

Statistische Woche 2009

Amt für Statistik Berlin-Brandenburg



Determinanten des Wahlverhaltens Auswahl und Bewertung ökologischer Variablen für eine Aggregatdatenanalyse - Ein Werkstattbericht -

Dienstag, 6. Oktober 2009, 11:10



Inhalt

- Theoretischer versus datengetriebener Ansatz
- Konzept- und empirische Variablen
- Daten- und Aussageebene
- Informationsgehalt der Variablen



Anforderungen und Bedingungen der Wahlkommentierung in der Wahlnacht: Äußerer Rahmen

- Vergleichbare Kommentierung für Berlin und Brandenburg (Stadtstaat/Kommune und Flächenland)
- Knappe Personalressourcen
- Fertigstellung Druckexemplare bis zur Pressekonferenz am Wahlmontag
- Belastbare Ergebnisse
- Wahrnehmung durch die Presse



Ausgangspunkt: Was ist verfügbar?

- Wahlausgang in den **Stimmbezirken**
(Berlin: 1984 Stimmbezirke und 516 Briefwahlbezirke)
- Sozialdemographische Variablen mit **Adressbezug**, die auf Briefwahlbezirke
(Urnen- und Briefwähler) aggregierbar sind
- Adressbezug:
Wichtige Einschränkung für die Verfügbarkeit von Daten



Theoretisches Wahlverhaltensmodell versus datengetriebener Ansatz

- Aufgabe/Problemstellung:
Wahlverhalten und sozialökonomische Umwelt in einer räumlichen Zusammenhangsanalyse darstellen
(=**Aggregatdatenanalyse** / Ökologische Wahlanalyse)

- Wahlverhaltensmodell:
Sozialstruktur → Wahlverhalten
(keine explizite Theorie, interpretatorisch-**datengetriebener Ansatz**)

- Statistisches Modell:
 - Entscheidung für eine **Zusammenhangsanalyse** mittels ökologischer Korrelationen mit Kontrollvariable
 - Auswahl der verwendeten **Aggregationsebene**
 - Auswahl der **ökologischen Variablen**
(Auswahl und Operationalisierung)



Auswahl der Aggregationsebene

- Basis der Analyse ist der Wahlausgang in den **Briefwahlbezirken** ($n=516$), da erst ab dieser Aggregationsebene **Urnen-** und **Briefwähler** gemeinsam nachgewiesen werden können und damit ein vollständiges Ergebnis vorliegt
- Sind die **516** Briefwahlbezirke den **78** Abgeordnetenhauskreisen vorzuziehen? Oder reichen auch die **12** Bezirke für die Untersuchung?
- Probleme ergeben sich durch Exklaven (ein Briefwahlbezirk kann mehr als ein Mitglied haben)



Auswahl und Operationalisierung der ökologischen Variablen

- Variablen müssen auf Ebene der räumlichen Aggregatzuschnitte (=ökologische Variablen) **verfügbar sein**.
Dadurch verringern sich die potentiell verfügbaren Merkmale
- Abgebildet werden sollen die **verschiedenen thematischen Bereiche** erklärender Einflüsse
- Auswahl und Operationalisierung möglicher Einflußfaktoren als **iterativer Prozeß** („interpretatorisch-datengetrieben“)



Auswahl und Operationalisierung ökologischer Variablen: Konzept- und empirische Variablen

Konzeptvariable	empirische Variable
Lebensphase	Alter
Geschlecht	Geschlecht
Attraktivität der Wohngegend	Einwohnerentwicklung, Wohnlage (Mietspiegel), Wohndauer, Ausländeranteil, Billeter-Altersmaß
Konfessionelle Bindung	Kirchenzugehörigkeit
Migrationsgeschichte	Migrationshintergrund
Wirtschaftliche und soziale Lage	SGB II-Empfänger, Kinder in Bedarfsgemeinschaften



Auswahl und Operationalisierung ökologischer Variablen: Anteile oder Quotienten

- Ökologischer Variablen: **Anteilvariablen** oder **Quotienten**
- **Basis Wahlberechtigte** bei der Prozentuierung von Partei- bzw. Wählersegmentanteilen, um den Effekt einer unterschiedlichen Mobilisierung des Elektorates bei einem Wahlpaarvergleich zu neutralisieren
- Basis der Indikatoren:
Wahlberechtigte oder eine zumindest umfangsähnliche Grundgesamtheit



Auswahl und Operationalisierung ökologischer Variablen: Unterscheidbarkeit, Zusammenfassung

- Welche Variablen sind unterscheidbar? Welche Ausprägungen könnten zusammengefaßt werden?

