STATISTIK FÜR DIE SOZIALWISSENSCHAFTEN

Zusammenhänge nominaler Variablen

Symmetrische und Asymmetrische Zusammenhänge nominaler Variablen beschreiben und testen

Meine Notizen:

Beispiel

	katholisch (x_1)	Lutherisch (x_2)	Ohne Konfession (x_3)	Gesamt
Linke (y_1)	14 0,0148 0,0569 0,1138	18 0,0190 0,0520 0,1463	91 0,0961 0,2563 0,7398	123 0,1299 0,1299 1
Grüne (y_2)	41 0,0433 0,1667 0,2828	55 0,0581 0,1590 0,3793	49 0,0517 0,1380 0,3379	145 0,1531 0,1531 1
SPD (y_3)	64 0,0676 0,2602 0,2162	128 0,1352 0,3699 0,4324	104 0,1098 0,2930 0,3514	296 0,3126 0,3126 1
CDU/CSU (y_4)	127 0,1341 0,5163 0,3316	145 0,1531 0,4191 0,3786	111 0,1172 0,3127 0,2898	383 0,4044 0,4044 1
Gesamt	246 0,2598 1 0,2598	346 0,3654 1 0,3654	355 0,3749 1 0,3749	947 1 1 1

Meine Notizen:

Beispiel: Erwartete Werte berechnen

Original-daten

	katholisch (x_1)	Lutherisch (x_2)	Ohne Konfession (x_3)	Gesamt
Linke (y_1)	14	18	91	123
Grüne (y_2)	41	55	49	145
SPD (y_3)	64	128	104	296
CDU/CSU (y_4)	127	145	111	383
Gesamt	246	346	355	947

Erwartungswerte nach Randverteilung:

$$e_{ij} = \frac{n_{i+} * n_{+j}}{n}$$

	katholisch (x_1)	Lutherisch (x_2)	Ohne Konfession (x_3)	Gesamt
Linke (y_1)	31,95	44,94	46,11	123
Grüne (y_2)	37,67	52,98	54,36	145
SPD (y_3)	76,89	108,15	110,96	296
CDU/CSU (y_4)	99,49	139,93	143,57	383
Gesamt	246	346	355	947

Meine Notizen:

Beispiel: standardisierte Residuen berechnen

Original-daten

	katholisch (x_1)	Lutherisch (x_2)	Ohne Konfession (x_3)	Gesamt
Linke (y_1)	14	18	91	123
Grüne (y_2)	41	55	49	145
SPD (y_3)	64	128	104	296
CDU/CSU (y_4)	127	145	111	383
Gesamt	246	346	355	947

	katholisch (x_1)	Lutherisch (x_2)	Ohne Konfession (x_3)	Gesamt
Linke (y_1)	-3,176	-4,019	6,611	0
Grüne (y_2)	0,543	0,278	-0,726	0
SPD (y_3)	-1,470	1,909	-0,661	0
CDU/CSU (y_4)	2,758	0,428	-2,719	0
Gesamt	0	0	0	0

Standardisierte Residuen:

$$r_{ij,std.} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}}$$

Standardisierte Residuen sind bei Unabhängigkeit von Zeilen- u. Spaltenvariable asymptotisch normalverteilt:

$\Rightarrow |r_{i,j,std}| > 2 \rightarrow$ signifikantes Resid. $\Rightarrow |r_{i,j,std}| > 3 \rightarrow$ höchstsignif. Resid.

Meine Notizen:

Beispiel: standardisierte Residuen: Interpretation

Standardis. Residuen

	katholisch (x_1)	Lutherisch (x_2)	Ohne Konfession (x_3)	Gesamt
Linke (y_1)	-3,176	-4,019	6,611	0
Grüne (y_2)	0,543	0,278	-0,726	0
SPD (y_3)	-1,470	1,909	-0,661	0
CDU/CSU (y_4)	2,758	0,428	-2,719	0
Gesamt	0	0	0	0

- Bei symmetrischer Beziehung → Interpretation standardisierter Residuen:

1. Die Tabelle weist drei größere negative und zwei größere positive standardisierte Residuen auf.
2. Bei Grünen und der SPD weichen die Konfessionen kaum von der Konfessionsverteilung insgesamt ab. Am relativ stärksten ist hier das annähernd (aber nicht ganz) überzufällige Zusammentreffen von SPD-Wahl und lutherischer Konfession.
3. Überzufällig häufig treten die Kombinationen Wahl der Linken und konfessionslos sowie Wahl der CDU/CSU und katholisch auf.
4. Überzufällig selten tritt dagegen Wahl der Linken und lutherisch oder katholisch und Wahl der CDU/CSU und konfessionslos auf.

