Statistik I - Sitzung 9

Bernd Schlipphak

Institut für Politikwissenschaft

Sitzung 9

1. Drittvariablenkontrolle
   - Theoretische Einführung
   - Drittvariablenkontrolle: Partialtabellen

Vorgriff - Multivariate Regression

- Die multivariate Regression enthält im Gegensatz zur bivariaten Regression **mehr** als eine unabhängige Variable.

$$y = \hat{y} + e = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + ... + e$$

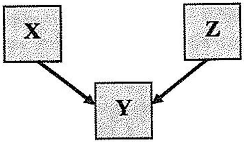
- Dadurch soll kontrolliert werden für die Möglichkeit
    - des Einflusses von zwei oder mehr unabhängigen Variablen
    - des Einflusses durch Kontrollvariablen
- Die multivariate Regression ist also EINE Möglichkeit der **Drittvariablenkontrolle**

Die Drittvariablenkontrolle

- **Drittvariablenkontrolle** = Überprüfung des Einflusses einer Variable auf einen bivariaten Zusammenhang
- Generell drei Modelle der und Begründungen für die Drittvariablenkontrolle
  - Grundlegend: Vorstellung von **Multikausalität** (Model 1)
  - Kontrolle: Vermeidung von **Scheinkausalität** (Model 2)
  - Einflussmoderation: Kontrolle von **Interaktionseffekten** (Model 3)

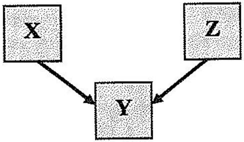
Die Drittvariablenkontrolle

- Drittvariablenkontrolle (Modell 1)
  - Multikausalität, d.h., die Erklärung einer abhängigen Variable durch *mehrere* unabhängige Variablen ist die Regel in den Sozialwissenschaften
  - Drittvariablenkontrolle = grundsätzliche Kontrolle, ob eine monokausale Beziehung zwischen unabhängiger und abhängiger Variable existiert



Die Drittvariablenkontrolle

- Drittvariablenkontrolle (Modell 1)
    - Durch Test für dritte Variable wird Effekt der unabhängigen auf die abhängige Variable oft kleiner => Messung eines bivariaten Zusammenhangs überschätzt die Wirkung einer einzelnen unabhängigen Variable
    - Korrektur der Überschätzung durch das Hinzufügen weiterer (theoretisch erklärungskräftiger Variablen) in das Modell!



Die Drittvariablenkontrolle

- Drittvariablenkontrolle (Modell 2)
  - Drittvariable kann sowohl die unabhängige als auch die abhängige Variable beeinflussen=> **Scheinkausalität**

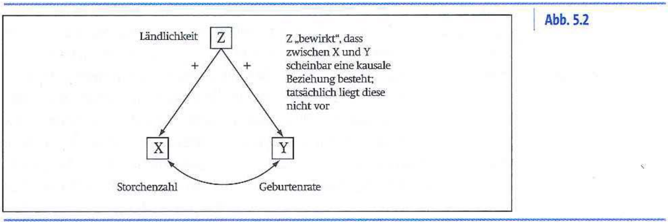


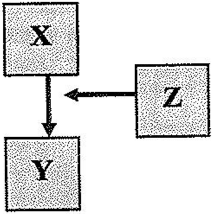
Abbildung: aus Diaz-Bone(2013): 115

Die Drittvariablenkontrolle

- Drittvariablenkontrolle (Modell 2)
  - ACHTUNG – selbst nach der Drittvariablenkontrolle kann ein statistischer Zusammenhang zwischen den kausal nicht verbundenen Variablen auftreten.
  - Dieser wird aber durch die Drittvariablenkontrolle deutlich geringer als vorher!

Die Drittvariablenkontrolle

- Drittvariablenkontrolle (Modell 3)
  - Die Drittvariable tritt als *interagierende* Variable auf



Die Drittvariablenkontrolle

- In diesem dritten Modell des Drittvariableneinflusses lassen sich zwei mögliche Formen unterscheiden
  - **Interaktion** = 'Eine Interaktion liegt vor, wenn je nach Ausprägung der Drittvariablen Z der statistische Zusammenhang [zwischen der abhängigen und der unabhängigen Variablen] verschieden ausfällt.' (Diaz-Bone 2006: 114)
  - **Suppression** = '[Die Suppression] besteht darin, dass eine Drittvariable Z den vorliegenden kausalen Zusammenhang zwischen zwei Variablen durch ihren Einfluss verdeckt.' (Diaz-Bone 2006: 117)

