

Statistik II - Sitzung 7

Lena Masch

Institut für Politikwissenschaft

Sitzung 7

Statistik II - Sitzung 7

1 Interaktionen

2 Beispiele

Einleitung

- Ziel: Einführung in Interaktionsanalysen in der linearen Regression
- Ziel: Anwendung und Hintergründe verstehen

Logik einer Interaktion

- Definition: Ein Interaktionsterm misst, wie der Effekt einer unabhängigen Variable auf die abhängige Variable durch eine andere Variable beeinflusst oder moderiert wird.
- Regressionsmodell mit Interaktion: Der Interaktionsterm ($X_1 \times X_2$) zeigt an, ob die Wirkung von X_1 sich ändert, wenn X_2 sich ändert (variiert).
- eine Interaktion wird teilweise auch mit dem Begriff der Moderation beschrieben
- Moderation: Eine Variable (Moderator) beeinflusst die Stärke oder Richtung des Zusammenhangs zwischen einer unabhängigen und der abhängigen Variable.
- es ist v.a. eine theoretische Überlegung, welche Variable als Moderator fungiert
- Interaktionsanalysen als Analysen komplexer Zusammenhänge unter Berücksichtigung von Bedingungen und Kontexten (Möglichkeiten der Regression)

Allgemeine Überlegungen zu Interaktionstermen

- Signifikanz prüfen: Es ist wichtig, die statistische Signifikanz von Interaktionstermen zu überprüfen.
- Interpretation: Ein signifikanter Interaktionsterm deutet darauf hin, dass der Effekt einer unabhängigen Variable durch eine andere beeinflusst wird.
- Multikollinearität: Das Einfügen von Interaktionstermen kann die Multikollinearität erhöhen (Thema Sitzung 8).
- Zentrierung der Variablen: Vor der Bildung von Interaktionstermen kann es hilfreich sein, kontinuierliche Variablen zu zentrieren, um die Interpretation zu erleichtern (z.B. durch eine Mittelwertzentrierung).

Regressionsgleichung mit Interaktion

- Regression mit Interaktionsterm:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) + \epsilon$$

- Interpretation der Koeffizienten:

- ▶ β_1 : Effekt von X_1 , wenn $X_2 = 0$.
- ▶ β_2 : Effekt von X_2 , wenn $X_1 = 0$.
- ▶ β_3 : Änderung des Effekts von X_1 auf Y , abhängig von der Höhe von X_2 .

Interaktionen

- Interaktionsterme können zwischen zwei metrischen und/oder kategoriellen Variablen untersucht werden
- Dabei wird das Produkt aus den Variablen gebildet, u.U. bevor die Variablen in die Regression aufgenommen werden (abhängig vom Statistikprogramm)
- Interaktionen sollten stets grafisch inspiziert werden

Beispiel zur Veranschaulichung: Einkommen, Alter und Geschlecht

- Daten aus der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 2018
- Forschungsfragen: 1) Inwiefern hängt das monatliche Nettoeinkommen vom Alter der Befragten ab?
- Der Einfluss des Alters auf das Einkommen könnte durch das Geschlecht unterschiedlich stark ausgeprägt sein. 2) Moderiert das Geschlecht den Einfluss des Alters auf das Einkommen?
- Einkommen als monatliches Nettoeinkommen (in €)
- Alter in Jahren (18-95), Mittelwert= 51.7, SD = 17.6
- binäres Geschlecht: männlich (0), weiblich (1)

Beispiel zur Veranschaulichung: Einkommen, Alter und Geschlecht

- Wiederholung: Bivariate Regression des Einkommens auf Alter

call:

```
lm(formula = di01a ~ age, data = za)
```

Residuals:

Min	1Q	Median	3Q	Max
-1860.8	-842.1	-270.7	481.7	16217.4

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1584.087	77.300	20.493	< 2e-16 ***
age	3.970	1.421	2.794	0.00525 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1278 on 2645 degrees of freedom
(830 observations deleted due to missingness)

