Florian Wickelmaier
Forschungsmethoden und Mathematische Psychologie
Fachbereich Psychologie
Universität Tübingen

Übungsaufgaben
Generalized linear models

Erzeugen Sie ein lauffähiges und kommentiertes R-Skript, in dem Sie die folgenden Aufgaben bearbeiten.

a) Der Datensatz `cars` in R enthält Geschwindigkeiten von 50 Autos, deren Bremswege gemessen wurden. Schätzen Sie zunächst das Regressionsmodell

$$dist_i = \beta_0 + \beta_1 speed_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \text{ i.i.d.} \tag{1}$$

Wie viel Varianz der Bremsdistanzen wird durch die Geschwindigkeit erklärt? Diskutieren Sie, ob die Modellannahmen erfüllt sind. Hinweis: Verwenden Sie die Plot-Methode des Objekts.

Schätzen Sie dann das Modell

$$dist_i = \beta_0 + \beta_1 speed_i + \beta_2 speed_i^2 + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \text{ i.i.d.} \tag{2}$$

Hinweis: Verwenden Sie `I(speed^2)` in der Modellformel. Wählen Sie eines der beiden Modelle und begründen Sie Ihre Entscheidung.

b) In einem psychophysischen Experiment werden einer Versuchsperson zwei LEDs präsentiert: Ein Standardreiz mit $40\,\mathrm{cd/m^2}$ und ein Vergleichsreiz mit variierenden Intensitäten. Die Aufgabe der Versuchsperson ist es zu entscheiden, ob der Vergleichsreiz heller als der Standardreiz ist. Jede Vergleichsintensität x wird 40 mal dargeboten. Die Anzahl der positiven Antworten ("Vergleichsreiz heller als Standard") können der folgenden Tabelle entnommen werden:

x ($\mathrm{cd/m^2}$)	37	38	39	40	41	42	43
y (positiv)	2	3	10	25	34	36	39

Schätzen Sie mit `glm()` die Parameter c und d der logistischen psychometrischen Funktion

$$p_{pos} = \frac{1}{1 + \exp(-\frac{x - c}{d})}.$$

Das Modell kann als GLM geschrieben werden: Es sei $logit(p_{pos}) = \beta_0 + \beta_1 x$, dann gilt $c = -\beta_0/\beta_1$ und $d = 1/\beta_1$.

Führen Sie einen Anpassungstest für das Modell durch ($\alpha = 0.05$).

Schätzen Sie die Intensität, für die $p_{pos} = 0.5$ gilt (Punkt der subjektiven Gleichheit).

Erstellen Sie eine Abbildung der Wahrscheinlichkeit einer positiven Antwort in Abhängigkeit von der Intensität des Vergleichsreizes. Zeichnen Sie in diese Abbildung die logistische psychometrische Funktion ein. Verwenden Sie dazu `predict()`, um die vorhergesagten Werte zu bestimmen. Zeichnen Sie mit `abline()` die Schätzung für den Parameter c ein.

c) Ein alternatives Modell für die Vergessensdaten von Myung (2003, siehe Folien) ist das Exponentialmodell:

$$p(w, t) = w_1 \exp(-w_2 t), \quad \text{mit } w_1, w_2 > 0.$$

Schätzen Sie dieses Modell mit Hilfe der Maximum-Likelihood-Methode. Verwenden Sie dafür `nlm()` oder `optim()`.

Bestimmen Sie approximative 95%-Konfidenzintervalle für die Parameter.

Beschreibt das Exponential- oder das Potenzmodell die Daten besser?

d) In einem Signal-Entdeckungs-Experiment werden einer Person Signal- und Rausch-Durchgänge dargeboten. Sie soll dabei jedes Mal entscheiden, ob sie ein Signal erkennen kann oder nicht:

```
        response
trial    yes  no
  noise   89 361
  signal 330 120
```

In der Signal-Entdeckungs-Theorie (SDT) wird das Sensitivitätsmaß d' definiert als die Differenz zwischen zwei internen Verteilungen (Signal- und Rauschverteilung). Das einfache SDT-Modell kann als Probit-Modell geschrieben werden: $\Phi^{-1}(p_{yes}) = \beta_0 + \beta_1 x$, wobei Φ^{-1} die Quantilfunktion der Standardnormalverteilung ist. In diesem Modell gilt $d' = \beta_1$.

Schätzen Sie d' mittels `glm(..., family = binomial(probit), ...)`.

Literatur

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47* (1), 90–100. doi: 10.1016/S0022-2496(02)00028-7