Universität Stuttgart 10.11.2018
Institut für Sozialwissenschaften, SOWI IV
Tutorium Statistische Modellbildung
Wintersemester 2018/19 Marco Brenner-Mancebo

1)

Standardisierungen werden immer dann durchgeführt, wenn verschiedene Werte miteinander verglichen werden sollen, die in ihrer ursprünglichen Form nicht miteinander verglichen werden können (z. B. wegen unterschiedlicher Skalenbreiten). Bei der Berechnung eines standardisierten Wertes wird daher stets auf die Standardabweichung zurückgegriffen. Nach der z-Transformation weist eine Variable den Mittelwert Null (durch Subtraktion mit dem Mittelwert) und eine Standardabweichung und Varianz von Eins auf (durch Division mit der Standardabweichung). Bei der Berechnung des standardisierten Regressionskoeffizienten b* wird ebenfalls auf die Standardabweichung zurückgegriffen. Durch Multiplikation des unstandardisierten, partiellen Regressionskoeffizienten b mit der Standardabweichung der unabhängigen Variablen dieses Koeffizienten und der anschließenden Division durch die Standardabweichung der abhängigen Variablen des Modells erhält man b*.

In einem bivariaten Regressionsmodell von Einkommen auf Alter bedeutet ein standardisierter Regressionskoeffizient b* von 1,5, dass die abhängige Variable Einkommen um 1,5 Standardabweichungen zunimmt, wenn die unabhängige Variable Alter um eine Standardabweichung zunimmt (während der unstandardisierte Koeffizient b Aussagen über den Anstieg empirischer Einheiten erlaubt). In einem multivariaten Regressionsmodell können die Effektstärken der verschiedenen unabhängigen Variablen unter Zuhilfenahme von b* miteinander verglichen werden, wenn die zugrundeliegenden Skalenweiten dieser Variablen sich voneinander unterscheiden. Wenn alle Variablen innerhalb eines Regressionsmodells vor dessen Schätzung z-transformiert wurden, gleichen die Regressionskoeffizienten b den standardisierten Regressionskoeffizienten b*.

2a)

alter_z und alter_0z weißen beide einen Mittelwert von Null und eine Standardabweichung von Eins auf. Da es keinen Einfluss auf die z-Transformation hat, ob die Variable zuvor auf 0 gesetzt wurde oder nicht, bzw. diese Eigenschaft der z-Transformation üblich ist, war dies zu erwarten.

Auch die Zahlenwerte der beiden neu gebildeten Variablen alter_z und alter_0z sind identisch. Auch das ist nicht verwunderlich, da bei der z-Transformation beider Variablen unterschiedliche Mittelwerte zugrunde gelegt wurden, da sich diese vor der z-Transformation voneinander unterscheiden, da eine der Variablen zuvor auf 0 gesetzt wurde.

2b)

Der Regressionskoeffizient b der Altersvariable in Modell 2 ist mit einem Wert von 0,72 deutlich höher als derjenige der Altersvariable in Modell 1 mit einem Wert von 0,04. Das kommt daher, dass der Range der Altersvariable in Modell 2 mit einem Umfang von 4,17 deutlich geringer ist als der Range der Altersvariable in Modell 1 mit einem Umfang von 73. Da der partielle, unstandardisierte Regressionskoeffizient b Auskunft über empirische Einheiten gibt, ist ein Sprung um eine empirische Einheit auf der Altersvariable in Modell 2 deutlich größer als auf der Altersvariable in Modell 1. Folglich ist auch der entsprechende b-Wert (der Anstieg der abhängigen Variablen in empirischen Einheiten) deutlich größer. Der b-Wert der Bildungsvariablen in Modell 2 ist mit einem Wert von 1,36 ebenfalls größer als der b-Wert aus Modell 1 mit einem Wert von 1,09. Die „Vergrößerung" ist hier allerdings deutlich geringer als bei der Altersvariablen, da der Range der Bildungsvariablen deutlich weniger schrumpft, wenn diese z-Transformiert wird. Der Range der unstandardisierten Bildungsvariablen beträgt 6, der der standardisierten beträgt 4,8.

2c)

Wenn ich den Tipp aus Aufgabe 2c berücksichtige und Einkommen z-transformiere, dann gleichen sich die unstandardisierten Regressionskoeffizienten b mit den Standardisierten Regressionskoeffizienten b*. Das liegt daran, dass die unabhängigen sowie die abhängige Variable nun eine Standardabweichung von 1 aufweisen. Der unstandardisierte Regressionskoeffizient b wird bei seiner Standardisierung folglich mit 1 multipliziert und anschließend durch 1 dividiert. Er ändert sich bei der Standardisierung somit nicht.

3)

Modell 3 siehe angehängtes R-Script. R^2 beträgt 0,98 und das korrigierte R^2 0,95

Ich habe mich dazu entschieden das Modell einfach mit möglichst vielen Variablen „vollzumüllen" und habe nach dem Trial&Error-Prinzip diejenigen Variablen im Modell belassen, die mein R^2 erhöhen, bzw. diejenigen die das nicht taten wieder herausgenommen und somit das Modell iterativ aufgebläht.