Formelsammlung

Silvan Hüsler

A James et al. 2013: Logistic Regression

$$
\begin{aligned}
&\Pr(\text{default} = \text{yes}|\text{balance}) \\
\text{p}(X) &= \beta_0 + \beta_1 X && \text{(4.1)} \\
\text{p}(X) &= \frac{e^{\beta_0+\beta_1 X}}{1 - e^{\beta_0+\beta_1 X}} && \text{(4.2: logistic function)} \\
\frac{\text{p}(X)}{1-\text{p}(X)} &= e^{\beta_0+\beta_1 X} && \text{(4.3)} \\
&= \frac{\text{p}(X)}{1-\text{p}(X)} && \text{(odds)} \\
\log\left(\frac{\text{p}(X)}{1-\text{p}(X)}\right) &= \beta_0 + \beta_1 X && \text{(4.4: lhs -> log-odds oder logit)}
\end{aligned}
$$

Wie wirkt sich eine Einheit mehr von X aus?

$$
\text{->}\ \frac{\text{p}(X)}{1-\text{p}(X)} \times e^{b_1}
$$

B West et al. 2014: Linear Mixed Models

3 Two-Level Models for Clustered Data: The Rat Pup Example

A General Model

$$
\begin{aligned}
\text{WEIGHT}_{ij} = \beta_0 &+ \beta_1 \times \text{TREAT1}_j + \beta_2 \text{TREAT2}_j + \beta_3 \times \text{SEX1}_{ij} + \\
&\beta_4 \times \text{LITSIZE}_j + \beta_5 \times \text{TREAT1}_j \times \text{SEX1}_{ij} + \\
&\beta_6 \times \text{TREAT2}_j \times \text{SEX1}_{ij} + \\
\\
&\text{u}_j + \varepsilon_{ij}
\end{aligned}
$$

Level 1 Model (Rat Pup)

$$
\text{WEIGHT}_{ij} = \beta_{0j} + \beta_{1j} \times \text{SEX1}_{ij} + \varepsilon_{ij}
$$

Level 2 Model (Litter)

$$
\begin{aligned}
\beta_{0j} &= \beta_0 + \beta_1 \times \text{TREAT1}_j + \beta_2 \text{TREAT1}_j + \beta_4 \text{LITSIZE}_j + \text{u}_j \\
\beta_{1j} &= \beta_3 + \beta_5 \times \text{TREAT1}_j + \beta_6 \text{TREAT2}_j
\end{aligned}
$$

6 Random Coefficient Models for Longitudinal Data: The Autism Example

A General Model

$$
\begin{aligned}
\mathrm{VSAE}_{ti} = \beta_0 + \beta_1 \times \mathrm{AGE2}_{ti} + \\
&\beta_2 \times \mathrm{AGE2SQ}_{ti} + \\
&\beta_3 \times \mathrm{SICDEGP1}_{i} + \\
&\beta_4 \times \mathrm{SICDEGP2}_{i} + \\
&\beta_5 \times \mathrm{SICDEGP1}_{i} \times \mathrm{AGE2}_{ti} + \\
&\beta_6 \times \mathrm{AGE2}_{ti} \times \mathrm{SICDEGP2}_{i} \times \beta_7 \times \mathrm{AGE2SQ} \times \mathrm{SICDEGP1}_{i} + \\
&\beta_8 \times \mathrm{AGE2SQ} \times \mathrm{SICDEGP2}_{i} +
\end{aligned}
$$

$$
\mathrm{u}_{0i} + \mathrm{u}_{1i} \times \mathrm{AGE2}_{ti} + \mathrm{u}_{2i} \times \mathrm{AGE2SQ}_{ti} + \varepsilon_{ti}
$$

Level 1 Model (Time)

$$
\mathrm{VSAE}_{ti} = \beta_{0i} + \beta_{1i} \times \mathrm{AGE2}_{ti} + \beta_{2i}\mathrm{AGE2SQ}_{ti} + \varepsilon_{ti}
$$

Level 2 Model (Child)

$$
\begin{aligned}
\beta_{0i} &= \beta_0 + \beta_3 \times \mathrm{SICDEGP1}_i + \beta_4 \times \mathrm{SICDEGP2}_i + \mathrm{u}_{0i} \\
\beta_{1i} &= \beta_1 + \beta_5 \times \mathrm{SICDEGP1}_i + \beta_6 \times \mathrm{SICDEGP2}_i + \mathrm{u}_{1i} \\
\beta_{2i} &= \beta_2 + \beta_7 \times \mathrm{SICDEGP1}_i + \beta_8 \times \mathrm{SICDEGP2}_i + \mathrm{u}_{2i}
\end{aligned}
$$

