Herzlich willkommen zur Vorlesung Statistik

Zusammenhänge zwischen nominalen (und/oder ordinalen) Merkmalen: **Kreuztabellenanalyse und Assoziationsmaße II: Signifikanztests und Maße der Assoziation**

Der Chi-Quadrat-Test nach Pearson (am Beispiel 2x2-Tabelle)

Problem: Können wir mit einiger Sicherheit annehmen, dass der Unterschied in den Anteilswerten in unserer Stichprobe auch in der Grundgesamtheit besteht?

Schritt 1: Formulierung der Hypothesen.

Nullhypothese: Es besteht kein Unterschied zwischen den Anteilswerten einer Zeile (bzw. – bei Zeilenprozentuierung – einer Spalte).

Alternativhypothese: Es besteht (irgend)ein Unterschied.

Schritt 1 des Chi-Quadrat-Tests

Werte, die bei Gültigkeit der Nullhypothese (Unabhängigkeit der Merkmale) zu erwarten wären:

(Spaltenprozent)	Arbeiter	Angestellter	gesamt
Hauptschule	252	378	630
	63%	63%	63%
RS/Gymnasium	148	222	370
	37%	37%	37%
gesamt	400	600	1000

Schritt 1 des Chi-Quadrat-Tests

Allgemeine Formulierung der Nullhypothese (hier bezogen auf Spaltenprozentuierung):

$$\pi_{1|X=1} = \pi_{1|X=2} = \pi_{1\bullet} \quad \text{bzw.} \quad \pi_{2|X=1} = \pi_{2|X=2} = \pi_{2\bullet}$$

Alternativhypothese:

$$\pi_{1|X=1} \neq \pi_{1|X=2} \neq \pi_{1\bullet} \quad \text{bzw.} \quad \pi_{2|X=1} \neq \pi_{2|X=2} \neq \pi_{2\bullet}$$

Schritt 2 des Chi-Quadrat-Tests: Prüfgröße und Teststatistik

Die Prüfgröße sind die (bedingten) Anteilswerte.

Eine geeignete Teststatistik bezieht sich auf den Vergleich der absoluten Häufigkeiten, die unter der Nullhypothese zu erwarten wären, mit den beobachteten Häufigkeiten.

Die Statistik

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij}-e_{ij}\right)^2}{e_{ij}}$$

e_{ij}: die unter H_0 erwarteten Häufigkeiten
I, J: Zahl der Ausprägungen der Variablen (hier: I = J = 2)

folgt einer Chi-Quadrat-Verteilung mit (I-1)•(J-1) Freiheitsgraden.

Freiheitsgrade

Bei gegebenen Randhäufigkeiten liegen alle Werte in den Zellen fest, sobald eine Zellhäufigkeit festliegt.

Es besteht daher 1 Freiheitsgrad für die Teststatistik Chi-Quadrat.

	Arbeiter	Angestellter	gesamt
Hauptschule	360	+270=	630
RS/Gymnasium	+40=	+330=	370
gesamt	400	600	1000

Schritt 2 des Chi-Quadrat-Tests: Prüfgröße und Teststatistik

Einfache Berechnung der erwarteten Werte aus der Randverteilung:

$$e_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}, \text{ z. B. } e_{21} = \frac{n_{2\bullet} \cdot n_{\bullet 1}}{n} = \frac{370 \cdot 400}{1000} = 148$$

	Arbeiter	Angestellter	gesamt
Hauptschule	252	378	630
	63%	63%	63%
RS/Gymnasium	**148**	222	370
	37%	37%	37%
gesamt	400	600	1000

Schritt 2 des Chi-Quadrat-Tests: Prüfgröße und Teststatistik, hier: Anwendungsvoraussetzung

- Der Chi-Quadrat-Test ist nur gültig, wenn gilt: $e_{ij} > 5$ für alle (oder: die meisten) e_{ij}.
- Außerdem soll evtl. bei $n < 60$ die sog. „Kontinuitätskorrektur" nach Yates verwendet werden (strittig). Bei sehr kleinen Fallzahlen ($n<30$) muss zu „exakten" Testverfahren gegriffen werden.
- In beiden Hinsichten ergeben sich keine Probleme → BAU (Business As Usual).

Schritt 3 des Chi-Quadrat-Tests: Irrtumswahrscheinlichkeit und Ablehnungsbereich

Üblicherweise wählen wir eine Irrtumswahrscheinlichkeit von α=0,05. Wäre ein „Irrtum“ (d.h. Ablehnung der Nullhypothese, obwohl sie zutrifft), besonders gravierend, wäre auch an α=0,01 oder gar α=0,001 zu denken. Das Risiko hierbei: Die Nullhypothese beizubehalten, obwohl sie falsch ist (Fehler 2. Art).

Der „kritische Wert“ der Chi-Quadratverteilung (hier: 1–0,05=0,95, 1 Freiheitsgrad) liegt bei 3,841. Erreicht oder übertrifft die Teststatistik diesen Wert, wird die Nullhypothese verworfen.

