Übungsaufgabe 1 – Milena Rapp

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?

Unter Auspartialisierung versteht man die Bereinigung jeder unabhängigen Variable in einer multiplen Regression um die Einflüsse anderer X-Variablen. So kann eine Regression der abhängigen Variablen auf die bereinigten unabhängigen Variablen durchgeführt werden. Die Varianz von Y wird also nur auf den Anteil der X-Variablen zurückgeführt, auf den die anderen im Modell vorhandenen X-Variablen keinen Einfluss mehr ausüben.

Es ist wichtig mehrere Prädiktorvariablen zu berücksichtigen, da Suppressoreffekte vorhanden sein können, die den Effekt einer X-Variablen verstärken können.Das heißt, dass die Supressorvariable selbst keinen Effekt auf Y aufweist, jedoch steigt der Effekt der anderen X-Variable unter Einbezug der Supressorvariable. Das heißt, dass eine X-Variable (X2) mit einer anderen unabhängigen Variablen (x1), nicht jedoch mit der abhängigen Variablen korreliert sind. Durch Auspartialisieren der Anteile von x2, die nicht mit Y korreliert sind, aus X1 kann dann der „wahre“ Effekt freigelegt werden. Die Supressorvariable selbst weist also keinen Effekt auf Y aus, jedoch steigt der Effekt der anderen X-Variable unter Einbezug der Supressorvariable. Außerdem kann durch Einbezug weiterer X-Variablen die Effektstärke verringert, die Signifikanz aufgehoben oder die Effektrichtung verändert werden

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

Bei der bivariaten Korrelation wird lediglich ein Zusammenhang geprüft, während andere unabhängige Variablen im Hintergrund diesen Zusammenhang beeinflussen.

Die in der bivariaten Korrelation nicht-signifikante X1-Variabel kann im multiplen Regressionsmodell einen Einfluss auf Y haben, wenn die X1-Variable aus der bivariaten Korrelation aufgeschlüsselt nach bestimmten Gruppen (X2-Variable) unterschiedliche Wirkungsverläufe aufweist. Erst unter Kontrolle der X2-Variablen kann der kontrollierte Effekt der X1-Variable sichtbar werden. Daher ist ein sequentielles Vorgehen bei der Modellbildung wichtig.

3.

Bivariate Regressionen von Einkommen auf Alter, Bildung bzw. Geschlecht

	Einkommen~Alter	Einkommen~Bildung	Einkommen~Geschlecht
b	0,019***	1,052***	3,474***
b*	0,068	0,255	0,350
se	0,005	0,072	0,168
N	3064	3040	3065
R^2	0,005	0,065	0,122
Korr. R^2	0,004***	0,065***	0,122***

Anmerkungen: Signifikanz: *** p≤0,001, ** p≤0,05, * p≤0,10

Im ersten bivariaten Modell wird Einkommen auf das Alter zurückgeführt. Es besteht lediglich ein sehr schwacher Zusammenhang (b*=0,068). Mit jedem Lebensjahr steigt das Einkommen auf der Skala um 0,019 Einheiten an. Das Modell ist ebenfalls hochsignifikant, besitzt jedoch eine schwache Erklärungskraft, nur 0,4% der Varianz im Einkommen kann durch Kenntnis des Alters ausgeschöpft werden (korr. R^2 = 0,004).

Im zweiten bivariaten Modell wird Einkommen auf die Bildung zurückgeführt. Es besteht ein eher schwacher Zusammenhang (b*=0,255). Mit jeder Einheit auf der Bildungsskala steigt das Einkommen auf der gruppierten Einkommens-Skala um 1,052 Einheiten an (b=1,052). Das Modell ist ebenfalls hochsignifikant und weist eine schwache Erklärungskraft auf, durch Kenntnis des Bildungsabschlusses kann 6,5% der Varianz im Einkommen ausgeschöpft werden (korr. R^2 = 0,065).

Im dritten bivariaten Modell wird das Einkommen auf das Geschlecht zurückgeführt. Es besteht ein mäßig starker Zusammenhang (b*=0,350). Der Regressionskoeffizient ist hochsignifikant (b=3,474), für Männer steigt das Einkommen auf der entsprechenden Skala um 3,474 Einheiten an. Das Modell ist ebenfalls hochsignifikant. Mit einem korrigierten R^2 von 0,112 weist das Modell eine eher schwache Erklärungskraft auf, da durch Kenntnis des Geschlechts 11,2% der Varianz im Einkommen ausgeschöpft werden kann.

Sequentielle Regression von Einkommen auf Alter, Bildung und Geschlecht

	Modell 1			Modell 2			Modell 3		
	b	se	b*	b	se	b*	b	se	b*
Alter	0,019***	0,005	0,068	0,039***	0,005	0,135	0,040***	0,005	0,140
Bildung				1,199***	0,074	0,291	1,244***	0,069	0,302
Geschlecht							3,558***	0,160	0,359
N	3064			3039			3039		
R^2	0,005			0,082			0,211		
Korr. R^2 Sig. Gesamtmodell	0,004***			0,081***			0,210***		

Anmerkungen: Signifikanz: *** p≤0,001, ** p≤0,05, * p≤0,10

Modell 1 entspricht dem ersten bivariaten Modell, wird in dieser Tabelle aufgrund der sequentiellen Regressionsanalyse nochmals aufgeführt.

In Modell 2 sind die Regressionskoeffizienten sowohl für Alter als auch für Bildung hochsignifikant (b=0,039; b=1,199). Mit jedem Lebensjahr steigt das Einkommen auf der Einkommensskala um 0,039 Einheiten, mit jeder Einheit auf der Bildungsskala steigt das Einkommen auf der Einkommensskala um 1,199 Einheiten. Die Bildungsvariable (b*=0,291) weist einen stärkeren Einfluss auf, als die Altersvariable (b*=0,135). Der Einfluss der Altersvariable verstärkt sich geringfügig in Modell 2 von b=0,019 auf b=0,039 unter Kontrolle der Bildungsvariable. Mit einem korrigierten R^2 von 0,081 weist das Modell eine geringe Erklärungskraft aus, ist jedoch hochsignifikant.

Im dritten Modell wird zudem das Geschlecht als Kontrollvariable miteinbezogen. Die b-Koeffizienten aller drei im Modell enthaltenen unabhängigen Variablen sind hochsignifikant. Mit jedem Lebensjahr steigt das Einkommen auf der Einkommensskala um 0,04 Einheiten an, mit jeder Einheit auf der Bildungsskala um 1,244 Einheiten. Männer weisen ein um 3,558 Einheiten auf der Einkommensskala höheres Einkommen auf wie Frauen. Die Geschlechtsvariable weist den stärksten Einfluss auf das Einkommen aus (b*=0,359), dicht gefolgt von der Bildungsvariable (b*=0,302). Die Altersvariable weist einen deutlich geringeren Einfluss auf (b*=0,140). Das Modell ist insgesamt hochsignifikant und bindet 21% der Varianz in der Einkommensvariable.

Das korrigierte R^2 steigt über die Modelle hinweg deutlich an. Die Hinzunahme der Kontrollvariablen Bildung und Geschlecht führt zu einer Erhöhung der Varianzerklärung der Einkommensvariable.