Die Drittvariablenkontrolle

- Die Interaktion

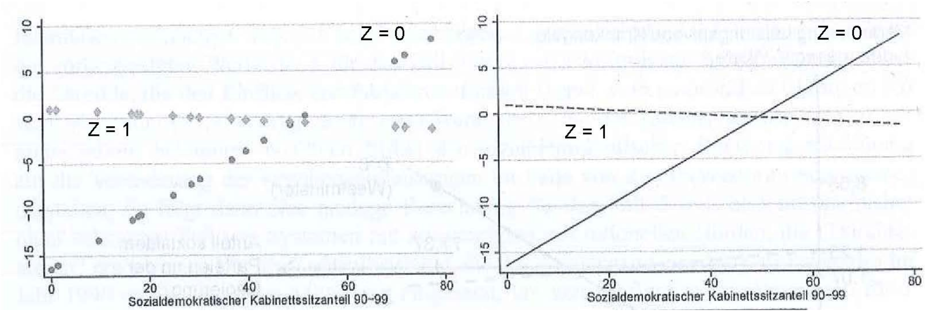
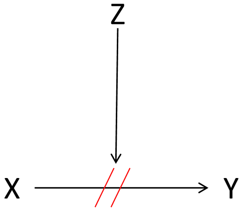


Abbildung: aus Wenzelburger et al. (2014): 48

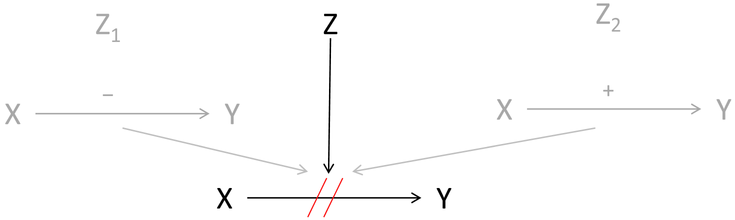
Die Drittvariablenkontrolle

- Die Suppression



Die Drittvariablenkontrolle

- Die Suppression



Die Drittvariablenkontrolle

- Unterscheidung der Durchführung (NICHT der Logik) der Drittvariablenkontrolle für nominal / ordinal skalierte Variablen einerseits und für metrisch skalierte Variablen andererseits
- Für nominale / ordinale Variablen: Berechnung sogenannter Partialtabellen und / oder entsprechender Zusammenhangsmaße getrennt für jede Ausprägung von Z
- Für metrische Variablen: Multivariate Regression (für Model 1 und 2) bzw. Regressionsmodelle mit Interaktionseffekten (für Model 3) => *Statistik II*

Die Drittvariablenkontrolle - Bsp. Scheinkorrelation

- Ein kurzes Einleitungsbeispiel:
  - Befragung zeigt, dass Lanz-Anhänger*innen auch die CDU überdurchschnittlich gut finden
  - Führt Vorliebe für Markus Lanz (=X) zur Befürwortung der CDU (=Y)?
  - Antwort: Vermutlich Drittvariable (Z), die dazu führt, dass Menschen sowohl Markus Lanz als auch CDU-Regierung gut finden

Die Drittvariablenkontrolle - Bsp. Scheinkorrelation

- Zur Überprüfung - Aufteilung der **Marginaltabelle** (Beziehung zwischen X und Y) in **Partialtabellen** (Beziehung zwischen X und Y **je nach Ausprägung von Z**)
  - Demzufolge gibt es immer so viele Partialtabellen wie Ausprägungen von Z!

Die Drittvariablenkontrolle - Bsp. Scheinkorrelation

	Lanz negativ (x_1)	Lanz positiv (x_2)	$\sum$
CDU negativ (y_1)	310(68.89%)	240 (43.64%)	550
CDU positiv (y_2)	140 (31.11%)	310 (56.36%)	450
$\sum$	450 (100%)	550 (100%)	**1000**

$$OR = \frac{Odds_{y_1}}{Odds_{y_2}} = \frac{\frac{310}{140}}{\frac{240}{310}} = \frac{2.21}{0.77} = 2.87 \Rightarrow Yules'\ Q = 0.48$$

Die Drittvariablenkontrolle - Bsp. Scheinkorrelation

- Da Yules' Q = 0.48, mittelstarker Zusammenhang zwischen X und Y $\Rightarrow$ Lanz-Liebhaber*innen finden auch die CDU besser!
    - Was passiert aber, wenn wir für die gesellschaftliche Vorlieben der Befragten kontrollieren?
    - Z = Drittvariable = gesellschaftliche Vorlieben
    - Ausprägungen: Z_1 = konservativ, Z_2 = progressiv

Die Drittvariablenkontrolle - Bsp. Scheinkorrelation

Konservativ (z_1)	Lanz negativ (x_1)	Lanz positiv (x_2)	$\sum$
CDU negativ (y_1)	30(30%)	120 (30%)	150
CDU positiv (y_2)	70 (70%)	280 (70%)	350
$\sum$	100 (100%)	400 (100%)	**500**

