

Sommersemester 2009 Statistik mit SPSS



Methodenzentrum Sozialwissenschaften

Überblick

Mehrfeldertabellen und Zusammenhangsmaße

- 1. Mehrfeldertabellen und Zusammenhangsmaße:
nominale Daten**
- 2. Mehrfeldertabellen und Zusammenhangsmaße:
ordinale Daten**

1. Skalenniveaus

1. Nominalskalenniveau

Die Zahlen haben nur klassifikatorischen Charakter. Unterschiedliche Zahlen bedeuten unterschiedliche Merkmalausprägungen, sie stehen nicht für ein Mehr oder Weniger (z.B. Familienstand, Religion, Parteizugehörigkeit).

2. Ordinalskalenniveau

Die Zahlen haben zusätzlich eine Rangordnung (Auskunft über ein Mehr oder Weniger). Gleiche Abstände zwischen den Zahlenwerten bedeuten nicht gleiche Abstände in der Realität (z.B. Zufriedenheit, soziale Schicht)

3. Intervallskalenniveau (metrisch)

Die Zahlen geben Informationen über die Abstände zwischen den gemessenen Ausprägungen (kein echter Nullpunkt) (z.B. Geburtsjahr).

4. Ratioskalenniveau (metrisch)

Es ermöglicht, das Verhältnis zweier Werte von Eigenschaftsmessungen zu interpretieren (Nullpunkt vorhanden) (z.B. Einkommen, Alter).

1. Nominale Daten

1.a. Mehrfeldertabellen: nominale Daten

1.b. Zusammenhangsmaße: nominale Daten

1.a. Mehrfeldertabellen: nominale Daten

```
*****
Beispiel 1: Kreuztabelle, recall x konfession, BTW05
*****

weight by owweight.

***Rekodierung.
recode s03 (1=1) (2,3=2) (6=0) into konfession.
val lab konfession 1 'katholisch' 2 'evangelisch' 0 'konfessionslos'.
var lab konfession 'Konfessionszugehörigkeit'.
fre konfession recall.

cro recall by konfession
/cells col.
```

1.a. Mehrfeldertabellen: nominale Daten

recall * konfession Konfessionszugehörigkeit Kreuztabelle

% von konfession Konfessionszugehörigkeit

		konfession Konfessionszugehörigkeit			Gesamt
		,00 konfessio nslos	1,00 katholisch	2,00 evangelisch	
recall	1,00 SPD	37,4%	31,8%	46,3%	38,9%
	2,00 CDU/CSU	21,6%	44,2%	29,1%	31,9%
	3,00 Bd90/Die Gruenen	10,7%	7,3%	9,7%	9,2%
	4,00 FDP	9,3%	13,1%	9,5%	10,6%
	5,00 Die Linke.PDS	21,0%	3,5%	5,4%	9,4%
Gesamt		100,0%	100,0%	100,0%	100,0%

1.a. Mehrfeldertabellen: nominale Daten

Standardisierte Residuen und Chi-Quadrat-Test

Beispiel 2: Residuen recall x konfession

cro recall by konfession

/cells col count s|resid

/stat chi.

Standardisierte Residuen und Chi-Quadrat-Test

Die **standardisierten Residuen** werden berechnet, indem man die Differenz aus beobachteter und erwarteter Häufigkeit durch die Quadratwurzel der erwarteten Häufigkeiten dividiert.

