

Everything hangs together: Zusammenhangsmaße

716408 | Sozialwiss. Methoden – How 2 do Things with
Numbers

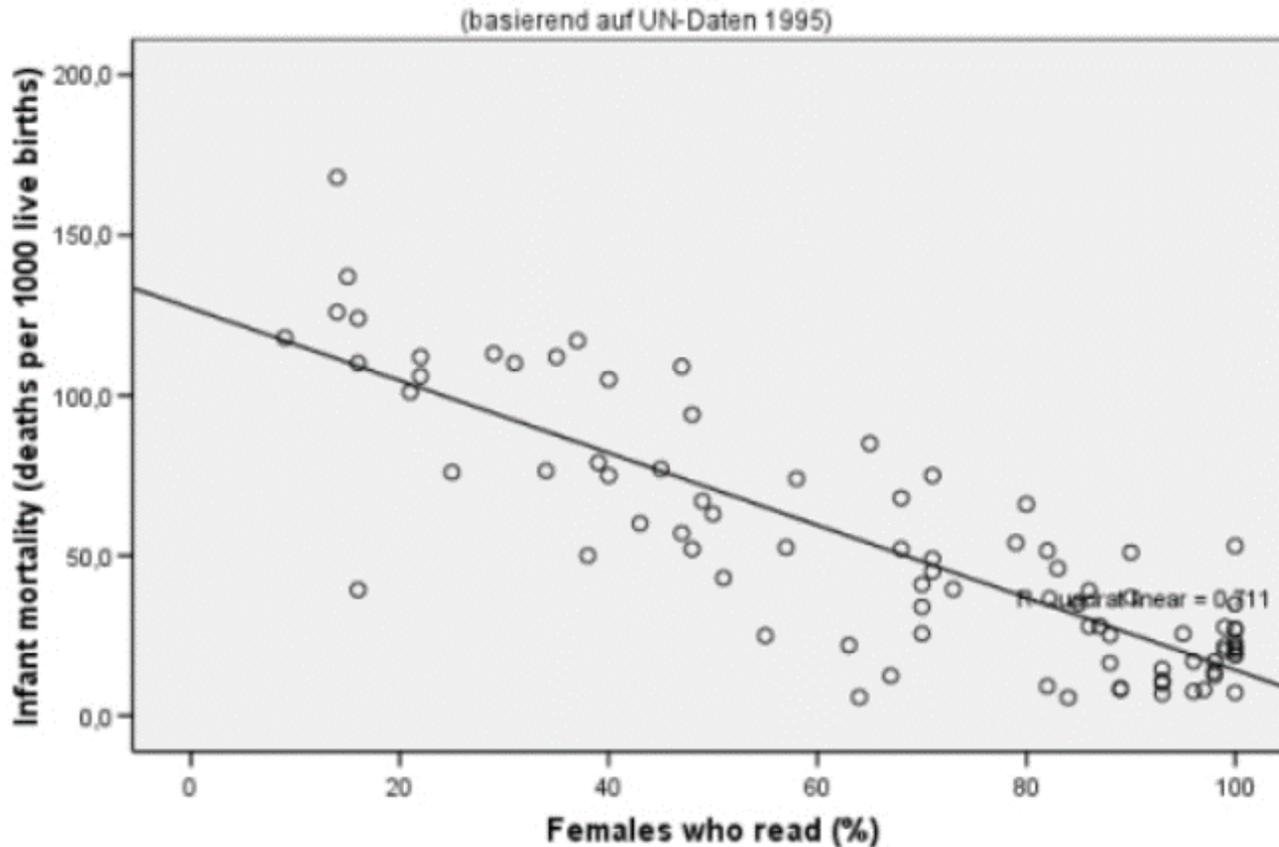
KMH
SS 22 (updated: 2022-05-25)



Everything hangs together?