Schritt 4 des Chi-Quadrat-Tests: Berechnung der Teststatistik und Entscheidung über H_0

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

$$\chi^2 = \frac{(360-252)^2}{252} + \frac{(40-148)^2}{148} +$$

$$\frac{(270-378)^2}{378} + \frac{(330-222)^2}{222} =$$

$$46{,}29 + 78{,}81 + 30{,}86 + 52{,}54 = 208{,}5$$

208,5 > 3,841: H_0 wird verworfen (mit α=0,05).

Probleme des Chi-Quadrat-Tests

- Wie bei allen Signifikanztests gilt auch hier: Der Stichprobenumfang entscheidet. Bei n= 1000 werden schon relativ kleine Unterschiede signifikant.
- Daraus folgt: Signifikanz sagt nichts über die Stärke des Zusammenhanges.
- Der Chi-Quadrat-Test ist ein „Omnibus-Test“: Er sagt nur, dass irgendwelche empirischen Werte von den erwarteten Werten abweichen. Er sagt aber nichts über die Richtung der Abweichung.

Größere Kreuztabellen

Beispiel: Zusammenhang zwischen Zahl der Wohnräume pro Person und Beurteilung der Wohnungsgröße (kursiv: Spaltenprozent; absolute Zahlen nur zum Nachrechnen)

Wohnräume p. P. → Beurteilung ↓	<1		1 bis <1,5		1,5 +		Gesamt	
zu klein	70	*56*	77	*25*	14	*5*	161	*23*
richtig	49	*39*	224	*73*	217	*82*	490	*70*
zu groß	7	*6*	7	2	35	*13*	49	*7*
Gesamt	126	*100*	308	*100*	266	*100*	700	*100*

Größere Kreuztabellen

Chi-Quadrat für vorstehende Tabelle: 142,4 (4 Freiheitsgrade); Wert liegt im Ablehnungsbereich
→ H_0 (kein Zusammenhang) wird verworfen.

Aber Achtung: Chi-Quadrat-Test ist gleichgültig gegenüber Richtung des Zusammenhanges – folgende Tabelle führt zu gleichem Wert!

Wohnräume p. P. → Beurteilung ↓	<1		1 bis <1,5		1,5 +		Gesamt	
zu klein	70	*56*	14	*5*	77	*25*	161	*23*
richtig	49	*39*	217	*82*	224	*73*	490	*70*
zu groß	7	*6*	35	*13*	7	*2*	49	*7*
Gesamt	126	*100*	266	*100*	308	*100*	700	*100*

Assoziationsmaße

Zusammenhänge in größeren Kreuztabellen lassen sich im Prinzip durch eine Vielzahl von Prozentsatzdifferenzen/Relativen Risiken/ Odds Ratios ausdrücken. Diese sind dann aber kaum mehr zusammenhängend zu interpretieren.

Assoziationsmaße sind Versuche, die Stärke des Zusammenhanges in einer einzigen Maßzahl auszudrücken. Solche Maßzahlen gibt es auch für Vier-Felder-Tabellen.

Die Vielzahl entsprechender Maßzahlen verbietet eine ausführliche Diskussion. Die folgende Übersicht gibt nur einige Hinweise.

Assoziationsmaße

Die wichtigste Unterscheidung betrifft das Messniveau (Skalenniveau): Maße für nominalskalierte Variablen (oder eine nominal- und eine ordinalskalierte Variable) sind von Maßen für zwei ordinalskalierte Variable zu unterscheiden.

Bei den erstgenannten unterscheiden wir zwischen Chi^2-basierten und sonstigen Maßen.

Assoziationsmaße für nominalskalierte Merkmale: Chi-Quadrat-basierte Maße

- In der Maßzahl Chi² drückt sich auch die Stärke des Zusammenhangs aus. Sie wird jedoch zusätzlich ganz wesentlich von der Fallzahl beeinflusst.
- Chi²-basierte Maße korrigieren daher den Chi²-Wert um die Fallzahl.
- Sie nehmen Werte zwischen 0 (kein Zusammenhang) und 1 (perfekter Zusammenhang) an (gilt nicht für C).

Chi-Quadrat-basierte Maße

φ (phi)	$\sqrt{\frac{\chi^2}{n}}$	Nur für 2x2-Tabellen geeignet; kann bei alternativer Berechnung (K&K, S. 336) auch Werte bis –1 annehmen.
V	$\sqrt{\frac{\chi^2}{n \cdot (\min(I,J)-1)}}$	Voller Name: Cramérs V I: Zahl der Werte von X J: Zahl der Werte von Y
C	$\sqrt{\frac{\chi^2}{\chi^2+n}}$	Voller Name: Kontingenzkoeffizient Max(C) < 1!

