

Multivariate Regressionsanalyse:

1a. Was ist unter Auspartialisierung zu verstehen und wieso ist es aufgrund der beteiligten Mechanismen wichtig immer mehrere Prädiktorvariablen zu berücksichtigen, auch wenn diese ggf. keinen Einfluss auf die abhängige Variable haben?

Die Auspartialisierung ist die Berechnung von einen Koeffizient und ergibt sich während die Partialisierung von multivariaten Modell unter anderem. Es ist ein sequenzielles Vorgehen der sequenziell Modelle nacheinander spezifiziert in dem zusätzliche Variable eingefügt werden. Die Idee bei der Auspartialisierung ist die Kontrolleffekte von Variablen zu identifizieren (im Sinne von Einfluss entweder verstärkt oder verliert) um die Wirkungszusammenhänge zwischen den Variablen sichtbar zu machen. Anders formuliert die Einflussstärke von jeder X-Variablen wird unabhängig von die Einflüsse von alle anderen unabhängige Variablen geschätzt in dem die X-Variable nur auf ihren Erklärungsbereich reduziert wird.

Die Auspartialisierung geschieht bei der Entstehung von mehrere Modelle, die immer weitere unabhängige Variablen integrieren, so genannte Prädiktorvariable. Es ist wichtig mehrere Prädiktorvariablen zu berücksichtigen, auch wenn diese ggf. keinen Einfluss auf die abhängige Variable haben, denn sie einen Suppressoreffekt ergeben. Es ist eine Möglichkeit um die Koeffizient der erste Variable zu spezifizieren.

NB: Prädiktorvariablen sind unabhängige Variable (X_i) und keine Antwortvariablen (Y).

1b. Wieso können unabhängige Variablen (x_i) im multiplen Regressionsmodell einen Einfluss auf Y haben, obwohl die bivariate Korrelation zwischen ihnen und Y nicht signifikant ist?

Unabhängige Variablen (x_i) im multiplen Regressionsmodell können einen Einfluss auf Y haben, obwohl die bivariate Korrelation zwischen ihnen und Y nicht signifikant ist. Denn die unabhängigen Variablen können untereinander korrelieren (also eine unabhängige Variable erklärt ein Teil der Effekt von eine andere abhängige Variable, sie partialisieren Anteile der UV) und damit Einfluss auf Y . Die Auspartialisierung zwischen unabhängige Variable verdeutlicht z.B. der Effekt von X_{i1} auf Y bei der Einführung von X_{i2} (wobei die bivariate Korrelation zwischen X_{i2} und Y nicht signifikant ist, aber signifikant bei der bivariate Korrelation zwischen X_{i1} und Y). Es ist den so genannten "Suppressoreffekt", der die "wahren" Effekte bzw. Anteile von eine unabhängige Variable identifiziert, die mit Y korrelieren. Es ergibt sich, dass die unabhängige Variable durch eine andere unabhängige Variable, auf den sie Einfluss hat, einen Einfluss letztendlich auf Y hat.

2. Bevor Sie die Analysen durchführen, suchen Sie im Codebuch (o. Variablenliste) Ihres Datensatzes (ALLBUS 2014) am besten Mittels STRG+F (aufrufen der „Suchenfunktion“ in

nahezuallen Programmen) die folgenden Variablen heraus: Alter, Geschlecht, Schulabschluss und individuelles Nettoeinkommen in der Fassung „Offene Angaben+Listeangaben“.

Kodieren Sie dann diese Variablen wie folgt:

- Alter: Startwert auf 0 setzen; 18=0, 48=30
- Schulabschluss- bzw. Schuldbildung: 5 Ausprägungen; 0=kein Schulabschluss, 1=HS, 2=RS, 3=FHR, 4=Abi; Rest=-1 bzw. Missing
- Geschlecht: 0=weiblich; 1=männlich

3 Berechnen Sie folgende (sequentielle) Regressionsmodelle:

- Modell a: Einkommen auf Alter;
- Modell b: Einkommen auf Bildung;
- Modell c: Einkommen auf Geschlecht;
- Modell ab: Einkommen auf Alter und Bildung;
- Modell abc: Einkommen auf Alter, Bildung und Geschlecht.