Auf den Zusammenhang kommt es an I

Zusammenhang zwischen Frauen-Alphabetisierung und Kindersterblichkeit



(<https://www.univie.ac.at/ksa/elearning/cp/quantitative/quantitative-113.html>)

Auf den Zusammenhang kommt es an II

Table 5. Percentage of Each Social Indicator Per Specific Hazard Zone

Hazard Zones	% Total Population	% Housing Units	% Mobile Homes	% Age <18	% Age >65	% Nonwhite	% Female	% Area	Average Mean House Value (\$)
Chemical release—fixed	49.23	42.09	34.34	51.67	48.41	48.90	49.81	21.97	55,402
Chemical release—rail	19.48	16.88	14.98	21.03	20.26	25.92	20.01	4.75	47,726
Chemical release—road	46.49	49.18	45.37	43.42	55.87	42.88	47.36	9.10	71,421
Drought	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	62,956

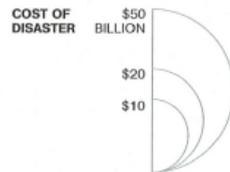
(Cutter, Mitchell & Scott 2000)

Auf den Zusammenhang kommt es an III

ALL U.S. WEATHER DISASTERS from 1980 to 2011 that caused at least one billion dollars* in damages are plotted by month and year; the size of each half circle represents the cost of the disaster. The ten most costly events are labeled.

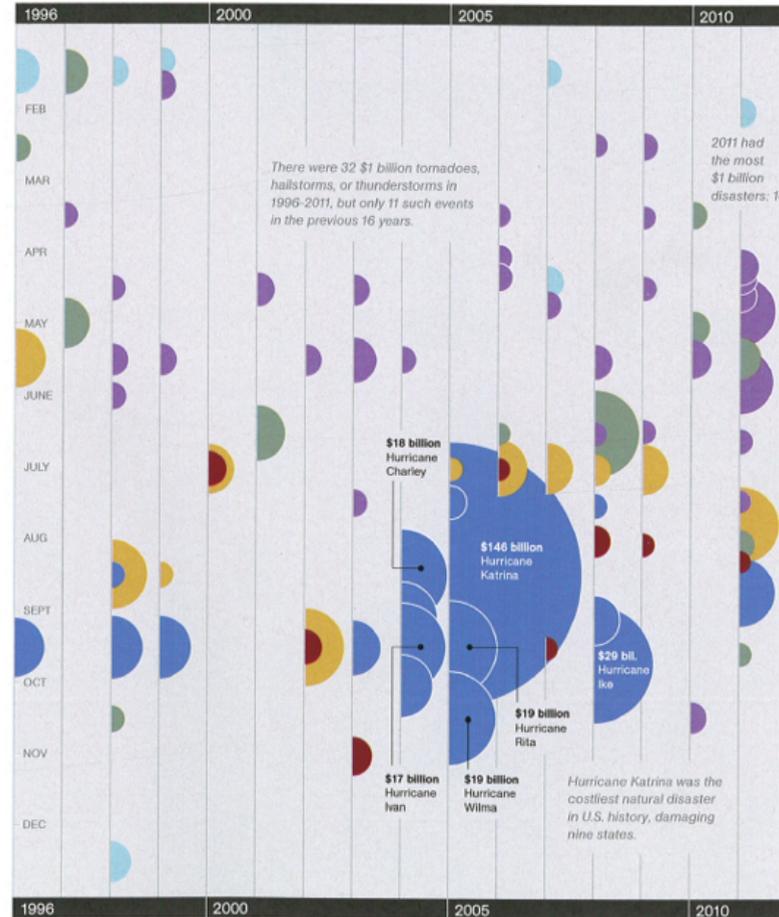
DISASTER TYPE

- Drought/heat wave
- Hurricane
- Wildfire
- Tornado/hailstorm/thunderstorm
- Flood
- Blizzard/ice storm/freeze