Alter in Jahren	Altersgruppe 1	Altersgruppe 2
18	Yellow	Cyan
19	Yellow	Cyan
20	Yellow	Cyan
21	Orange	Cyan
22	Orange	Cyan
23	Orange	Cyan
24	Orange	Cyan
25	Yellow	Cyan
26	Yellow	Cyan
27	Yellow	Cyan
28	Yellow	Cyan
29	Yellow	Cyan
30	Orange	Blue
31	Orange	Blue
32	Orange	Blue
33	Orange	Blue
34	Orange	Blue
35
...		
75
76		
77		
78		
79		
80		
...		

Kriterien:

- inhaltliche
- statistische



Auswahl und Operationalisierung ökologischer Variablen: Unterscheidbarkeit, Zusammenfassung

- Wie kategoriale oder quasi stetige Variablen gruppieren?
- Besonderheit der Aggregatdaten:
 - Ein **Individuum** kann nicht gleichzeitig 18 und 25 Jahre alt sein: distinkt
 - Bei **Aggregatdaten** sind die Kategorien (Anteilsvariablen) korreliert

Hilfsmittel:

- Korrelationsmatrix
- Image/Antiimage-Matrix
- Hauptkomponentenanalyse

Auswahl und Operationalisierung ökologischer Variablen: Unterscheidbarkeit, Zusammenfassung

Korrelationmatrix

	p18U25	p25U30	p30U35	p35U40	p40U45	p45U50	p50U55	p55U60	p60U65
p18U25	1,00								
p25U30	,55	1,00							
p30U35	,26	,85	1,00						
p35U40	,00	,56	,79	1,00					
p40U45	-,11	,08	,32	,57	1,00				
p45U50	,07	-,28	-,28	-,13	,33	1,00			
p50U55	-,04	-,40	-,48	-,43	-,05	,55	1,00		
p55U60	-,32	-,59	-,61	-,55	-,26	,25	,60	1,00	
p60U65	-,49	-,67	-,67	-,57	-,33	,01	,33	,71	1,00
p60plus	-,51	-,63	-,67	-,64	-,49	-,20	,03	,34	,58

Komponentenmatrix (Ladungen)

	Komponente		
	1	2	3
p18U25	,90	-,23	,12
p25U30	,80	,22	-,41
p30U35	,57	,54	-,48
p35U40	,27	,81	-,38
p40U45	-,02	,90	,18
p45U50	,04	,25	,85
p50U55	-,10	-,17	,85
p55U60	-,49	-,36	,55
p60U65	-,72	-,42	,22
p60plus	-,70	-,57	-,15

Auswahl und Operationalisierung ökologischer Variablen: Faktorenstruktur der Variablen/Indikatoren

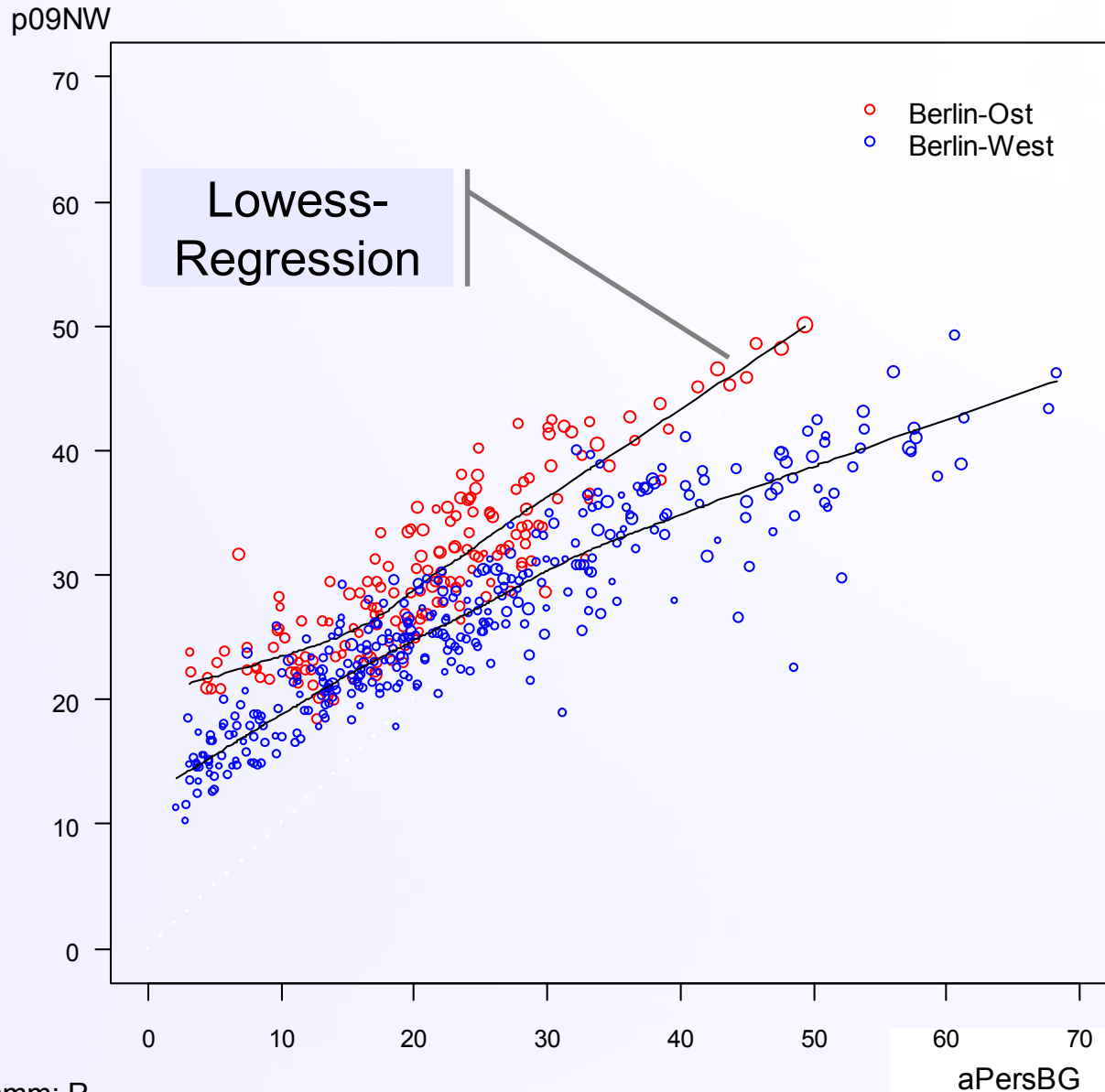
- Messen alle Variablen das gleiche?

