1 Kodierung qualitativer Einflussgrößen

Aufgabe 1

Man betrachte ein Regressionsproblem mit einer kategorialen Einflussgröße $A \in \{1, 2, \ldots, J\}$. Bestimmen Sie, ausgehend von der Strukturannahme

$$\mu = E(y|A) = \beta_0 + \beta_1 x_{A(1)} + \ldots + \beta_{J-1} x_{A(J-1)},$$

die Parameter β_0, $\beta_1, \ldots, \beta_{J-1}$ in Abhängigkeit der Erwartungswerte $\mu_j = E(y|A = j)$ für

(a) Dummykodierung,

(b) Effektkodierung.

Wie kann man jeweils sinnvoll einen Parameter β_J definieren?

Aufgabe 2

Für das Regressionsmodell ohne Konstante

$$E(y|A) = \beta_1 x_{A(1)} + \ldots + \beta_J x_{A(J)}$$

soll eine Kodierung entwickelt werden, deren Parameter β_j sich als Zuwächse beim Übergang von Kategorie $j - 1$ nach Kategorie j interpretieren lassen.

(a) Überlegen Sie sich, wie die $\mu_j = E(y|A = j)$ in Abhängigkeit der β_j geschrieben werden können.

(b) Geben Sie nun die Kodierung an.

(c) Für welche Art von Einflussgrößen erscheint eine solche Kodierung besonders geeignet? Wie lassen sich in diesem Fall insbesondere Tests bezüglich $H_0 : \beta_j = 0$ interpretieren?

Aufgabe 3

Wir betrachten ein lineares Modell, dessen linearer Prädiktor η die einfache Form $E(y|x) = \eta = \beta_0 + \beta_1 x_1$ besitzt. Nun wird eine zweite Kovariable x_2 eingeführt, die wie folgt definiert ist: $x_2 = 1.8 \cdot x_1 + 32$. Welche Auswirkungen hätte die gleichzeitige Aufnahme von x_1 und x_2 in das Modell? Begründen Sie Ihre Antwort.