Weitere Assoziationsmaße für nominalskalierte Merkmale I

Yules Q (für 2x2-Tabellen): Nimmt Wert 1 (oder –1) an, wenn eine einzige Zelle die Häufigkeit Null aufweist. Beispielsweise nimmt Q in beiden folgenden Tabellen den Wert 1 an:

	Arb.	Ang.
HS	400	590
RS/Gymn.	0	10

	Arb.	Ang.
HS	400	0
RS/Gymn.	0	600

Weitere Assoziationsmaße für nominalskalierte Merkmale II

λ (lambda): Nimmt Wert 0 an, wenn die Modalwerte pro Spalte alle in der gleichen Zeile auftreten. So beträgt der Wert von Lambda in beiden folgenden Tabellen 0:

	Arb.	Ang.		Arb.	Ang.
HS	210	310		400	310
RS/Gymn.	190	290		0	290

Alternatives Maß mit ähnlicher Logik: Goodmans und Kruskals Tau (nicht mit den nachfolgenden diskutierten Tau-a, -b und -c zu verwechseln).

Fazit: Assoziationsmaße für nominalskalierte Merkmale

Ein rundherum befriedigendes Assoziationsmaß für nominalskalierte Merkmale existiert nicht. (Der Unsicherheitskoeffizient – siehe K&K – ist eine brauchbare, aber nicht sehr eingängige Alternative.) In der Praxis werden oft Phi (für 2x2-Tabellen) oder Cramérs V (für größere Tabellen) verwendet, aber nicht alle halten diese für die besten Maßzahlen (u.A. wegen fehlender inhaltlicher Interpretation und fehlender Differenzierung zwischen unabhängiger und abhängiger Variablen).

Assoziationsmaße für ordinalskalierte Merkmale

Die wichtigsten Maße sind:

- Kendalls τ (tau) (in drei Versionen: τ_a, τ_b, τ_c); am häufigsten verwendet: τ_b.
- Goodmans und Kruskals Gamma (wird gerne von Angebern verwendet, weil große Werte; leidet an ähnlichem Problem wie Yules Q).
- Somers' d (unterscheidet zwischen unabhängiger und abhängiger Variablen); sollte häufiger verwendet werden.

Signifikanztest ordinalskalierter Merkmale

Für die genannten Assoziationsmaße lassen sich Standardfehler und auf deren Grundlage stan-dardnormalverteilte Teststatistiken berechnen.
Es kann von einem signifikanten Zusammenhang ($\alpha<0{,}05$) ausgegangen werden, wenn gilt:

$$\left|\frac{Koeff.}{S.E.}\right| > 1{,}96 \text{ für } \mathrm{H}_1: \text{Koeff.} \neq 0$$

$$\frac{Koeff.}{S.E.} > 1{,}645 \text{ für } \mathrm{H}_1: \text{Koeff.} > 0$$

$$\frac{Koeff.}{S.E.} < -1{,}645 \text{ für } \mathrm{H}_1: \text{Koeff.} < 0$$

Signifikanztest ordinalskalierter Merkmale

Hat man es mit ordinalskalierten Merkmalen zu tun und nimmt man einen gerichteten Zusammenhang an (je ... desto [weniger]), ist der Signifikanztest für das verwendete Assoziations-maß dem Chi-Quadrat-Test vorzuziehen, da letzterer auf Abweichungen von den erwarteten Werten in beliebiger Richtung reagiert.

Es kann daher leicht geschehen, dass der Chi²-Test einen signifikanten Zusammenhang andeutet, obwohl der angenommene Zusammenhang keineswegs signifikant ist.

Signifikanztest ordinalskalierter Merkmale

Beispiel (fiktiv): Der Zusammenhang zwischen Salz im Mensaessen und Geschmack ist höchst signifikant (Chi^2=218, krit. Wert 12,59 [α<0,05]) – aber die Assoziationsmaße für ordinalskalierte Merkmale betragen 0 (und unterscheiden sich somit a fortiori nicht signifikant von 0).

Salz →	kein	wenig	mittel	viel	n
lecker	80	10	10	80	180
sehr lecker	10	10	10	10	40
sensationell	10	80	80	10	180
n	100	100	100	100	400

Signifikanztest ordinalskalierter Merkmale

Auch das umgekehrte ist möglich: Bei Tabellen mit vielen Zellen und nicht sehr großen Fallzahlen kann es geschehen, dass auch bei erwartungsgemäßem Zusammenhang der Chi²-Test nicht signifikant ausfällt, der Test eines geeigneten Assoziationsmaßes jedoch schon.

Wenn tatsächlich ein Zusammenhang der genannten Art (je ... desto) vermutet wurde, ist auch in dieser Situation dem Test des Assoziationsmaßes mehr Glauben zu schenken.

Signifikanztest ordinalskalierter Merkmale

Beispiel (fiktiv), 2. Version: Der Zusammenhang zwischen Salz im Mensaessen und Geschmack ist nach Chi²-Test nicht signifikant (Chi²=7,945, krit. Wert 12,59 [α<0,05]) – aber die Assoziationsmaße für ordinalskalierte Merkmale sind sämtlich positiv und signifikant von 0 verschieden.

Salz →	kein	wenig	mittel	viel	n
lecker	10	8	6	4	28
sehr lecker	8	10	6	4	28
sensationell	4	6	8	10	28
n	22	24	20	18	84