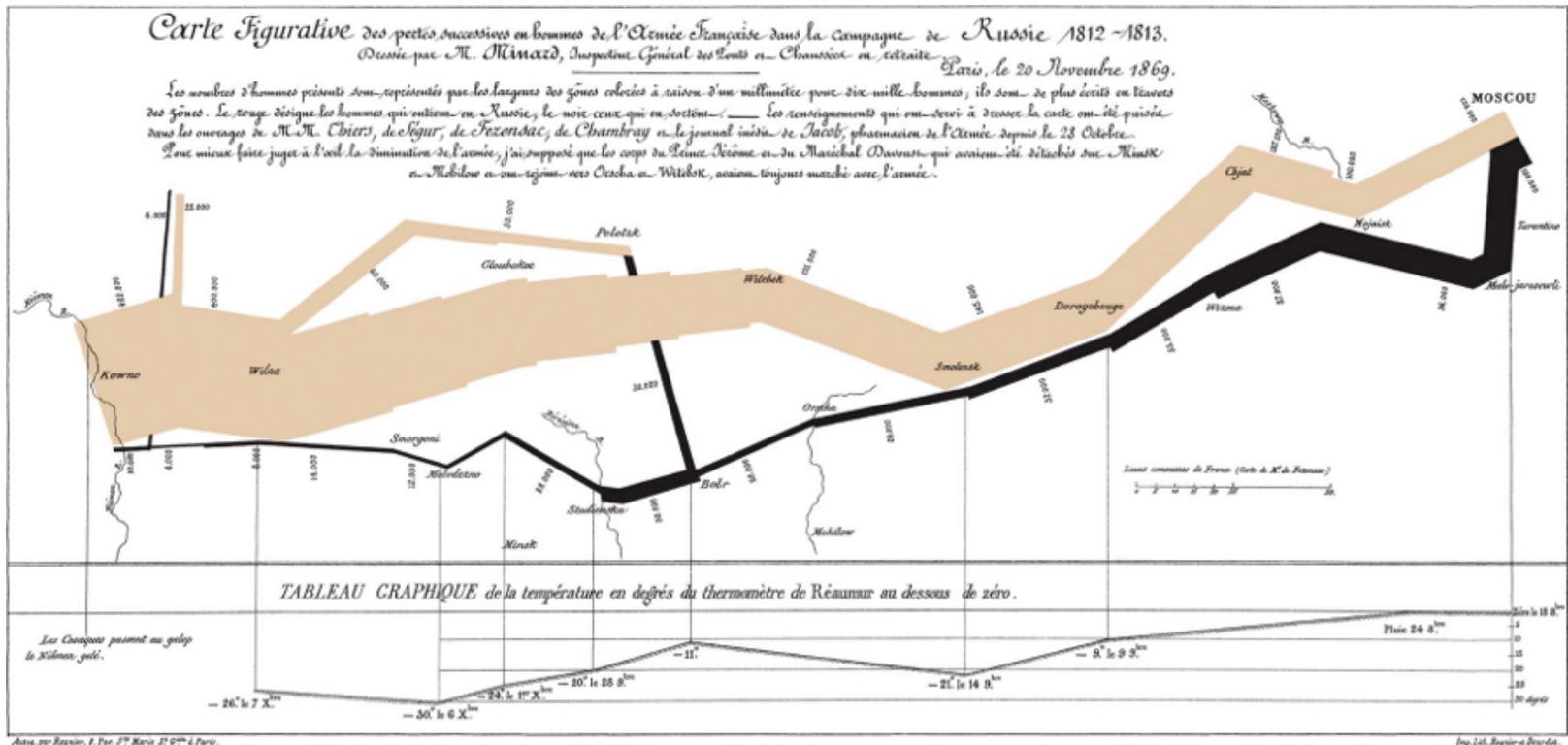


87 DISASTERS

Causing at least \$1 billion in damages, 1996-2011



Auf den Zusammenhang kommt es an IV



Recap: Skalenniveaus quantitativer Daten

Skalentyp	Eigenschaften				Erlaubte Rechenoperationen	Beispiele
	Nullpunkt	Abstände	Ränge	Identität		
Nominalskala				X	=, ≠	Familienstand, Telefonnummern, Krankheitsklassifikation
Ordinalskala			X	X	=, ≠, >, <	Zufriedenheit, militärische Ränge, Windstärke
Intervallskala		X	X	X	=, ≠, >, <, +, -	Temperatur °, Kalenderzeit
Ratioskala	X	X	X	X	=, ≠, >, <, +, -, *, /	Längen- & Gewichtsmessung

(vgl. Bortz & Döring 2009:69; Reuber & Pfaffenbach 2005:94)

