

Analyse zeitabhängiger Daten

Kumulierte (gepoolte)
Querschnittsdaten

Teil II: Kohortenanalyse

Warum geht es in den folgenden Sitzungen?

	Datum	Vorlesung
	13.04.05	Einführung
	20.04.05	Schätzverfahren für Regressionsmodelle
	27.04.05	Zeitreihenanalyse I
	04.05.05	Zeitreihenanalyse II
Kumulierte Querschnitte (Trend-, Kohortenanalyse)	11.05.05	Kumulierte Querschnittsdaten I
	18.05.05	Pfingstferien
	25.05.05	K. Querschnittsdaten II (Kohortenanalyse)
	01.06.05	Paneldaten kontinuierlicher Zielvariablen I
	08.06.05	Paneldaten kontinuierlicher Zielvariablen II
	15.06.05	Paneldaten kategorialer Zielvariablen I
	22.06.05	Paneldaten kategorialer Zielvariablen I
	29.06.05	Ereignisdaten
	06.07.05	Quasi-experimentelle Untersuchungen
	13.07.05	Trunkierte Variablen und Zählvariablen
	20.07.05	Missings, einflußreiche Fälle, komplexe Stp

Gliederung

1. Querschnittsdaten und sozialer Wandel
2. Kohortenanalyse: Was ist das?
3. Probleme der Kohortenanalyse
4. Beispiel: Religiösität in den Niederlanden

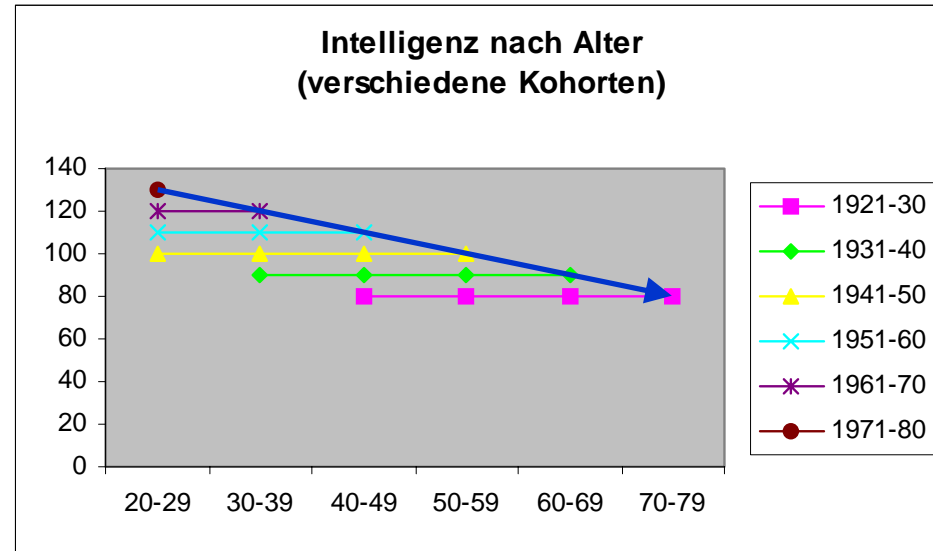
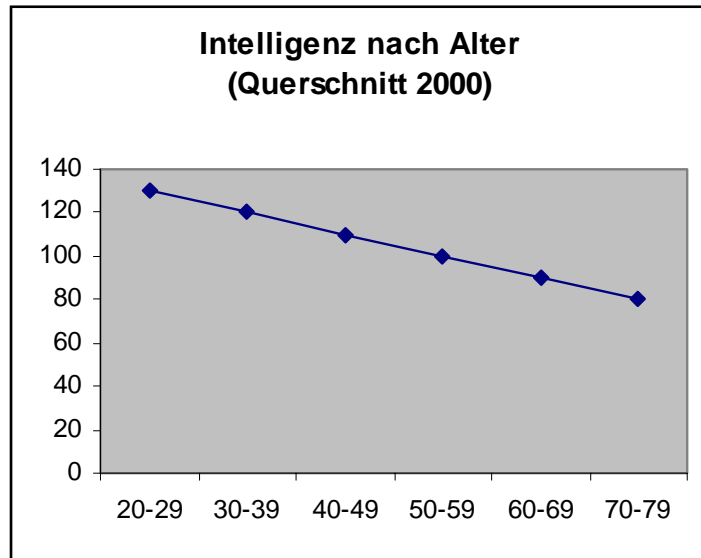
Gliederung

1. Querschnittsdaten und sozialer Wandel
2. Kohortenanalyse: Was ist das?
3. Probleme der Kohortenanalyse
4. Beispiel: Religiösität in den Niederlanden

Querschnittsdaten ungeeignet ...