$$sr_{ij} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}}$$

1.a. Mehrfeldertabellen: nominale Daten

Standardisierte Residuen

recall * konfession Konfessionszugehörigkeit Kreuztabelle

			konfession Konfessionszugehörigkeit			Gesamt
			,00 konfessionlos	1,00 katholisch	2,00 evangelisch	
recall	1,00 SPD	Anzahl	210	199	335	744
		% von konfession Konfessionszugehörigkeit	37,4%	31,8%	46,3%	38,9%
		Standardisierte Residuen	-6	-2,9	3,2	
	2,00 CDU/CSU	Anzahl	121	277	211	609
		% von konfession Konfessionszugehörigkeit	21,6%	44,2%	29,1%	21,8%
		Standardisierte Residuen	-4,3	5,5	-1,3	
	3,00 Bd90/Die Gruenen	Anzahl	60	46	70	176
		% von konfession Konfessionszugehörigkeit	10,7%	7,3%	9,7%	9,2%
		Standardisierte Residuen	1,2	-1,5	,4	
	4,00 FDP	Anzahl	52	82	69	203
		% von konfession Konfessionszugehörigkeit	9,3%	13,1%	9,5%	10,6%
		Standardisierte Residuen	-1,0	1,9	-,9	
	5,00 Die Linke.PDS	Anzahl	118	22	39	179
		% von konfession Konfessionszugehörigkeit	21,0%	3,5%	5,4%	9,4%
		Standardisierte Residuen	9,0	-4,8	-3,5	

Standardisierte Residuen:
 ≤ -2 bzw. $\geq +2$.

Die standardisierten Residuen deuten auf überzufällige Abweichungen von bei Unabhängigkeit erwarteten Werten hin.

1.a. Mehrfeldertabellen: nominale Daten

Chi-Quadrat-Test

- Hypothesentest, der mit Hilfe der Teststatistik χ^2 durchgeführt wird.
- Chi-Quadrat-Statistik wird in der Tabellenanalyse als Teststatistik zur Prüfung von Hypothesen über die Unabhängigkeit zweier Variablen verwendet.
- Chi-Quadrat-Statistik fasst alle Abweichungen zwischen beobachteten und erwarteten Häufigkeiten in einer IxJ-Tabelle zusammen. Je größer die Abweichungen sind, desto größer ist der Wert der Teststatistik.
- Wenn in der Population ein Zusammenhang zwischen den Variablen besteht (H_0 unzutreffend), hat die Teststatistik größere Werte als bei Gültigkeit der Nullhypothese.
- Bei einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$ ist die Nullhypothese abzulehnen, wenn der Wert der Teststatistik größer ist als der Wert des $1 - \alpha = 95\%$ -Quantils der Chi-Quadratverteilung mit $df = (I-1) \cdot (J-1)$ Freiheitsgraden

1.a. Mehrfeldertabellen: nominale Daten

Chi-Quadrat-Test

Chi-Quadrat-Statistik:
$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

n_{ij} = beobachtete Häufigkeit
 e_{ij} = erwartete Häufigkeit

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	194,934 ^a	8	,000
Kontinuitätskorrektur			
Likelihood-Quotient	181,999	8	,000
Zusammenhang linear-mit-linear	56,297	1	,000
Anzahl der gültigen Fälle	1911		

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 51,67.

1.b. Zusammenhangsmaße: nominale Daten

```
[/STATISTICS=[CHISQ] [LAMBDA] [BTAU] [GAMMA] [ETA ]]  
[PHI ] [UC ] [CTAU] [D ] [CORR]  
[CC ] [KAPPA ] [RISK] [MCNEMAR] [CMH [(value)]]  
[ALL ] [NONE]
```

Zwei Arten von Zusammenhangsmaßen für **nominale** Daten sind in SPSS verfügbar:

1. **Symmetrische Maße: Phi, Cramer's V und Kontingenzkoeffizient (basieren auf dem Chisq – Konzept)**
2. **Richtungsmaße: Lambda und Unsicherheitskoeffizient (basieren auf dem Konzept der proportionalen Fehlerreduktion)**

1.b. Zusammenhangsmaße: nominale Daten

Beispiel 2, Zusammenhangsmaße recall x konfession

```
cro recall by konfession  
  /cells col  
  /stat phi lambda cc uc.
```



**Berechnet folgende
Zusammenhangsmaße:
Phi,
Cramer's V (phi)
Lambda, Tau (lambda)
Kontingenzkoeffizient (cc)
Unsicherheitskoeffizient (uc)**