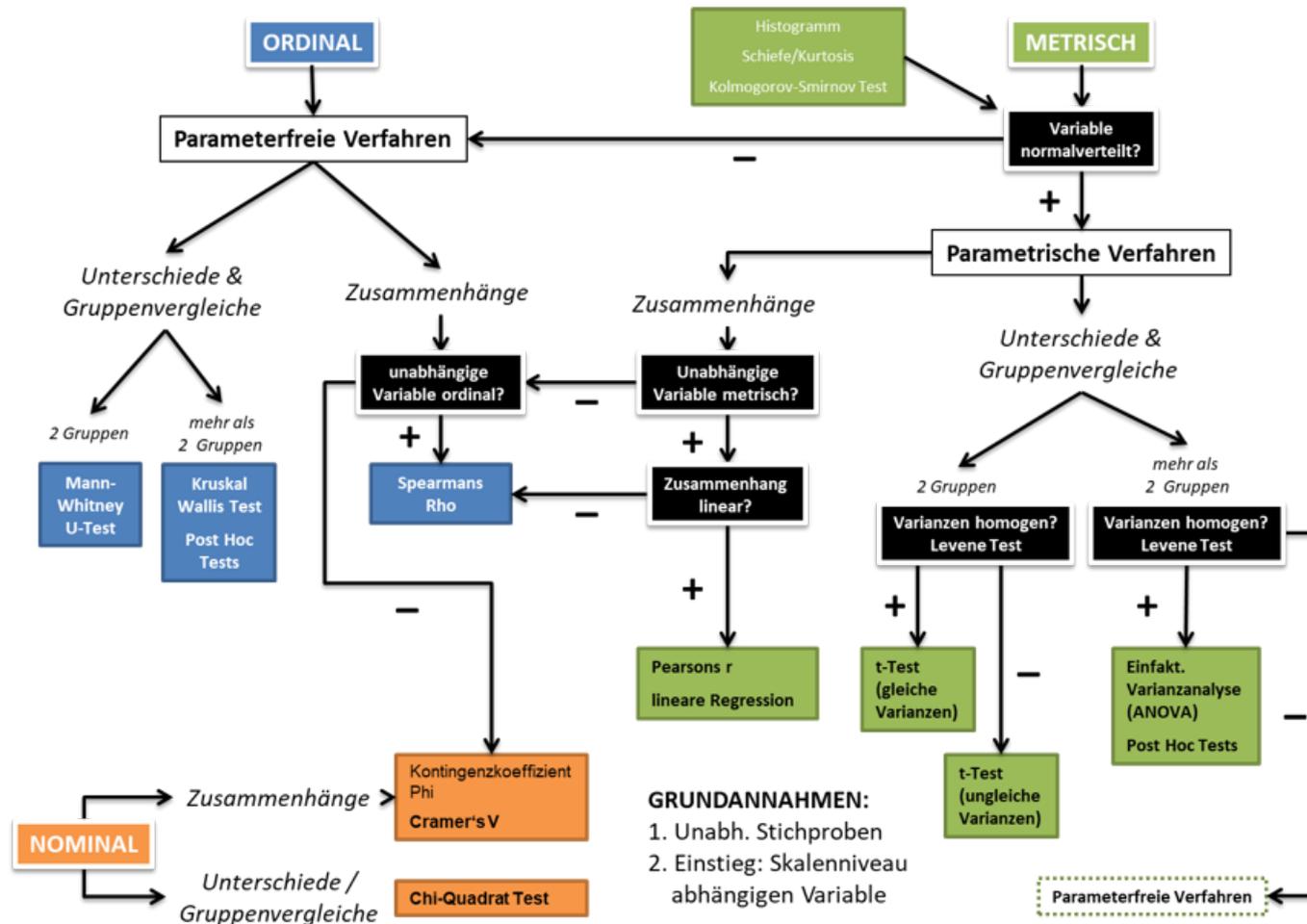
Zusammenhänge erkennen

- **2 Möglichkeiten** bivariater Analyse:
 - Explorativ:
 - Grafisch („Streudiagramm“)
 - Numerisch (Kontingenztabellen)
 - Tests auf Signifikanz:
 - Tests auf Zusammenhang (👉 Richtung)
 - Messung der Stärke eines Zusammenhangs
- **Prinzipiell:** je Skalenniveau → **unterschiedliche Maße**



Zusammenhänge - aber wie
messen?

