Statistik I - Sitzung 10

Bernd Schlipphak

Institut für Politikwissenschaft

Sitzung 10

1. Wiederholung und Weiterführung: Drittvariablenkontrolle
   - Wiederholung: Theoretische Einführung
   - Drittvariablenkontrolle: Multivariate Regression

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 für die Möglichkeit der gleichzeitigen Effekte zweier unabhängiger Variablen ODER für die Effekte durch dritte (Kontroll-)Variablen auf einen bivariaten Zusammenhang kontrolliert werden.
- Die multivariate Regression ist also EINE Möglichkeit der **Drittvariablenkontrolle**

Die Drittvariablenkontrolle

- **Drittvariablenkontrolle** = Überprüfung des Einflusses einer Variablen 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)
  - Einflussmediation: Kontrolle von **Interaktionseffekten** (Model 3)

Die multivariate Regression

- Eine multivariate Regression kontrolliert, ob unser Effekt zwischen zwei Variablen tatsächlich weiter auftritt, wenn wir andere Variablen mit in die Regressionsgleichung (d.h. in unser theoretisches Modell) aufnehmen

Die multivariate Regression

- Die daraus entstehende multivariate Gleichung lautet dann:

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + b_4 * x_4 + e$$

- Die generelle multivariate Gleichung ist

$$y = b_0 + b_1 * x_1 + b_k * x_k + e \text{ mit } k = \text{Anzahl der Variablen}$$

Die multivariate Regression

- Für die multivariate Regression gilt, dass die Herleitung und Interpretation des Determinationskoeffizienten R^2 und der Regressionskoeffizienten $b_1, b_2, b_3, ..., b_k$ gleich bleibt
- Zusätzlich gibt es aber
  - die standardisierten Regressionskoeffizienten (β oder Beta-Koeffizienten)
  - Probleme mit den Anwendungsvoraussetzungen der linearen Regression ($\Rightarrow$ Statistik II)

Standardisierte Regressionskoeffizienten

- In der multivariaten Regression unterscheidet man zwischen den unstandardisierten Regressionskoeffizienten $b_1, b_2, b_3, ..., b_k$ und den standardisierten Regressionskoeffizienten $\beta_1, \beta_2, \beta_3, ..., \beta_k$
- Die standardisierten Regressionskoeffizienten sind in ihrer Stärke untereinander vergleichbar
  - Über die Standardisierung wird die Einflussstärke der Koeffizienten auf den Mittelwert $= 0$ und eine Standardabweichung $= 1$ standardisiert.
  - Das bedeutet: Hat eine Variable X_1 einen höheren (positiven ODER negativen) Koeffizienten (β_1) als die Variable X_2, so übt Variable X_1 den stärkeren Einfluss auf die abhängige Variable (Y) aus ($X_1 > X_2$ weil $\beta_1 > \beta_2$)

Standardisierte Regressionskoeffizienten

- Die unstandardisierten Regressionskoeffizienten sind hingegen nicht direkt miteinander vergleichbar, drücken aber den Grad des individuellen Einflusses der Variable auf Y aus
  - Wenn b_1 für $X_1 = 0.5$, dann verändert sich Y für jede Einheit von X_1 um eine halbe (0.5) Einheit
  - Wenn b_1 für $X_1 = 0.3$, dann verändert sich Y für jede Einheit von X_1 um 0.3 Einheiten
  - Beispiel - Beide Variablen weisen $\%$ als Einheit auf: Wenn b_1 für $X_1 = 0.3$, dann verändert sich Y für jeden 1$\%$-Anstieg von X_1 um 0.3$\%$

Standardisierte Regressionskoeffizienten

- Geht es also in der Überprüfung eines theoretischen Arguments darum, die Stärke eines Einfluss einer bestimmten Variable zu überprüfen, so nutzt man den unstandardisierten Regressionskoeffizienten (=> X-Zentrierung)
- Will man hingegen herausfinden, in welcher Rangfolge eine abhängige Variable Y durch viele verschiedene Variablen erklärt wird, so nutzt man die standardisierten Regressionskoeffizienten (=> Y-Zentrierung)

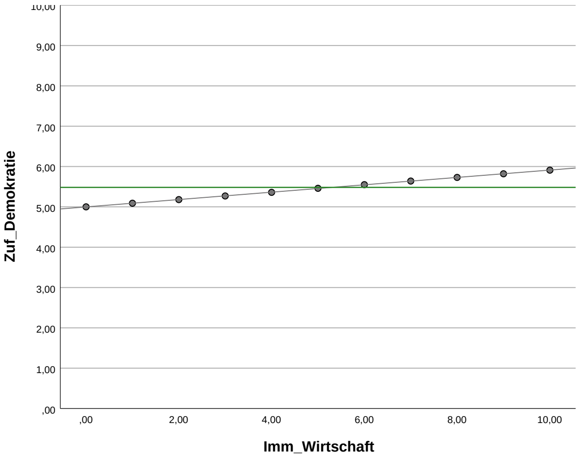
Ein Beispiel - Ausschnitt

Zufriedenheit mit Demokratie in D	b (unstandardisiert)	β (standardisiert)
Zufriedenheit mit eigenem Leben	.086	.082
Zufriedenheit mit Regierung	.335	.330
Wahrnehmung: Immigration gut für Wirtschaft	.091	.097
Religiosität	.049	**.069**
Hohe Bildung	.024	**.006**
Alter	.001	**.024**
Weiblich	-.032	**-.007**
Modellgüte (R^2)	49%	
Quelle: ESS 2012. Eigene Berechnung		

Ein Beispiel - Interpretation

- Die Variable "Wahrnehmung: Immigration gut für Wirtschaft" hat einen unstandardisierten Koeffizienten von .091. Der Effekt dieser Variable auf die AV beträgt also .091. Was bedeutet das?
  - Für jede Einheit, um die sich die X-Variable verändert, verändert sich der Y-Wert um .091 Einheiten
  - Y hat 11 Werte (von 0 bis 10), X hat ebenfalls 11 Werte (von 0 bis 10)
  - Für jeden Anstieg von X um einen Punkt auf einer elfstufigen Skala steigt Y um .091 Punkte auf einer elfstufigen Skala
  - Vergleicht man einen Befragten B1 mit dem X-Wert 0 (*Minimum*) mit einer Befragten B2 mit dem X-Wert 10 (*Maximum*), so weist B2 (im Durchschnitt) einen Y-Wert auf, der **fast einen ganzen Skalenpunkt (.9) höher** ist als jener von B1!

Ein Beispiel - Interpretation



Ein Beispiel - Weiterführung

- In Statistik II lernen wir dann mehr darüber
  - worin sich standardisierte und unstandardisierte Koeffizienten unterscheiden
  - warum wir für die Interpretation unstandardisierter Koeffizienten mehr Informationen brauchen, als das für die standardisierten Koeffizienten gilt
  - wie sich standardisierte Koeffizienten berechnen lassen