Rotierte Komponentenmatrix (Hauptkomponentenanalyse, Varimax)

Indikatoren		Komponente			
		1	2	3	4
Alter	junge Erwachsene, 18U30	0,64	0,16	-0,61	0,20
	mittlere Altersgruppe, 30U60	0,83	-0,05	0,10	-0,01
	Ältere, 60+	-0,90	-0,07	0,30	-0,12
Altersmaß	Billeter (je jünger, desto höher)	0,87	0,19	-0,30	0,15
Verhältnis F/M	Frauen je 100 junge Männer	0,57	0,10	0,23	-0,08
Einwohnerentw.	2008 / 2003	0,61	-0,44	-0,05	0,26
Kirchenzugeh.		-0,22	0,12	0,73	0,05
Migration	Dt mit MH	-0,03	0,92	0,16	0,14
	Ausländer	0,47	0,72	0,19	0,15
Soziale Lage	SGBII-Quote	0,04	0,81	-0,48	0,20
	Kinder in BGs	0,05	0,78	-0,55	0,15
Wohnlage	einfach	0,11	0,29	-0,42	0,78
	mittel	-0,16	-0,17	-0,35	-0,87
	gut	0,04	-0,14	0,83	0,04
Wohndauer	5 Jahre und länger	-0,80	-0,11	0,32	-0,19