Insgesamt ist zumindest bei Linkspartei und CDU/CSU eine deutliche Assoziation zu Konfessionsgruppen festzustellen!

Meine Notizen:

Zusammenfassendes symmetrisches Zusammenhangsmaß

- Da die standardisierten Residuen positiv und negativ sein können, ist die einfache Summe dieser nicht zielführend.
- Lösung: Es werden die Quadrate der standardisierten Residuen aufsummiert.
- Diese Summe ist abhängig von der Größe der Tabelle und der Gesamtfallzahl n.
- Der Statistiker Cramér entwickelte daher folgende Formel:

$$Cram\acute{e}rs\ V = \sqrt{\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij}-\frac{n_{i+}*n_{+j}}{n}\right)^2}{\frac{n_{i+}*n_{+j}}{n}}}{n*\min(I-1;J-1)}}$$

Meine Notizen:

Cramérs V am Beispiel

	katholisch (x_1)	Lutherisch (x_2)	Ohne Konfession (x_3)	Gesamt
Linke (y_1)	14	18	91	123
Grüne (y_2)	41	55	49	145
SPD (y_3)	64	128	104	296
CDU/CSU (y_4)	127	145	111	383
Gesamt	246	346	355	947

$$Cramérs\ V = \sqrt{\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij}-\frac{n_{i+}*n_{+j}}{n}\right)^2}{\frac{n_{i+}*n_{+j}}{n}}}{n*\min(I-1;J-1)}} = \sqrt{\frac{\sum_{i=1}^{4}\sum_{j=1}^{3}\frac{\left(n_{ij}-\frac{n_{i+}*n_{+j}}{947}\right)^2}{\frac{n_{i+}*n_{+j}}{947}}}{947*\min(4-1;3-1)}} = \sqrt{\frac{92{,}26}{947*2}}$$

$$= 0{,}2207$$

Interpretation:

V = 0	kein Zusammenhang
0 < V < 0,05	vernachlässigbarer Zusammenhang
0,05 ≤ V < 0,25	geringer Zusammenhang
0,25 ≤ V < 0,5	mittlerer Zusammenhang
0,50 ≤ V <1	starker Zusammenhang
V = 1	perfekter Zusammenhang

Meine Notizen:

Symmetrische Zusammenhänge Nominaler Variablen testen

Meine Notizen:

Symmetrische Zusammenhänge testen

- Für den Test auf einen symmetrischen Zusammenhang zweier Nominalskalierter Variablen kann die bekannte Logik von Chiquadrat genutzt werden:

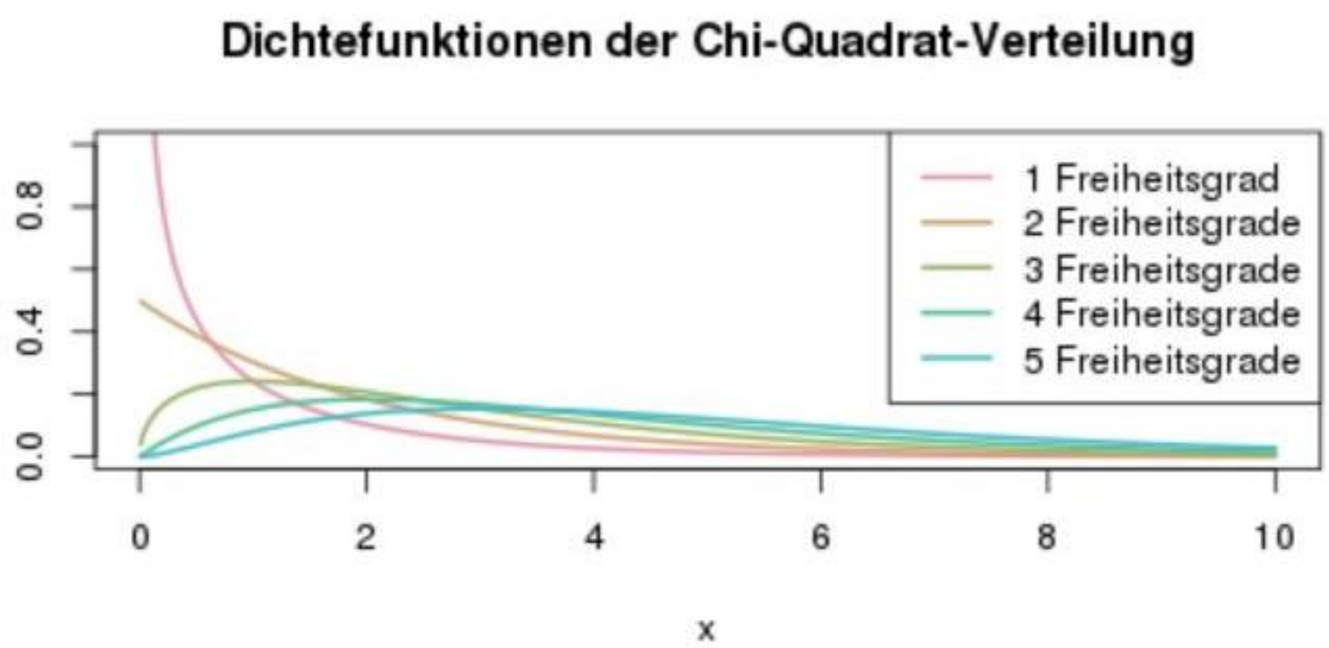
$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij}-e_{ij}\right)^2}{e_{ij}} = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij}-\frac{n_{i+}*n_{+j}}{n}\right)^2}{\frac{n_{i+}*n_{+j}}{n}} = V^2*n*\min(I-1;J-1)$$

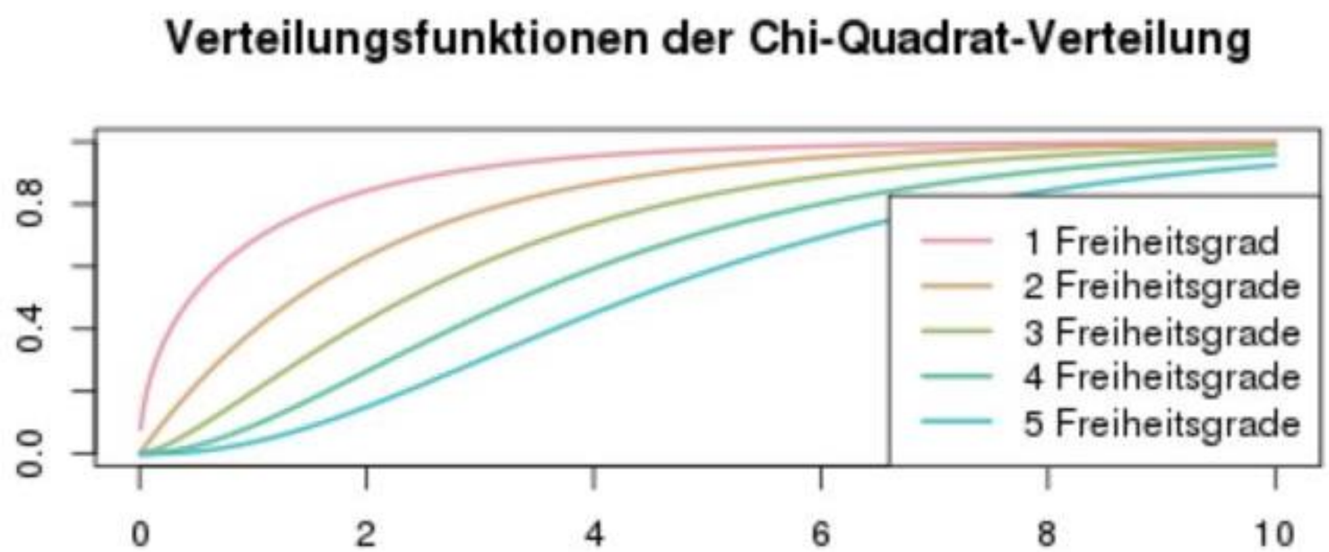
- Die Freiheitsgrade des Tests berechnen sich nach

$$df = (I-1)*(J-1)$$

Meine Notizen:

Chiquadrat-Verteilungen





df/p	0,10	0,05	0,025	0,01	0,005
1	2,71	3,84	5,02	6,64	7,88
2	4,61	5,99	7,38	9,21	10,60
3	6,25	7,82	9,35	11,35	12,84
4	7,78	9,49	11,14	13,28	14,86
5	9,24	11,07	12,83	15,09	16,75
6	10,65	12,59	14,45	16,81	18,55
7	12,02	14,07	16,01	18,48	20,28
8	13,36	15,51	17,54	20,09	21,96
9	14,68	16,92	19,02	21,67	23,59
10	15,99	18,31	20,48	23,21	25,19
11	17,28	19,68	21,92	24,73	26,76
12	18,55	21,03	23,34	26,22	28,30
13	19,81	22,36	24,74	27,69	29,819
14	21,06	23,69	26,12	29,14	31,32
15	22,31	25,00	27,49	30,58	32,80
16	23,54	26,30	28,85	32,00	34,27
17	24,77	27,59	30,19	33,41	35,72
18	25,99	28,87	31,53	34,81	37,16
19	27,20	30,14	32,85	36,19	38,58
20	28,41	31,41	34,17	37,57	40,00

Meine Notizen:

Hypothesentest: Schritt 1: Hypothesen formulieren

- Forschungshypothese:

  Es besteht ein Zusammenhang zwischen der Religionsmitgliedschaft und der Wahlpräferenz.

- Gegenhypothese:

  Es besteht kein Zusammenhang zwischen der Religionsmitgliedschaft und der Wahlpräferenz.

- Mathematisierung:

$H_1\colon \pi_{ij} \neq \pi_{i+} * \pi_{+j}$

für mindestens ein i und j

$H_0\colon \pi_{ij} = \pi_{i+} * \pi_{+j}$

für alle i und j

Meine Notizen:

Hypothesentest: Schritt 2: Teststatistik berechnen

- Die Teststatistik entspricht der Berechnung von Chiquadrat:

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - e_{ij}\right)^2}{e_{ij}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - \frac{n_{i+} * n_{+j}}{n}\right)^2}{\frac{n_{i+} * n_{+j}}{n}} = V^2 * n * \min(I-1; J-1)$$

$$= 0{,}221^2 * 947 * 2 = 92{,}26$$

Meine Notizen:

Schritt 3: Kritischen Wert bestimmen

(1) Irrtumswahrscheinlichkeit festlegen:	$\alpha = 0{,}01$
(2) Geeignete Verteilung wählen:	Chiquadrat-Verteilung
(3) Freiheitsgrade festlegen:	$df = (I-1)*(J-1) = 3*2 = 6$
(4) Kritischen Wert aus der Tabelle ablesen.	$Chi^2_{0,99;6} = 16{,}81$

Meine Notizen:

Hypothesentest: Schritt 4: Entscheidung treffen

- Die Teststatistik muss rechts des kritischen Wertes liegen, um die H_0 ablehnen zu können.
- Im Beispiel ist die Teststatistik rechts des kritischen Wertes, da 16,81 < 92,26 .
- Die Nullhypothese kann damit verworfen werden. Die Stichprobe ist auf einem 99%-Level geeignet zu zeigen, dass ein Zusammenhang zwischen Wahlabsicht und Religionsmitgliedschaft besteht.

Meine Notizen:

Hypothesentest: Schritt 5: Anwendungsvoraussetzungen

- Die Anwendungsvoraussetzungen sind erfüllt, wenn

1. $e_{i,j} > 5$ für alle Tabellenzellen bei df=1
2. $e_{i,j} > 1$ für alle Tabellenzellen und $e_{i,j} > 5$ für mind. 80% aller Zellen bei df > 1

Diese Anwendungsvoraussetzungen sind hier erfüllt.