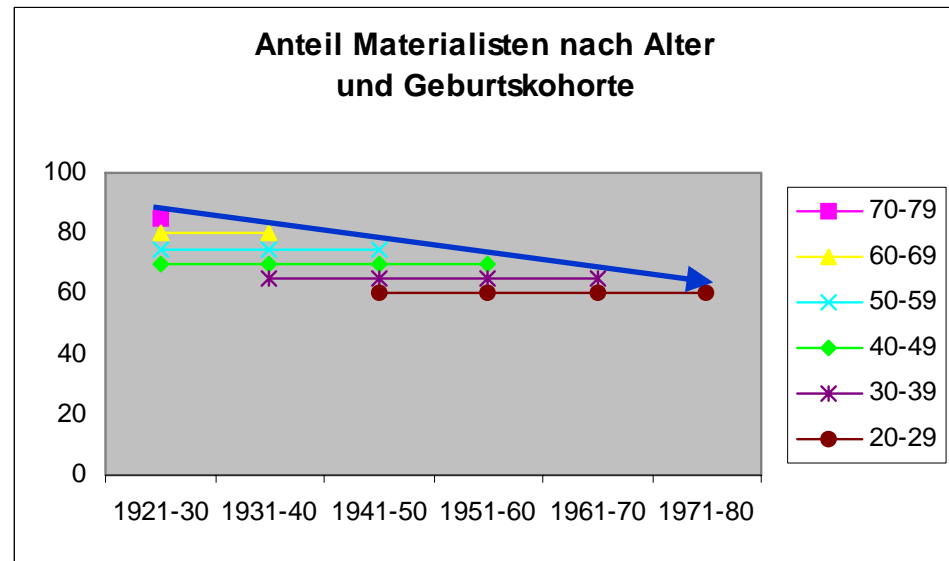
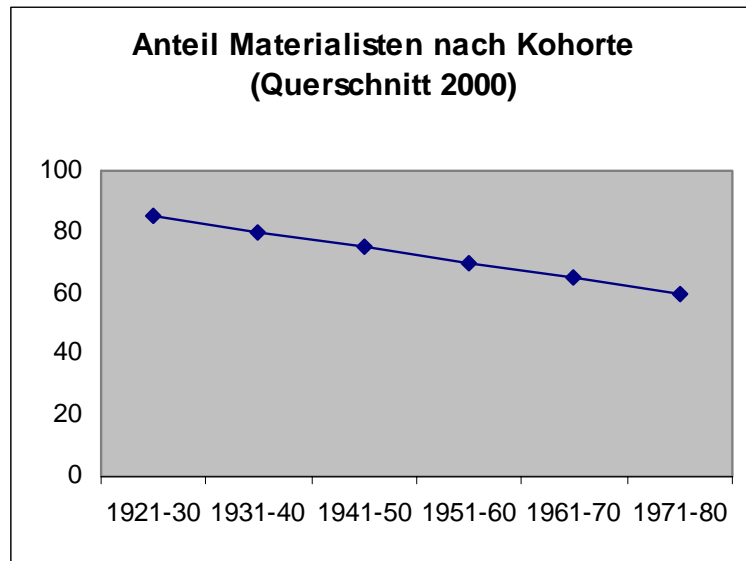
- ... für Analyse des sozialen Wandels!
- dennoch häufig verwendet
 - Altersunterschiede \Rightarrow Lebenszyklus
 - Unterschiede nach Geburtsjahr \Rightarrow Generation
- Alters- und Generationsunterschiede sind miteinander vermengt
 - Lebenszyklus-Fehlschluss
 - Kohorten-Fehlschluss

Lebenszyklus-Fehlschluß



- Im Querschnitt nimmt Intelligenz mit Alter ab.
- Intelligenz jüngerer Kohorten nimmt zu (Bildung).
- Innerhalb der Kohorten keine Alterseffekte.
- Kohortenunterschiede erzeugen Querschnittsergebnis.
- Quelle: Hunt, M. (1991): *Die Praxis der Sozialforschung. Reportagen aus dem Alltag einer Wissenschaft*. Frankfurt / New York: Campus

Kohorten-Fehlschluss



- In jeder Generation sinkt der Anteil der Materialisten.
- Materialismus älterer Personen ist größer (materielle Sicherheit).
- Innerhalb jeder Altersgruppe keine Kohorteneffekte.
- Altersunterschiede erzeugen Querschnittsergebnis.
- Quelle: Inglehart, R. (1977): *The silent revolution. Changing values and political styles among western publics*. Princeton, NJ: Princeton University Press

