Galiane Lavisse
Wintersemester 18-19

SM II Abgabe 2

1. Wozu werden Standardisierungen durchgeführt und wie wird dabei vorgegangen? Erläutern Sie exemplarisch wozu b* benutzt wird und wie man diesen inhaltlich interpretiert!

Standardisierungen beziehen sich auf die empirische Rohwerte der X- und Y- Variable innerhalb eines Modell. Die Variablen werden durch die Standardisierung auf einer gleiche Skala zugeordnet.
Bei einer Standardisierung werden die arithmetische Mittelwerte aller empirischen Variablen Werte berechnet. Folgedessen werden die jeweilige Mittelwerte von jedem einzelnen Variablenwert subtrahiert. Ein weiteren Schritt der Standardisierung ist die Standardabweichung (Wurzel aus der Varianz) für jeder erhaltene Wert durchzuführen: $Sx= \sqrt{(\sum(Xi- \bar{x})^2/N}$ .
Aus der Standardisierung kommen standardisierte Werte, die sich in den Wertbereich -1 und +1 befinden, die die Gleichung und Berechnung des standardisierten Regressionskoeffizient ermöglichen:
b^* = b * Sz/Sy. Die Standardisierung zuschreibt einem jeden Variablenwert ein fester Platz auf einer Standardskala. Das heisst, dass die Werte nicht mehr abhängig von Wertbereich oder der Skalenbreite sind (was für unstandardisierten Regressionsschätzung nicht der Fall ist). Da alle Koeffizienten (b*) auf einer gleiche Skala zugeordnet werden, werden sie miteinander vergleichbar bzw. b* wird benutzt um die Effektstärke von unterschiedliche skalierten X-Variablen in einer Schätzung zu vergleichen. Die inhaltliche Interpretation soll heissen, dass wenn sich X um eine Einheit verändert, ändert sich Y um b*-Einheiten.

**2.Führen Sie eine z-Standardisierung für die Originalaltersvariable (alter_z) und die auf Nullgesetzte Altersvariable (alter_0z) sowie für „unsere" Bildungsvariable (0 bis 4). [Daten: ALLBUS2014]**

**a) Vergleichen Sie die Zahlenwerte, Mean und die Standardabweichung von alter_z und alter_0z und erklären Sie Ihre „Beobachtung".**

Das Vergleichen der Zahlenwerte Mean und Standardabweichung ergibt, dass Mean in beide Modelle gleich 0 und die Standardabweichung (SD) gleich 0 ist. Dieses lässt sich dadurch erklären, dass die X-Variablen aus einer z-Standardisierung entstehen. Doch die standardisierte Koeffizienten bezeichnen das Ausmaß der Veränderung einer abhängigen Y-Variablen auf einer Standardskala von einer Mittelwert die gleich 0 ist, und die eine Standardabweichung die gleich 1 ist hat.

**b) Führen Sie eine Regression von Einkommen auf alter_0 und bildung (Modell 1) und eine Regression von einkommen_z auf alter_0z und bildung_z (Modell 2) durch und vergleichen Sie die b-Koeffizienten.**

| | Model 1 | Model 2 |
|---|---|---|
| (Intercept) | 6.21 *** | 0.00 |
| | (0.34) | (0.02) |
| alter0 | 0.04 *** | |
| | (0.01) | |
| bildung | 1.09 *** | |
| | (0.07) | |
| alter_0z | | 0.14 *** |
| | | (0.02) |
| bildung_z | | 0.29 *** |
| | | (0.02) |
| R^2 | 0.07 | 0.08 |
| Adj. R^2 | 0.07 | 0.08 |
| Num. obs. | 3062 | 3039 |
| RMSE | 4.78 | 0.96 |

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

Das Vergleich von den zwei Regressionsmodelle führt zu der Schlussfolgerung, dass die b-Koeffizienten in den zwei Modell nicht die gleiche Werte haben, aber die Einflussrichtung und die Signifikanz bleibt unverändert. Die Veränderung zwischen der nicht standardisiert Koeffizient des Alters und der standardisierter Koeffizient ist, dass die Werte der Alter grösser wird: von 4% zu 14%. Anders formuliert zeigt den standardisierten Koeffizient, dass Alter mehr Einflussstärke im Modell 2 hat.

Die Standardisierung der Variable Bildung ändert den Koeffizient ziemlich viel, da er größer als 1 war. Da die Werte von standardisierte Koeffiziente sich zwischen -1 und +1 befinden, war eine Veränderung zu erwarten. Der Einfluss von Bildung sinkt von 1,09 auf 0,29.

Dabei ist es wichtig zu sehen, dass die zwei Variablen Alter und Bildung im Modell 2 weniger Abstand in deren Einflussstärke haben auf Einkommen.

**c) Wie erklären Sie die Werte b und “b^*” in Modell 2? TIPP: Verwenden Sie bei Modell 2 das z-transformierte Einkommen als abhängige Variable**

**3. Erstellen Sie ein multivariates Regressionsmodell mit Y=Einkommen. Versuchen Sie dabei den R2-Wert so groß wie nur irgendwie möglich zu bekommen. Jeder schmutzige Trick der Sozialforschung ist erlaubt (und in diesem Fall erwünscht). Fügen Sie die entsprechenden Teile des SPSS-Outputs in Ihre Abgabe ein.**

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
|---|---|---|---|---|---|---|---|
| (Intercept) | 11.14 *** | 11.17 *** | 16.42 *** | 16.16 *** | 16.16 *** | 13.89 *** | 0.00 *** |
| | (0.09) | (0.09) | (0.25) | (0.32) | (0.32) | (0.30) | (0.00) |
| alter_0z | 0.34 *** | 0.68 *** | 0.70 *** | 0.32 ** | 0.32 ** | 0.12 | 0.00 |
| | (0.09) | (0.09) | (0.08) | (0.10) | (0.10) | (0.09) | (0.00) |
| bildung_z | | 1.45 *** | 1.51 *** | 1.31 *** | 1.31 *** | 1.11 *** | 0.00 * |
| | | (0.09) | (0.08) | (0.10) | (0.10) | (0.09) | (0.00) |
| geschl | | | -3.56 *** | -4.28 *** | -4.28 *** | -3.10 *** | -0.00 ** |
| | | | (0.16) | (0.19) | (0.19) | (0.17) | (0.00) |
| V489 | | | | 0.00 *** | 0.00 *** | -0.00 ** | -0.00 |
| | | | | (0.00) | (0.00) | (0.00) | (0.00) |
| V419 | | | | | | 0.00 *** | 0.00 *** |
| | | | | | | (0.00) | (0.00) |
| V418 | | | | | | | 1.00 *** |
| | | | | | | | (0.00) |
| R^2 | 0.00 | 0.08 | 0.21 | 0.36 | 0.36 | 0.50 | 1.00 |
| Adj. R^2 | 0.00 | 0.08 | 0.21 | 0.36 | 0.36 | 0.50 | 1.00 |
| Num. obs. | 3064 | 3039 | 3039 | 1904 | 1904 | 1904 | 18 |
| RMSE | 4.95 | 4.74 | 4.40 | 4.03 | 4.03 | 3.57 | 0.00 |

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