Zusammenhangsanalyse: Ökologische Korrelation mit Kontrollvariable



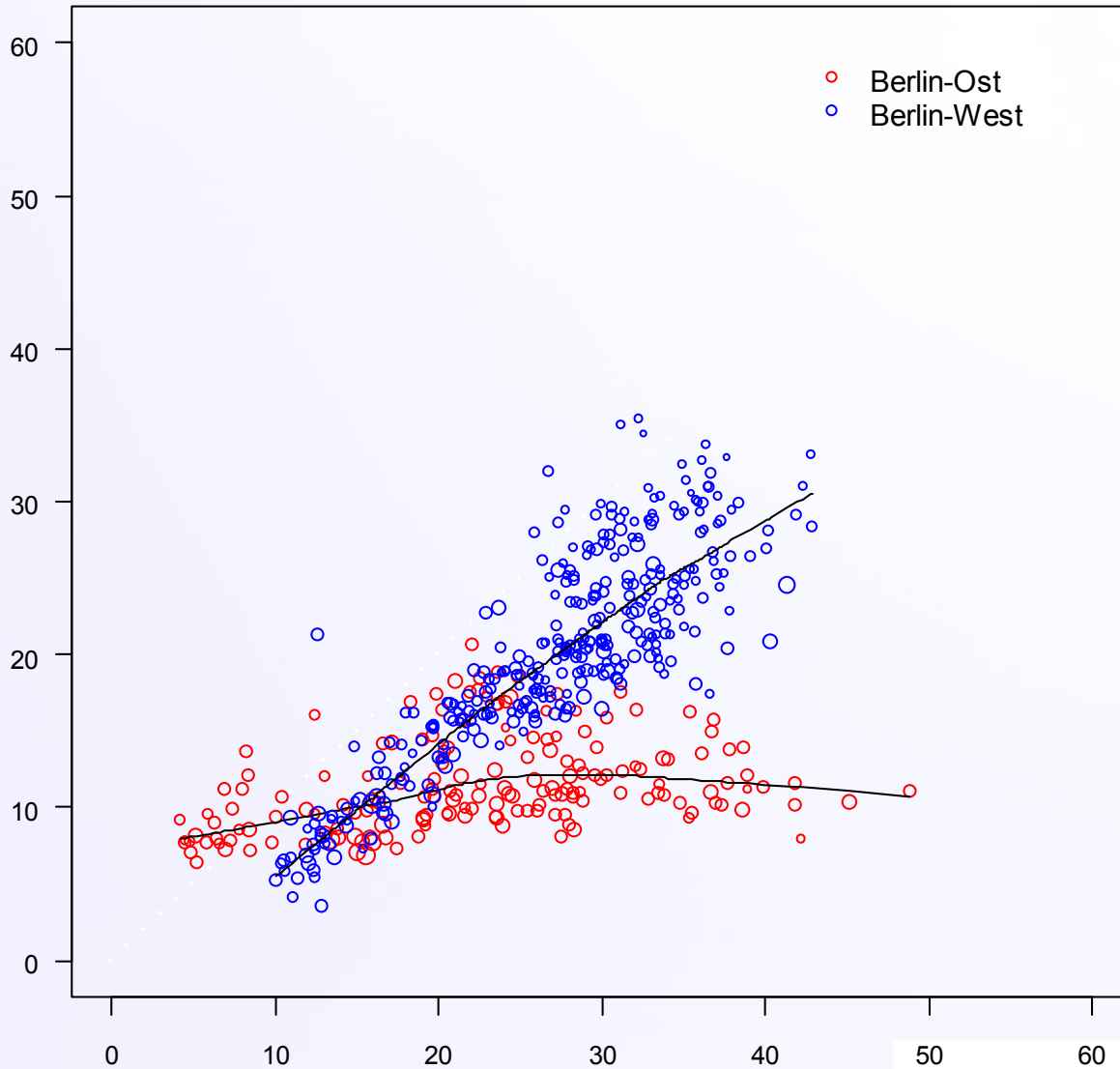
**Regional-
faktor
Ost/West:**

**Niveau-
unterschied:
Versatz**



Zusammenhangsanalyse: Ökologische Korrelation mit Kontrollvariable

p09CDU



**Regional-
faktor
Ost/West:**

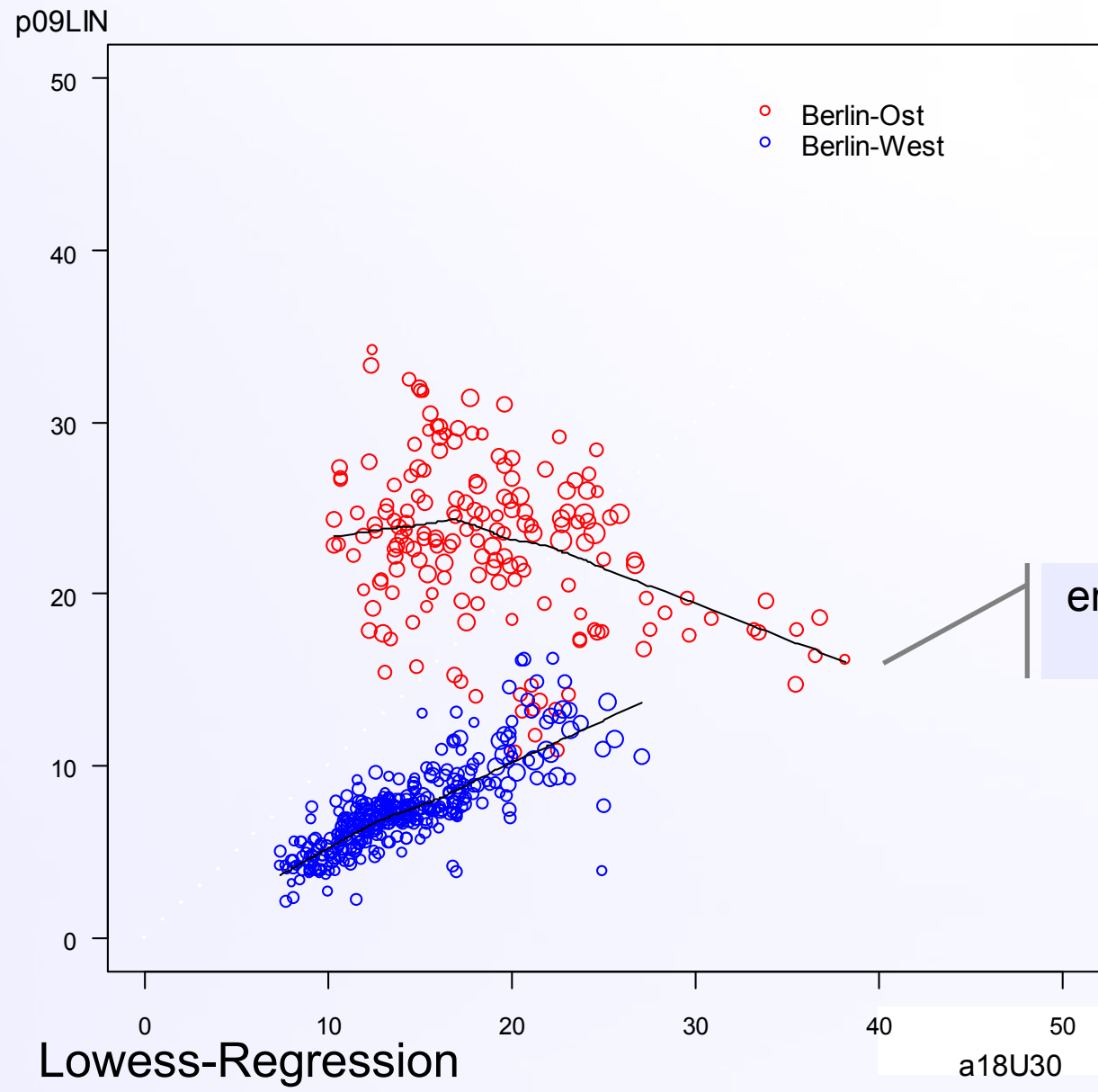
Berliner
Parteien-
system
inner-
städtisch
geteilt

