Methods of Applied Statistics I

STA2101H F LEC9101

Week 8

October 29 2020

Today

Start Recording

1. HW2 due November 5
2. Measures of risk
3. Modelling with binomial data (FELM §2.4–2.11)
4. Generalized linear models (FELM Ch. 6)
5. HW2 Questions

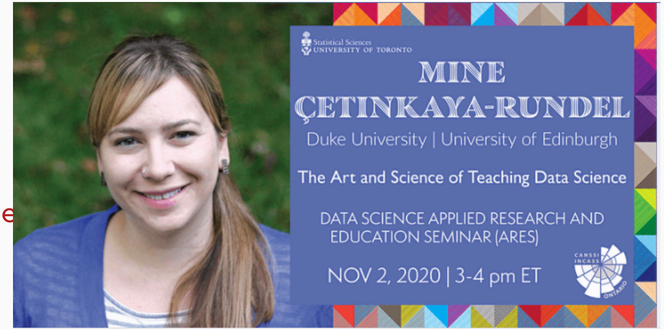
- November 2 3.00 – 4.00 Mine Çetinkaya-Rundel
- `https://canssiontario.utoronto.ca/?mec-events=ares_cetinkaya-rundel_mine`
- "The art and science of teaching data science"

Today

Start Recording

1. HW2 due November 5
2. Measures of risk
3. Modelling with binomial data (FELM §2.4–2.11)
4. Generalized linear models (FELM Ch. 6)
5. HW2 Questions

- November 2 3.00 – 4.00 Mine Çetinkaya-Rundel
- `https://canssiontario.utoronto.ca/?mec-eve`
- "The art and science of teaching data science"



Statistical Sciences
UNIVERSITY OF TORONTO

MINE
ÇETINKAYA-RUNDEL

Duke University | University of Edinburgh

The Art and Science of Teaching Data Science

DATA SCIENCE APPLIED RESEARCH AND EDUCATION SEMINAR (ARES)

NOV 2, 2020 | 3-4 pm ET

Friday

Register

HALLOWEEN GGPLOT WORKSHOP! 🎃

Want to learn how to make compelling data visualizations with the powerful and flexible ggplot2 package in R?

Want an excuse to dress up for Halloween even if you're not leaving the house?

If your answer to one or both of those questions is "YES!" come join us for a very spooooooky workshop.

Friday, Oct 30, 12:00–2:00 p.m. ET

Register here.

Recap

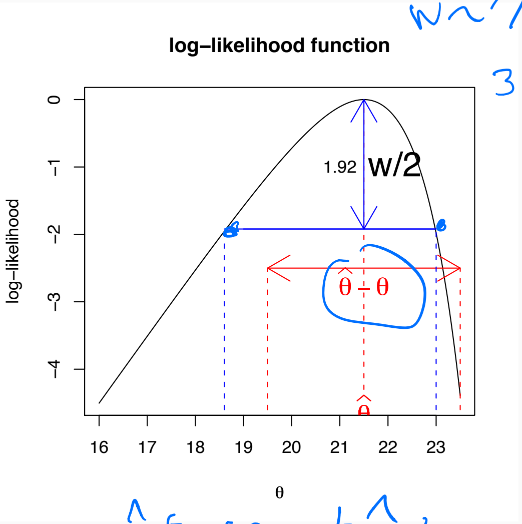
FLM-2 Ch. 5

- Regression for explanation
- observational data vs designed experiment; causality

- likelihood inference
- standardized maximum likelihood estimate (Wald te
- likelihood ratio test

Wald test

- modelling and inference for binary/binomial data
- saturated model and residual deviance
- interpretation of coefficients
- variable selection, residuals, diagnostics



$W \sim \chi^2_1$

3.84

$$\frac{\hat\theta - \theta}{\hat\sigma} \dot\sim N(0,1)$$

$= \text{Wald st.}$

$\hat\theta \pm 1.96 \cdot \widehat{s.e.}$

Measures of risk

- see posted handout on case-control studies
- consider for simplicity binomial responses with a single binary covariate:

$$\log \frac{p_i}{1-p_i} = \text{logit}(p_i) \sim \beta_0 + \beta_1 z_i, \quad i = 1, \dots, n \qquad y_i \sim Bin(n_i, p_i)$$

$$\hookrightarrow = \begin{cases} \beta_0 & z_i = 0 \text{ "control"} \\ \beta_0 + \beta_1 & z_i = 1 \text{ "trt"} \end{cases}$$

