

Regorz Statistik

Nachhilfe – Beratung – Tutorials

Klausurtraining: Umrechnung Effektcodierung in Dummycodierung (Regression)

Arndt Regorz (B.Sc.Psychologie & Dipl. Kaufmann)

Stand: 09.05.2018

Mit diesem Klausurtraining soll das Bearbeiten von Aufgaben zu kategorialen Prädiktoren im Rahmen der multiplen Regression geübt werden.

Konkret geht es um Aufgaben folgenden Typs:

Ausgangspunkt ist eine Effektcodierung, deren Regressionsgewichte sind gegeben.

Es gibt Fragen zu den Gruppenmittelwerten.

Es gibt Fragen zu den Gewichten, wenn man das Gleiche in Dummycodierung abbildet.

Zunächst wird an zwei Lösungsbeispielen gezeigt, wie man aus einer Effektcodierung die Gruppenmittelwerte errechnet und wie man daraus anschließend für die gleichen Daten die Gewichte bei Dummycodierung ermittelt.

Anschließend gibt es drei Übungsaufgaben, um das Erlernte auszuprobieren: zunächst noch mit Hilfestellungen zu den einzelnen Schritten, dann nur noch mit Aufgabe und Lösung.

Lösungsbeispiel 1

Es wird der Einfluss des Bildungsgrades auf Krankheitsausfall untersucht. Dabei hat der Bildungsgrad drei Stufen: Bis Mittlere Reife, Fachabitur/Abitur, FH-/Universität. Effektcodierung, FH/Uni ist Referenzkategorie. Im Regressionsmodell x_1 für den kategorialen Prädiktor bei mittlerer Reife, Modell:
 $y = 4 - x_1 + 2 x_2$

Aufgabe A: Ermitteln Sie die Mittelwerte der drei Gruppen und die Regressionsgleichung bei Dummycodierung (mit der gleichen Referenzkategorie).

In einer MC-Klausur würde es eher Fragen zu den Gruppenmittelwerten und zu Gewichten der Dummycodierung geben, aber diese kann man problemlos beantworten, wenn man die Mittelwerte und die Dummycodierung ermittelt hat, z.B.:

Aufgabe B: Bei Dummycodierung ist das Regressionsgewicht b_2 kleiner als bei Effektcodierung

Lösungsschritt 1:

Zuordnung der Stufen zu den Variablen herauschreiben

FH/Uni ist Referenzkategorie, also keiner Variable zugeordnet

x_1 : mittlere Reife

damit bleibt nur noch:

x_2 : Fachabitur/Abitur

Lösungsschritt 2:

Mittelwerte der Gruppen ausrechnen

a) Für mittlere Reife (zugehörige Variable x_1):

$x_1=1$ und $x_2=0$ in die Gleichung einsetzen

$$y = 4 - 1 * 1 + 2 * 0 = 3$$

b) Für Fachabitur/Abitur (zugehörige Variable x_2):

$x_1=0$ und $x_2=1$ in die Gleichung einsetzen

$$y = 4 - 0 * 1 + 2 * 1 = 6$$

c) Für FH/Uni (Referenzkategorie)

Sowohl für $x_1 = -1$ als auch für $x_2 = -1$ einsetzen

$$y = 4 - 1 * (-1) + 2 * (-1) = 3$$

Lösungsschritt 3:

Parameter bei Dummycodierung bestimmen

a) Intercept (b_0): Mittelwert der Referenzkategorie

$$b_0 = 3 \text{ (FH/Uni)}$$

b) b_1 : Differenz zwischen Mittelwert von Gruppe der Variable x_1 und Referenzkategorie

Mittelwert mittlere Reife: 3

Mittelwert FH/Uni (Referenzkat.): 3

$$b_1 = 3 - 3 = 0$$

c) b_2 : Differenz zwischen Mittelwert von Gruppe der Variable x_2 und Referenzkategorie

Mittelwert Fachabitur/Abitur: 6

Mittelwert FH/Uni (Referenzkat.): 3

$$b_2 = 6 - 3 = 3$$

Lösungsschritt 4:

Regressionsgleichung aufstellen und damit ggf. MC-Fragen beantworten

Aufgabe A

Mittelwert Gruppe mittlere Reife 3

Mittelwert Fachabitur/Abitur 6

Mittelwert FH/Uni 3

Regressionsgleichung mit den unter 4. ermittelten Parametern:

$$y = 3 + 0 \cdot x_1 + 3 \cdot x_2$$

Aufgabe B, das Item war:

Bei Dummycodierung ist das Regressionsgewicht b_2 kleiner als bei Effekcodierung

Bei Effekcodierung war das Gewicht $b_2 = 2$, bei Dummycodierung ist $b_2 = 3$, also ist das Item falsch und nicht anzukreuzen.

