1 Kodierung kategorialer 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

Der Datensatz `lesen` untersucht den Zusammenhang zwischen dem Verhalten von Grundschülern und der Anzahl an Fehler bei einem Lesetest und enthält (unter anderem) folgende Variablen:

`Fehlerzahl`	Anzahl an Fehler beim Lesetest
`sex`	Geschlecht (1: männlich, 0: weiblich)
`Jahrgang`	Klassenstufe (1: 3.Klasse, 0: 4.Klasse)
`Lesezeitmin`	Lesezeit in der Schule
`WieoftLesen`	Wie oft wird gelesen? (1: oft, ..., 5: fast nie)

Laden Sie den Datensatz von der Vorlesungshomepage herunter und lesen Sie diesen in `R` ein. Fitten Sie ein Modell mit den angegebenen Kovariablen und der `Fehlerzahl` als abhängiger Variable. Vergleichen Sie dabei für die Kovariable `WieoftLesen` die verschiedenen Kodierungsarten aus den Aufgaben 1 und 2 und interpretieren Sie die Koeffizienten.