

Sommersemester 2009, Statistik mit SPSS



Überblick Kreuztabellen

1. Ziel der Kreuztabellierung
2. Übersicht CROSSTABS - Syntax
3. 2x2 Kreuztabellen
 - a) Kontingenztabelle: Beobachtete Häufigkeiten und Spaltenprozente
 - b) Indifferenztabelle: Erwartete Häufigkeiten bei statistischer Unabhängigkeit
4. Tests und Zusammenhangsmaße

Ziel der Kreuztabellierung

- In einer Kreuztabelle wird die gemeinsame Verteilung zweier (oder auch mehrerer) Variablen abgebildet. Damit können Zusammenhänge zwischen nominalskalierten Variablen (bzw. ordinalen/ metrischen Variablen mit wenigen Ausprägungen) dargestellt werden.
- Mit SPSS können zwei- und mehrdimensionale Kreuztabellen erstellt und div. Tests und Zusammenhangsmaße berechnet werden. Außerdem können Balkendiagramme erstellt werden.
- Typische Fragestellungen:
 - Welche Zusammenhänge sind zu erkennen?
 - In welche Richtung gehen die Zusammenhänge?
 - Wie stark sind die Zusammenhänge?
 - Sind die Zusammenhänge signifikant?

Crosstabs Command Syntax

General mode:

```
CROSSTABS [TABLES=]varlist BY varlist [BY...]  
          [/varlist...]
```

```
[/MISSING={TABLE**}]  
      {INCLUDE}
```

```
[/FORMAT={TABLES**}{AVALUE**}]  
      {NOTABLES}{DVALUE }
```

```
[/COUNT = [{ASIS  }] [{ROUND **}] ]  
      {CASE  } {TRUNCATE}  
      {CELL**}
```

```
[/CELLS=[COUNT**][ROW  ][EXPECTED][SRESID  ]]  
      [COLUMN][RESID  ][ASRESID]  
      [TOTAL ][ALL  ][NONE  ]]
```

```
[/WRITE[={NONE**  }]  
      {CELLS  }]
```

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

```
[/METHOD={MC [CIN({99.0  })] [SAMPLES({10000})]}]++  
      {value}          {value}  
      {EXACT [TIMER({5  })]          }  
      {value}
```

```
[/BARCHART]
```

a) Kontingenztabelle, 2 x 2 Kreuztabelle, absolute Häufigkeiten

```
****Beispiel 1: Allbus 2006, Kreuztabelle mit absoluten Häufigkeiten,  
"Politiker desinteressiert an einfachen Leuten" x Ost/West.  
  
weight by v735.  
  
fre v10.  
  
cross v10 by v4.
```

SPSS Processor is ready

Ln 9

Kreuztabellen-Ausgabe

Abhängige Variable in der Zeile, Zeilenvariable

Unabhängige Variable im Tabellenkopf, Spaltenvariable

v10 POLITIKER UNINTERESSIERT AN EINF.LEUTEN * v4 WEST - OST

Count

		v4 ERHEBUNGSGEBIET: WEST - OST		Total
		1 ALTE BUNDESL AENDER	2 NEUE BUNDESL AENDER	
v10 POLITIKER UNINTERESSIERT AN EINF.LEUTEN	1 BIN DERS.MEINUNG	1740	959	2699
	2 BIN ANDERER MEINUNG	438	138	576
Total		2178	1097	3275

Randverteilungen:

- In der rechten Spalte steht die Häufigkeitsverteilung der AV (v10, Politiker desint.)
- In der untersten Zeile steht die Häufigkeitsverteilung der UV (v4, West/Ost)
- Diese Informationen erhält man auch, wenn man sich eine einfache Häufigkeitstabelle ausgeben lässt (sofern keine fehlenden Werte vorliegen)

zusätzlich Spaltenprozente anfordern

```
***Zusätzlich Spaltenprozente anfordern.  
cross v10 by v4  
  /cells count column.
```