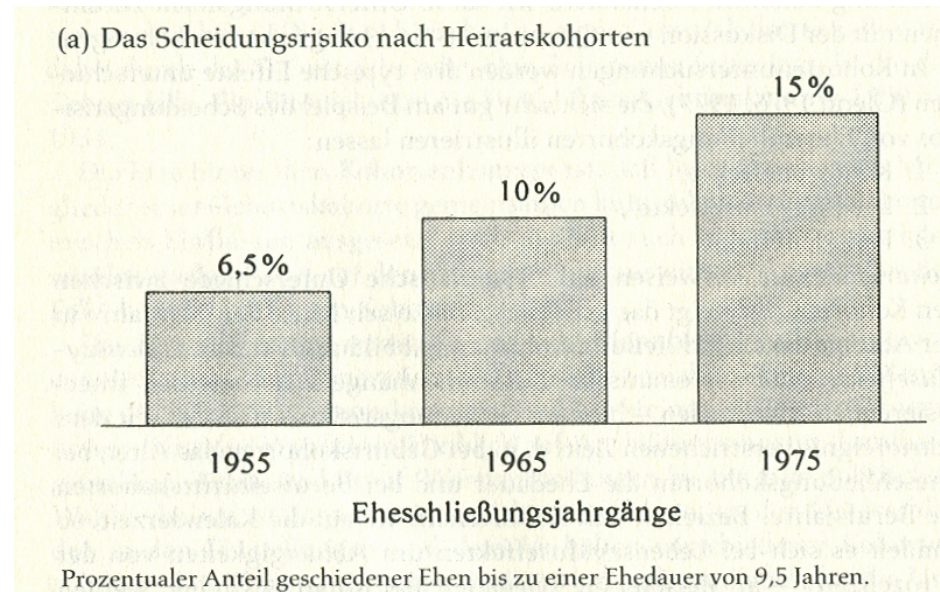
Gliederung

1. Querschnittsdaten und sozialer Wandel
2. Kohortenanalyse: Was ist das?
3. Probleme der Kohortenanalyse
4. Beispiel: Religiösität in den Niederlanden

Definition und Analyseziele

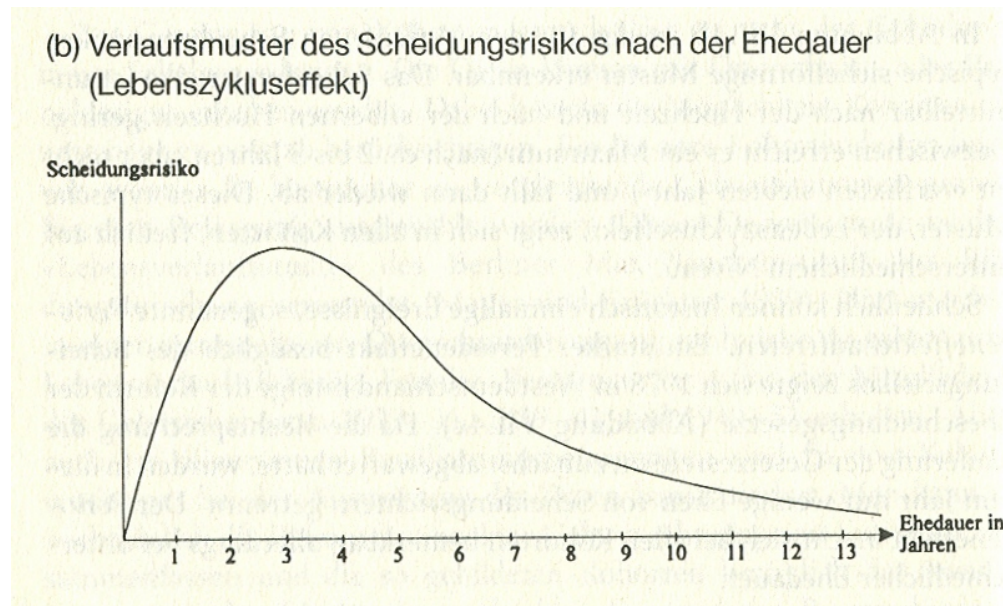
- Kohorte: römische Militäreinheit
 - “im Gleichschritt marschieren”
- Kohorte: eine Menge von Untersuchungseinheiten, die durch einen gemeinsamen Startpunkt gekennzeichnet sind:
 - Geburtskohorte, Eheschließungskohorte, Berufsanfängerkohorte, Markteintrittskohorte usw.
- Ziel: Trennung von Alters-, Kohorten- und Periodeneffekten
 - Alter (Zeit seit Startpunkt): Unterschiede nach Lebensalter, Ehedauer, Tätigkeitsdauer, Überlebensdauer (allgemein: Lebenszyklus)
 - Kohorte (chronologische Zeit): Unterschiede nach Generation
 - Periode (chronologische Zeit): Unterschiede nach Erhebungszeitpunkt
- Generationen als Träger sozialen Wandels (Karl Mannheim)
 - K. Mannheim (1928/29): Das Problem der Generationen. Kölner Vierteljahreshefte zur Soziologie 7: 157-185, 309-330

Beispiel 1: Scheidung



- Die Scheidungsquote späterer Ehekohorten ist höher.
⇒ Kohorteneffekt

Beispiel 1: Scheidung



- Das Scheidungsrisiko verändert sich mit der Ehedauer.
⇒ Alterseffekt (in diesem Fall besser: Ehedauereffekt)

Beispiel 1: Scheidung



- Die Anzahl der Scheidungen zeigt im Jahr 1978 wg. der Eherechtsreform einen Einbruch.
⇒ Periodeneffekt (in diesem Fall besser: Zeitpunkteffekt)