1.b. Zusammenhangsmaße: nominale Daten

Symmetrische Zusammenhangsmaße

Symmetrische Maße				
		Wert	Asymptotischer Standardfehler ^a	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,319		,000
	Cramer-V	,226		,000
	Kontingenzkoeffizient	,304		,000
Anzahl der gültigen Fälle		1911		

Interpretation von Cramérs V erfolgt analog zu Phi, d.h. Cramérs V = 0,226 = mittlerer Zusammenhang

Interpretation von Kontingenzkoeffizienten C: i.d.R. größer als Cramérs V. Maximalwert von C hängt von der Zahl der Ausprägungen der Variablen ab (der Spalten und der Zeilen) und ist kleiner als 1. Vergleiche zwischen verschiedenen Kontingenzkoeffizienten nur sinnvoll, wenn die Tabellengröße jeweils identisch ist. Aus diesem Grund ist Cramérs V vorzuziehen.

1.b. Zusammenhangsmaße: nominale Daten

Symmetrische Zusammenhangsmaße

		Symmetrische Maße		
		Wert	Asymptotischer Standardfehler ^a	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,319		,000
	Cramer-V	,226		,000
	Kontingenzkoeffizient	,304		,000
Anzahl der gültigen Fälle		1911		

Bei einer Irrtumswahrscheinlichkeit von $\leq 0,1\%$ ($p \leq 0,001$) wird die Nullhypothese abgelehnt.

1.b. Zusammenhangsmaße: nominale Daten

Asymmetrische Zusammenhangsmaße (Richtungsmaße)

- Möchte man die Realisation einer abhängigen Variable voraussagen, so kann man dabei Fehler machen.
- Die Anzahl der Fehler lässt sich verringern, wenn die AV mit einer UV zusammenhängt und man die Werte der UV bei den einzelnen Fällen kennt, man also Zusatzinformationen hat.

Das Ausmaß, indem sich die Fehler bei Kenntnis einer UV reduzieren, ergibt sich über folgende Formel

PRE (proportional reduction in error) =
$$\frac{E_0 - E_1}{E_0} = 1 - \frac{E_1}{E_0}$$

1.b. Zusammenhangsmaße: nominale Daten

Asymmetrische Zusammenhangsmaße (Richtungsmaße) Lambda

		Richtungsmaße				
			Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Lambda	Symmetrisch	,100	,017	5,567	,000
		recall abhängig	,067	,018	3,587	,000

Interpretation von Lambda:

Bei Kenntnis der Konfessionszugehörigkeit lässt sich die Wahlentscheidung mit einer um 6,7 % geringeren Fehlerquote vorhersagen als ohne Kenntnis der Konfessionszugehörigkeit

1.b. Zusammenhangsmaße: nominale Daten

Asymmetrische Zusammenhangsmaße (Richtungsmaße)

Lambda

Stärke des Zusammenhangs

praktisch kein Zusammenhang	$\lambda < 0.0025$
geringer Zusammenhang	$0.0025 < \lambda < 0.01$
mittlerer Zusammenhang	$0.01 < \lambda < 0.0625$
Starker Zusammenhang	$0.0625 \leq \lambda$

2. Ordinale Daten

2.a. Mehrfeldertabellen: ordinale Daten

2.b. Zusammenhangsmaße: ordinale Daten

2.a Mehrfeldertabellen: ordinale Daten

```
*****  
Beispiel 3: Bildung x Politisches Interesse  
*****  
****Rekodierung.  
recode s05 (2,3=1) (4=2) (5=3) (6=4) into schule.  
val lab schule 1 'höchstens Hauptschule'  
                2 'Mittlere Reife'  
                3 'Abi, Fachabi'  
                4 'Hochschulabschluss'.  
  
cro f005 by schule  
    /cells col.
```