Lösungsbeispiel 2

Es wird der Einfluss der Wohnregion auf Lebenszufriedenheit untersucht. Dabei gibt es für die Region drei Stufen: Süd, Nord-West, Ost. Effektcodierung, Ost ist Referenzkategorie.

Im Regressionsmodell steht x_1 für den kategorialen Prädiktor bei Region Süd

Modell:

$$y = 2 + 0 \cdot x_1 - 1 \cdot x_2$$

Aufgabe: Ermitteln Sie die Mittelwerte der drei Gruppen und die Regressionsgleichung bei Dummycodierung (mit der gleichen Referenzkategorie).

Lösungsschritt 1:

Zuordnung der Stufen zu den Variablen herausschreiben

Ost ist Referenzkategorie, also keiner Variable zugeordnet

x_1 : Süd

damit bleibt:

x_2 : Nord-West

Lösungsschritt 2:

Mittelwerte der Gruppen ausrechnen

a) Für Süd (zugehörige Variable x_1):

$x_1=1$ und $x_2=0$ in die Gleichung einsetzen

$$y = 2 + 0 \cdot 1 - 1 \cdot 0 = 2$$

b) Nord-West (zugehörige Variable x_2):

$x_1=0$ und $x_2=1$ in die Gleichung einsetzen

$$y = 2 + 0 \cdot 0 - 1 \cdot 1 = 1$$

c) Ost (Referenzkategorie)

Sowohl für $x_1 = -1$ als auch für $x_2 = -1$ einsetzen

$$y = 2 + 0 \cdot (-1) - 1 \cdot (-1) = 3$$

Lösungsschritt 3:
Parameter bei Dummycodierung bestimmen

a) Intercept (b_0): Mittelwert der Referenzkategorie

$$b_0 = 3 \text{ (Ost)}$$

b) b_1 : Differenz zwischen Mittelwert von Gruppe der Variable x_1 und Referenzkategorie

Mittelwert Süd: 2

Mittelwert Ost (Referenzkat.): 3

$$b_1 = 2 - 3 = -1$$

c) b_2 : Differenz zwischen Mittelwert von Gruppe der Variable x_2 und Referenzkategorie

Mittelwert Nord-West: 1

Mittelwert Ost (Referenzkat.): 3

$$b_2 = 1 - 3 = -2$$

Lösungsschritt 4:
Regressionsgleichung aufstellen

Regressionsgleichung mit den unter 4. ermittelten Parametern:

$$y = 3 - x_1 - 2 x_2$$

Ab hier kommen Aufgaben, an denen Sie sich selbst versuchen können.

Bei der ersten Aufgabe noch mit Hilfestellungen zu den einzelnen Schritten, später dann nur noch die Aufgabe und die Lösung.

Aufgabe 1

Es wird der Einfluss des Studienfachs auf die Anzahl der Fehlversuche untersucht. Dabei drei Stufen, Medizin, Psychologie, Germanistik; Effektcodierung, Germanistik Referenzkategorie. Im Regressionsmodell x_1 für den kategorialen Prädiktor bei Medizin
Modell

$$y = 2 + 3x_1 - 2x_2$$

Aufgabe: Ermitteln Sie die Mittelwerte der drei Gruppen und die Regressionsgleichung bei Dummycodierung (mit der gleichen Referenzkategorie).

