29.11.12@20.00

4 GLMs

Summary 1 *binomial data, logistic function, logit transform, odds, logit regression, probit, latent variables, marginal effects, grouped data, identifiablility, overparametrization.*

4.1 Binary Regression

Transformational (e.g. logit) vs. latent variables (e.g. probit)

data: $\begin{bmatrix} y_1 & \cdots & y_n \\ x_1 & \cdots & x_n \end{bmatrix}$, $y_1, \dots y_n$ binary, $x_1, \dots, x_n$ on interval scale

general link function

$$\mathbb{P}\left(Y = 1 |\; x\right) = 1 - \mathbb{P}\left(Y = 0 |\; x\right) = F\left(\alpha + \beta x\right)$$

where F : sigmoidFunktion: monoton, stetig, $F\left(-\infty\right) = 0$, $F\left(\infty\right) = 1$

cases:

logistic distribution:	$F\left(x\right) = \frac{e^x}{1+e^x}$	$\Longrightarrow$ logit-model
standard normal distribution:	$F\left(x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right)\; du$	$\Longrightarrow$ probit-model

4.1.1 Logit, introduction

derivation by transformational approach: **logitRegMod**: $F\left(x\right) = \frac{e^x}{1+e^x} \Longrightarrow$

$$\mathbb{P}\left(Y = 1 \mid x\right) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}$$
$$\mathbb{P}\left(Y = 0 \mid x\right) = \frac{1}{1 + e^{\alpha+\beta x}}$$

Interpetation coefficients: odds ratio

notation: $\mathbb{P}\left(Y = 1 \mid x\right) = p\left(x\right)$
odds: $\frac{p(x)}{1-p(x)}$

1.**oddsFactor**

$$\frac{p(x+\Delta x)}{1-p(x+\Delta x)} = e^{\alpha+\beta(x+\Delta x)} = e^{\alpha+\beta x}e^{\beta\Delta x} = \frac{p(x)}{1-p(x)}e^{\beta\Delta x}$$

2. **%-change odds ratio**

$$\frac{\frac{p(x+\Delta x)}{1-p(x+\Delta x)} - \frac{p(x)}{1-p(x)}}{\frac{p(x)}{1-p(x)}} = \frac{\frac{p(x+\Delta x)}{1-p(x+\Delta x)}}{\frac{p(x)}{1-p(x)}} - 1 = e^{\beta\Delta x} - 1$$

3. logit: **log odds** linear

$$\text{logit}(p(x)) = \log\frac{p(x)}{1-p(x)} = \log e^{\alpha+\beta x} = \alpha + \beta x$$

NB:

$$e^{\beta\Delta x} - 1 \approx \beta\Delta x \text{ for } \Delta x \to 0$$

4.1.2 Probit, introduction

Interpretation via *latent variables approach*:

$$y^* = \alpha + \beta x + \epsilon$$

example: random utility function

observation

$$y = \begin{cases} 1 & \text{if } \ y^* > 0 \\ 0 & \text{if } \ y^* \le 0 \end{cases}$$

therefore

$$\epsilon \sim \mathcal{N}(0,1) \implies \mathbb{P}(Y=1\ |x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\alpha+\beta x} e^{-\frac{1}{2}t^2}dt \implies \text{probit}$$

Proof:

$$\mathbb{P}(Y=1\ |x) = \mathbb{P}(Y^*>0\ |x) = \mathbb{P}(\alpha+\beta x+\epsilon>0\ |x) = \mathbb{P}(\epsilon > -\alpha-\beta x)$$
$$= \frac{1}{\sqrt{2\pi}}\int_{-\alpha-\beta x}^{\infty} e^{-\frac{1}{2}t^2}dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\alpha+\beta x} e^{-\frac{1}{2}t^2}dt$$

Note: the case
$\mathbb{P}(y=1) = \mathbb{P}(y^*>a)$ where $y^* \sim \mathcal{N}(0,\sigma^2)$ can be reduced to $a=0,\ \sigma^2=1$