Lowess-Regression

a60plus



Zusammenhangsanalyse: Ökologische Korrelation mit Kontrollvariable



Region
Ost/West ist
„Drittvariable“

entgegengesetzte
Vorzeichen



Ökologische Wahlanalyse vs. ökologische Inferenz: Daten- und Aussageebene

- Daten liegen auf der Ebene der Aggregate vor. **Aussagen** können sich auf die **Aggregatebene** beziehen oder auf die **Individualebene** schließen.



Ökologische Wahlanalyse vs. ökologische Inferenz: Daten- und Aussageebene

- Nicht alle Aggregatvariablen haben eine Entsprechung auf Individualebene

Konzeptvariable	empirische Variable	Typ
Lebensphase	Alter	individual
Geschlecht	Geschlecht	individual
Attraktivität der Wohngegend	Einwohnerentwicklung, Wohnlage (Mietspiegel), Wohndauer, Ausländeranteil, Billeter-Altersmaß	Kontext
Konfessionelle Bindung	Kirchenzugehörigkeit	individual
Migrationsgeschichte	Migrationshintergrund	individual
Wirtschaftliche und soziale Lage	SGB II-Empfänger, Kinder in Bedarfsgemeinschaften	individual Kontext



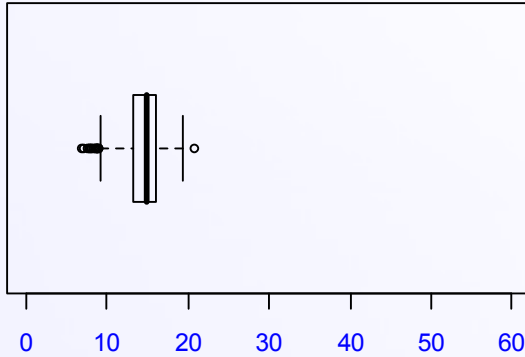
Ökologische Wahlanalyse vs. ökologische Inferenz: Aussagen auf Aggregat- und Individualebene

- Hat die LINKE in Gebieten mit einem hohen SGB II-Anteil **überdurchschnittlich abgeschnitten**?
- Ebene der Aggregate
- Haben SGB II-Empfänger LINKE gewählt?
Wie groß ist die **Wahrscheinlichkeit** $p(\text{LINKE} \mid \text{SGBII})$?
- Individualebene: Ökologische Inferenz
- Ökologischer Fehlschluß (*ecological fallacy*, Robinson 1950)
- Durchgängiges Problem:
Aggregatbefunde werden, wenn eine Entsprechung vorhanden ist, „gerne“ als Individualaussagen interpretiert.

