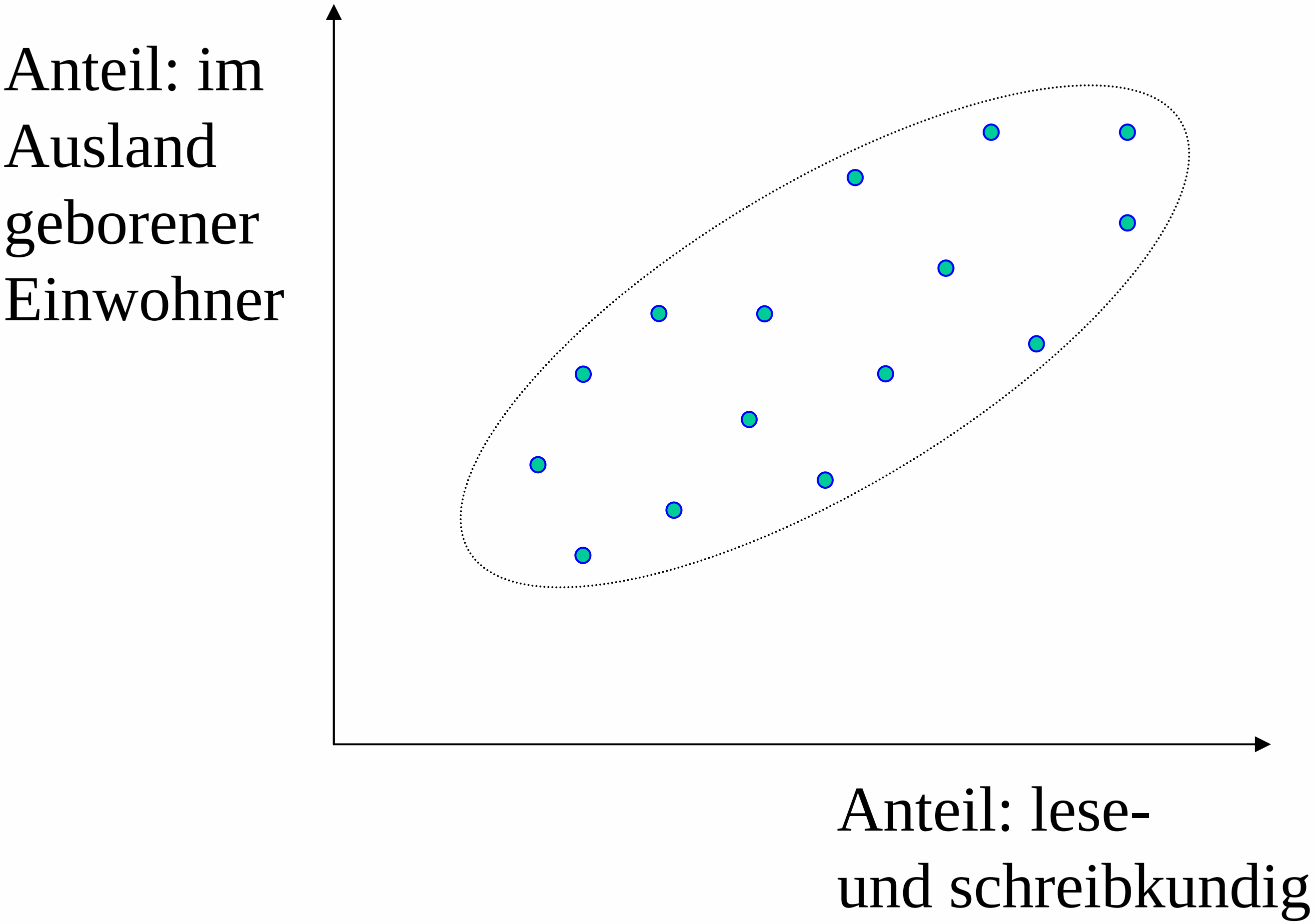
Ökologischer Fehlschluss

Inwieweit können wir von statistischen Beziehungen zwischen **Aggregat**merkmalen auf die Beziehungen zwischen den korrespondierenden **Individual**merkmalen schließen?

Kann z.B. aus einer Korrelation, die **„im Aggregat“** zwischen zwei Merkmalen besteht, auf die Korrelation zwischen den korrespondierenden **Individual**merkmalen geschlossen werden?



W.S. Robinson‘s (1950) Beispiel

W.S.Robinson (1950) klassisches Beispiel

„foreign-born“ $\xrightarrow{?}$ „literate in English“

Wenn es sich um verschiedene Gruppen handelt,
die sich z.B. in großen Städten konzentrieren,
dann befinden sich in Aggregateinheiten
mit vielen „foreign-born“ Bewohnern auch viele Bewohner,
die des Lesens und Schreibens mächtig sind

	„foreign-born“ residents	„literate in English“
Agg.Einheit 1	n_{11}	n_{12}
Agg.Einheit 2	n_{21}	n_{22}
Agg.Einheit 3	n_{31}	n_{32}
.	.	.
.	.	.

Analyseebenen

Aggregatebene(n)	**Kollektive** Eigenschaften	Aggregatdaten-analyse
	[Inklusion]	
Individualebene	**Individuelle** Eigenschaften	Individualdaten-analyse

Landtagswahl Bremen 2007

(Ortsteilebene)

Die Analyse beruht auf den Daten des vorl. amtl. Wahl-ergebnisses, wie es im Weserkurier (WK) vom 14. Mai 2007, Seite 14, ausge-wiesen worden war. Außerdem wurden Daten zum Stand 31.12.2005 aus „Bremen kleinräumig“ verwendet.

Aufgrund fehlender Werte schwankt die Fallzahl. Außerdem waren die Daten für zwei Ortsteile im WK aus Datenschutzgründen einem anderen Ortsteil zugeschlagen worden.

Beispiel einer Aggregatdatenanalyse

Korrelationen

		mig	hlu	alo	Single_HH
Beteiligung	Korrelation nach Pearson	-,681	-,846	-,871	-,081
	Signifikanz (2-seitig)	,000	,000	,000	,456
	N	86	84	83	86
SPD2007	Korrelation nach Pearson	,288	,682	,646	-,264
	Signifikanz (2-seitig)	,007	,000	,000	,014
	N	86	84	83	86
CDU2007	Korrelation nach Pearson	-,246	-,396	-,524	-,441
	Signifikanz (2-seitig)	,023	,000	,000	,000
	N	86	84	83	86
Gruene2007	Korrelation nach Pearson	-,301	-,403	-,302	,634
	Signifikanz (2-seitig)	,005	,000	,005	,000
	N	86	84	83	86
FDP2007	Korrelation nach Pearson	-,316	-,605	-,642	-,174
	Signifikanz (2-seitig)	,003	,000	,000	,109
	N	86	84	83	86
DVU2007	Korrelation nach Pearson	,659	,527	,612	-,319
	Signifikanz (2-seitig)	,000	,000	,000	,003
	N	86	84	83	86
Linke2007	Korrelation nach Pearson	,318	,188	,489	,455
	Signifikanz (2-seitig)	,003	,088	,000	,000
	N	86	84	83	86

mig Anteil von Personen mit Migrationshintergrund an allen Personen
hlu Anteil von Empfängern mit lfd. Hilfe zum Lebensunterhalt an allen Personen
alo Anteil der Arbeitslosen an allen Personen im Alter von 18 – 65
Single_HH Anteil von Singlehaushalten an allen Haushalten
Beteiligung Wahlbeteiligung
SPD2007 … Linke2007 : Prozentsätze der auf die jeweilige Partei entfallenen Stimmen

Cross-Level Inferenz

Inwieweit können wir von statistischen Beziehungen zwischen **Aggregat**merkmalen auf die Beziehungen zwischen den korrespondierenden **Individual**merkmalen schließen?

Kann z.B. aus einer Korrelation, die **„im Aggregat“** zwischen zwei Merkmalen besteht, auf die Korrelation zwischen den korrespondierenden **Individual**merkmalen geschlossen werden?

Beispiel	Anteil 1 [z.B.Arbeits-losenrate]	Anteil 2 [z.B. Anteil pol. extremer Wahlen]
Aggregat-Einheit 1	0,20	0,30
Aggregat-Einheit 2	0,30	0,40
Aggregat-Einheit 3	0,40	0,50
Aggregat-Einheit 4	0,50	0,60
.	.	.
.	.	.
.	.	.
Aggregat-Einheit k	.	.

[Fiktive Zahlen]

Agg.-Einheit 1

		X = 1	X = 2	
Y	1	?	?	0,30
	2	?	?	0,70
		0,20	0,80	

Agg.-Einheit 2

		X = 1	X = 2	
Y	1	?	?	0,40
	2	?	?	0,60
		0,30	0,70	

Agg.-Einheit 3

		X = 1	X = 2	
Y	1	?	?	0,50
	2	?	?	0,50
		0,40	0,60	

Agg.-Einheit 4

		X = 1	X = 2	
Y	1	?	?	0,60
	2	?	?	0,40
		0,50	0,50	

		X: 1	X: 2	
Y	1	$p_{1\|1}$	$p_{1\|2}$	p_{1+}
	2	$p_{2\|1}$	$p_{2\|2}$	p_{2+}
		p_{+1}	p_{+2}	1,0

		X: 1	X: 2	
Y	1	0,30	0,40	0,38
	2	0,70	0,60	0,62
		0,20	0,80	1,0

$$p_{1+} = p_{1|1} \cdot p_{+1} + p_{1|2} \cdot p_{+2}$$

$$p_{1+} = 0{,}3 \cdot 0{,}2 + 0{,}4 \cdot 0{,}8$$

Struktur einer 4-Felder-Tafel (mit Spaltenprozenten)

[Fiktive Zahlen]

		X 1	X 2	
Y	1	$p_{1\|1}$	$p_{1\|2}$	p_{1+}
	2	$p_{2\|1}$	$p_{2\|2}$	p_{2+}
		p_{+1}	p_{+2}	1,0

		X 1(AK)	X 2 (MK)	
Y	1 (D)	$p_{1\|1} = ?$	$p_{1\|2} = ?$	0,6
	2 (R)	$p_{2\|1} = ?$	$p_{2\|2} = ?$	0,4
		0,9	0,1	1,0

$$p_{1+} = p_{1|1} \cdot p_{+1} + p_{1|2} \cdot p_{+2}$$

$$p_{1+} = p_{1|1} \cdot 0{,}9 + p_{1|2} \cdot 0{,}1$$

$$p_{1|1} = \frac{p_{1+}}{p_{+1}} - \frac{p_{+2}}{p_{+1}} \cdot p_{1|2}$$

$$p_{1|1} = \frac{0{,}6}{0{,}9} - \frac{0{,}1}{0{,}9} \cdot p_{1|2}$$

[2 Unbekannte in 1 Gleichung; Gl. unteridentifiziert]

$$p_{1|1} = \frac{0{,}6}{0{,}9} - \frac{0{,}1}{0{,}9} \cdot p_{1|2}$$

$$wenn: p_{1|2} = 1{,}0$$

$$p_{1|1} = \frac{0{,}6}{0{,}9} - \frac{0{,}1}{0{,}9} \cdot 1{,}0 = 0{,}56$$

$$wenn: p_{1|2} = 0{,}0$$

$$p_{1|1} = \frac{0{,}6}{0{,}9} - \frac{0{,}1}{0{,}9} \cdot 0{,}0 = 0{,}67$$

Y \ X	AK	MK	
D	$p_{1\|1} = ?$	$p_{1\|2} = ?$	0,6
R	$p_{2\|1} = ?$	$p_{2\|2} = ?$	0,4
	0,9	0,1	1,0

AK Arbeiterklasse

MK Mittelklasse

D Wahl der Demokraten

R Wahl der Republikaner

$$p_{1|2} = \frac{0{,}6}{0{,}1} - \frac{0{,}9}{0{,}1} \cdot p_{1|1}$$

$$wenn: p_{1|1} = 1{,}0$$

$$p_{1|2} = \frac{0{,}6}{0{,}1} - \frac{0{,}9}{0{,}1} \cdot 1{,}0 = 0{,}0$$

$$wenn: p_{1|1} = 0{,}0$$

$$p_{1|2} = \frac{0{,}6}{0{,}1} - \frac{0{,}9}{0{,}1} \cdot 0{,}0 = 1{,}0$$

X

Y		AK	MK	
	D	$p_{1\|1} = ?$	$p_{1\|2} = ?$	0,6
	R	$p_{2\|1} = ?$	$p_{2\|2} = ?$	0,4
		0,9	0,1	1,0

$$p_{1+} = p_{1|1} \cdot p_{+1} + p_{1|2} \cdot p_{+2}$$

$$p_{1|2} = \frac{p_{1+}}{p_{+2}} - \frac{p_{+1}}{p_{+2}} \cdot p_{1|1}$$

Landtagswahl Bremen 2007
(Ortsteilebene; hier N=83 Ortsteile)

Beispiel: Wahlbeteiligung
und Erwerbsstatus

$p_{1|1}$ ist in keinem Ortsteil auf ein kleineres als das maximale 0 – 1 Intervall eingrenzbar.

$p_{1|2}$ Alle unteren und oberen Grenzwerte liegen innerhalb des für Anteilswerte validen 0 – 1 Wertebereichs.

Durchschnittliche *Differenz* zwischen oberem und unterem Grenzwert: 0,13

Bei einer Standardabweichung von 0,06

		X: arbeitslos	X: nicht arbeitslos	
Y	w	$p_{1\|1} = ?$	$p_{1\|2} = ?$	p_{1+}
	nw	$p_{2\|1} = ?$	$p_{2\|2} = ?$	p_{2+}
		p_{+1}	p_{+2}	1,0

w an Wahl beteiligt
nw nicht an Wahl beteiligt

Goodman's „ökologische Regression"

		X 1	X 2	
Y	1	$p_{1\|1(j)}$	$p_{1\|2(j)}$	$p_{1+(j)}$
	2	$p_{2\|1(j)}$	$p_{2\|2(j)}$	$p_{2+(j)}$
		$p_{+1(j)}$	$p_{+2(j)}$	1,0

Annahmen: $p_{1|1(j)} = p_{1|1(.)} \qquad p_{1|2(j)} = p_{1|2(.)}$

[$p_{1|1}$ und $p_{1|2}$ bzw. deren Erwartungswerte sind konstant über Aggregateinheiten hinweg]

$$\hat{p}_{1+(j)} = p_{1|1(.)} \cdot p_{+1(j)} + p_{1|2(.)} \cdot p_{+2(j)}$$ Grundgleichung

$$= p_{1|1(.)} \cdot p_{+1(j)} + p_{1|2(.)} \cdot \left(1 - p_{+1(j)}\right)$$ $p_{+2(j)} = 1 - p_{+1(j)}$

$$= p_{1|1(.)} \cdot p_{+1(j)} + p_{1|2(.)} \cdot 1 - p_{1|2(.)} \cdot p_{+1(j)}$$ Kl. auflösen

$$= p_{1|2(.)} + p_{1|1(.)} \cdot p_{+1(j)} - p_{1|2(.)} \cdot p_{+1(j)}$$ $p_{1|2(.)}$ •1 vorziehen, • 1 weglassen

$$= \underbrace{p_{1|2(.)}}_{\text{„Intercept“ a (Konstante)}} + \underbrace{\left(p_{1|1(.)} - p_{1|2(.)}\right)}_{\text{„slope“ b (Regressionskoeffizient)}} \cdot p_{+1(j)}$$ $p_{+1(j)}$ aus-klammern

Nach: Achen, Ch. H.; W. Ph. Shively (1995) Cross-Level Inference. Chicago, p. 13

$$\hat{p}_{1+(j)} = p_{1|2(.)} + \left(p_{1|1(.)} - p_{1|2(.)}\right) \cdot p_{+1(j)}$$

Regressionskonstante → $a = p_{1|2(.)}$

Regressionskoeffizient → $b = p_{1|1(.)} - p_{1|2(.)}$

$$b + a = p_{1|1(.)} - p_{1|2(.)} + p_{1|2(.)} = p_{1|1(.)}$$

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten B	Nicht standardisierte Koeffizienten Standardfehler	Standardisierte Koeffizienten Beta	T	Signifikanz
1	(Konstante)	,764	,012		64,158	,000
	alo	-1,580	,099	-,871	-15,967	,000

a. Abhängige Variable: bet

Wahlbeteiligung (abh. Var.)

$$p_{1|2} = 0{,}764$$

$$p_{1|1} = -1{,}580 + 0{,}764 = -0{,}816$$

Die Anwendung der Berechnungsvorschriften erbringt nur für p1|2(.) einen Wert im legitimen Wertebereich, nicht jedoch für p1|1(.). Es ist daher anzunehmen, dass die in die Berechnung eingegangenen Annahmen unzutreffend sind. Wären sie zutreffend, so müsste die Summe von a und b und somit p1|1 positiv werden, erst dann könnte b auch als Differenz (p1|1 – p1|2) zwischen zwei Anteilswerten aufgefasst werden.

Mit den Daten des vorl. amtl. Ergebnisses der Landtagswahl für Bremen 2007 berechnet.

Klein und Friedman‘s „Nachbarschaftsmodell“

		X 1	2	
Y	1	$p_{1\|1(j)}$	$p_{1\|2(j)}$	$p_{1+(j)}$
	2	$p_{2\|1(j)}$	$p_{2\|2(j)}$	$p_{2+(j)}$
		$p_{+1(j)}$	$p_{+2(j)}$	1,0

Annahmen: $p_{1|1(j)} = p_{1|2(j)}$

$$p_{1+(j)} = p_{1|1(j)} \cdot p_{+1(j)} + p_{1|2(j)} \cdot p_{+2(j)}$$

$$p_{1+(j)} = p_{1|1(j)} \cdot p_{+1(j)} + p_{1|1(j)} \cdot p_{+2(j)}$$

mit $p_{1|1(j)} = a + bp_{+1(j)}$

(Goodman‘s und Klein/Friedman‘s Modell empirisch nicht unterscheidbar)

Nach: Achen, Ch. H.; W. Ph. Shively (1995) Cross-Level Inference. Chicago, p. 14