$\mathbb{P}(y^*>a) = \mathbb{P}(\alpha+\beta x+\epsilon>a) = \mathbb{P}(\epsilon > a-\alpha-\beta x)$

$= \mathbb{P}\left(\frac{1}{\sigma}\epsilon > \frac{1}{\sigma}\left(a - \alpha - \beta x\right)\right) = \mathbb{P}\left(\epsilon^* > -\alpha^* - \beta^* x\right)$

interpretation coefficients by **marginal effects**:

$$\frac{d}{dx}\mathbb{P}\left(Y = 1 \mid x\right) = \beta \ f\left(\alpha + \beta x\right)$$

4.2 ML-estimation for binary reg mods

Likelihood

$$L\left(\mathbf{y} \mid \ \alpha, \beta\right) = \prod_{i}^{n} \mathbb{P}\left(Y_i = y_i \mid \alpha, \beta\right)$$

$\implies$ log-likelihood

$$\begin{aligned} l\left(\mathbf{y} \mid \alpha, \beta\right) &= \sum_{i}^{n} \log \mathbb{P}\left(Y_i = y_i \mid \alpha, \beta\right) \\ &= \sum_{i}^{n} \left(y_i \log F\left(\alpha + \beta x\right) + \left(1 - y_i\right) \log F\left(\alpha + \beta x\right)\right) \end{aligned}$$

mle

$$\frac{\partial}{\partial \beta} l\left(\mathbf{y} \mid \widehat{\alpha}, \widehat{\beta}\right) = 0$$
$$\frac{\partial}{\partial \alpha} l\left(\mathbf{y} \mid \widehat{\alpha}, \widehat{\beta}\right) = 0$$

Solve be Newton-method:

1) compute observed information

$$\widehat{\mathbb{I}}_n\left(\alpha, \beta\right) = -\begin{bmatrix} \frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \beta} \\ \frac{\partial^2 l}{\partial \alpha \partial \beta} & \frac{\partial^2 l}{\partial \beta^2} \end{bmatrix}$$

2) iterate:

$$\begin{bmatrix} \alpha^{(new)} \\ \beta^{(new)} \end{bmatrix} = \begin{bmatrix} \alpha^{(old)} \\ \beta^{(old)} \end{bmatrix} + \widehat{\mathbb{I}}_n^{-1}\left(\alpha^{(old)}, \beta^{(old)}\right) \begin{bmatrix} \frac{\partial}{\partial \alpha} l\left(\mathbf{y} \mid \alpha^{(old)}, \beta^{(old)}\right) \\ \frac{\partial}{\partial \beta} l\left(\mathbf{y} \mid \alpha^{(old)}, \beta^{(old)}\right) \end{bmatrix}$$

ML-estimation for binary logit

likelihood

$$L(\mathbf{y} \mid \alpha, \beta) = \prod_i^n \mathbb{P}(Y_i = y_i \mid x_i,\ \alpha, \beta) = \prod_i^n \left(\frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}\right)^{y_i} \left(\frac{1}{1 + \exp(\alpha + \beta x_i)}\right)^{1-y_i}$$
$$= \prod_i^n \frac{(\exp(\alpha + \beta x_i))^{y_i}}{1 + \exp(\alpha + \beta x_i)}$$

log-likelihood

$$l(\mathbf{y} \mid \alpha, \beta) = \sum_i^n (\ y_i(\alpha + \beta x_i) - \log(1 + \exp(\alpha + \beta x_i))\ )$$

likelihood equations

$$\frac{\partial}{\partial \alpha} l(\mathbf{y} \mid \alpha, \beta) = \sum_i^n \left(y_i - \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}\right) = 0$$
$$\frac{\partial}{\partial \beta} l(\mathbf{y} \mid \alpha, \beta) = \sum_i^n x_i \left(y_i - \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}\right) = 0$$

$\Longrightarrow$ solution: mle $\widehat{\alpha}, \widehat{\beta}$. solving by Newton's method.