Beispiel 2: Religiösität

Anteil der Frauen ohne Konfession (Niederlande, 1899-1969)								
Alter	Zeitpunkt (Periode)							
	1899	1909	1919	1929	1939	1949	1959	1969
20-30	1.90%	4.78%	7.27%	13.87%	16.00%	17.40%	18.01%	23.93%
30-40	1.61%	4.22%	7.25%	14.14%	16.54%	18.20%	19.12%	22.28%
40-50	1.13%	3.13%	5.61%	11.94%	14.88%	17.18%	18.61%	22.04%
50-60	0.85%	2.46%	4.17%	9.10%	12.23%	15.05%	17.62%	21.39%
60-70	0.65%	1.88%	3.11%	6.72%	8.96%	11.90%	15.63%	19.41%
70+	0.48%	1.30%	2.12%	4.67%	6.36%	8.24%	11.25%	15.25%

- Quelle: Andreß, H.J./Hagenaars, J./Kühnel, St. (1997): Analyse von Tabellen und kategorialen Daten. Log-lineare Modelle, latente Klassenanalyse, logistische Regression und GSK-Ansatz. Berlin / Heidelberg: Springer
- Datenbasis: Volkszählungen der Jahre 1899, 1909, 1920, 1930, 1947, 1960 und 1971 (linear interpoliert)

Inhaltliche Hypothesen

- Altersunterschiede
 - unterschiedliche soziale Rollen im Lebenslauf
 - sich verändernde “Sicht der Welt” im Lebenslauf
- Kohortenunterschiede
 - religiöse Sozialisation unterschiedlicher Generationen
- Periodenunterschiede
 - punktuelle Zeitumstände
 - Krisen (Weltkrieg, wirtschaftliche Depression der 30er Jahre)

Standard-Kohortentabelle

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8

- Abstände zwischen Zeitpunkten (1909, 1919, 1929 usw.) müssen der Breite der Altersklassen (30-40, 40-50, 50-60 usw.) entsprechen
- damit in der Diagonalen immer die gleiche Kohorte erscheint

Spalte: Alterseffekte

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8

- Altersunterschiede zum Zeitpunkt P3
- auch Lebenszykluseffekte genannt
- Zeitdefinition: Lebenszeit

Zeile: Periodeneffekte

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8

- Veränderungen im Zeitablauf für die Altersgruppe A3
- Zeitdefinition: chronologische Zeit

Diagonale: Kohorteneffekte

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8

- Veränderungen zwischen den Kohorten für die Altersgruppe A3
- auch Generationenunterschiede genannt
- Zeitdefinition: chronologische Zeit

Daten für Kohortenanalysen

- kumulierte (gepoolte) Querschnittsdaten
 - synthetische Kohorten
- retrospektive Kohortenstudie
 - Retrospektivbefragung eines Querschnitts
- prospektive Kohortenstudie
 - Paneldaten

Kumulierte Querschnittsdaten

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8

- Beispiel: kumulierter ALLBUS
- statistisch unabhängige Beobachtungen
- synthetische Kohorten

Retrospektive Kohortenstudie

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8

- Beispiel: Lebensverlaufsstudie des Max-Planck Instituts für Bildungsforschung (K.U. Mayer)
- Querschnittserhebung zum Zeitpunkt P6, Restrospektivbefragung der Geburtskohorten K6 bis K11 über ihren Lebensverlauf seit dem 20. Lebensjahr
- statistisch abhängige Beobachtungen: jeweils mehrere Meßwerte für die gleichen Personen
- Problematik der Selektivität (Altersmortalität) und der Validität (Erinnerungsprobleme)

Prospektive Kohortenstudie

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8

- Beispiel: Sozio-ökonomisches Panel
- Panel: Querschnittserhebung in P2, Wiederholungsbefragung in P3-P6
- Analyse von Berufseintrittskohorten (aller 20jährigen der jeweiligen Welle)
- Alternativ: zu jedem Zeitpunkt P2-P6 Querschnittserhebung von Berufseinsteigern, die dann in Folgejahren erneut befragt werden
- statistisch abhängige Beobachtungen: Wiederholungsbefragung
- Problematik der Selektivität (Panelmortalität) und der Validität (Gewöhnungseffekt)

Im Folgenden betrachten wir nur ...

- kumulierte Querschnittsdaten
- das heißt:
 - statistisch unabhängige Beobachtungen
 - keine Notwendigkeit der Kontrolle der seriellen Abhängigkeit

Gliederung

1. Querschnittsdaten und sozialer Wandel
2. Kohortenanalyse: Was ist das?
3. Probleme der Kohortenanalyse
4. Beispiel: Religiösität in den Niederlanden

Probleme

- Vergleichbarkeit im Zeitablauf
 - Ist der Anteil der Konfessionslosen im Zeitablauf ein valider (vergleichbarer) Indikator der Religiösität?
- Selektivität der Kohorten
 - Lebensverlaufsstudie MPI: Waren die Überlebenden der Geburtskohorte 1929-31 noch für die gesamte Kohorte repräsentativ, als sie 1981-83 befragt wurden?
- Identifikationsproblem
 - Alters-, Kohorten- und Periodeneffekte lassen sich nicht unabhängig voneinander schätzen.