$y = 1$ "success"

0 "failure"

$$\beta_1 \text{ effect of "trt" on } p_i = \log\left(\frac{p_1}{1-p_1}\right) - \log\left(\frac{p_0}{1-p_0}\right)$$

$$= \log\left(\frac{p_1}{1-p_1} \Big/ \frac{p_0}{1-p_0}\right) \quad \boxed{e^{\underline{\underline{\hat\beta_1}}}} = \frac{\hat p_1(1-\hat p_0)}{\hat p_0(1-\hat p_1)} \hookrightarrow \text{odds ratio}$$

- no difference between groups $\iff$ odds-ratio $= 1$

… Measures of risk

- we might be interested in risk ratio $\frac{p_1}{p_0}$ instead of odds ratio $\frac{p_1(1-p_0)}{p_0(1-p_1)}$
- also called relative risk

... Measures of risk

- we might be interested in risk ratio $\frac{p_1}{p_0}$ instead of odds ratio $\frac{p_1(1-p_0)}{p_0(1-p_1)}$
- also called relative risk
- if p_1 and p_0 are both small, ($y = 1$ is rare), then

$$\frac{p_1}{p_0} \approx \frac{p_1(1-p_0)}{p_0(1-p_1)}$$

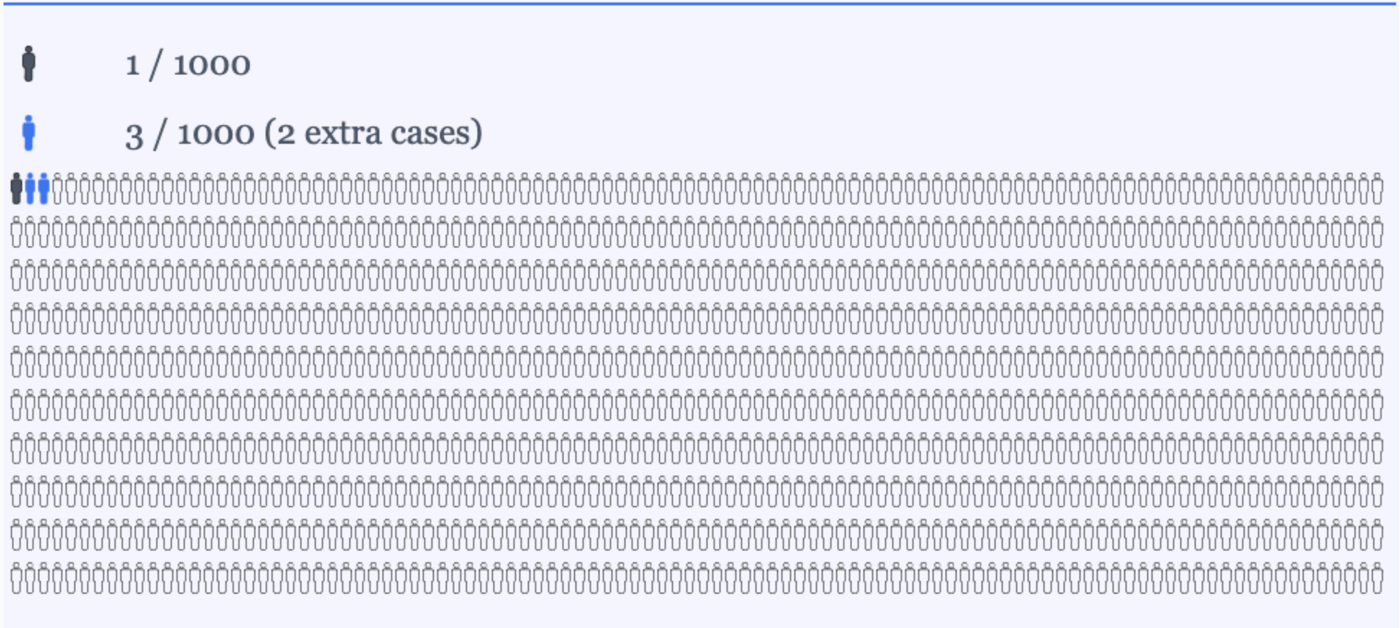
- sometimes p_1/p_0 can be large but if p_1 and p_0 are both small the difference $p_1 - p_0$ might also be very small