Progressiv (z_2)	Lanz negativ (x_1)	Lanz positiv (x_2)	$\sum$
CDU negativ (y_1)	280 (80%)	120 (80%)	400
CDU positiv (y_2)	70 (20%)	30 (20%)	100
$\sum$	350 (100%)	150 (100%)	**500**

Die Drittvariablenkontrolle - Bsp. Scheinkorrelation

- $OR_{kons} = \dfrac{Odds_{y_1}}{Odds_{y_2}} = \dfrac{\frac{30}{70}}{\frac{120}{280}} = \dfrac{0.43}{0.43} = 1$
- $OR_{prog} = \dfrac{Odds_{y_1}}{Odds_{y_2}} = \dfrac{\frac{280}{70}}{\frac{120}{30}} = \dfrac{4.00}{4.00} = 1$
- Da OR $=$ 1 kein Zusammenhang zwischen Lanz-Vorliebe und CDU-Befürwortung!

Die Drittvariablenkontrolle - Bsp. Scheinkorrelation

- Kontrolliert man also den Zusammenhang zwischen X und Y für die Drittvariable Z, so zeigt sich, dass der Zusammenhang verschwindet
- Stattdessen zeigt ein Blick auf die Partialtabellen, dass es unter Konservativen sehr viel mehr Lanz-Befürworter*innen und CDU-Befürworter*innen gibt als unter Progressiven.
- Wir können also annehmen, dass die gesellschaftliche Vorliebe beide anderen Variablen bedingt!

Drittvariablenkontrolle - Bsp. Interaktion

- Existiert ein Zusammenhang zwischen Alter und (traditionellen) Vorstellungen von Geschlechterrollen? (Diaz-Bone 2006: 115)
- Argument: Mit zunehmendem Alter steigt die Zustimmung zu traditionellen Positionen.
- Test: Aufteilung der Fälle in Vierfeldtabelle nach Alter: < 40 und > 40 und nach Position: traditionell und progressiv

Drittvariablenkontrolle - Bsp. Interaktion

	< 40 Jahre	> 40 Jahre	$\sum$
Traditionell	65 (14.41%)	221 (32.79%)	286
Progressiv	386 (85.59%)	453 (67.21%)	839
$\sum$	451 (100%)	674 (100%)	**1125**

- $OR_{gesamt} = \frac{Odds_{y_1}}{Odds_{y_2}} = \frac{\frac{65}{386}}{\frac{221}{453}} \approx 0.35 \Rightarrow$ Yule's Q ≈ -0.48
- Mittelstarker Zusammenhang!

Drittvariablenkontrolle - Bsp. Interaktion

- Zusätzliches Argument: Aufgrund (zumindest vordergründiger) Gleichstellung der Geschlechter in DDR schwächerer Zusammenhang in Ost als in West (Diaz-Bone 2006: 115)
- Überprüfung: Berechnung desselben Zusammenhangs getrennt für Fälle aus Westdeutschland (Z_1) und aus Ostdeutschland(Z_2)
  - 1. Schritt: Erstellung der Kreuztabelle für Fälle getrennt nach West und Ost
  - 2. Schritt: Berechnung der OR / Yules' Q für die beiden neuen Kreuztabellen

Drittvariablenkontrolle - Bsp. Interaktion

West	< 40 Jahre	> 40 Jahre	$\sum$
Traditionell	57 (15.28%)	192 (36.23%)	249
Progressiv	316 (84.72%)	338 (63.77%)	654
$\sum$	373 (100%)	530 (100%)	**903**

- $OR_{West} = \frac{Odds_{y_1}}{Odds_{y_2}} = \frac{\frac{57}{316}}{\frac{192}{338}} \approx 0.32 \Rightarrow$ Yule's Q $\approx$ -0.52
- Starker Zusammenhang!

Drittvariablenkontrolle - Bsp. Interaktion

Ost	<40 Jahre	>40 Jahre	$\sum$
Traditionell	8 (10.26%)	29 (20.14%)	37
Progressiv	70 (89.74%)	115 (79.86%)	185
$\sum$	78 (100%)	144 (100%)	**222**

- $OR_{Ost} = \frac{Odds_{y_1}}{Odds_{y_2}} = \frac{\frac{8}{70}}{\frac{29}{115}} \approx 0.44 \Rightarrow$ Yule's Q $\approx$ -0.39
- Mittelstarker / Mittlerer Zusammenhang!