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	10.18*** (0.27)	8.05*** (0.25)	16.27*** (0.26)	5.48*** (0.41)	10.51*** (0.45)
ALTER	0.02*** (0.01)			0.04*** (0.01)	0.04*** (0.00)
ALLGEMEINERSCHULABSCHLUSS		0.93*** (0.07)		1.09*** (0.07)	1.14*** (0.07)
GESCHLECHT			-3.47*** (0.17)		-3.56*** (0.16)
R ²	0.00	0.05	0.12	0.07	0.20
Adj. R ²	0.00	0.05	0.12	0.07	0.20
Num. obs.	3064	3063	3065	3062	3062
RMSE	4.95	4.83	4.65	4.78	4.44

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 1: Statistical models

3a Vergleichen Sie die Regressionskoeffizienten über die Modelle und erläutern Sie was hier festzustellen ist!

Die Regressionskoeffizienten können über die Modelle verglichen werden, denn es handelt sich nicht um standardisierte Koeffizienten.

In jeder Modell sind die Effekte aller Variablen signifikant.

Die Regressionskoeffizienten der Variable Alter verändern sich von einem Modell zur anderen nur minimal mit einen Wert von 0,02 für Modell A Einkommen auf Alter zu 0,04 für die Modelle AB Einkommen auf Alter und Bildung und ABC Einkommen auf Alter, Bildung und Geschlecht. Diese Koeffiziente erklären der Effekt von Alter auf Einkommen für jeden zusätzlichen Jahr, also auf eine Skala von zehn Jahre ist die Werte von Alter größer 0,2 und 0,4. Der Effekt von Alter auf Einkommen ist positiv und immer höchst signifikant ($p < 0,001$). Der Effekt von Alter wird umso stärker wenn die Kontrollvariable Allgemeiner Schulabschluss im Modell hinzugefügt wird. Anders formuliert um so älter um so gebildeter.

Für die unabhängige Variable Allgemeiner Schulabschluss bzw. Bildung sind die Effekte stark, positiv auf den Einkommen. Bildung ist höchst signifikant ($p < 0,001$). Je gebildeter eine Person ist, desto höher ist den Einkommen. Außerdem, sind die Effekte von Allgemeiner Schulabschluss umso stärker, wenn die Kontrollvariable Geschlecht hinzugefügt wird.

Die Regressionskoeffizienten von der Variable Geschlecht haben einen sehr großen Effekt und sind höchst signifikant. In der Tabelle stehen negative Werte, die Interpretation soll so lauten, dass der Effekt von Geschlecht auf Einkommen negativ ist, wenn die Person eine Frau ist. Andersrum wäre der Effekt von Geschlecht auf Einkommen positiv, wenn die Person einen Mann ist. Der Wert von Geschlecht nimmt im Modell 5 bzw. ABC (Einkommen auf Alter, Bildung und Geschlecht) zu. Eine Interpretation wäre, dass je älter und gebildeter eine Person ist desto stärker ist der Effekt von Geschlecht auf Einkommen. Der negativen Effekt wirkt auf Frauen im Sinne, dass der Effekt von Geschlecht immer stärker wird über die Jahre auf dem Einkommen.

3b Vergleichen Sie R^2 über die Modelle und erläutern Sie was hier festzustellen ist!

Das Vergleichen von R^2 erklärt die Varianz, die ausgeschöpfte Varianz wird durch die beobachtende Varianzen erklärt. Diese Beobachtung ermöglicht die Analyse der Zusammenhang zwischen die unabhängige Variablen X mit der abhängige Variable Y.

In den bivariaten Modelle A, B und C zeigt R^2 , dass die unabhängige Variablen die Varianz vom Einkommen nicht in den gleichen Ausmaß erklären. R^2 im Model Einkommen auf Alter und Einkommen Einkommen auf Bildung ist kleiner als 0,1. Im Model Einkommen auf Geschlecht ist R^2 gleich 0,12, das Geschlecht erklärt sehr viel Varianzen vom Einkommen im Modell 3.

ist im Modell ABC am wichtigsten, denn R^2 steigt umso mehr Variablen im Modell eingefügt werden. Die Varianz von der abhängige Variable Einkommen wird immer besser erklärt. Nichtsdestotrotz beobachtet man einen Differenz von 0,13 für R^2 zwischen den Modell AB ($R^2 = 0,07$) und ABC ($R^2 = 0,20$). Das Einfügen von Geschlecht im Multi Regressionsmodell ermöglicht wieder eine bessere Erklärung der Varianz von Einkommen.