Universität Stuttgart 31.10.2018

Institut für Sozialwissenschaften, SOWI IV

Tutorium Statistische Modellbildung

Wintersemester 2018/19 Marco Brenner-Mancebo

1a)

Variablen besitzen häufig gemeinsame Varianzanteile. Dies führt dazu, dass der Regressionskoeffizient b der unabhängigen Variablen sich durch Hinzunahme weiterer solcher Variablen mit gemeinsamem Varianzanteil verändert. Inhaltlich bedeutet das, dass der Einfluss einer unabhängigen Variablen (z. B. die Haupterklärungsvariable, die für Bestätigung/Widerlegung der Forschungshypothese maßgeblich ist) auf ihren alleinigen Anteil in Bezug auf die „Erklärungskraft“ der abhängigen Variablen isoliert wird. Der Einfluss von X auf Y ist im multivariaten Fall nicht mehr durch gemeinsame Varianzanteile der Variablen beeinträchtigt („verschmutzt“), sofern keine bedeutenden Kontrollvariablen vergessen wurden. Es ist daher wichtig immer mehrere Prädiktorvariablen zu berücksichtigen. Selbst wenn diese keinen Einfluss auf die abhängige Variable haben, sind sie wichtig, da eine solche Kontrollvariable denjenigen Varianzanteil anderer unabhängiger Variablen „auffängt“, der keinen Einfluss auf die abhängige Variable ausübt.

1b)

Beim bivariaten Fall könnte man simplifiziert sagen, dass die unabhängige Variable zwei Varianzanteile besitzt. Derjenige der einen systematischen Einfluss auf die abhängige hat und derjenige, der ihn nicht hat. Im bivariaten Fall wird es nun keine Korrelation geben, da der zweite Varianzanteil, der keinen systematischen Einfluss hat, den ersten bei der Berechnung des Zusammenhangskoeffizienten neutralisiert. Im multiplen Regressionsmodell kann die besagte unabhängige Variable nun einen Einfluss auf die abhängige Variable haben, sofern eine der Kontrollvariablen denjenigen Varianzanteil, der zuvor den Zusammenhang neutralisiert hat, auspartialisiert.

2)

Zuerst werden die Variablen herausgesucht:

* Alter: V84
* Geschlecht: V81
* Schulabschluss: V86
* Individueller Nettoeinkommen: V419

3a)

Das R-Script befindet sich in der .R-Datei im Anhang

Es fällt zuerst auf, dass alle Regressionskoeffizienten über alle Modelle hinweg hochsignifikant sind. Der Einfluss von Alter ist praktisch nicht vorhanden (Modell a). Bei Hinzunahme von Schulabschluss wird er marginal stärker, ist jedoch immer noch nahe Null (Modell ab). Der Einfluss des Schulabschlusses wird etwas stärker, wenn man um den Effekt von Alter kontrolliert (Modell ab). Dies liegt vermutlich darin begründet, dass jüngere häufiger Abitur und seltener einen Hauptschulabschluss haben (siehe Folien von SM II am 25.10.). Durch Hinzunahme der Variable Geschlecht wird der Einfluss von Schulabschluss erneut etwas stärker. Allerdings ist eine Zunahme von 0,04 ebenfalls nur sehr gering (Modell abc).

3b)

Das bivariate Modell a weist ein R² von Null auf (zumindest auf zwei Dezimalstellen gerundet). Das stärkste R² unter den bivariaten Modellen weist Modell c auf. In Modell ab wird deutlich, dass durch die Hinzunahme von Alter der Wert von R² um 0,01 steigt verglichen mit Modell b. Alter hat folglich ein wenig Varianzaufklärungspotential, dieses scheint in Modell a demnach der Rundung zum Opfer zu fallen. Das größte R² weist Modell abc mit einem Wert von 0,21 auf. Wie für die Berechnung des nicht angepassten R² üblich, steigt es mit zunehmender Anzahl unabhängiger Variablen an. Da die unabhängigen Variablen der bivariaten Modelle nur geringe gemeinsame Varianzanteile besitzen (vgl. die Änderung der partiellen Regressionskoeffizienten b) ist das unangepasste R² in Modell abc in diesem speziellen Fall relativ ähnlich der Summe der drei Einzelwerte von R² der bivariaten Modelle. Die Varianzanteile die sich gegenseitig auspartialisieren sind somit relativ gering.