Zusätzlich werden die Spaltenprozente ausgegeben

SPSS Processor is ready

→ Die Spaltenprozente sind die bedingten relativen Häufigkeiten der Zeilenvariable bezüglich der Spaltenvariable

Interpretation der Spaltenprozent

v10 POLITIKER UNINTERESSIERT AN EINF.LEUTEN * v4 ERHEBUNGSGEBIET: WEST - OST Cross tabulation

			v4 ERHEBUNGSGEBIET: WEST - OST		Total
			1 ALTE BUNDESL AENDER	2 NEUE BUNDESL AENDER	
v10 POLITIKER UNINTERESSIERT AN EINF.LEUTEN	1 BIN DERS.MEINUNG	Count	1740	858	2698
		% within v4 WEST - OST	79,9%	87,4%	82,4%
	2 BIN ANDERER MEINUNG	Count	438	138	576
		% within v4 WEST - OST	20,1%	12,6%	17,6%
Total		Count	2178	1097	3275
		% within v4 WEST - OST	100,0%	100,0%	100,0%

Zusammenhänge zwischen den Variablen lassen sich aufdecken, indem man die Prozentwerte der Zeilen vergleicht:

Insgesamt sind **82,4,6%** der Befragten der Meinung, dass Politiker desinteressiert an einfachen Leuten sind. Von den westdeutschen Befragten sind **79,9%** dieser Meinung. Von den ostdeutschen Befragten sind sogar **87,4%** dieser Meinung ($d\%=7,5$).

Interpretation der Spaltenprozent

v10 POLITIKER UNINTERESSIERT AN EINF.LEUTEN * v4 ERHEBUNGSGEBIET: WEST - OST Cross tabulation

			v4 ERHEBUNGSGEBIET: WEST - OST		Total
			1 ALTE BUNDESL AENDER	2 NEUE BUNDESL AENDER	
v10 POLITIKER UNINTERESSIERT AN EINF.LEUTEN	1 BIN DERS.MEINUNG	Count	1740	850	2699
		% within v4 WEST - OST	79,9%	87,4%	82,4%
	2 BIN ANDERER MEINUNG	Count	438	138	576
		% within v4 WEST - OST	20,1%	12,6%	17,6%
Total		Count	2178	1097	3275
		% within v4 WEST - OST	100,0%	100,0%	100,0%

Man betrachtet hier also den Zusammenhang zwischen beiden Variablen **aus der Perspektive der Spaltenvariable - der Region**. Oder anders ausgedrückt: man vergleicht die bedingten relativen Häufigkeiten der Zeilenvariable bezüglich der Spaltenvariable (spaltenweise Prozentuierung).

b) Indifferenztabelle

Kreuztabelle, die die theoretischen Häufigkeiten beinhaltet, die sich bei statistischer Unabhängigkeit ergeben würden.

```
***Indifferenztabelle.  
cross v10 by v4  
  /cells count expected.
```

Es werden neben den beobachteten auch die erwarteten Häufigkeiten ausgegeben

Processor area SPSS Processor is ready

b) Indifferenztafel

Kreuztafel, die die theoretischen Häufigkeiten beinhaltet, die sich bei statistischer Unabhängigkeit ergeben würden.

v10 POLITIKER UNINTERESSIERT AN EINF.LEUTEN * v4 ERHEBUNGSGEBIET: WEST - OST Crosstabulation

			v4 ERHEBUNGSGEBIET: WEST - OST		Total
			1 ALTE BUNDESL AENDER	2 NEUE BUNDESL AENDER	
v10 POLITIKER UNINTERESSIERT AN EINF.LEUTEN	1 BIN DERS.MEINUNG	Count	1740	959	2699
		Expected Count	1794,9	904,1	2699,0
	2 BIN ANDERER MEINUNG	Count	438	138	576
		Expected Count	383,1	192,9	576,0
Total		Count	2178	1097	3275
		Expected Count	2178,0	1097,0	3275,0

→ 138 Befragte aus den neuen Ländern sind nicht der Meinung, dass Politiker uninteressiert sind an einfachen Leuten.