Multiple R-squared: 0.002942, Adjusted R-squared: 0.002565

F-statistic: 7.804 on 1 and 2645 DF, p-value: 0.00525

Beispiel zur Veranschaulichung: Einkommen, Alter und Geschlecht

- Wiederholung: Bivariate Regression des Einkommens auf Alter (Mittelwertzentrierung)

call:

```
lm(formula = di01a ~ agem, data = za)
```

Residuals:

Min	1Q	Median	3Q	Max
-1860.8	-842.1	-270.7	481.7	16217.4

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1789.248	24.837	72.040	< 2e-16 ***
agem	3.970	1.421	2.794	0.00525 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1278 on 2645 degrees of freedom
(830 observations deleted due to missingness)

Multiple R-squared: 0.002942, Adjusted R-squared: 0.002565

F-statistic: 7.804 on 1 and 2645 DF, p-value: 0.00525

Beispiel zur Veranschaulichung: Einkommen, Alter und Geschlecht

- Wiederholung: Multivariate Regression des Einkommens durch Alter und Geschlecht

call:

```
lm(formula = di01a ~ age + gender, data = za)
```

Residuals:

Min	1Q	Median	3Q	Max
-2265.4	-738.2	-204.8	450.5	15832.1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1932.212	76.042	25.410	< 2e-16 ***
age	4.713	1.348	3.495	0.000481 ***
genderweiblich	-813.044	47.188	-17.230	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1212 on 2644 degrees of freedom
(830 observations deleted due to missingness)

Multiple R-squared: 0.1036, Adjusted R-squared: 0.1029

F-statistic: 152.8 on 2 and 2644 DF, p-value: < 2.2e-16

Beispiel zur Veranschaulichung: Einkommen, Alter und Geschlecht

- Interaktion: Multivariate Regression des Einkommens durch Alter moderiert durch Geschlecht

Call:

```
lm(formula = di01a ~ gender * age, data = za)
```

Residuals:

Min	1Q	Median	3Q	Max
-2346.7	-744.9	-196.5	458.3	15835.3

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1766.533	99.014	17.841	< 2e-16 ***
genderweiblich	-449.476	147.133	-3.055	0.00227 **
age	7.963	1.835	4.340	1.48e-05 ***
genderweiblich:age	-7.048	2.702	-2.609	0.00914 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1210 on 2643 degrees of freedom

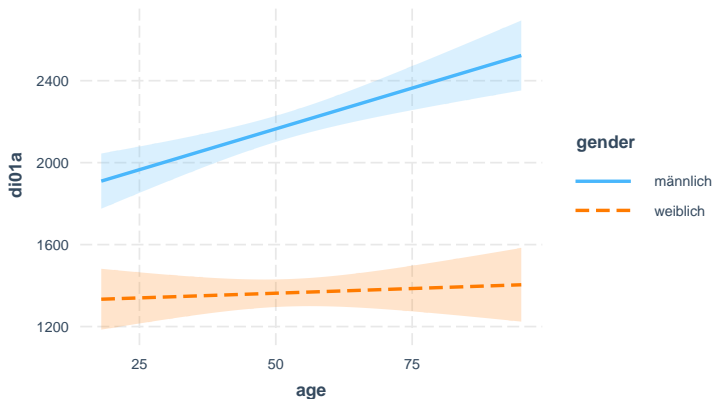
(830 observations deleted due to missingness)

Multiple R-squared: 0.1059, Adjusted R-squared: 0.1049

F-statistic: 104.3 on 3 and 2643 DF, p-value: < 2.2e-16

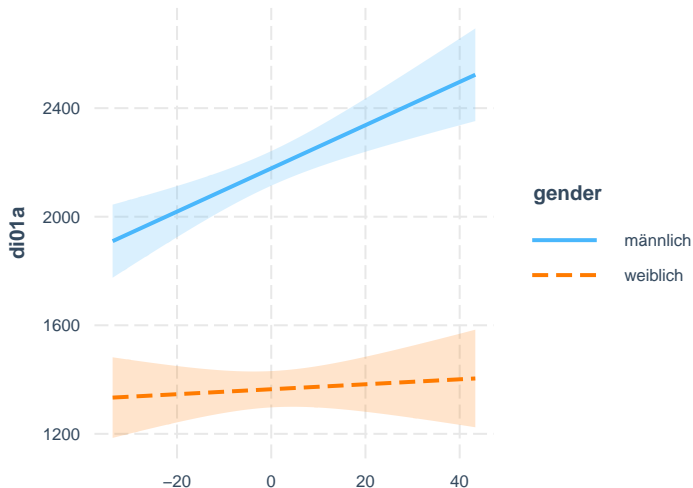
Beispiel zur Veranschaulichung: Einkommen, Alter und Geschlecht