Drittvariablenkontrolle - Bsp. Interaktion

- Interpretation: Der Zusammenhang zwischen Alter und traditionellen Rollenbildern ist also in den westlichen Bundesländern stärker als in den östlichen Bundesländern
  - Das negative Vorzeichen bedeutet: Mit einem höheren Wert für X (also das Alter) wird der höhere Y-Wert (progressiv, y_2) unwahrscheinlicher!
  - Das Ergebnis bedeutet: Das Alter einer Person lässt uns im Westen besser vorhersagen / erklären, welche Einstellung diese Person gegenüber traditionellen Rollenbildern hat!

Drittvariablenkontrolle - Bsp. Suppression

- Forschungsfrage: Hat Bildung von Befragten einen Einfluss auf die Einschätzung ihrer Partizipationsmöglichkeiten im politischen System?
- Argument: Je höher die Bildung, desto höher die Wahrnehmung politischer Einflussmöglichkeiten
- Datensatz: Hypothetisch

Drittvariablenkontrolle - Bsp. Suppression

	Niedrige Bildung	Hohe Bildung	$\sum$
Wenig Partizipationsmöglichkeit	2400 (60%)	1200 (60%)	3600
Viel Partizipationsmöglichkeit	1600 (40%)	800 (40%)	2400
$\sum$	4000 (100%)	2000 (100%)	**6000**

- $OR_{gesamt} = \frac{Odds_{y_1}}{Odds_{y_2}} = \frac{\frac{2400}{1600}}{\frac{1200}{800}} = 1$
- Kein Zusammenhang!
- ?

Drittvariablenkontrolle - Bsp. Suppression

- Mögliche Erklärung: Regime, in dem Befragte/r lebt, hat Einfluss auf den Zusammenhang
- **Autoritäre** Staaten: Höhere Bildung ⇒ Wahrnehmung von **weniger** Partizipationsmöglichkeiten
- **Demokratische** Staaten: Höhere Bildung ⇒ Wahrnehmung von **mehr** Partizipationsmöglichkeiten

Drittvariablenkontrolle - Bsp. Suppression

Autoritäre Staaten	Niedrige Bildung	Hohe Bildung	$\sum$
Wenig Partizipationsmöglichkeit	800 (40%)	900 (90%)	1700
Viel Partizipationsmöglichkeit	1200 (60%)	100 (10%)	1300
$\sum$	2000 (100%)	1000 (100%)	**3000**

- $OR_{autoritaer} = \frac{Odds_{y_1}}{Odds_{y_2}} = \frac{\frac{800}{1200}}{\frac{900}{100}} = 0.07 \Rightarrow$ Yule's Q ≈ -0.87
- Sehr starker Zusammenhang!

Drittvariablenkontrolle - Bsp. Suppression

Demokratische Staaten	Niedrige Bildung	Hohe Bildung	$\sum$
Wenig Partizipationsmöglichkeit	1600 (80%)	300 (30%)	1900
Viel Partizipationsmöglichkeit	400 (20%)	700 (70%)	1100
$\sum$	2000 (100%)	1000 (100%)	**3000**

- $OR_{demokratisch} = \frac{Odds_{y_1}}{Odds_{y_2}} = \frac{\frac{1600}{400}}{\frac{300}{700}} = 9.30 \Rightarrow$ Yule's Q ≈ 0.81
- Sehr starker Zusammenhang!

Drittvariablenkontrolle - Bsp. Suppression

- Bildung hat also einen Einfluss auf die Wahrnehmung von Partizipationsmöglichkeiten, dieser variiert aber in der Richtung je nach Kontext (des politischen Systems)
- Ohne Drittvariablenkontrolle kein Aufschluss über diesen Zusammenhang!

Drittvariablenkontrolle - ACHTUNG!

- Bislang Durchführung sehr vereinfachter Drittvariablenkontrollen
  - Drittvariablen mit mehr Ausprägungen ⇒ wesentlich mehr Partialtabellen
  - Mehr Tabellen ⇒ weniger überschaubar!
- Scheinkausalität / Suppression kommen in Realität selten so eindeutig vor!
- Zudem: oft Kontrolle nicht nur für eine, sondern für mehrere Drittvariablen als potentielle Einflussfaktoren wünschenswert!

Drittvariablenkontrolle - ACHTUNG!

- Berechnung bislang nur für nominalskalierte Dummy-Variablen
- Für ordinal skalierte Variablen: Berechnung der Zusammenhangsmaße jeweils für die einzelnen Kategorien von Z
  - D.h., für alle Werte von Z ($z_1, z_2, \dots$) Berechnung der jeweiligen Zusammenhänge zwischen X und Y anhand von Goodman und Kruskals Gamma, Kendalls tau b etc.
- In der Praxis daher meist direkt Anwendung multivariater Regressionsverfahren für nominal- oder ordinalskalierte (abhängige) Variablen
  - logistische Regression
  - ordered logit-Regression
  - Aber: das alles erst in $\Rightarrow$ Statistik II !