Merke: Wichtiger Aspekt bezüglich dem Verständnis der Modelle ist Kenntnis, wie Formeln den Datensatz transformieren!

```
form <- vsae ~
        age.2 +
        I(age.2^2) +
        sicdegp2.f +
        age.2*sicdegp2.f +
        I(age.2^2)*sicdegp2.f +
        (age.2 + I(age.2^2) | childid)

model_matrix(autism.updated, form) %>% glimpse

## Observations: 610
## Variables: 10
## $ (Intercept)                      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ age.2                            <dbl> 0, 1, 3, 7, 11, 0, 1, 3, 7, 1...
## $ I(age.2^2)                       <dbl> 0, 1, 9, 49, 121, 0, 1, 9, 49...
## $ sicdegp2.f1                      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ sicdegp2.f2                      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ age.2 + I(age.2^2) | childidTRUE <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ age.2:sicdegp2.f1                <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ age.2:sicdegp2.f2                <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ I(age.2^2):sicdegp2.f1           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
```

```
## $ I(age.2^2):sicdegp2.f2          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
```

C Steele 2005: Event History Analysis

2. Descriptive Event History Analysis

2.1 Definitions: hazard and survivor functions

2.2 Estimation of the hazard and survivor functions

$$
\begin{aligned}
\hat{h}_t &= \frac{d_t}{r_t - w_t/2} \qquad (2.1)\\
\hat{S}_1 &= 1\\
\hat{S}_t &= (1-\hat{h}_1)(1-\hat{h}_2)\ldots(1-\hat{h}_{t-1}) \text{ for } t > 1 \qquad (2.2)\\
&= \prod_{j=1}^{t-1}(1-\hat{h}_j)
\end{aligned}
$$

3. Continuous-time Event History Models

3.1 Introduction

3.2 Proportional hazards and accelerated life models

3.2.1 Proportional hazards model

$$
\begin{aligned}
h(t;\mathbf{x}_i) &= h_0(t)g(\mathbf{x}_i) \qquad (3.1)\\
\frac{h(t;\mathbf{x}_1)}{h(t;\mathbf{x}_2)} &= \frac{g(\mathbf{x}_1)}{g(\mathbf{x}_2)}
\end{aligned}
$$

3.3 Cox proportional hazards model

$$
\begin{aligned}
h(t;\mathbf{x}_i) &= h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}_i) \qquad (3.2)\\
\log h(t;\mathbf{x}_i) &= \log h_0(t) + \boldsymbol{\beta}'\mathbf{x}_i \qquad (3.3)
\end{aligned}
$$

4. Discrete-time Event History Models

4.1 The discrete-time approach

4.2 Data structure for a discrete-time analysis

$$
y_{ti} = \begin{cases} 0 & t < y_i \\ 0 & t = y_i,\ \delta_i = 1 \\ 1 & t = y_i,\ \delta_i = 0. \end{cases}
$$

4.3 Discrete-time models

$$\operatorname{logit}(h_{ti}) = \log\left(\frac{h_{ti}}{1-h_{ti}}\right) = \alpha(t) + \boldsymbol{\beta}'\mathbf{x}_{ti} \tag{4.1}$$

5. Unobserved Heterogeneity

5.1 The problem and implications of unobserved heterogeneity

5.2 Incorporating unobserved heterogeneity

5.2.1 The Cox model with frailty

$$h(t; \mathbf{x}_i, u_i) = h_0(t) u_i \exp\left(\boldsymbol{\beta}'\mathbf{x}_i\right) \tag{5.1}$$

5.2.2 The discrete-time logit model with frailty

$$\operatorname{logit}(h_{ti}) = \alpha(t) + \boldsymbol{\beta}'\mathbf{x}_{ti} + u_i \tag{5.2}$$

5.3 Example: age at first partnership

6. Repeated Events

6.1 Examples

6.2 Issues in the analysis of repeated events

6.3 A discrete-time model for repeated events

$$\operatorname{logit}(h_{tij}) = \alpha(t) + \boldsymbol{\beta}'\mathbf{x}_{tij} + u_i \tag{6.1}$$

7. Competing Risks

7.1 Introduction

7.2 Definitions

$$h^{(r)}(t) = \lim_{dt \to 0} \frac{\Pr(t \leq T < t + dt, R = r | T \geq t)}{dt}.$$