$$\frac{\partial}{\partial \alpha} l\left(\mathbf{y} \mid \mathbf{x}, \widehat{\alpha}, \widehat{\beta}\right) = 0$$
$$\frac{\partial}{\partial \beta} l\left(\mathbf{y} \mid \mathbf{x}, \widehat{\alpha}, \widehat{\beta}\right) = 0$$

observed Information:

$$\widehat{\mathbb{I}}_n(\alpha, \beta) = -\begin{bmatrix} \frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \beta} \\ \frac{\partial^2 l}{\partial \alpha \partial \beta} & \frac{\partial^2 l}{\partial \beta^2} \end{bmatrix} = \begin{bmatrix} \sum_i^n x_i^2 \frac{e^{\alpha+\beta x_i}}{\left(e^{\alpha+\beta x_i}+1\right)^2} & \sum_i^n x_i \frac{e^{\alpha+\beta x_i}}{\left(e^{\alpha+\beta x_i}+1\right)^2} \\ \sum_i^n x_i \frac{e^{\alpha+\beta x_i}}{\left(e^{\alpha+\beta x_i}+1\right)^2} & \sum_i^n \frac{e^{\alpha+\beta x_i}}{\left(e^{\alpha+\beta x_i}+1\right)^2} \end{bmatrix}$$

note: in this case $\widehat{\mathbb{I}}_n(\alpha, \beta) = \mathbb{E}\widehat{\mathbb{I}}_n(\alpha, \beta) = \mathbb{I}_n(\alpha, \beta)$ hence Newton-method = Fisher scoring

solve by **Newton method:** with $\alpha = \alpha^{(old)},\ \beta = \beta^{(old)}$

$$\begin{bmatrix}\alpha^{(new)}\\ \beta^{(new)}\end{bmatrix} = \begin{bmatrix}\alpha\\ \beta\end{bmatrix} + \mathbb{I}_n^{-1}(\alpha,\beta)\begin{bmatrix}\sum_i^n \left(y_i - \frac{\exp(\alpha+\beta x_i)}{1+\exp(\alpha+\beta x_i)}\right)\\ \sum_i^n x_i\left(y_i - \frac{\exp(\alpha+\beta x_i)}{1+\exp(\alpha+\beta x_i)}\right)\end{bmatrix}$$

It follows from **ML-theory:** Asymptotic distribution:

$$\mathbb{I}_n^{1/2}\left(\begin{bmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{bmatrix} - \begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right) \to \mathcal{N}(0,\ I)$$

with estimated asymptotic variance

$$\begin{bmatrix}\mathrm{var}\widehat{\beta} & \mathrm{cov}\left(\widehat{\alpha},\widehat{\beta}\right)\\ \mathrm{cov}\left(\widehat{\alpha},\widehat{\beta}\right) & \mathrm{var}\widehat{\alpha}\end{bmatrix} \approx \mathbb{I}_n^{-1}\left(\widehat{\alpha},\widehat{\beta}\right)$$

ML-estimation for binary probit

$$\begin{aligned} l(\mathbf{y} \mid \alpha,\beta) &= \sum_i^n \left(y_i \log F(\alpha+\beta x_i) - (1-y_i)\log(1-F(\alpha+\beta x_i))\right)\\ &\sim \sum_i^n \left(y_i \log \int_{-\infty}^{\alpha+\beta x_i} e^{-\frac{1}{2}t^2}dt - (1-y_i)\log\left(\int_{\alpha+\beta x_i}^{\infty} e^{-\frac{1}{2}t^2}dt\right)\right)\\ &= \sum_i^n \left(y_i \log \Phi(\alpha+\beta x_i) - (1-y_i)\log(1-\Phi(\alpha+\beta x_i))\right)\end{aligned}$$

this is numerically more cumbersome (especially in higher dimensions!)