… Measures of risk

- we might be interested in risk ratio $\frac{p_1}{p_0}$ instead of odds ratio $\frac{p_1(1-p_0)}{p_0(1-p_1)}$
- also called relative risk
- if p_1 and p_0 are both small, ($y = 1$ is rare), then

$$\frac{p_1}{p_0} \approx \frac{p_1(1-p_0)}{p_0(1-p_1)}$$

- sometimes p_1/p_0 can be large but if p_1 and p_0 are both small the difference $p_1 - p_0$ might also be very small
- in order to estimate the risk difference we need to know the baseline risk p_0
- bacon sandwiches www.youtube.com/watch?v=4szyEbU94ig
- risk calculator realrisk.wintoncentre.uk/p8



Odds ratio 2.91; baseline risk 1/1000

Biostats secret sauce

Whether we sample prospectively or retrospectively, the odds ratio is the same

	Lung cancer	
	1	0
	cases	controls
smoke = 1 (yes)	688	650
smoke = 0 (no)	21	59
	709	709

#

$p_1/(1-p_1)$

“smok/not”

$$\text{retro: } OR = \frac{(688/709)/(21/709)}{(650/709)/(59/709)} = \frac{688 \times 59}{650 \times 21} = \boxed{2.97}$$

p_1

Lung c. vs not

$$\text{prosp: } OR = \frac{\{688/(688+650)\}/\{650/(688+650)\}}{21/(21+59)/\{59/(21+59)\}} = \frac{688 \times 59}{650 \times 21} = 2.97$$

see “case-control”, FELM §2.5,6, SM §10.4.2

Link function

```
glm(cbind(r, m-r) ~ age + weight, data = mydata,
          family = binomial, link = logit )
```

```
?family
link
```

a specification for the model link function. This can be a name/expression, a literal character string, a length-one character vector, or an object of class ‘‘link-glm’’ ...

$\lg(x_i^T\beta)$

def. def.

The gaussian family accepts the links (as names) identity, log and inverse; the binomial family the links logit, probit, cauchit, (corresponding to logistic, normal and Cauchy CDFs respectively) log and cloglog (complementary log-log)

Tolerance distributions

hidden
latent

$$\longrightarrow \quad z_i = x_i^T\gamma + \epsilon_i, \quad y_i = I(z_i \geq 0) \qquad \varepsilon_i \sim N(0,\sigma^2)$$

see also FELM Fig 2.3

$$y_i = 1 \quad \text{if} \quad z_i \geq 0$$

$$\boxed{e^{-e^{-x_i^T\beta}}} \longleftarrow p_i$$

$$p_i = \Pr(z_i \geq 0) = 1 - \Phi\left(\frac{-x_i^T\gamma}{\sigma}\right) \qquad \varepsilon_i \geq -x_i^T\gamma$$