Meine Notizen:

Bivariate Zusammenhänge nach Skalenniveaus

Symmetrisch

	Dichotome	Nominalskala	Ordinalskala	Metrische Skala
Metrische Skala				Korrelation
Ordinalskala				
Nominalskala	Cramers V	Cramers V		
Dichotome V.	Phi, α			

Asym-metrisch

AV \ UV	Dichotome	Nominalskala	Ordinalskala	Metrische Skala
Metrische Skala	Mittelwert-vergleich, Regression			Regression
Ordinalskala				
Nominalskala	jetzt	jetzt		
Dichotome V.	Prozentsatz-differenz, α	jetzt		

- Asymmetrische Zusammenhänge gehen von einer Wirkung einer unabhängigen Variablen (UV) auf eine abhängige Variable (AV) aus.
- Symmetrische Zusammenhänge sind geeignet, wenn keine (einseitige) kausale Wirkung unterstellt werden soll oder Wirkung in beide Richtungen angenommen werden muss.

Meine Notizen:

Asymmetrische Zusammenhänge Nominaler Variablen

Meine Notizen:

Beispiel

	katholisch (x_1)	Lutherisch (x_2)	Ohne Konfession (x_3)	Gesamt
Linke (y_1)	14 0,0148 0,0569 0,1138	18 0,0190 0,0520 0,1463	91 0,0961 0,2563 0,7398	123 0,1299 0,1299 1
Grüne (y_2)	41 0,0433 0,1667 0,2828	55 0,0581 0,1590 0,3793	49 0,0517 0,1380 0,3379	145 0,1531 0,1531 1
SPD (y_3)	64 0,0676 0,2602 0,2162	128 0,1352 0,3699 0,4324	104 0,1098 0,2930 0,3514	296 0,3126 0,3126 1
CDU/CSU (y_4)	127 0,1341 0,5163 0,3316	145 0,1531 0,4191 0,3786	111 0,1172 0,3127 0,2898	383 0,4044 0,4044 1
Gesamt	246 0,2598 1 0,2598	346 0,3654 1 0,3654	355 0,3749 1 0,3749	947 1 1 1

Meine Notizen:

Interpretation der Spaltenprozentuierung

	katholisch (x_1)	Lutherisch (x_2)	Ohne Konfession (x_3)	Gesamt
Linke (y_1)	0,0569	0,0520	0,2563	0,1299
Grüne (y_2)	0,1667	0,1590	0,1380	0,1531
SPD (y_3)	0,2602	0,3699	0,2930	0,3126
CDU/CSU (y_4)	0,5163	0,4191	0,3127	0,4044
Gesamt	1	1	1	1

1. Zeile: Während der Anteil der Wähler der Linken bei Katholiken und Lutheranern bei 5.7% bzw. 5.2% liegt, beträgt der Anteil bei Konfessionslosen 25.6%.

2. Zeile: Bei Katholiken ist der Anteil der Wähler der Grünen mit 16.7% geringfügig größer als bei den Lutheranern mit 15.9% und Konfessionslosen mit 13.8%.

3. Zeile: Bei Katholiken und Konfessionslosen ist der Anteil der SPD-Wähler mit 26.0% und 29.3% geringer als bei Lutheranern mit 37.0%.

4. Zeile: Bei Katholiken ist der Anteil der CDU-Wähler mit 51.6% deutlich größer bei Lutheranern mit 41.9% oder bei Konfessionslosen mit nur 31.3%.

Meine Notizen:

Interpretation der Zeilenprozentuierung

	katholisch (x_1)	Lutherisch (x_2)	Ohne Konfession (x_3)	Gesamt
Linke (y_1)	0,1138	0,1463	0,7398	1
Grüne (y_2)	0,2828	0,3793	0,3379	1
SPD (y_3)	0,2162	0,4324	0,3514	1
CDU/CSU (y_4)	0,3316	0,3786	0,2898	1
Gesamt	0,2598	0,3654	0,3749	1

1. Spalte: Unter Wählen der Linken sind 11.4% katholisch, unter denen der SPD 21.6%, unter denen der Grünen 28.3% und unter denen der CDU 33.2%.

2. Spalte: Unter den Wählern der Linken sind 14.6% lutherisch, unter denen der CDU 37.9%, unter den Wählern der Grünen ebenfalls 37.9% und unter Wählern der SPD 43.2%.

3. Spalte: Unter den Wählern der Linken sind 74.0% konfessionslos, unter denen der SPD 35.1%, unter denen der Grünen 33.8% und unter den Wählern der CDU 29.0%.