- Interaktion: Multivariate Regression des Einkommens auf Alter moderiert durch Geschlecht



Beispiel zur Veranschaulichung: Einkommen, Alter und Geschlecht

- Interaktion: Multivariate Regression des Einkommens durch Alter (zentriert) moderiert durch Geschlecht



Beispiel zur Veranschaulichung: Einkommen, Alter und Geschlecht

- Regressionstabelle (mit und ohne Mittelwertzentrierung des Alters)

	<i>Dependent variable:</i>	
	di01a	
	(1)	(2)
genderweiblich	-449.476*** (147.133)	-813.717*** (47.137)
age	7.963*** (1.835)	
genderweiblich:age	-7.048*** (2.702)	
agem		7.963*** (1.835)
genderweiblich:agem		-7.048*** (2.702)
Constant	1,766.533*** (99.014)	2,178.046*** (32.504)
Observations	2,647	2,647
R ²	0.106	0.106
Adjusted R ²	0.105	0.105
Residual Std. Error (df = 2643)	1,210.462	1,210.462
F Statistic (df = 3; 2643)	104.339***	104.339***

kurze Pause

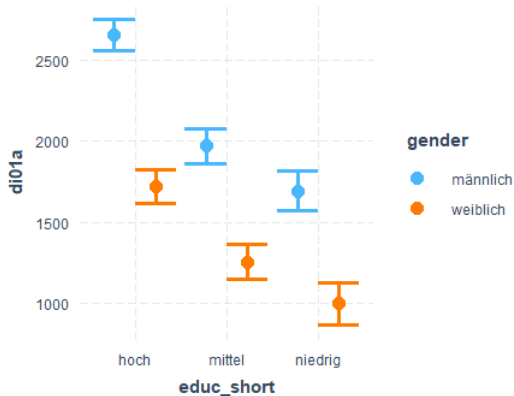
- 5min Pause!

Beispiel zur Veranschaulichung: Einkommen, Geschlecht und Bildung

- Wie beeinflussen Geschlecht und Bildung das Einkommen? Wird der Einfluss der Bildung durch das Geschlecht moderiert?
- Überspitzt: Lohnt sich Bildung für beide Geschlechter finanziell gleichermaßen?

Beispiel zur Veranschaulichung: Einkommen, Geschlecht und Bildung

- Wie beeinflussen Geschlecht und Bildung das Einkommen?



Beispiel zur Veranschaulichung: Einkommen, Geschlecht und Bildung

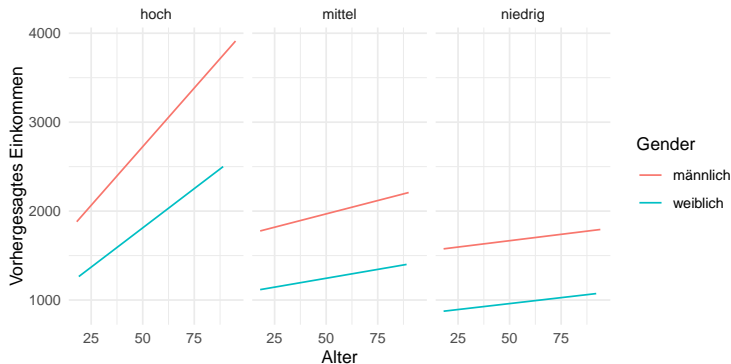
- Wie beeinflussen Geschlecht und Bildung das Einkommen?

