

Aufgabenblatt 15 (8.7.2010)

1. Durch Y wird die Lohnhöhe, durch X das Geschlecht erfasst ($X = 0$ bei Männern, $X = 1$ bei Frauen). Ein lineares Regressionsmodell liefert: $M(Y|X = x) = 1300 - 300x$.
 - a) Welches Regressionsmodell würde sich ergeben, wenn man die Kodierung der X -Variablen umkehrt (Männer = 1, Frauen = 0)?
 - b) Welches Regressionsmodell würde sich ergeben, wenn man die Kodierung der X -Variablen folgendermaßen verändert: Männer = 1, Frauen = 2?
 - c) Welches Regressionsmodell würde sich ergeben, wenn man die Kodierung der X -Variablen folgendermaßen verändert: Männer = -1, Frauen = +1?
2. Durch Y wird die Lohnhöhe, durch X das Geschlecht und durch Z das Alter erfasst. Bilden Sie zwei lineare Regressionsmodelle für $M(Y|X = x, Z = z)$, einmal ohne, einmal mit einem Interaktionseffekt der beiden Regressorvariablen. Erklären Sie, warum die Einbeziehung eines Interaktionseffekts sinnvoll sein kann.
3. Es sei X eine qualitative Variable mit den Werten 1 (vollzeit beschäftigt), 2 (teilzeit beschäftigt), 3 (arbeitslos), 4 (nicht mehr im Erwerbsleben). Außerdem wird durch Y das Einkommen, durch Z das Geschlecht ($M = 0$, $F = 1$) erfasst.
 - a) Definieren und beschreiben Sie die Dummy-Variablen, durch die man X erfassen kann.
 - b) Es seien folgende Mittelwerte gegeben:

X	$M(Y X = x)$
1	2400
2	1900
3	1200
4	1600

Formulieren Sie ein lineares Regressionsmodell für $M(Y|X = x)$ mit Hilfe der Dummy-Variablen und geben Sie die Parameterwerte an.

Tabelle 1 Datensatz zur Erwerbsbeteiligung verheirateter Frauen in 1971 (Arminger, 1983).

X	Y	Z	N	M	X	Y	Z	N	M
1	1	450	32	16	3	2	1000	14	9
1	2	450	96	52	3	3	1000	15	13
1	3	450	57	43	1	1	1500	207	100
2	1	450	16	5	1	2	1500	1246	927
2	2	450	35	13	1	3	1500	1126	1022
2	3	450	26	17	2	1	1500	617	178
1	1	700	383	132	2	2	1500	2036	1581
1	2	700	1155	640	2	3	1500	2420	2118
1	3	700	793	607	3	1	1500	23	10
2	1	700	217	47	3	2	1500	51	39
2	2	700	461	260	3	3	1500	109	95
2	3	700	364	265	1	1	2000	32	16
3	1	700	3	1	1	2	2000	162	143
3	3	700	1	0	1	3	2000	153	147
1	1	1000	845	329	2	1	2000	199	106
1	2	1000	4398	2925	2	2	2000	820	722
1	3	1000	3359	2838	2	3	2000	960	908
2	1	1000	913	242	3	1	2000	23	16
2	2	1000	2926	1874	3	2	2000	102	94
2	3	1000	2877	2384	3	3	2000	217	209
3	1	1000	13	1					

- c) Bilden Sie zwei lineare Regressionsmodelle für $M(Y|X = x, Z = z)$, wobei X durch Dummy-Variablen erfasst wird, einmal ohne, einmal mit einem Interaktionseffekt der Regressorvariablen.
4. Die Variablen für die Daten in Tabelle 1 haben folgende Bedeutung: X Schulbildung (1 = nur Volksschule, 2 = mittlere Ausbildung, 3 = höhere Ausbildung); Y Kinder (1 = keine Kinder, 2 = Kinder älter als 6 Jahre, 3 = Kinder jünger als 6 Jahre); Z Einkommen des Mannes. N gibt die Gesamtzahl der Frauen, M die Anzahl der davon erwerbstätigen Frauen an. Weiterhin wird eine Variable B verwendet, die den Wert 1 hat, wenn eine Frau erwerbstätig ist, und die andernfalls den Wert 0 hat.
 - a) Berechnen und interpretieren Sie $P(B = 1|X = 2, Y = 3, Z = 1000)$.
 - b) Berechnen und interpretieren Sie $P(B = 1|X = 2, Y = 3, Z = 1000) - P(B = 1|X = 2, Y = 2, Z = 1000)$.
 - c) Bilden (definieren und beschreiben) Sie für X drei Dummy-Variablen.
 - d) Bilden (definieren und beschreiben) Sie für Y drei Dummy-Variablen.

Tabelle 2 Parameterwerte für das Logitmodell.

Variable	Parameter	Wert
	$\hat{\alpha}$	1.2030
X_2	$\hat{\beta}_1$	0.1638
X_3	$\hat{\beta}_2$	-0.0089
Y_2	$\hat{\beta}_3$	-1.4531
Y_3	$\hat{\beta}_4$	-2.4119
Z_2	$\hat{\beta}_5$	-0.0679
Z_3	$\hat{\beta}_6$	-0.4847
Z_4	$\hat{\beta}_7$	-0.9392
Z_5	$\hat{\beta}_8$	-1.7951

- e) Bilden (definieren und beschreiben) Sie für Z fünf Dummy-Variablen.
f) Betrachten Sie ein Logit-Modell

$$\hat{P}(B = 1 | \dots) = \frac{\exp(v)}{1 + \exp(v)}$$

wobei v durch

$$v = \alpha + x_2\beta_1 + x_3\beta_2 + y_2\beta_3 + y_3\beta_4 + x_2\beta_5 + z_3\beta_6 + z_4\beta_7 + z_5\beta_8$$

definiert ist. Tabelle 2 zeigt Parameterwerte für das mit den Daten aus Tabelle 1 geschätzte Modell.

- g) Berechnen Sie einerseits mit dem Modell und andererseits mit den Daten in Tabelle 1 Werte für die Anteile erwerbstätiger Frauen für folgende Konstellationen der Regressorvariablen: $X=1, Y=1, Z=1$; und $X=2, Y=2, Z=1, 2, 3, 4, 5$; und $X=2, Y=3, Z=1, 2, 3, 4, 5$. Interpretieren Sie die Ergebnisse.
h) Erklären Sie, warum sich $P(B = 1 | \dots)$ und $\hat{P}(B = 1 | \dots)$ bei gleichen Werten der Regressorvariablen unterscheiden.