Meine Notizen:

Weitere asymmetrische Interpretationen

- Eine durchaus verwendete Möglichkeit der Interpretation größerer Tabellen ist es, sich viele einzelne Vierfeldertafeln herauszuschneiden und für diese dann die bekannten Maße zu berechnen.
- Dieses führt aber im Zweifel zu einer großen Zahl von Zusammenhangsmaßen, die schwer überschaubar sind.
- Die Methode der Odds-Ratios kann aber – wie schon die Zeilen- und Spaltenprozentuierung – direkt auf größere Tabellen übertragen werden.
- Die Odds müssen dann ebenfalls zeilen- oder spaltenweise gebildet werden. Sie sind daher asymmetrische Parameter.

Meine Notizen:

Beispiel: Multiplikative Sicht mit unabh. Spalten

Berichtetes Wahlverhalten	Konfession (X) kathol.	luther.	ohne K.	Summe
Linke	14	18	91	123
Grüne	41	55	49	145
SPD	64	128	104	296
CDU/CSU	127	145	111	383
Summe	246	346	355	947

Daten: Allbus 2014: ISSP

Odds der Parteien nach Konfession

Berichtetes Wahlverhalten	Konfession (X) kathol.	Luther.	Ohne K.	Summe
Linke	$\boxed{16.6^{-1}}$	18.2^{-1}	2.9^{-1}	6.7^{-1}
Grüne	5.0^{-1}	5.3^{-1}	6.2^{-1}	5.5^{-1}
SPD	2.8^{-1}	1.7^{-1}	2.4^{-1}	2.2^{-1}
CDU/CSU	1.1	1.4^{-1}	2.2^{-1}	1.5^{-1}
Summe	(246)	(346)	(355)	(947)

Daten: Allbus 2014: ISSP

Beispiel: $14/(246-14) = 0.0603 = 1/16.6 = 16.6^{-1}$

Interpretationsbeispiele:

Das Odds der Wähler der Linken zu Wählern anderer Parteien beträgt insgesamt 1 zu 6.7, bei Katholiken 1 zu 16.6 bei Lutheranern 1 zu 18.2 und bei Konfessionslosen 1 zu 2.9.

Das Odds der CDU-Wähler beträgt 1 zu 1.5, unter Katholiken 1.1 zu 1, unter Lutheranern 1 zu 1.4 und unter Konfessionsl. 1 zu 2.2.

Meine Notizen:

Asymmetrische Zusammenhänge

- Alle bisherigen Parameter beschreiben Zusammenhänge in einem asymmetrischen Sinne, sind aber nicht in der Lage, die asymmetrische Wirkung einer nominalskalierten Variablen auf eine Andere in einem einzigen Wert zu messen.
- Ihre Interpretation ist daher aufwendig, unübersichtlich und schwer darzustellen.
- Um eine Messung in einem Parameter zu erhalten, erinnern wir uns an die Logik der PRE-Maße, die wir für den Determinationskoeffizienten R^2 kennengelernt haben.

Meine Notizen:

PRE-Maße

- PRE steht für Proportional Reduction of Errors.
- Leitend ist die Frage: Um welchen Anteil kann ich meinen Vorhersagefehler in der Prognose von Y reduzieren, wenn ich die Information aus X hinzuziehe?
- Anders ausgedrückt: Der Prognosefehler in der Randverteilung (E_Y) lässt sich spalten in einen Prognosefehler in der bedingten Verteilungen ($E_{Y|X}$) und einen durch X aufgeklärten Teil.
- PRE-Maße berechnen sich allgemein:

$$PRE = \frac{E_Y - E_{Y|X}}{E_Y} = 1 - \frac{E_{Y|X}}{E_Y}$$

Meine Notizen:

PRE-Maß am Beispiel

Berichtetes Wahlverhalten	Konfession (X) kathol.	luther.	ohne K.	Summe
Linke	14	18	91	123
Grüne	41	55	49	145
SPD	64	128	104	296
CDU/CSU	127	145	111	383
Summe	246	346	355	947

Daten: Allbus 2014: ISSP

Abhängig: Wahlverhalten

E_Y: 947 – 383 = 564

$E_{Y|X}$: 246 – 127 + 346 – 145 + 355 – 111
= 947 – 127 –145 – 111 = 564

$\Rightarrow \lambda_{YX} = 1 - 564/564 = 0$

Abhängig: Konfession

E_X: 947 – 355 = 592

$E_{X|Y}$: 947 – 91 – 55 – 128 – 145 = 528

$\Rightarrow \lambda_{XY} = 1 - 528/592 = 0.1081$

$\lambda_{YX} = 0 \rightarrow$ keine Prognoseverbesserung der Parteienwahl bei Kenntnis der Konfession;

λ_{XY} = 10.81% Prognoseverbesserung der Konfession bei Kenntnis der Parteienwahl.