<i>Dependent variable:</i>	
di01a	
genderweiblich	-934.354*** (71.807)
educ_shortmittel	-684.721*** (73.422)
educ_shortniedrig	-960.494*** (78.458)
genderweiblich:educ_shortmittel	220.472** (105.344)
genderweiblich:educ_shortniedrig	237.688** (115.685)
Constant	2,653.087*** (48.458)
Observations	2,621
R ²	0.180
Adjusted R ²	0.178
Residual Std. Error	1,160.972 (df = 2615)
F Statistic	114.710*** (df = 5; 2615)
Note:	*p<0.1; **p<0.05; ***p<0.01

Exkurs: Dreier-Interaktion: Alter, Geschlecht, Bildung

- Wird der Effekt des Alters auf das Einkommen durch Bildung und Geschlecht beeinflusst?

Dreier-Interaktion: Gender, Bildung und Alter

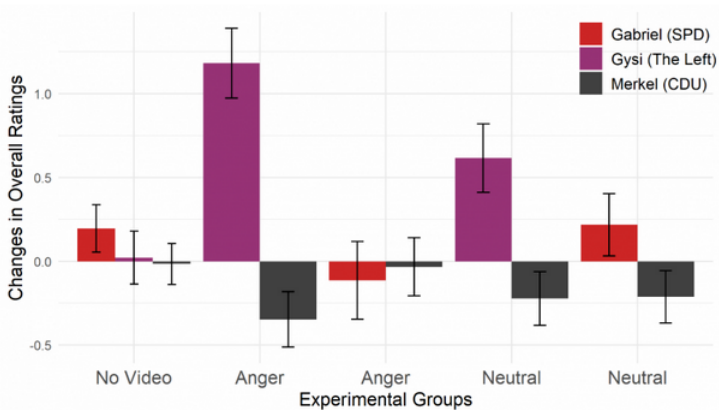


Diskussion

- Was sollten wir noch berücksichtigen oder testen?
- Überzeugt Sie die Operationalisierung des Alters?

Exkurs: Beispiel Masch 2020

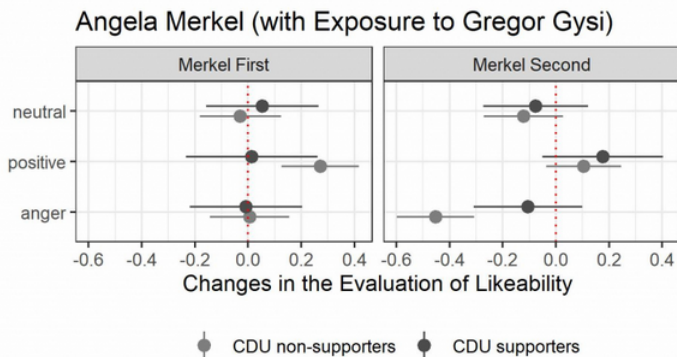
Figure 1: Changes in overall ratings of three German politicians by experiment group



Note: The chart is divided into different experimental groups. The first group was not shown any video material. The second group was shown Gysi and Merkel displaying anger. The third group was shown Gabriel and Merkel expressing anger. The fourth group was shown Gysi and Merkel with neutral expressions, while the final group was shown Gabriel and Merkel with neutral expressions. The bars indicate how impressions changed of each politician before and after the videos shown.

Exkurs: Beispiel Masch 2020

Figure 2: Changes in Merkel's likeability ratings according to treatment order and party identification



Note: The first box shows how impressions toward Angela Merkel changed when the video material of Merkel was shown before that of Gregor Gysi. The second box shows how impressions toward Merkel changed when video material of Gregor Gysi was shown first. In the second box there is a significant decrease in Merkel's likeability ratings which is not apparent in the first box.

Fazit

- Interaktionen ermöglichen die Untersuchung komplexer Zusammenhänge
- vielfältige Anwendungen in den Sozialwissenschaften
- Interaktionen helfen dabei zu verstehen, wie eine Variable den Effekt einer anderen Variable auf die abhängige Variable beeinflusst
- Die Zentrierung der metrischen Variablen kann die Interpretation erleichtern
- Die Interpretation sollte sorgfältig erfolgen und grafisch dargestellt werden

Ausblick

- Möglichkeiten und Limitationen der linearen Regression
- Annahmen der linearen Regression prüfen
- Konfidenzintervalle der Koeffizienten verstehen