$$= \Phi\{x_i^T\underbrace{(\tfrac{\gamma}{\sigma})}_{\beta}\}$$

$$-\log(-\log(p_i))$$
$$= x_i^T\beta$$

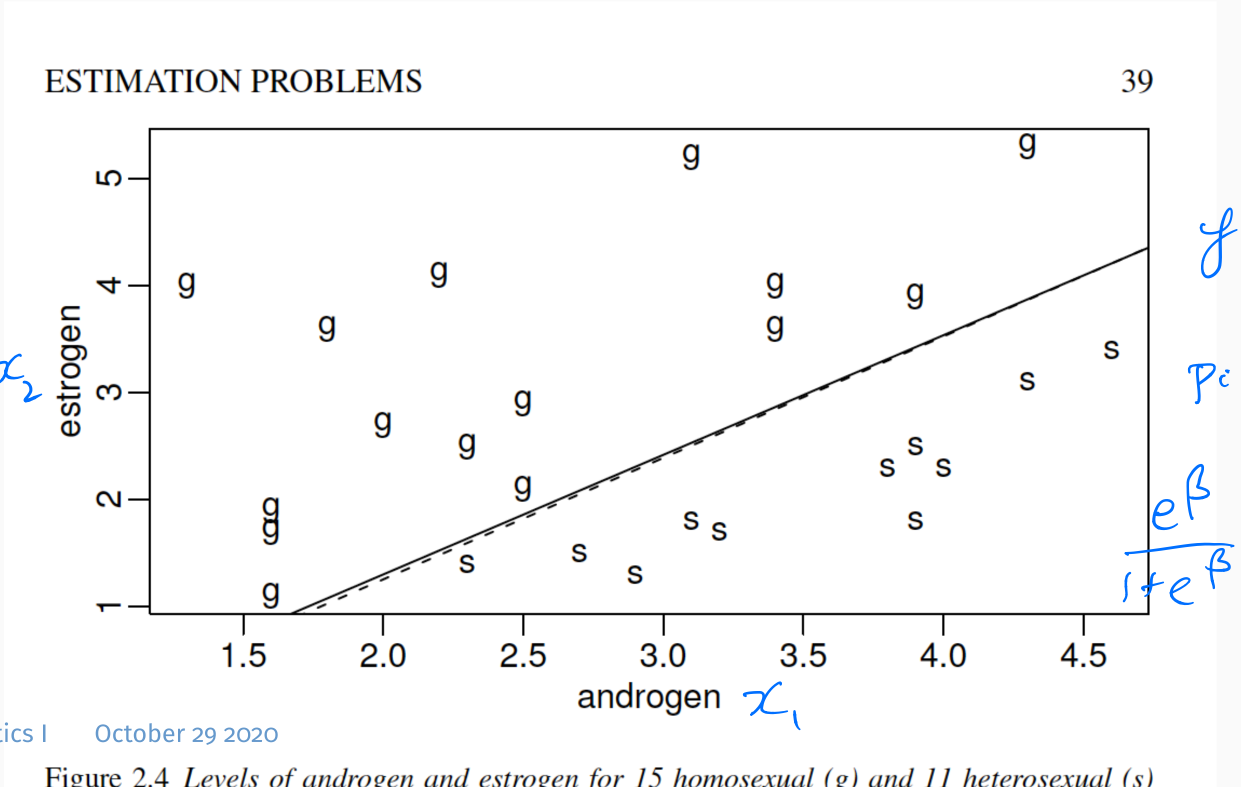
$$\sum_i x_i^T\beta y_i + \ldots$$

$$L(\beta) = \prod_{i=1}^{n} \binom{n_i}{y_i} \Phi(x_i^T\beta)^{y_i}\{1 - \Phi(\quad)\}^{n_i - y_i}$$

$$\ell(\beta) = \sum \log\{ \qquad \swarrow \qquad \uparrow$$

Linear separability

FELM §2.8



ESTIMATION PROBLEMS 39

Figure 2.4 *Levels of androgen and estrogen for 15 homosexual (g) and 11 heterosexual (s)*

x_2

x_1

$y = \begin{cases} g \\ s \end{cases}$

$p_i = \begin{matrix} 0 \\ 1 \end{matrix}$

$\frac{e^{\beta}}{1+e^{\beta}} = 0, 1$

$$x_{1/2} = -\frac{\beta_0}{\beta_1} \qquad \hat{x}_{-1/2} = -\frac{\hat\beta_0}{\hat\beta_1} \qquad \boxed{\underset{\sim}{x}^{*T}\underset{\sim}{\hat\beta}}$$

$\beta = (\beta_0, \beta_1)$

m.v.