Effekte sind linear verknüpft

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8

- $\text{Geburtsjahr} = \text{Erhebungszeitpunkt} - \text{Alter}$
- bei klassierten Daten wie oben:
 - $\text{Kohorte} = 6 + \text{Periode} - \text{Alter}$

Zur Erinnerung: OLS Annahmen

Annahme	Definition	Funktion
Datengenerierender Prozess	einfache Zufallsstichprobe	Statistische Inferenz
Modell linear in Parametern	$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + u_i$	Art des Schätzverfahrens
Unabhängigkeit des Fehlerterms	$E(u_i \mathbf{X}) = 0, i = 1, \dots, n$	Erwartungstreue
Keine perfekte Kollinearität	x_{ki} keine Konstante, keine lineare Funktion	Berechenbarkeit
Homoskedastizität	$Var(u_i \mathbf{X}) = \sigma^2, i = 1, \dots, n$	Effizienz
Unkorrelierte Fehlerterme	$E(u_i, u_j \mathbf{X}) = 0$, für alle $i \neq j$	Effizienz
Normalverteilte Fehlerterme	$u_i \sim N(0, \sigma^2)$	Art der statistischen Tests

Es ist unmöglich, *gleichzeitig* Alters-, Perioden- und Kohorteneffekte zu schätzen!

Siehe Regression der Anteilswerte im Handout.

Altersunterschiede zum Zeitpunkt P3

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8

... können auch Kohortenunterschiede sein!

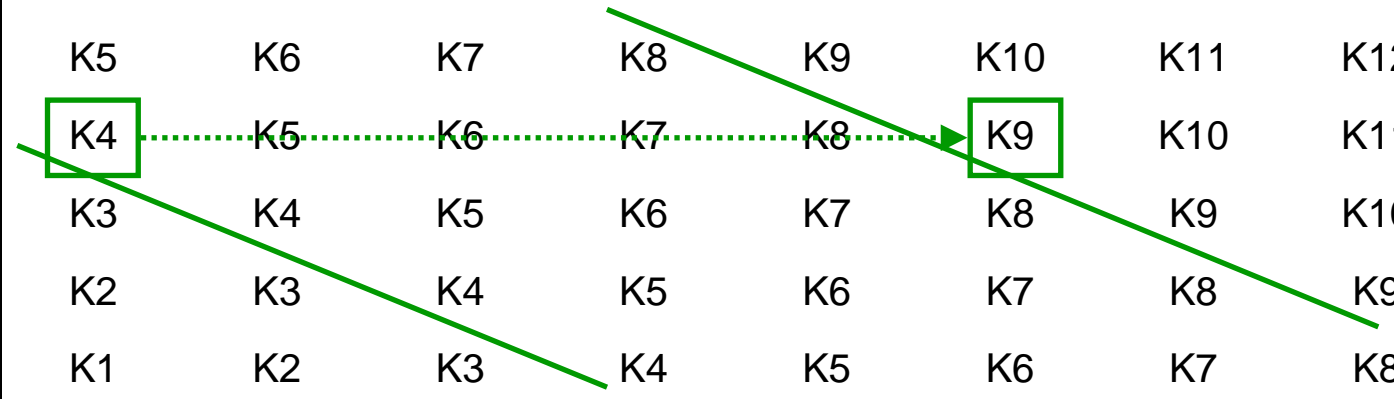
Veränderungen im Zeitablauf für die Altersgruppe A3

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8

... können auch Kohortenunterschiede sein!

Generationenunterschiede für die Altersgruppe A3

Alter	Zeitpunkt (Periode)							
	P1	P2	P3	P4	P5	P6	P7	P8
A1	K6	K7	K8	K9	K10	K11	K12	K13
A2	K5	K6	K7	K8	K9	K10	K11	K12
A3	K4	K5	K6	K7	K8	K9	K10	K11
A4	K3	K4	K5	K6	K7	K8	K9	K10
A5	K2	K3	K4	K5	K6	K7	K8	K9
A6	K1	K2	K3	K4	K5	K6	K7	K8



... können auch Periodeneffekte sein!

Gliederung

1. Querschnittsdaten und sozialer Wandel
2. Kohortenanalyse: Was ist das?
3. Probleme der Kohortenanalyse
4. Beispiel: Religiösität in den Niederlanden