Ein Überblick auf ausgewählte Maße



(Eigene Überarbeitung 2016 von Hager, 2011)

Ausgewählte (kategoriale) Zusammenhangsmaße

Unabhängige Variable	Abhängige Variable		
	nominal	ordinal	metrisch
nominal	χ^2 , Cramers V	←	← Eta, Eta ²
ordinal		Spearman's Rho, Kendalls Tau	Eta, Eta ²
metrisch			Bravais-Pearson-Korrelation

(Eigene Erstellung 2018; adaptiert von: Heger & Prust 2009)



Kategoriale Zusammenhänge erkunden

Zusammenhang zwischen zwei kategorialen Variablen

- Klassiker: **Kreuz- bzw. Kontingenztabelle**
 - Absolute / relative Häufigkeiten der Kombination von Merkmalsausprägungen → nominale & ordinale Daten
 - Inhaltliche Auswertung → Randsummen
 - BSP:

Verkehrsmittelwahl Anreise						
Geschlecht	Auto	Bus	Bahn	Rad	zu Fuß	Summe
weiblich	110	120	20	30	20	300
männlich	90	60	30	10	10	200
Summe	200	180	50	40	30	500

Zufall, oder nicht? Der Chi²-Test

- Suche nach „signifikanten“ ($\alpha < 0,05$) Zusammenhängen

Merkmal x	Merkmal y						Zeilen- summen n_i
	1	2	...	j	...	r	
1							n_1
2							n_2
...							...
i							n_i
...							...
m							n_m
Spaltensummen n_j	n_1	n_2	...	n_j	...	n_r	n

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^m \frac{(f_{b(i,j)} - f_{e(i,j)})^2}{f_{e(i,j)}}$$

$f_{b(i,j)}$... beobachteter Wert in Zeile i & Spalte j

$f_{e(i,j)}$... erwarteter Wert in Zeile i & Spalte j

r ... Anzahl der Zeilen (= Kategorien Variable x)

m ... Anzahl der Spalten (= Kategorien Variable y)

$$f_{e(i,j)} = \frac{\text{Zeilensumme } i * \text{Spaltensumme } j}{n}$$

(Kuckartz et al. 2010:87ff.)

- Wichtige Voraussetzung:** $f_{b(i,j)} \geq 5$

Der Chi²-Test: Ein Beispiel

Beobachtete Häufigkeiten

Geschlecht	Verkehrsmittelwahl Anreise					Summe
	Auto	Bus	Bahn	Rad	zu Fuß	
weiblich	110	120	20	30	20	300
männlich	90	60	30	10	10	200
Summe	200	180	50	40	30	500

Erwartete Häufigkeiten

Geschlecht	Verkehrsmittelwahl Anreise					Summe
	Auto	Bus	Bahn	Rad	zu Fuß	
weiblich	120	108	30	24	18	300
männlich	80	72	20	16	12	200
Summe	200	180	50	40	30	500

Der Chi²-Test: Ein Beispiel

BSP: Zusammenhang Geschlecht & Verkehrsmittelwahl

- **Nullhypothese H₀:** Es besteht KEIN ein statistisch signifikanter Zusammenhang zwischen Geschlecht & Verkehrsmittelwahl**
 - **Einshypothese H₁:** Es besteht ein statistisch signifikanter Zusammenhang zwischen Geschlecht & Verkehrsmittelwahl**
-
- Excel (CHIU.TEST) bzw. R (chisq.test) → $p=0,001203623$
 - **Da p kleiner ist als α (0,05) 🖱️ Ablehnung der Nullhypothese H₀**
 - Es besteht ein statistisch signifikanter Zusammenhang zwischen Geschlecht & Verkehrsmittel

Die Stärke des Zusammenhangs zwischen kat. Variablen

- χ^2 bei Unabhängigkeit = 0 (beobachtet = erwartet)
 - Kann beliebig groß werden
 -  **Normierung notwendig** um Aussage zu Stärke des Zusammenhangs [0 bis +1] zu treffen