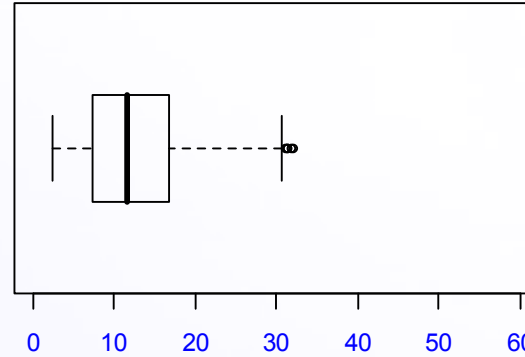


Welche Informationen sind vorhanden? Streuung Parteianteile auf der Aggregatebene

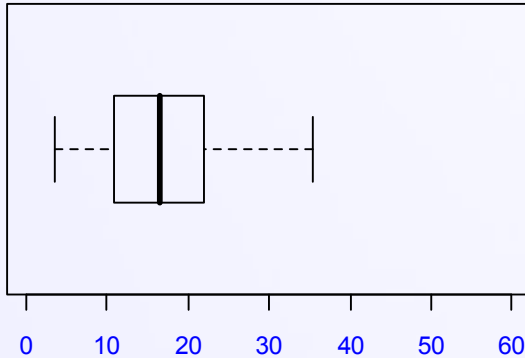
p09SPD



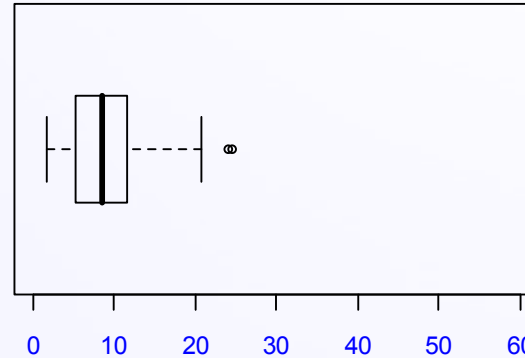
p09GRU



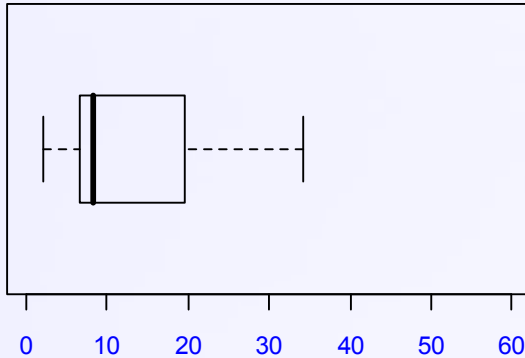
p09CDU



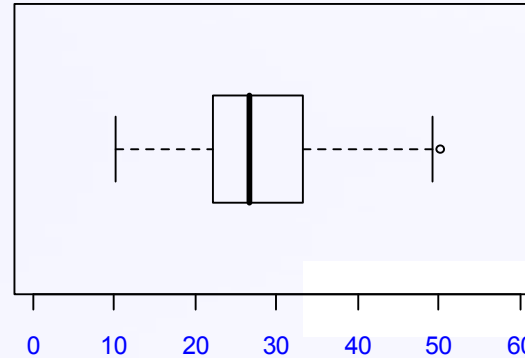
p09FDP



p09LIN



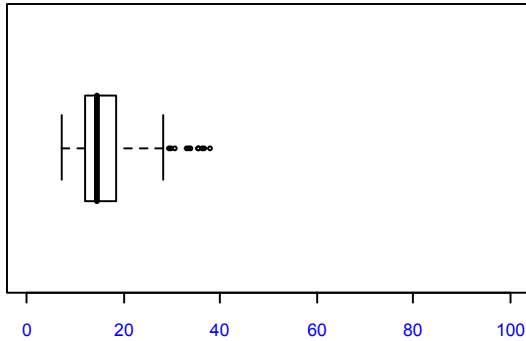
p09NW



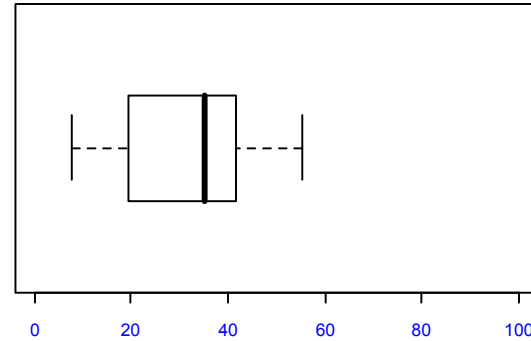


Welche Informationen sind vorhanden? Streuung Strukturvariablen auf der Aggregatebene

