDESLAE NDER	
MEHRHEIT UNINTERESSIERT AN MITMENSCHEN	BIN DERS.MEINUNG BIN ANDERER MEINUNG	1604	833	2437
Gesamtsumme		2297	1099	3396

optimales Maß:
Prozentsatzdifferenz

$$d_{YX}\%$$

auch möglich:

$$\lambda_{YX}$$

Übungsaufgabe 5b: Lösung

Anzahl

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
MEHRHEIT UNINTERESSIERT AN MITMENSCHEN	BIN DERS.MEINUNG	1604	833	2437
	BIN ANDERER MEINUNG	693	266	959
Gesamtsumme		2297	1099	3396

- Berechnung:

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

- $d_{yx}\% = 100 * \left(\frac{1604}{2297} - \frac{833}{1099} \right) = -5,96$

Übungsaufgabe 5b: Lösung II

% in ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
MEHRHEIT UNINTERESSIERT AN MITMENSCHEN	BIN DERS.MEINUNG	69,8%	75,8%	71,8%
	BIN ANDERER MEINUNG	30,2%	24,2%	28,2%
Gesamtsumme		100,0%	100,0%	100,0%

• Interpretation:

- Die Prozentsatzdifferenz zwischen West- und Ostdeutschland liegt bei -5,96 Prozentpunkten, ist also relativ gering.
- Die Ostdeutschen sind tendenziell häufiger der Meinung, dass die Mehrheit uninteressiert an den Mitmenschen ist, als die Westdeutschen.

Übungsaufgabe 5b: Alternativlösung

Anzahl

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
MEHRHEIT UNINTERESSIERT AN MITMENSCHEN	BIN DERS.MEINUNG BIN ANDERER MEINUNG	1604 693	833 266	2437 959
Gesamtsumme		2297	1099	3396

- Berechnung:

- $\lambda_{YX} = PRE = \frac{E_0 - E_1}{E_0}$
- $E_0 = 3396 - 2437 = 959$
- $E_1 = (2297 - 1604) + (1099 - 833) = 959$
- $\lambda_{YX} = PRE = \frac{959 - 959}{959} = 0$

- Interpretation:

- Durch Kenntnis des Erhebungsgebietes lässt sich die Prognose der Zustimmung zu der Frage, ob die Mehrheit uninteressiert an Mitmenschen ist, um 0 %, also gar nicht verbessern. Dies kommt zu Stande, weil sowohl für die Westdeutschen als auch für die Ostdeutschen der beste Tipp Zustimmung zu der These ist. Das Erhebungsgebiet liefert also keinen Erklärungsbeitrag.

Übungsaufgabe 6: nicht-jugendfreie Filme

Ein US-Wissenschaftler möchte untersuchen, ob das Geschlecht des Befragten einen Einfluss auf den Konsum von nicht-jugendfreien Filmen gibt. Mit Hilfe des GSS 2012 erhält er folgende Kreuztabelle:

Anzahl

		RESPONDENTS SEX		Gesamtsumme
		MALE	FEMALE	
SEEN X-RATED MOVIE IN LAST YEAR	YES	215	117	332
	NO	382	620	1002
Gesamtsumme		597	737	1334

- Welches Zusammenhangsmaß erscheint Ihnen geeignet?
- Berechnen Sie das Zusammenhangsmaß und interpretieren Sie dieses.

Übungsaufgabe 6a: Analyse

asymmetrische
Fragestellung:
Einfluss Wohngebiet (X)
auf Einstellung (Y)

Vierfeldertafel mit zwei
dichotomen, nominalen
Variablen

geeignetes Maß:
Prozentsatzdifferenz

$$d_{YX}\%$$

suboptimal:

$$\lambda_{YX}$$

Ein US-Wissenschaftler möchte untersuchen, ob das Geschlecht des Befragten einen Einfluss auf den Konsum von nicht-jugendfreien Filmen gibt.