Meine Notizen:

Das PRE-Maß Lambda

- Einfaches PRE-Maß für nominalskalierte Variablen: Lambda
- Der Modus ist der beste Vorhersagewert auf Nominalskalenniveau.
- Abweichungen vom Modus sind Prognosefehler E_Y
- bei multimodalen Verteilungen kann ein beliebiger Modalwert verwendet werden

→ bei Zeilenvariable abhängig: $E_Y = n - \max(n_{1+}; n_{2+}; \ldots; n_{I+})$

→ bei Spaltenvariable abhängig: $E_X = n - \max(n_{+1}; n_{+2}; \ldots; n_{+J})$

$$\lambda_{YX} = 1 - \frac{n - \sum_{j=1}^{J} \max(n_{1j}; \ldots; n_{Ij})}{n - \max(n_{1+}; \ldots; n_{I+})}$$

$$\lambda_{XY} = 1 - \frac{n - \sum_{i=1}^{I} \max(n_{i1}; \ldots; n_{iJ})}{n - \max(n_{+1}; \ldots; n_{+J})}$$

Meine Notizen:

Streuung als Fehlermaß

- Fehler als Abweichungen vom Modalwert sind oft ungünstig, da sie informationsarm sind. Wir haben das am Beispiel gesehen.
- Alternativ lässt sich die Streuung einer Variablen als Vorhersagefehler betrachten:

$$PRE = 1 - \frac{Summe\ bedingte\ Streuungen}{unbedingte\ Streuung}$$

- Das gängige Streuungsmaß nominalskalierter Variablen ist die Devianz

$$D_Y = -2 * \sum_{i=1}^{I} n_{i+} * \ln\left(\frac{n_{i+}}{n}\right)$$

$$D_{Y|X} = -2 * \sum_{j=1}^{J} \sum_{i=1}^{I} n_{ij} * \ln\left(\frac{n_{ij}}{n_{+j}}\right)$$

Meine Notizen:

Das Zusammenhangsmaß McFaddens Pseudo R²

$$PRE = 1 - \frac{Summe\ bedingte\ Streuungen}{unbedingte\ Streuung}$$

$$McFaddens\ Pseudo\ R^2 = R'_{YX} = 1 - \frac{-2 * \sum_{j=1}^{J} \sum_{i=1}^{I} n_{ij} * \ln\left(\frac{n_{ij}}{n_{+j}}\right)}{-2 * \sum_{i=1}^{I} n_{i+} * \ln\left(\frac{n_{i+}}{n}\right)}$$

Meine Notizen:

Devianzanteil berechnen:

$$-2 * n_{i,j} * ln\left(\frac{n_{i,j}}{n_{+,j}}\right)$$

McFaddens Pseudo R² am Beispiel

	katholisch	lutherisch	ohne Konf.	Summe
Linke	14	18	91	123
Grüne	41	55	49	145
SPD	64	128	104	296
CDU/CSU	127	145	111	383
Summe	246	346	355	947
Allbus 2014				

	katholisch	lutherisch	ohne Konf.	Summe
Linke	80,26	106,42	247,75	502,11
Grüne	146,92	202,30	194,07	544,20
SPD	172,35	254,57	255,37	688,46
CDU/CSU	167,93	252,21	258,09	693,43
Summe	567,46	815,50	955,28	2428,21

1. Für jede Zelle der Tabelle wird die bedingte Devianz berechnet (siehe oben).
2. Aufsummieren der bedingten und unbedingten Devianzen:
    1. $D_{Y|X}$ = 567,46 + 815,5 + 955,28 = 2338,24
    2. D_Y = 2428,21
3. Berechnung von McFaddens Pseudo R^2:

$$R'_{YX} = 1 - \frac{2338,24}{2428,21} = 0{,}0371 = 3{,}7\%$$

Meine Notizen:

Interpretation von McFaddens Pseudo R^2

- Bei Kenntnis der Konfession reduziert sich die Devianz des Wahlverhaltens um 3.7%.
- Faustregeln:

• $R' = 0$	kein Effekt
• $0 < R' < 0{,}0025$	vernachlässigbarer Effekt
• $0{,}0025 \leq R' < 0{,}0625$	geringer Effekt
• $0{,}0625 \leq R' < 0{,}25$	mittlerer Effekt
• $0{,}25 \leq R' < 1$	starker Effekt
• $R' = 1$	perfekter Effekt

- Im Beispiel besteht eine geringe Wirkung der Konfession auf das Wahlverhalten.