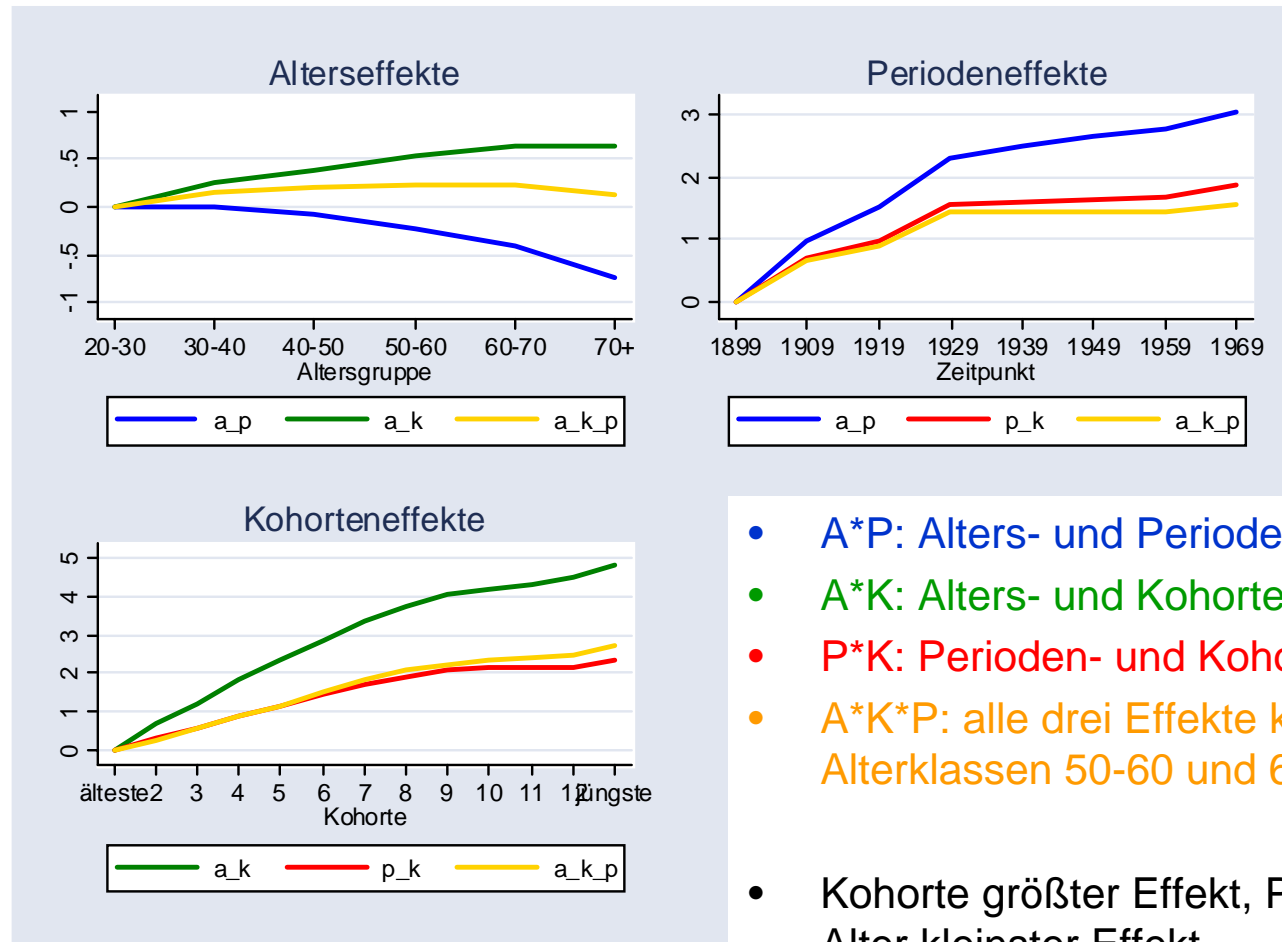
Daten für Beispiel 2

- Annahme statistisch unabhängiger Beobachtungen eigentlich nicht zutreffend
 - wg. Totalerhebung (Volkszählung) werden gleiche Personen erfasst
 - serielle Abhängigkeit wird im Folgenden der Einfachheit halber ignoriert
- Analyse dichotomer Zielvariablen
 - Konfession: 1=ohne, 0=mit
 - alle Variablen kategorial: Konfession, Alter, Periode, Kohorte
 - ⇒ vereinfachte Dateneingabe: aggregierte (gruppierte) Daten statt Individualdaten (s. Handout)
 - logistische Regression (mit gruppierten Daten)

Empfohlene Analysestrategie

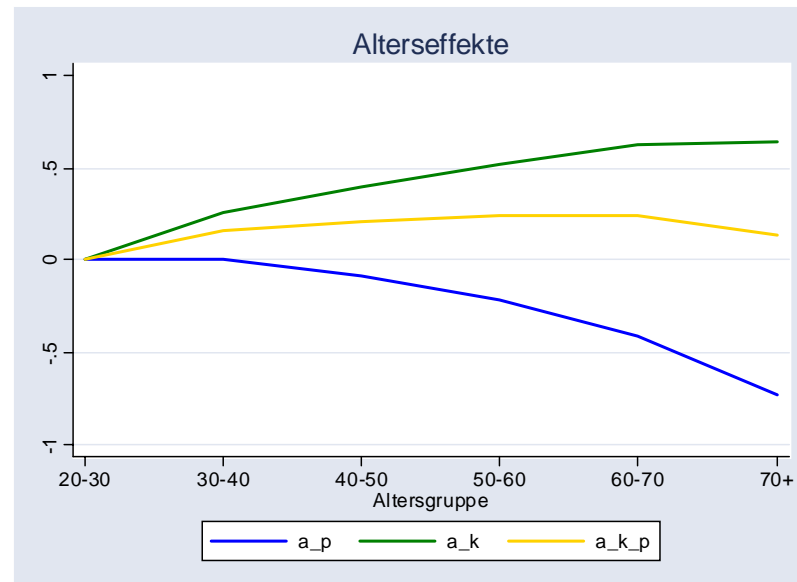
- gleichzeitige Betrachtung von A, P und K unmöglich
- beginne mit jeweils zwei Effekten: A*P, P*K und A*K
 - Ergeben sich substantiell unterschiedliche Alters-, Perioden- und Kohorten-Effekte je nach Modell?
- löse das Identifikationsproblem
 - ersetze A, P oder K durch das eigentlich interessierende Konstrukt (z.B. Alter durch Lebensphase)
 - alternativ: teste verschiedene Identifikationsrestriktionen
 - zwei gleich gesetzte Effekte genügen zur Identifikation
 - aber Schätzergebnisse abhängig davon, welche Restriktion gewählt
 - wenn die Ausprägungen einer alternativen Variablen (z.B. Lebensphase) jeweils mehrere Altersklassen zusammenfassen, ist das identisch mit einer Gleichheitsrestriktion