7.3 Models for competing risks

7.3.1 Separate models for each type of transition

$$\delta_i^{(r)} = \begin{cases} 1 & \text{if } r_i \neq r \\ 0 & \text{if } r_i = r. \end{cases}$$

$$\log h^{(r)}(t; \mathbf{x}_i^{(r)}(t)) = \log h_0^{(r)}(t) + \boldsymbol{\beta}^{(r)\prime}\mathbf{x}_i^{(r)}(t), r = 1, \cdots, k \tag{7.1}$$

$$y_{ti}^{(r)} = \begin{cases} 0 & t < y_i \\ 0 & t = y_i,\ r_i = r \\ 1 & t = y_i,\ r_i \neq r. \end{cases}$$

$$\begin{aligned} h_{ti}^{(r)} &= \Pr(y_{ti}^{(r)} = 1 | \mathrm{y}_{si} = 0, s < t) \\ logit(h_{ti}^{(r)}) &= \alpha^{(r)}(t) + \boldsymbol{\beta}^{(r)\prime}\mathbf{x}_{ti}^{(r)}, r = 1, \cdots, k \end{aligned} \tag{7.2}$$

7.3.2 Modelling event types simultaneously: the multinomial logit model

$$h_{ti}^{(0)} = \Pr(y_{ti} = 0 | y_{si} = 0,\ s < t) = 1 - \sum_{r=1}^{k} h_{ti}^{(r)} \tag{7.3}$$

$$\log\left(\frac{h_{ti}^{(r)}}{h_{ti}^{(0)}}\right) = \alpha^{(r)}(t) + \boldsymbol{\beta}^{(r)\prime}\mathbf{x}_{ti}^{(r)} \quad r = 1, \cdots, k. \tag{7.4}$$

$$h_{ti}^{(r)} = \frac{\exp[\alpha^{(r)}(t) + \beta^{(r)}{}_1 \mathrm{x}_{ti}^{(r)}]}{1 + \sum_{lr=1}^{k} \exp[\alpha^{(l)}(t) + \boldsymbol{\beta}^{(l)}{}_1 \mathbf{x}_{ti}^{(l)}]}, \quad r = 1, \cdots k, \tag{7.5}$$

$$\hat{S}_{1i} = 1; \hat{S}_{ti} = \prod_{j=1}^{t-1}(1 - \hat{h}_{ji}), t > 1, \tag{7.6}$$

$$\text{where } \hat{h}_{ji} = \sum_{r=1}^{k} \hat{h}_{ji}^{(r)}.$$

$$\hat{f}_{ti}^{(r)} = \hat{h}_{ti}^{(r)} \hat{S}_{ti} \tag{7.7}$$

$$\hat{\mathrm{F}}_{ti}^{(r)} = \sum_{j}^{t} \hat{f}_{ji}^{(r)} \tag{7.8}$$

D Durrant & Müller 2013: Modeling Call Record Data: Examples From Cross-Sectional and Longitudinal Surveys

12.3.2.1 Modeling the Outcome at a Particular Call

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta' x_i \tag{Model 1}$$

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta' x_{ij} + u_j \tag{Model 2}$$

12.3.2.2 Modeling the Outcome Across Calls (Discrete-Time Event History Analysis)

$$\log\left(\frac{\pi_{ti}}{1-\pi_{ti}}\right) = \alpha_t + \beta' x_{ti} \tag{Model 3}$$

12.3.2.3 Modeling the Outcome Across Calls for Repeated Events (Discrete-Time Event History Analysis)

$$\log\left(\frac{\pi_{ti}}{1-\pi_{ti}}\right) = \alpha_t + \beta' x_{ti} + h_i \tag{Model 4}$$

$$\log\left(\frac{\pi_{tij}^{(s)}}{\pi_{tij}^{(4)}}\right) = \alpha_t^{(s)} + \beta'^{(s)} x_{tij}^{(s)} + h_{ij}^{(s)} + u_j^{(s)} \tag{Model 5}$$

$$\log\left(\frac{\pi_{tij}^{(s)}}{\pi_{tij}^{(4)}}\right) = \alpha_t^{(s)} + \beta^{(s)} x_{tij}^{(s)} + \lambda^{(s)} h_{ij} + \gamma^{(s)} u_j \tag{Model 5$'$}$$

Literatur

Durrant, Gabriele B., Julia D'arrigo, and Gerrit Müller. 2013. "Modeling Call Record Data: Examples from Cross-Sectional and Longitudinal Surveys." In *Improving Surveys with Paradata*, edited by Frauke Kreuter, 281–308. Hoboken, New Jersey: John Wiley & Sons, Inc. doi:10.1002/9781118596869.ch12.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Springer New York. doi:10.1007/978-1-4614-7138-7.

Steele, F. 2005. "Event History Analysis." NCRM Methods Review Papers: NCRM/004.

West, Brady T., Kathleen B. Welch, and Andrzej T. Galecki. 2014. *Linear Mixed Models: A Practical Guide Using Statistical Software, Second Edition*. 2. ed. Hoboken: Taylor and Francis.