Lösungsschritt 1:

Zuordnung der Stufen zu den Variablen herausschreiben

-> Lösung für diesen Schritt auf der nächsten Seite

Germanistik ist Referenzkategorie, also keiner Variable zugeordnet
x1: Medizin
damit bleibt:
x2: Psychologie

weiter mit

Lösungsschritt 2:

Mittelwerte der Gruppen ausrechnen

-> Lösung für diesen Schritt auf der nächsten Seite

a) Für Medizin (zugehörige Variable x_1):
 $x_1=1$ und $x_2=0$ in die Gleichung einsetzen

$$y = 2 + 3 * 1 - 2 * 0 = \mathbf{5}$$

b) Psychologie (zugehörige Variable x_2):
 $x_1=0$ und $x_2=1$ in die Gleichung einsetzen

$$y = 2 + 3 * 0 - 2 * 1 = \mathbf{0}$$

c) Germanistik (Referenzkategorie)
Sowohl für $x_1 = -1$ als auch für $x_2 = -1$ einsetzen

$$y = 2 + 3 * (-1) - 2 * (-1) = \mathbf{1}$$

weiter mit

Lösungsschritt 3:

Parameter bei Dummycodierung bestimmen

-> Lösung für diesen Schritt auf der nächsten Seite

a) Intercept (b_0): Mittelwert der Referenzkategorie

$b_0 = 1$ (Germanistik)

b) b_1 : Differenz zwischen Mittelwert von Gruppe der Variable x_1 und Referenzkategorie

Mittelwert Medizin: 5

Mittelwert Germanistik (Referenzkat.): 1

$b_1 = 5 - 1 = 4$

c) b_2 : Differenz zwischen Mittelwert von Gruppe der Variable x_2 und Referenzkategorie

Mittelwert Psychologie: 0

Mittelwert Germanistik (Referenzkat.): 1

$b_2 = 0 - 1 = -1$

weiter mit

Lösungsschritt 4:

Regressionsgleichung aufstellen

-> Lösung für diesen Schritt auf der nächsten Seite

Regressionsgleichung mit den unter 4. ermittelten Parametern:

$$y = 1 + 4 x_1 - x_2$$

weiter, jetzt ohne Hilfestellung, mit

Aufgabe 2:

Es wird der Einfluss des Lieblingsgetränks auf die Reaktionszeit untersucht. Dabei drei Stufen, Mineralwasser, Wein, Bier; Effektcodierung, Bier Referenzkategorie

Im Regressionsmodell x_1 für den kategorialen Prädiktor bei Mineralwasser Modell

$$y = 2 - 1 x_1 + 0 x_2$$

Aufgabe: Ermitteln Sie die Mittelwerte der drei Gruppen und die Regressionsgleichung bei Dummycodierung (mit der gleichen Referenzkategorie)

Lösung Aufgabe 2

Mittelwert Mineralwasser = 1
Mittelwert Wein = 2
Mittelwert Bier = 3

Regressionsgleichung mit Dummycodierung:
 $y = 3 - 2x_1 - x_2$

weiter mit
Aufgabe 3:

Es wird der Einfluss des Altersgruppe von Schülern am Gymnasium auf das Ausmaß der Störungen im Unterricht untersucht. Die Stufen sind:

Unterstufe, Mittelstufe, Oberstufe. Effektcodierung, Oberstufe ist Referenzkategorie

Im Regressionsmodell x_1 für den kategorialen Prädiktor der Unterstufe

Modell

$$y = 3 - 2x_1 + 4x_2$$

Aufgabe: Ermitteln Sie die Mittelwerte der drei Gruppen und die Regressionsgleichung bei Dummycodierung (mit der gleichen Referenzkategorie)

Lösung Aufgabe 3

Mittelwert Unterstufe = 1
Mittelwert Mittelstufe = 7
Mittelwert Oberstufe = 1

Regressionsgleichung mit Dummycodierung:
 $y = 1 + 0 \cdot x_1 + 6 \cdot x_2$

Impressum:

Arndt Regorz
Alemannenstraße 6
44793 Bochum
mail@regorz-statistik.de
www.regorz-statistik.de

Wie kann ich Sie weiter unterstützen?

Nachhilfe & Prüfungsvorbereitung Statistik

Statistik kann man umständlich und formel-lastig erklären, wie es viele Hochschulen leider tun. Und man kann Statistik so erklären, dass es verständlich ist. Wenn Ihnen mein Erklärungs-Stil liegt und Sie Nachhilfe in Statistik benötigen, finden Sie auf meiner Seite zu [Statistik-Nachhilfe](#) weitere Infos.

Beratung für Datenauswertung bei Bachelorarbeit oder Masterarbeit

Welche Auswertungen sind für Ihre Fragestellung richtig und was müssen Sie dabei beachten? Schon in einer Stunde (Telefon/Skype/vor Ort) kann man viele Fragen klären. Auf meiner Seite zu [Statistik-Beratung](#) finden Sie weitere Informationen.