

3 Generalisierte lineare Modelle (III)

Aufgabe 1

Man betrachte wieder die `fakesoep`-Daten von Blatt 4 und ein GLM unter Annahme einer Gamma-Verteilung und Einschluss aller Prädiktoren (Haupteffekte).

- (a) Führen Sie einen LQ-Test zur Überprüfung der beiden Hypothesen $H_0 : \beta_{\text{alter}} = 0$ und $H_0 : \beta_{\text{deutsch}} = 0$ durch. Dabei sollen die Hypothesen zunächst einzeln und im Anschluss gemeinsam getestet werden.
- (b) Leiten Sie allgemein die Wald-Teststatistik zur Überprüfungen der Hypothese $H_0 : \beta_j = 0$ her. Überprüfen Sie im Anschluss $H_0 : \beta_{\text{deutsch}} = 0$ durch einen Wald-Test.

Aufgabe 2

Der Datensatz `leafblotch` (McCullagh & Nelder, 1989; Download von der Veranstaltungshomepage) untersucht die Auswirkungen einer Blattkrankheit. Er enthält folgende Variablen:

<code>blotch</code>	Anteil des Befalls an der Oberfläche der Blätter
<code>site</code>	Anbau-Ort (kategorial, 9 Orte)
<code>variety</code>	Gerstensorte (kategorial, 10 Sorten)

Die Response-Variable `blotch` ist nicht binomial, aber auf das Intervall $[0, 1]$ beschränkt, da es sich um Anteile handelt. Daher erscheint eine Modellierung mit einem Quasi-Likelihood-Ansatz sinnvoll, bei dem die Strukturannahme für den Erwartungswert über den Logit-Link spezifiziert wird.

- (a) Fitten Sie ein Haupteffektmodell mit den Prädiktoren `site` und `variety` über einen Quasi-Likelihood-Ansatz mit Logit-Link, der die gleiche Varianzstruktur wie das Logit-Modell verwendet (also $v(\mu) = \mu(1 - \mu)$). Was fällt bei der Betrachtung des Outputs auf?
- (b) Verdeutlichen Sie anhand eines Plots, dass die in (a) gewählte Varianzfunktion nicht geeignet ist. Was hat die (offensichtliche) Fehlspezifikation der Varianzfunktion für Konsequenzen?
- (c) Versuchen Sie die Varianz Ihres Parameter-Schätzers aus (a) zu schätzen. Fassen Sie hierzu die Varianzfunktion $v(\mu) = \mu(1 - \mu)$ lediglich als Arbeits-Varianz auf.
- (d) Eine alternative Wahl für die Varianzfunktion wäre $v(\mu) = \mu^2(1 - \mu)^2$. Diese ist im R-Package `gnm` implementiert (setze `family=wedderburn`). fitten Sie das entsprechende Modell und betrachten Sie den Output sowie einen Plot analog zu (b). Halten Sie die Varianzfunktion $v(\mu) = \mu^2(1 - \mu)^2$ für besser geeignet als $v(\mu) = \mu(1 - \mu)$ (Begründung)?
- (e) Zeigen Sie für das Modell aus (d) unter der Annahme, dass es sich bei $v(\mu) = \mu^2(1 - \mu)^2$ um die wahre Varianzfunktion handelt: Die Varianzen der einzelnen Komponenten von $\hat{\beta}$, die sich aus der asymptotischen Verteilung von $\hat{\beta}$ ergeben, sind unabhängig vom wahren Koeffizientenvektor β .
- (f) Sind auch die geschätzten Standardfehler, wie sie im Output in (d) zu finden sind, unabhängig von der Schätzung $\hat{\beta}$?