Geschätzte Effekte auf die Odds, keiner Konfession anzugehören



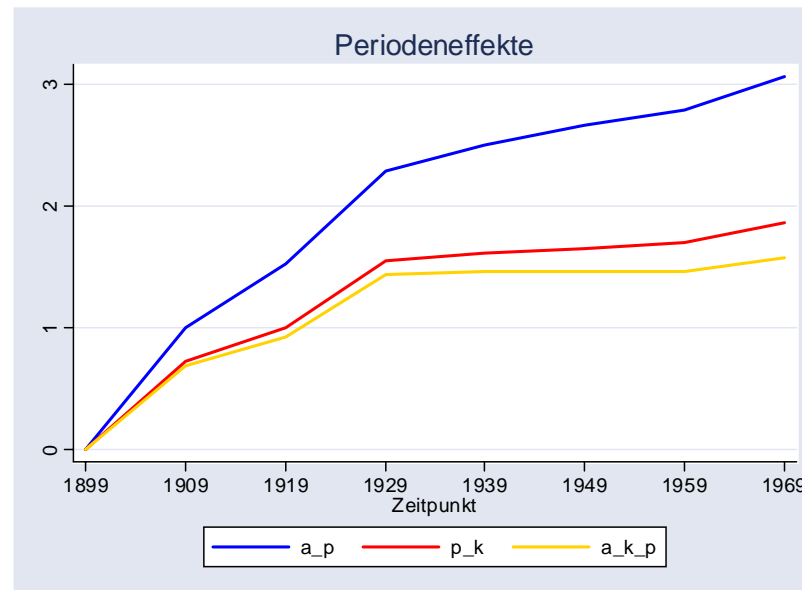
- A*P: Alters- und Periodeneffekte kontrolliert
- A*K: Alters- und Kohorteneffekte kontrolliert
- P*K: Perioden- und Kohorteneffekte kontrolliert
- A*K*P: alle drei Effekte kontrolliert, Effekte der Altersklassen 50-60 und 60-70 gleich gesetzt
- Kohorte größter Effekt, Periode mittlerer Effekt, Alter kleinster Effekt

Alterseffekte



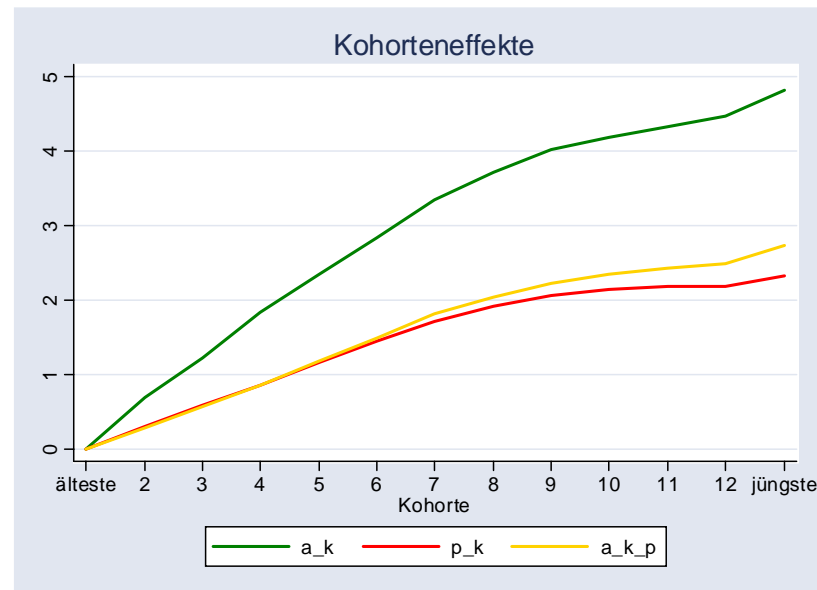
- **A*P:** ältere Personen sind religiöser (Odds ohne Konfession werden geringer gegenüber jüngster Altersklasse)
- **A*K:** ältere Personen sind weniger religiös (Odds ohne Konfession werden größer gegenüber jüngster Altersklasse)
- **A*K*P:** Personen in den mittleren Altersklassen sind weniger religiös, im Alter nimmt Religiösität wieder zu, aber nicht bis auf das Niveau der jüngsten Altersklasse