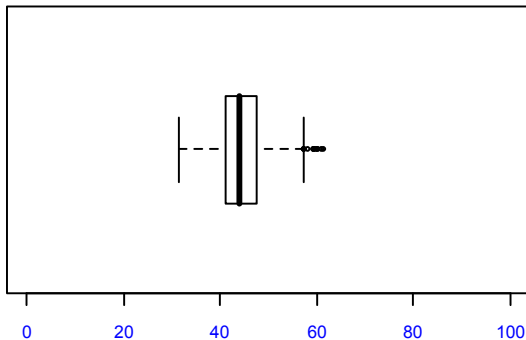
a18U30



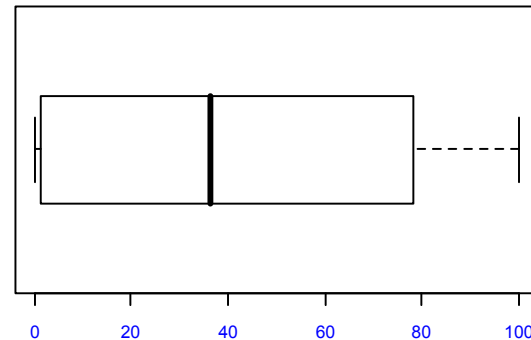
akonf



a30U60

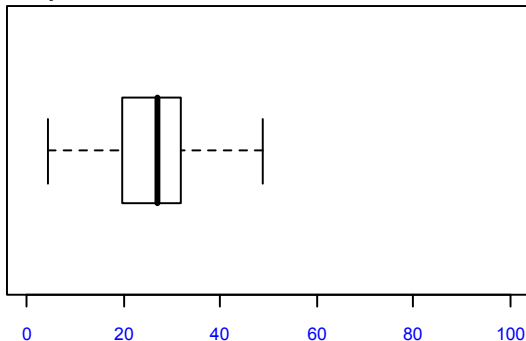


awmit

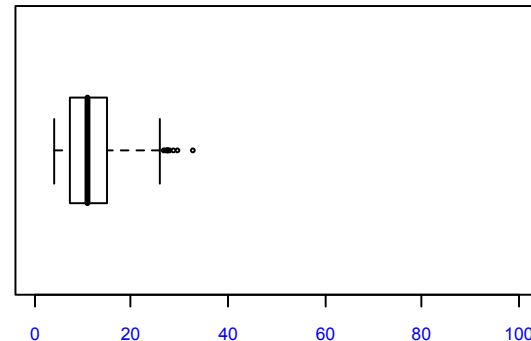


Mittlere Wohnlage:
Extremgruppen
vorhanden

a60plus



aDmMH





Welche Informationen sind vorhanden? Individualvarianz und Aggregatvarianz

- **Wieviel Varianz der Individualebene ist auf der Aggregatebene vorhanden?**
- Für kategoriale (dichotome) Variablen läßt sich die **Individualniveau-Varianz** errechnen: $V_i = \mu_k (1 - \mu_k)$
- Die **Aggregatniveau-Varianz** ist die gewichtete Varianz der Anteilsvariablen
- Der **Varianzquotient V_i/V_A** (Aggregatvarianz/ Individualvarianz) können zwei Informationen entnommen werden:
 - Wieviel Information der Individualdaten ist auf dem Aggregatniveau noch vorhanden?
 - Lassen sich die Kategorien einer Variablen auf der Aggregatebene unterscheiden oder sind sie so gleichmäßig gestreut, dass sie auf Aggregatebene kaum zu unterscheiden sind?



Welche Informationen sind vorhanden?

Varianzquotient ($V_I V_A$)

Variablen		$V_I V_A$ in %
Konfession		7,9
Alter	18U30	2,6
	30U60	1,7
	60+	5,1
Migrations- status	Dt ohne MH	11,0
	Dt mit MH	3,0
	Ausländer	8,6
Pers in BGs		9,7
Wohnlage	einfach	66,3
	mittel	56,9
	gut	75,1
Wähler- segmente	SPD	0,5
	CDU	3,4
	LINKE	5,7
	GRÜNE	4,5
	FDP	2,1
	Sonstige	0,1
Nichtwähler		3,1

Es stehen nur 1,7% der Individualvarianz zur Verfügung.
Entsprechend unsicher sind Befunde.

Es stehen 9,7% der Individualvarianz zur Verfügung.

$$N_{\text{Agg}} = 516$$



Welche Informationen sind vorhanden?