Meine Notizen:

Asymmetrische Zusammenhänge Nominaler Variablen testen

Meine Notizen:

Asymmetrische Zusammenhänge testen

- Für den Test auf einen asymmetrischen Zusammenhang zweier Nominalskalierter Variablen können die berechneten (bedingten) Devianzen genutzt werden:

$$L^2 = D_Y - D_{Y|X}$$

- Die Freiheitsgrade des Tests berechnen sich nach

$$df = (I - 1) * (J - 1)$$

Meine Notizen:

Hypothesentest: Schritt 1: Hypothesen formulieren

- Forschungshypothese:

  Es besteht ein Zusammenhang zwischen der Religionsmitgliedschaft und der Wahlpräferenz.

- Gegenhypothese:

  Es besteht kein Zusammenhang zwischen der Religionsmitgliedschaft und der Wahlpräferenz.

- Mathematisierung:

$$H_1\colon \pi_{Y=i|X=j} \neq \pi_{Y=i} \qquad H_0\colon \pi_{Y=i|X=j} = \pi_{Y=i}$$

$$f\ddot{u}r\ alle\ i, j$$

Meine Notizen:

Hypothesentest: Schritt 2: Teststatistik berechnen

- Die Teststatistik entspricht der der Differenz von unbedingter und bedingter Devianz:

$$L^2 = D_Y - D_{Y|X} = 2428{,}21 - 2338{,}24 = 89{,}97$$

Meine Notizen:

Schritt 3: Kritischen Wert bestimmen

(1) Irrtumswahrscheinlichkeit festlegen: $\alpha = 0{,}01$

(2) Geeignete Verteilung wählen: Chiquadrat-Verteilung

(3) Freiheitsgrade festlegen: $df = (I-1)*(J-1) = 3*2 = 6$

(4) Kritischen Wert aus der Tabelle ablesen. $Chi^2_{0,99;6} = 16{,}81$

Meine Notizen:

Hypothesentest: Schritt 4: Entscheidung treffen

- Die Teststatistik muss rechts des kritischen Wertes liegen, um die H_0 ablehnen zu können.
- Im Beispiel ist die Teststatistik rechts des kritischen Wertes, da $16{,}81 < 89{,}97$.
- Die Nullhypothese kann damit verworfen werden. Die Stichprobe ist auf einem 99%-Level geeignet zu zeigen, dass eine Wirkung der Konfession auf die Wahlabsicht besteht.

Meine Notizen:

Hypothesentest: Schritt 5: Anwendungsvoraussetzungen

- Die Anwendungsvoraussetzungen sind erfüllt, wenn

1. $e_{i,j} > 5$ für alle Tabellenzellen bei df=1
2. $e_{i,j} > 1$ für alle Tabellenzellen und $e_{i,j} > 5$ für mind. 80% aller Zellen bei df > 1

Diese Anwendungsvoraussetzungen sind hier erfüllt.

Meine Notizen:

Was Sie am Ende der Woche können sollten

- **Kern**: Sie analysieren zwei nominale Variablen auf Zusammenhänge und interpretieren diese für Stichprobe und Grundgesamtheit.
- Sie bestimmen erwartete Werte und Residuen in Kreuztabellen.
- Sie bestimmen und interpretieren standardisierte Residuen und Cramérs V.
- Sie testen zwei nominalskalierte Variablen auf symmetrische Zusammenhänge.
- Sie bestimmen und interpretieren Odds für polytome Kreuztabellen.
- Sie berechnen und interpretieren Lambda und legen dessen Schwächen dar.
- Sie berechnen und interpretieren R´.
- Sie testen, ob ein asymmetrischer Zusammenhang vorliegt.

Meine Notizen: