

3 Generalisierte lineare Modelle

Aufgabe 5

Hypothesentests über den (p -dim.) Koeffizientenvektor β eines GLM sind meist linear und können in der Form

$$H_0 : \mathbf{C}\beta = \mathbf{d} \quad \text{gegen} \quad H_1 : \mathbf{C}\beta \neq \mathbf{d}$$

dargestellt werden ($\mathbf{C} : r \times p$, $r \leq p$, $\text{rg}(\mathbf{C}) = r$, $\mathbf{d} : r \times 1$).

- (a) Sei nun $\beta = (\beta_0, \dots, \beta_7)^T$. Wie lauten \mathbf{C} und \mathbf{d} , wenn folgende Hypothesen getestet werden sollen?
- (i) $H_0 : \beta_2 = 0$
 - (ii) $H_0 : \beta_2 = \beta_4 = \beta_6 = 0$
 - (iii) $H_0 : \beta_1 = \beta_3$
 - (iv) $H_0 : 2\beta_1 = \beta_2, 3\beta_1 = \beta_3, 4\beta_1 = \beta_4$
- (b) Nehmen Sie an bei β_1, \dots, β_4 handelt es sich um die Koeffizienten einer dummy-kodierten ordinalen Einflussgröße mit Levels $k = 0, 1, \dots, 4$, wobei Kategorie 0 als Referenz-Kategorie verwendet wird. Welcher Modellierungsansatz wird dann in Teilaufgabe (a)(iv) überprüft?
- (c) Geben Sie die drei Teststatistiken, die in der GLM-Theorie verwendet werden können, um die Hypothesen aus Teilaufgabe (a) zu überprüfen, in allgemeiner Form an. In welcher Situation empfiehlt sich die Verwendung jeweils welcher Statistik?

Im Folgenden wird wieder der Datensatz `fakesoep` von Übungsblatt 4 und ein GLM mit gammaverteiltem Response, allen Kovariablen (Haupteffekte) und log-Link betrachtet.

- (d) Leiten Sie für dieses Modell die LR-Teststatistik, die Wald-Teststatistik und die Score-Teststatistik zur Überprüfung der Hypothese in Teilaufgabe (a)(i) explizit (so weit wie möglich) her.
- (e) Führen Sie die drei Tests aus Teilaufgabe (d) in `R` durch.
- (f) Führen Sie in `R` jeweils einen LR-Test zur Überprüfung der Hypothesen in Teilaufgabe (a)(ii) und (a)(iii) durch.