Aggregationsebenen

- Je stärker die Daten aggregiert werden, desto kleiner wird der Wert des Quotienten

Variablen		V _I V _A nach Aggregationsebenen in %			
		Bez (12)	AbgH (78)	Briefw (516)	Stimmbez (1984)
Alter	18U30	0,9	1,5	2,6	2,7
	30U60	0,2	0,6	1,7	1,8
	60+	1,4	2,9	5,1	6,0
Migrations- status	Dt mit MH	1,3	2,0	3,0	4,3
	Ausländer	5,1	7,9	8,6	10,8
Wohnlage	einfach	26,6	46,7	66,3	84,5
	mittel	11,7	30,7	56,9	79,1
	gut	34,2	56,3	75,1	88,5

(..) Anzahl der Aggregate



Wie gut sind die Kategorien der Individualvariablen unterscheidbar? Reliabilitätsindizes (RAI)

- Die Partei X ist da stark, wo auch Y stark ist: Lassen sich die beiden Parteien auf Aggregatebene überhaupt **verlässlich unterscheiden**, sind die Anteilsvariablen reliabel?
- $V_I V_A$ (=RAI1)
- RAI2 und RAI3 (Prais & Aitchison)
- Geben Aufschluß über die Verlässlichkeit des Schlusses von der Aggregat- auf die Individualebene

Variablenblöcke	$V_I V_A$	RAI2	RAI3
Konfession	0,079	0,079	0,127
Älter	0,031	0,001	0,022
Migration	0,083	0,001	0,022
SGBII	0,097	0,097	0,146
Wohnlage	0,641	0,022	0,052
Wählersegmente	0,030	0,000	0,002



Ökologische Wahlanalyse vs. ökologische Inferenz

- Aus den Ergebnissen zur Unterscheidbarkeit haben wir den Schluß gezogen, uns auf die Verwendung ökologischer Korrelationen zu beschränken
- Statt von Determinanten ist es angemessener von Indikatoren zu sprechen, da keine Untersuchung zu Gesetzmäßigkeiten angestrebt ist

Ökologische Korrelationen Wählersegmente * Indikatoren

Indikatoren	SPD	CDU	DIE LINKE	GRÜNE	FDP	Sons-tige	Darunter		Nicht-wähler
							NPD	Piraten	
Berlin-Ost									
?									
Altersgruppen?									
junge Erwachsene, 18 bis unter 30 Jahre?	-0,35	-0,66	-0,34	0,33	-0,33	0,72	-0,23	0,74	0,25
mittlere Altersgruppen, 30 bis unter 60 Jahre?	0,08	0,01	-0,72	0,69	0,47	0,13	-0,44	0,43	-0,37
Ältere, 60 Jahre und älter?	0,22	0,35	0,70	-0,61	-0,13	-0,51	0,34	-0,65	0,03
Altersmaß ?									
Billeter (je jünger, desto höher der Wert)?	-0,12	-0,30	-0,76	0,64	0,18	0,52	-0,36	0,66	-0,09
Verhältnis Frauen / Männer?									
Frauen je 100 Männer unter jungen Erwachsenen?	0,35	-0,27	-0,56	0,74	0,10	0,18	-0,61	0,56	-0,39
Einwohnerentwicklung?									
2007 / 2002 ?	0,49	0,10	-0,63	0,71	0,41	0,28	-0,57	0,61	-0,60
Kirchenzugehörigkeit?									
evangelisch, römisch-katholisch?	0,53	-0,01	-0,82	0,93	0,44	0,10	-0,77	0,63	-0,63
Deutsche mit Migrationshintergrund?									
Ausländer?	-0,37	-0,30	0,31	-0,26	-0,36	-0,31	0,10	-0,29	0,42
SGB II-Quote?									
Kinder in Bedarfsgemeinschaften?	-0,77	-0,71	0,30	-0,43	-0,76	0,20	0,46	-0,16	0,89
Wohnlage nach Mietspiegel?									
einfach?	-0,02	-0,14	-0,34	0,28	0,03	0,37	-0,23	0,44	-0,05
mittel?	-0,11	-0,05	0,40	-0,30	-0,22	-0,25	0,26	-0,38	0,18
gut?	0,26	0,38	-0,12	0,04	0,38	-0,25	-0,09	-0,12	-0,25
Wohndauer?									
5 Jahre oder länger?	0,10	0,50	0,64	-0,61	0,02	-0,50	0,39	-0,68	0,01
Berlin-West									
?									



Ökologische Korrelation Wählersegmente * Indikatoren

- Eigenständigkeit ökologischer Erklärungen
- Sind kein Ersatz für Individualkorrelationen



Vielen Dank!



Literaturhinweise

- Jan-Bernd Lohmöller (1984): LVPLS 1.6 Program Manual – Latent Variables Path Analysis with Partial Least-Squares Estimation. Köln: ZA für empirische Sozialforschung.
- Jan-Bernd Lohmöller / Hartmut Bömermann (1993): Ökologische Inferenz – Eine anwendungsorientierte Einführung (=Berliner Arbeitshefte und Berichte zur sozialwissenschaftlichen Forschung); Freie Universität Berlin, Zentralinstitut für sozialwissenschaftliche Forschung.
- S. J. Prais / J. Aitchison (1954): The Grouping of Observations in Regression Analysis. Review of the International Statistical Institut, Vol. 22, No. 1/3, pp. 1-22.