4.3 binary regression and cross tabs

the problem of **overparametrization**

X : factor, p levels, β_i : effect of i^{th} level. define effect of X : $\beta_1 \mathrm{I}(X = level1) + \cdots + \beta_p \mathrm{I}(X = level\mathrm{p})$.

model:

$$\mathbb{P}(Y=1 \mid x) = 1 - \mathbb{P}(Y=0 \mid x) = F(\alpha+\beta_i)$$

$\Longrightarrow$

$$\mathbb{P}(Y=1 \mid x) = F\left(\alpha + \sum_{i=1}^p \beta_i \mathrm{I}(x = level\mathrm{i})\right)$$

this model is *over parametrized:*

$$F\left(\alpha + \sum_{i=1}^{p} \beta_i \mathrm{I}\left(x = level\mathrm{i}\right)\right) = F\left(\alpha^* + \sum_{i=1}^{p} \beta_i^* \mathrm{I}\left(x = level\mathrm{i}\right)\right)$$
$$\text{with } \alpha^* = \alpha - c,\ \beta_i^* = \beta_i + c,\ i = 1, \ldots, p$$

because

$$\sum_{i=1}^{p} \beta_i^* \mathrm{I}\left(x = level\mathrm{i}\right) = 1$$

$\Longrightarrow$ restriction: define reference group (**baseline**)

$$\mathbb{P}\left(Y = 1 \mid x\right) = F\left(\alpha + \beta_i\right),\ \beta_1 = 0$$

Example: dependence of graduation on social factors (Powers & Xie)

Table 3.1: High School Graduates by Race, Sex, and Family Structure

	Intact				Nonintact			
	Male		Female		Male		Female	
Race	y	n	y	n	y	n	y	n
White	843 (86%)	982	864 (93%)	931	168 (69%)	243	161 (77%)	208
Black	346 (78%)	441	360 (88%)	410	231 (69%)	337	225 (80%)	283
Hispanic	237 (78%)	305	208 (80%)	259	82 (64%)	128	78 (80%)	98

model

$$\mathbb{P}(Grad_{ijk} = yes) = p_{ijk} = F\left(\alpha + \beta_i + \gamma_j + \delta_k\right)$$
$$\beta_1 = \gamma_1 = \delta_1 = 0$$

ML-estimation for logit:

constant	black	hispanic	female	nonintact
1.764 (0.074)	-0.322(0091)	-0.561(0.105)	0.575(0.081)	-0.747(0.083)

interpret: the odds for graduation for nonintact are by a factor of exp(-0.747) smaller than for intacts (controlling for race and gender)

Note: likelihood-ratio tests against the full model:

$$p_{ijk} = F\left(\alpha + \beta_i + \gamma_j + \delta_k + (\beta\gamma)_{ij} + (\beta\delta)_{ik} + (\gamma\delta)_{jk} + (\beta\gamma\delta)_{ijk}\right)$$

reps against the null model and other nested submodels can readily be derived.

Note: Likelihood: Depending on the data (observations)

1) **person-level data**

$$y_{ijk\#} = \begin{cases} 1 & \text{if person \# does} \quad \text{graduate} \\ 0 & \text{if person \# does not} \quad \text{graduate} \end{cases}$$

likelihood

$$l\left(\alpha, \beta, \gamma, \delta \mid \mathbf{y}\right) = \sum_{i,j,k,\#} y_{ijk\#} \log p_{ijk\#} + (1 - y_{ijk\#}) \log\left(1 - p_{ijk\#}\right)$$

2) **group-level data**: introduce $\sum_{\#} y_{ijk\#} = y_{ijk\cdot}$ and cell totals $n_{ijk\cdot}$

$$\begin{cases} y_{ijk\cdot} & \text{number of persons who do} \quad \text{graduate} \\ n_{ijk\cdot} - y_{ijk\cdot} & \text{number of persons who not not} \quad \text{graduate} \end{cases}$$

likelihood

$$l\left(\alpha, \beta, \gamma, \delta \mid \mathbf{y}\right) = \sum_{i,j,k} y_{ijk} \log p_{ijk} + (n_{ijk} - y_{ijk}) \log\left(1 - p_{ijk}\right)$$

It is seen that in this case ML inference is identical on the person-level and group-level since $p_{ijk\#} = p_{ijk}$ for all persons (#)