- Die Lösung: **Cramers V [0 bis +1]:**

$$V = \sqrt{\frac{\chi^2}{n * (\min(m, r) - 1)}}$$

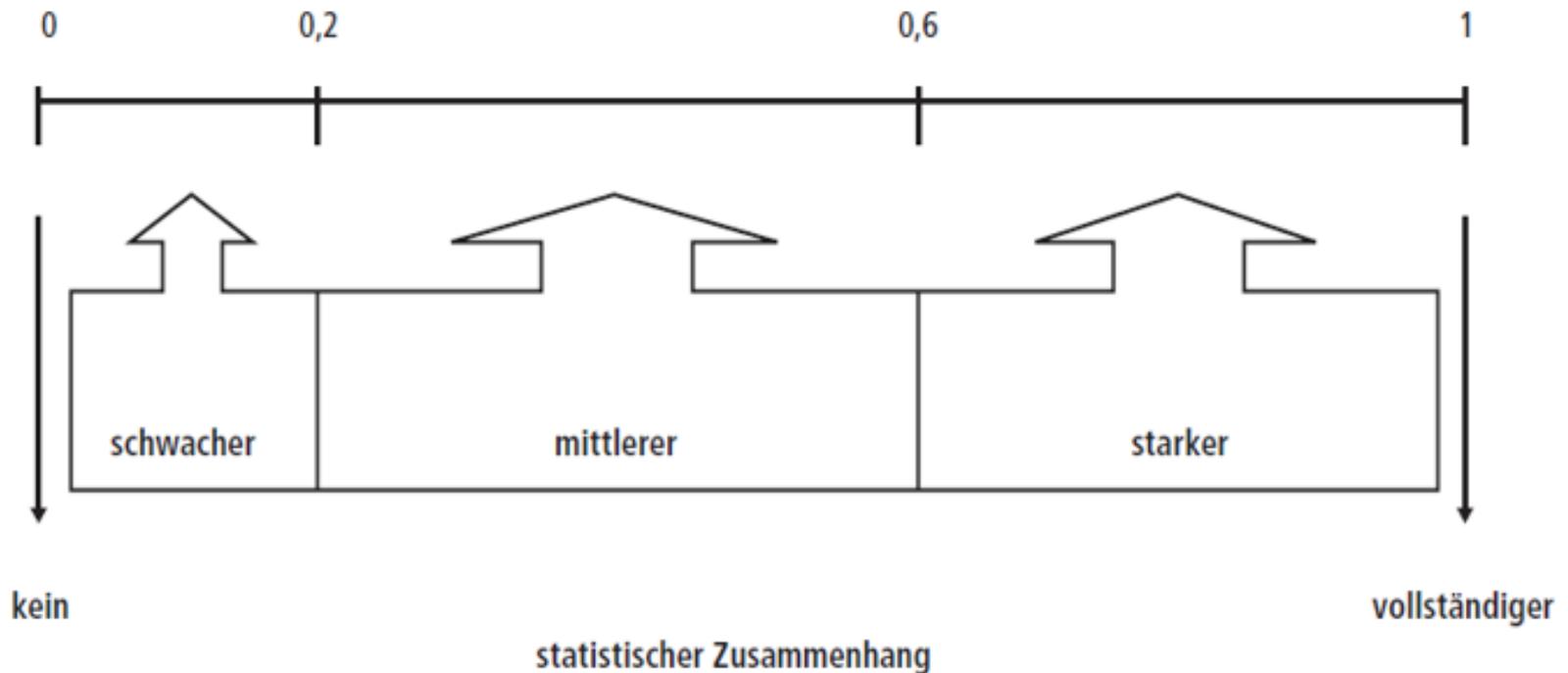
Merkmal x	Merkmal y						Zeilen- summen n_i
	1	2	...	j	...	r	
1							n_1
2							n_2
...							...
i							n_i
...							...
m							n_m
Spaltensummen n_j	n_1	n_2	...	n_j	...	n_r	n

(Quatember 2007:66)

Die Stärke des Zusammenhangs zwischen kat. Variablen

- **Cramers V [0 bis +1]:**

- Je **größer** V ist  **stärker** der stat. der Zusammenhang



(Quatember 2007:66)

Cramers V in Action

Beobachtete Häufigkeiten

Geschlecht	Verkehrsmittelwahl Anreise					Summe
	Auto	Bus	Bahn	Rad	zu Fuß	
weiblich	110	120	20	30	20	300
männlich	90	60	30	10	10	200
Summe	200	180	50	40	30	500

Erwartete Häufigkeiten

Geschlecht	Verkehrsmittelwahl Anreise					Summe
	Auto	Bus	Bahn	Rad	zu Fuß	
weiblich	120	108	30	24	18	300
männlich	80	72	20	16	12	200
Summe	200	180	50	40	30	500

Cramers V in Action

- **BSP: wie bei χ^2 -Test (Geschlecht – Verkehrsmittelwahl)**

$$\gg \chi^2 = \frac{(110-120)^2}{120} + \dots + \frac{(10-12)^2}{12} = 18,0556$$

$$\gg V = \sqrt{\frac{18,0556}{500 \cdot (2-1)}} = 0,19$$

→ schwacher stat. Zusammenhang zwischen den Variablen
Geschlecht & Verkehrsmittelwahl