2.a Mehrfeldertabellen: ordinale Daten

Stärke politisches Interesse * schule Schulbildung, dreistufig Kreuzt

% von schule Schulbildung, dreistufig

		schule Schulbildung, dreistufig				Gesamt
		1,00 höchstens Hauptschule	2,00 Mittlere Reife	3,00 Abi, Fachabi	4,00 Hochschul abschluss	
f005 Stärke politisches Interesse	0 ueberhaupt nicht	7,8%	3,3%	1,7%	1,6%	4,5%
	1 wenig	20,3%	11,5%	6,1%	3,6%	13,1%
	2 mittel	43,4%	41,2%	39,0%	23,7%	39,3%
	3 ziemlich stark	19,9%	31,9%	34,1%	40,1%	28,5%
	4 sehr stark	8,7%	12,2%	19,1%	32,0%	14,6%
Gesamt		100,0%	100,0%	100,0%	100,0%	100,0%

2.b. Zusammenhangsmaße: ordinale Daten

```
[/STATISTICS=[CHISQ] [LAMBDA] [BTAU] [GAMMA] [ETA ]]  
[PHI ] [UC ] [CTAU] [D ] [CORR]  
[CC ] [KAPPA ] [RISK] [MCNEMAR] [CMH [(value)]]  
[ALL ] [NONE]
```

Folgende Zusammenhangsmaßen für **ordinale sind in SPSS verfügbar:**

- 1. Symmetrische Maße: Gamma, Tau B, Tau C, Spearmans rho**
- 2. Asymmetrisches Richtungsmaß: Somers' d**

2.b. Zusammenhangsmaße: ordinale Daten

Beispiel 3: Bildung x Politisches Interesse

***Rekodierung.

recode s05 (2,3=1) (4=2) (5=3) (6=4) into schule.

val lab schule 1 'höchstens Hauptschule'

2 'Mittlere Reife'

3 'Abi, Fachabi'

4 'Hochschulabschluss'.

cro f005 by schule

/cells col

/stat gamma d btau.

**Berechnet folgende
Zusammenhangsmaße:
Gamma (gamma)
Somers's D (d)
btau (btau)**



2.b. Zusammenhangsmaße: ordinale Daten

Symmetrische Zusammenhangsmaße

Symmetrische Maße					
		Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Ordinal- bzgl. Ordinalmaß	Kendall-Tau-b	,283	,016	17,710	,000
	Gamma	,394	,021	17,710	,000
Anzahl der gültigen Fälle		2511			

Interpretation von Kendall-Tau-b: kann Werte zwischen 0 (kein Zusammenhang) und ± 1 annehmen. Bei symmetrischen Beziehungen wird Tau-b am meisten verwendet.

Interpretation von Gamma: kann Werte zwischen 0 (kein Zusammenhang) und ± 1 annehmen. Werte sind i.d.R deutlich größer als beim Tau-Koeffizienten. Gamma wird eher selten verwendet, da dieses Maß die Tendenz hat, hohe Werte anzunehmen.

2.b. Zusammenhangsmaße: ordinale Daten

Asymmetrische Zusammenhangsmaße (Richtungsmaße)

Richtungsmaße

			Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Ordinal- bzgl. Ordinalmaß	Somers-d	Symmetrisch f005 Staerke	,283	,016	17,710	,000
		politisches Interesse abhängig	,288	,016	17,710	,000
		schule Schulbildung, dreistufig abhängig	,278	,015	17,710	,000



Interpretation: Es besteht eine positive Beziehung, d.h. je höher die Schulbildung ist, desto stärker ist das politische Interesse. Die Koeffizienten sind bei einer Irrtumswahrscheinlichkeit von kleiner als 0,1% höchst signifikant.

Übersicht: Zusammenhangsmaße bei SPSS

Symmetrische Maße

Nominale Daten:

Phi (zwei dichotome Variablen)

Cramers V (polytome Var.)

Kontingenzkoeffizient (polyt. var.)

Ordinale Daten:

Gamma

Tau b

Tau c

Asymmetrische Maße

Nominale Daten:

Lambda

Unsicherheitskoeffizient

Ordinale Daten:

Somers D