→ Statistische Unabhängigkeit läge vor, wenn es 192,9 Befragte wären.

Unterbefehl „cells“

```
CROSSTABS [TABLES=]varlist BY varlist [BY...]  
          [/varlist...]  
  
          [/CELLS=[COUNT**] [ROW    ] [EXPECTED] [SRESID  ]]  
                    [COLUMN] [RESID   ] [ASRESID]  
                    [TOTAL  ] [ALL    ] [NONE    ]
```

Unterbefehl „CELLS“:

Mit dem Unterbefehl „CELLS“ kann man bestimmen, wie die Fälle in SPSS angegeben werden. Möglich sind u. a. folgende Angaben:

- Mit „**COUNT**“ fordert man eine Assoziations- bzw. Kontingenztabelle mit absoluten Häufigkeiten an. Dies wird auch gemacht, wenn man den Unterbefehl „CELLS“ weglässt.
- Mit „**TOTAL**“ fordert man eine Assoziations- bzw. Kontingenztabelle mit relativen Häufigkeiten an.
- Mit „**ROW**“ fordert man „Zeilenprozent“ an, also die bedingten relativen Häufigkeiten der Spaltenvariable bezüglich der Zeilenvariable.
- Mit „**COLUM**“ fordert man „Spaltenprozent“ an, also die bedingten relativen Häufigkeiten der Zeilenvariable bezüglich der Spaltenvariable: Die Zahl der Fälle in jeder Zelle wird ausgedrückt als Anteil an allen Fällen der jeweiligen Spalte.
- Mit „**EXPECTED**“ fordert man die erwarteten Häufigkeiten an, also die Zahl der Fälle, die in einer Zelle zu erwarten wären, wenn die beiden Variablen unabhängig voneinander wären.
- Mit „**RESID**“ fordert man die Residuen an – also die Differenz zwischen erwarteten und absoluten Häufigkeiten.
- Mit „**ALL**“ fordert man alle oben aufgelisteten Zellenformate an.

Tests und Zusammenhangsmaße

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

```
***Zusammenhangsmaß phi.  
|  
cross v10 by v4  
  /cells count column  
  /stat phi.
```

Zusammenhangsmaß Phi

Symmetric Measures

		Value	Approx. Sig.
Nominal by	Phi	-,093	,000
Nominal	Cramer's V	,093	,000
N of Valid Cases		3275	

Symmetrisches
Zusammenhangsmaß $\Phi = -0,093$: geringer Zusammenhang

Bei einer
Irrtumswahrscheinlichkeit von $\leq 5\%$ ($p \leq 0,05$) wird die
Nullhypothese abgelehnt

Die Interpretation erfolgt analog zur Prozentsatzdifferenz:

$-.05 < \Phi < +.05$	praktisch kein Zusammenhang;
$.05 \leq \Phi < .10$	geringer Zusammenhang
$.10 \leq \Phi < .25$	mittlerer Zusammenhang, darüber: starker Zusammenhang

Weitere Beispiele Allbus 2004

Irrtumswahrscheinlichkeit

<u>Irrtumswahrscheinlichk.</u>	<u>Bedeutung</u>	<u>Symbol</u>	
$P > 0,05$	(>5%)	nicht signifikant	n.s.
$P \leq 0,05$	(<= 5%)	signifikant	*
$P \leq 0,01$	(<= 1%)	sehr signifikant	**
$P \leq 0,001$	(<= 0,1%)	höchst signifikant	***

Morgen

1. Mehrfeldertabelle und Zusammenhangsmaße für nominale und ordinale Daten
2. Streudiagramm und Korrelation
3. Probeklausur