$$\text{var}\, \underset{=}{g}(\underset{\sim}{\hat\beta}) \doteq \underline{\underset{1\times 2}{g'(\hat\beta)^T}\ \underset{2\times 2}{\text{cov}(\hat\beta)}\ \underset{2\times 1}{g'(\hat\beta)}}$$

$$g'(\beta) = \begin{pmatrix} \frac{\partial}{\partial\beta_0} g \\ \frac{\partial}{\partial\beta_1} g \end{pmatrix}$$

$$g(\beta) = -\frac{\beta_0}{\beta_1} \qquad g'(\beta) = \begin{pmatrix} -\frac{1}{\beta_1} \\ \frac{\beta_0}{\beta_1^2} \end{pmatrix}$$

vcov

$$\widehat{\text{var}}(\hat{x}_{1/2}) =$$

$$g(\beta) \simeq g(\beta_0) + \underbrace{(\beta - \beta_0)\, g'(\beta_0)}$$

Prediction and prediction intervals

FELM §2.10, 3, 4

$$\hat{p}(x^*) = \frac{e^{x^{*T}\hat\beta}}{1+e^{x^{*T}\hat\beta}}$$

ilogit

$$\mathrm{var}(x^{*T}\hat\beta) = x^{*T}\{\mathrm{cov}(\hat\beta)\}x^*$$

$$x^{*T}\hat\beta \;\dot\sim\; N(x^{*T}\beta, \underbrace{\underset{\sim}{x}^{*T}\hat{\mathrm{cov}}(\hat\beta)\underset{\sim}{x}^*}_{\text{vcov}})$$

$$95\%\text{ CI for } x^{*T}\beta: \quad x^{*T}\hat\beta \pm 1.96\sqrt{\quad} = (L, U)$$

$$95\%\text{ CI for } \hat p(x^*) = (\mathrm{ilogit}(L), \mathrm{ilogit}(U)) \in (0,1)$$

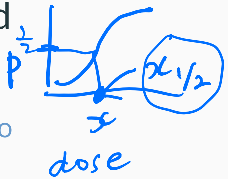
$$\hat\beta = \hat\beta(y)$$

m.l.e

$$\beta_0 + \beta_1 x = \mathrm{logit}\{p(x)\} \qquad y_i | x_i \sim \mathrm{Bin}\{n_i, p(x_i)\}$$

ED50 and delta method

$\beta_1 > 0$



$$p = \tfrac{1}{2} \qquad \log\left(\frac{1/2}{1-1/2}\right) = 0$$

$$\beta_0 + \beta_1 x_{1/2} = 0 \qquad \boxed{x_{1/2} = -\frac{\beta_0}{\beta_1}}$$

Overdispersion

ELM §2.11, SM 10.6

- $Y_i \sim Bin(n_i, p_i) \Rightarrow \mathrm{E}(Y_i) = n_i p_i, \quad \mathrm{Var}(Y_i) = n_i p_i(1 - p_i)$

$$\text{residual deviance } 2\{\ell(\overset{\text{sat.'d}}{\tilde{p}_i}) - \underbrace{\ell(p_i(\hat{\beta}))}_{\text{reg. model}}\}$$

$p = (\beta_1 \cdots \beta_p)$

$$= \cdots \sum_{i=1}^{n} \{$$

$$\text{var}(Y_i) = \phi\, n_i p_i (1 - p_i)$$

ϕ "fudge factor"

overdispersion par.

$$\tilde{\phi} = \frac{\text{Res. dev.}}{n-p}$$ estimated " " $\sim N \cdots$

quasi-likelihood $\left(\hat{\beta}_q, \ \widehat{s.e.}_q \hat{\beta}_q \right)$

Quasibinomial

overdisp.Rmd; overdisp.html

Generalized linear models: theory

- $\left\{\begin{array}{l}\text{p.m.f. for} \\ \text{Bin , Po} \\ \text{pdf for} \\ \text{N G iG}\end{array}\right\}$

$$f(y_i; \mu_i, \phi_i) = \underline{\underline{\exp}}\{\underbrace{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i)}\}$$

glm

binom -

\- norm

Poiss -

\- Gamma

(- inv. G.)

quasi-bin

quasi-P.

