Within-cases analysis of binary responses[1]

STA442/2101 Fall 2017

[1] This slide show is an open-source document. See last slide for copyright information.

The idea

- There are several binary responses for each case.
- Like was the person employed right after graduation, 6 months after, one year after . . . Yes or No
- Or did the consumer purchase at least one computer in 2020, 2021, 2022 . . .
- Binary choices in laboratory studies can be repeated measures.
- Model: Logistic regression with a random shock for case, pushing all the log odds values for that case up and down by the same amount.
- Random shock is added to the regression equation for the log odds.
- Usually the random shock is normal — what else?

A random intercept model

For $i = 1, \ldots, n$ and $j = 1, \ldots, m$

- $B_1, \ldots, B_n \overset{i.i.d.}{\sim} N(0, \sigma^2)$
- Conditionally on $B_i = b_i$ for $i = 1, \ldots, n$, binary responses y_{ij} are independent with

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = (\beta_0 + b_i) + \beta_1 x_{ij1} + \ldots + \beta_k x_{ijk}$$

where $\pi_{ij} = P\{y_{ij} = 1\}$.

Some of the x_{ij} could be dummy variables for time period or treatment, different for $j = 1, \ldots, m$ within case i.

Law of Total Probability

Formula sheet: $Pr(A) = \sum_{j=1}^{k} Pr(A|B_j)Pr(B_j)$

$$Pr\{\mathbf{Y}_i = \mathbf{y}_i\} = \int_{-\infty}^{\infty} Pr\{\mathbf{Y}_i = \mathbf{y}_i | B_i = b_i\} f_{\sigma^2}(b_i)\, db_i$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \left(\prod_{j=1}^{m} Pr\{Y_{ij} = y_{ij} | B_i = b_i\} \right) f_{\sigma^2}(b_i)\, db_i \\
&= \int_{-\infty}^{\infty} \left(\prod_{j=1}^{m} \left(\frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+b_i}}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+b_i}} \right)^{y_{ij}} \left(1 - \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+b_i}}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+b_i}} \right)^{1-y_{ij}} \right) f_{\sigma^2}(b_i)\, db_i \\
&= \int_{-\infty}^{\infty} \frac{e^{\sum_{j=1}^{m} y_{ij}\mathbf{x}'_{ij}\boldsymbol{\beta}+b_i}}{\prod_{j=1}^{m}\left(1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+b_i}\right)} f_{\sigma^2}(b_i)\, db_i \\
&= \int_{-\infty}^{\infty} \frac{e^{mb_i+\sum_{j=1}^{m} y_{ij}\mathbf{x}'_{ij}\boldsymbol{\beta}}}{\prod_{j=1}^{m}\left(1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+b_i}\right)} f_{\sigma^2}(b_i)\, db_i
\end{aligned}$$

The Likelihood Function

$\prod_{i=1}^{n} Pr\{\mathbf{Y}_i = \mathbf{y}_i\}$ as a function of the model parameters

$$\ell(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \frac{e^{mb_i + \sum_{j=1}^{m} y_{ij} \mathbf{x}'_{ij} \boldsymbol{\beta}}}{\prod_{j=1}^{m} \left(1 + e^{\mathbf{x}'_{ij} \boldsymbol{\beta} + b_i}\right)} f_{\sigma^2}(b_i)\, db_i$$

Maximum likelihood

Numerical, of course

- In principle, this is mostly straightforward.
- It's all classical likelihood stuff.
- We just have a random intercept in this class.
- But the model can be extended to

$$\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$$

- Where **w** is a vector of log odds.
- That's what the `glmer` function in the `lme4` package does.

There are problems

- Nobody can do the integral.
- It's really brutal for multivariate normal **b** and complicated designs.
- The approximate solutions are imperfect.
- There are numerical issues, even in our simple case.
- For the general case, it's easy to specify models whose parameters are not identifiable.
- This does not apply to us, but there is massive confusion in the user community.

The `glmer` function in the `lme4` package

- Syntax is like `lmer` for linear models.
- And like `glm` for generalized linear models with fixed effects.
- We are going to keep it simple.
- Just add `+(1|Subject)` for the random shock (intercept).
- Use effect coding (`contr.sum`) if there are interactions between factors.
- `Anova(model,type='III')` from the `car` package to test each effect controlling for all others.
- For follow-up tests, fit a no-intercept model on a combination variable and test contrasts on the categories of the combination variable using the `linearHypothesis` function from the `car package`.

Copyright Information