Anzahl

		RESPONDENTS SEX		Gesamtsumme
		MALE	FEMALE	
SEEN X-RATED MOVIE IN	YES	215	117	332
LAST YEAR	NO	382	620	1002
Gesamtsumme		597	737	1334

Übungsaufgabe 6b: Lösung

Anzahl

		RESPONDENTS SEX		Gesamtsumme
		MALE	FEMALE	
SEEN X-RATED MOVIE IN LAST YEAR	YES	215	117	332
	NO	382	620	1002
Gesamtsumme		597	737	1334

- Berechnung:

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

- $d_{yx}\% = 100 * \left(\frac{215}{597} - \frac{117}{737} \right) = 20,13$

- Interpretation:

- Die Differenz zwischen Männern und Frauen beträgt hier 20,13 Prozentpunkte. Da das Vorzeichen positiv ist, dominiert die AD-Diagonale. Männliche Befragte konsumieren also tendenziell eher mehr nicht-jugendfreie Filme als weibliche.

Übungsaufgabe 6b: Alternativlösung

Anzahl

		RESPONDENTS SEX		Gesamtsumme
		MALE	FEMALE	
SEEN X-RATED MOVIE IN LAST YEAR	YES	215	117	332
	NO	382	620	1002
Gesamtsumme		597	737	1334

- Berechnung:

- $\lambda_{YX} = PRE = \frac{E_0 - E_1}{E_0}$
- $E_0 = 1334 - 1002 = 332$
- $E_1 = (2297 - 1604) + (1099 - 833) = 332$
- $\lambda_{YX} = PRE = \frac{332 - 332}{332} = 0$

- Interpretation:

- Durch Kenntnis des Geschlechts lässt sich die Prognose des Konsums von nicht jugendfreien Filmen um 0% verbessern, da sowohl bei Männern und Frauen der beste Tipp kein Konsum ist. Lambda berücksichtigt nur die Modalwerte, nicht die Abstände. Betrachtet man die Prozentsatzdifferenz, so wird ersichtlich, dass ein Einfluss besteht, aber die Vorhersage mit der Information über das Geschlecht nicht verbessert werden kann.

Übungsaufgabe 7: Waffengesetzgebung (USA)

Eine Kriminologin möchte wissen, ob es in den USA einen Zusammenhang zwischen dem Besitz einer Waffe und der Zustimmung bzw. Ablehnung von zwingenden Waffenscheinen gibt. Hinsichtlich der Wirkrichtung hat sie sich keine Gedanken gemacht. Mithilfe des GSS 2012 erhält sie folgende Kreuztabelle:

Anzahl

		HAVE GUN IN HOME		Gesamtsumme
		YES	NO	
FAVOR OR OPPOSE	FAVOR	266	667	933
GUN PERMITS	OPPOSE	172	152	324
Gesamtsumme		438	819	1257

- Welches Zusammenhangsmaß erscheint Ihnen geeignet?
- Berechnen Sie das Zusammenhangsmaß und interpretieren Sie dieses!

Übungsaufgabe 7a: Analyse

symmetrische
Fragestellung:
Zusammenhang ohne
Richtungsaussage

Vierfeldertafel mit zwei
dichotomen, nominalen
Variablen

geeignete Maße:
 χ^2 -basierte Maße,
insbesondere: ϕ , auch
möglich: Cramers V,
Yules Q

Eine Kriminologin möchte wissen, ob es in den USA einen Zusammenhang zwischen dem Besitz einer Waffe und der Zustimmung bzw. Ablehnung von zwingenden Waffenscheinen gibt. Hinsichtlich der Wirkrichtung hat sie sich keine Gedanken gemacht.

Anzahl

		HAVE GUN IN HOME		Gesamtsumme
		YES	NO	
FAVOR OR OPPOSE	FAVOR	266	667	933
GUN PERMITS	OPPOSE	172	152	324
Gesamtsumme		438	819	1257

Übungsaufgabe 7b: Lösung

- Berechnung:

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

- $$\phi = \frac{266 \cdot 152 - 667 \cdot 172}{\sqrt{(933) \cdot (324) \cdot (438) \cdot (819)}}$$

- $$\phi = -0,23$$

- Interpretation:

- Phi beträgt -0,23. Es besteht ein mittlerer Zusammenhang zwischen dem Waffenbesitz und der Zustimmung bzw. Ablehnung.
- Das negative Vorzeichen deutet auf eine Dominanz der BC-Diagonale hin. Leute die keine Waffe besitzen sind eher für einen zwingend notwendigen Waffenschein als solche, die eine Waffe besitzen.

Anzahl

		HAVE GUN IN HOME		Gesamtsumme
		YES	NO	
FAVOR OR OPPOSE	FAVOR	266	667	933
GUN PERMITS	OPPOSE	172	152	324
Gesamtsumme		438	819	1257