Generalized linear models: theory

- $$f(y_i;\mu_i,\phi_i) = \exp\{\frac{\overbrace{y_i}\theta_i - b(\theta_i)}{\phi_i} + c(y_i;\phi_i)\}$$

$\mu_i = E(y_i)$

$\nwarrow \int y_i f(y_i)\,dy_i$

- $\underbrace{E(y_i) \cancel{| x_i)} = \boxed{b'(\theta_i) = \mu_i}}$ defines $\underbrace{\mu_i \text{ as a function of } \theta_i}$

Generalized linear models: theory

$f(y_i \mid x_i; \mu_i, \phi_i)$

- $$f(y_i; \mu_i, \phi_i) = \exp\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i)\}$$
- $E(y_i \mid x_i) = b'(\theta_i) = \mu_i$ defines μ_i as a function of θ_i
- $\underline{g(\mu_i)} = \underline{x_i^T}\beta = \eta_i$ links the n observations together via $\underline{\text{covariates}}$

$\mu_i = E(y_i \mid x_i)$

$p_i = E(y_i)$ $\quad \log \frac{p_i}{1-p_i}$

Generalized linear models: theory

- $$f(y_i; \mu_i, \phi_i) = \exp\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i)\}$$
- $E(y_i \mid x_i) = b'(\theta_i) = \mu_i$ defines μ_i as a function of θ_i
- $g(\mu_i) = \underbrace{x_i^T \underset{\sim}{\beta} = \boxed{\eta_i}}$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_i is the linear predictor

θ_i

μ_i $i=1,\dots,n$

η_i

$\underset{\sim}{x}_i^T$

$\underset{\sim}{x}_i$ $p\times 1$

$\underset{\sim}{\beta}$ $p\times 1$

~~d~~ $\boxed{p<n}$

β_j ~~A~~ $j=1,\dots,p$

Generalized linear models: theory

- $$\boxed{f(y_i; \mu_i, \phi_i) = \exp\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i)\}}$$

def^n glm

- $E(y_i \mid x_i) = \underline{b'(\theta_i)} = \mu_i$ defines μ_i as a function of θ_i

$E(y_i) = \int y_i f(y_i)\, dy_i$

- $g(\mu_i) = x_i^T\beta = \eta_i$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_i is the linear predictor
- $\underline{\text{Var}(y_i \mid x_i)} = \phi_i \underline{b''(\theta_i)} = \phi_i \underline{V(\mu_i)}$

variance f^n

prop's that follow

y_i ind't

extendible

$g\{E(y_i|x_i)\} = x_i^T\beta$

$g(\mu_i) = x_i^T\beta$

$$\boxed{var(y_i|x_i) = \phi_i V(\mu_i)}$$

Generalized linear models: theory

- $$f(y_i;\mu_i,\phi_i) = \exp\{\frac{\boxed{y_i\theta_i} - b(\theta_i)}{\phi_i} + c(y_i;\phi_i)\}$$
- $E(y_i \mid x_i) = b'(\theta_i) = \mu_i$ defines μ_i as a function of θ_i
- $g(\mu_i) = x_i^T\beta = \eta_i$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_i is the linear predictor
- $\mathrm{Var}(y_i \mid x_i) = \phi_i b''(\theta_i) = \phi_i V(\mu_i)$
- $V(\cdot)$ is the variance function