Periodeneffekte



- **A*P, P*K:** für alle Modelle gilt, dass die Religiosität im Zeitablauf abnimmt (Odds ohne Konfession werden größer verglichen mit dem Jahr 1889)
- **A*K*P:** im Modell mit allen drei Effekten ist die Religiosität seit 1929 praktisch konstant, aber niedriger ist als in den Jahren vor 1929

Kohorteneffekte



- **A*K, P*K, A*K*P**: für alle Modelle gilt, dass die Religiösität in den jüngeren Geburtskohorten immer weiter abnimmt (Odds ohne Konfession werden größer gegenüber der ältesten Geburtskohorte)

Fazit

- Die Ergebnisse verschiedener Modellspezifikationen führen nicht immer zu eindeutigen Ergebnissen.
 - hier: Alterseffekt – je nach Modell – positiv oder negativ
 - theoretisch plausibel ist der umgekehrt u-förmige Verlauf im Modell $A*K*P$
- Zwischen verschiedenen Interpretationen muss man durch Hintergrundwissen oder durch theoretische Argumente entscheiden.
- Ausblick
 - Die Betrachtung von Alter, Periode und Kohorte als kategoriale Variablen (Verwendung von Dummies) erlaubt nicht-lineare Beziehungen.
 - Lineare und kurvilineare Beziehungen lassen sich durch entsprechende Kodierung von Alter, Periode und Kohorte prüfen.

Zum Schluss

Zusammenfassung

Querschnittsdaten	<ul style="list-style-type: none">• Lebenszyklus-Fehlschluss• Kohorten-Fehlschluss
Kohorte	Menge von Untersuchungseinheiten mit gemeinsamem Startpunkt
Analyseziel	Kohortenanalyse untersucht Unterschiede: <ul style="list-style-type: none">• nach Alter: Lebenszyklus• nach Kohorte: Generation• nach Periode: Erhebungszeitpunkt
Daten	<ul style="list-style-type: none">• kumulierte (gepoolte) Querschnittsdaten• retrospektive Kohortenstudie• prospektive Kohortenstudie
Probleme	Vergleichbarkeit, Selektivität, Identifikation
Identifikation	<ul style="list-style-type: none">• Alternativindikatoren• Gleichheitsrestriktionen

Wichtige Fachausdrücke

Deutsch	Englisch	Deutsch	Englisch
Kohorten- analyse	cohort analysis	Standard- Kohortentabelle	standard cohort table
Alterseffekte	age effects	Fehlschluss	fallacy
Kohorteneffekte	cohort effects	Gleichheits- restriktion	equality restriction
Periodeneffekte	period effects		

Weiterführende Literatur

- Wooldridge (2003)
 - Kohortenanalyse wird bei Wooldridge nicht diskutiert – wahrscheinlich, weil sich hier keine neuen statistischen Probleme ergeben, abgesehen von dem eher inhaltlichen Identifikationsproblem.
- Glenn, N.D. (1977): Cohort analysis. Quantitative Applications in the Social Sciences 5. Beverly Hills et al.: Sage
 - Dieser ältere Text kann als Einführung gelesen werden, vermittelt aber keine statistischen Modelle.
- Andreß, H.J./Hagenaars, J./Kühnel, St. (1977): Analyse von Tabellen und kategorialen Daten. Log-lineare Modelle, latente Klassenanalyse, logistische Regression und GSK-Ansatz. Berlin / Heidelberg: Springer
 - Kapitel 7 zeigt eine log-lineare Analyse der Anzahl niederländischer Frauen ohne Konfession und gibt darüber hinaus einige weiterführende Literaturtipps zur Kohortenanalyse. Es lässt sich zeigen, dass die dort gefundenen passenden (log-linearen) Modelle mit den hier verwendeten logistischen Regressionsmodellen identisch sind.

Stata-Befehle

<code>table zeile spalte, c(mean y)</code>	Berechne eine Statistik der Zielvariablen y (hier: das arithmetische Mittel) für die durch die Kombination der Variablen zeile und spalte definierten Subgruppen und stelle das Ergebnis in einer Tabelle dar.
<code>xi: reg y i.x1 i.x2 i.x3</code>	Rechne eine Regression und verwandele die (kategorialen) unabhängigen Variablen x1-x3 in Dummies.
<code>quietly: reg y x1 x2 x3</code>	Berechne Regression, unterdrücke Output.
<code>estimates store name1</code>	Speicher Schätzergebnisse unter dem Namen name1.
<code>estimates table name1 name2</code>	Stelle die unter name1 und name2 gespeicherten Schätzergebnisse in Tabelle dar.
<code>estimates table name1 name2, eform</code>	Stelle die Antilogarithmen der geschätzten Regressionskoeffizienten dar.
<code>blogit n1 n x1 x2 x3</code>	logistische Regression mit gruppierten Daten: n1 von n Beobachtungen der jeweiligen Gruppe haben y=1.