$b(\mu_i) = + \frac{1}{2}\mu_i^2$

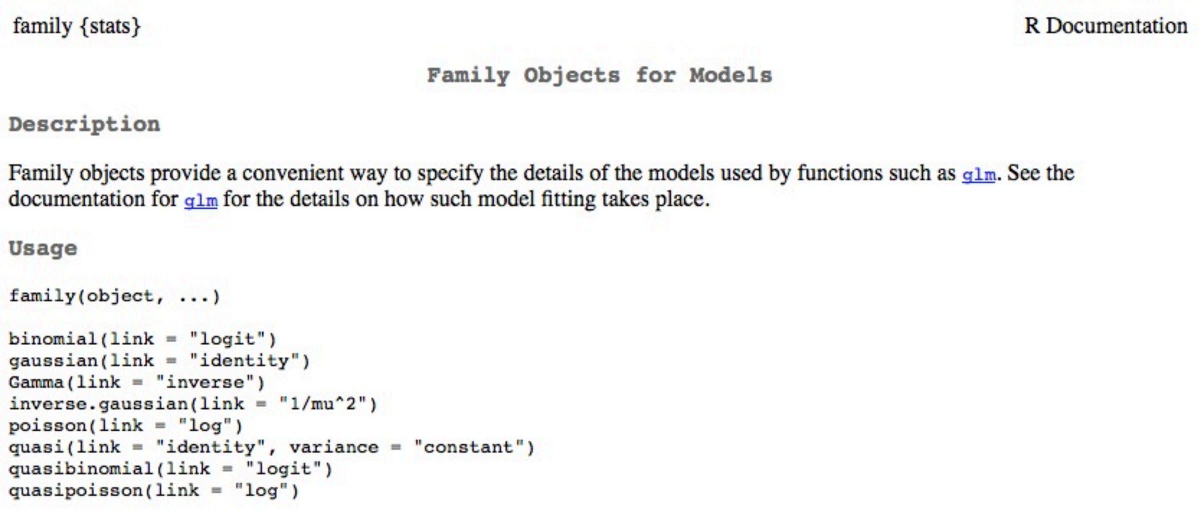
$b' = \mu_i \quad b'' = 1$

$\phi_i b''(\mu_i)$

$\mathrm{Var} = \sigma^2.$

Examples

- Normal
- Binomial
- Poisson
- Gamma/E
- Inverse Ga



family {stats}   R Documentation

Family Objects for Models

Description

Family objects provide a convenient way to specify the details of the models used by functions such as glm. See the documentation for glm for the details on how such model fitting takes place.

Usage

```
family(object, ...)

binomial(link = "logit")
gaussian(link = "identity")
Gamma(link = "inverse")
inverse.gaussian(link = "1/mu^2")
poisson(link = "log")
quasi(link = "identity", variance = "constant")
quasibinomial(link = "logit")
quasipoisson(link = "log")
```

Examples

- Normal: $f(y_i;\mu_i,\sigma^2) = \boxed{\frac{1}{\sqrt{(2\pi)}\sigma}\exp\{-\frac{1}{2\sigma^2}(y_i-\mu_i^{\cancel{2}})^{2}\}}$

$$e^{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i;\phi_i)}$$

$$= \boxed{\exp\{\frac{y_i\mu_i - (1/2)\mu_i^2}{\sigma^2} - (1/2)\log\sigma^2 - y_i^2/2\sigma^2 - (1/2)\log\sqrt{(2\pi)}\}}$$

$$\phi_i = \sigma^2, \quad \theta_i = \mu_i, \quad b(\mu_i) = \mu_i^2/2\sigma^2$$

note $b''(\mu_i) = 1$

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i-\mu_i)^2}$$

$$-\frac{1}{2\sigma^2}y_i^2 + \frac{1}{\sigma^2}y_i\mu_i - \frac{1}{2\sigma^2}\mu_i^2 - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\log 2\pi$$

$$= e$$

$$= e^{\frac{(y_i\overset{\theta_i}{\mu_i} - \overset{b}{\mu_i^2/2})}{\sigma^2 \Leftarrow \phi}} \left[-\frac{1}{2\sigma^2}y_i^2 - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\ln 2\pi\right]$$

$\uparrow c$

$E(y_i) = \mu_i = x_i^T\beta$

$var(y_i) = \sigma^2 = \phi_i$

$\phi_i \equiv \sigma^2$ $\quad \boxed{\theta_i \equiv \mu_i}$ identity

$b'(\theta_i) = b'(\mu_i) = 2\mu_i/2 = \mu_i = E(y_i)$

Examples

- Normal: $f(y_i; \mu_i, \sigma^2) = \frac{1}{\sqrt{(2\pi)}\sigma} \exp\{-\frac{1}{2\sigma^2}(y_i - \mu_i^2)\}$

(μ_i, σ^2) unk.

$$= \exp\{\frac{y_i\mu_i - (1/2)\mu_i^2}{\sigma^2} - (1/2)\log\sigma^2 - y_i^2/2\sigma^2 - (1/2)\log\sqrt{(2\pi)}\}$$

$$\boxed{\phi_i = \sigma^2,} \quad \theta_i = \mu_i, \quad b(\mu_i) = \mu_i^2/2\sigma^2$$

overdisp.

$\phi_i = \phi/m_i$

quasi

note $b''(\mu_i) = 1$

- Binomial: $f(r_i; p_i) = \binom{m_i}{r_i} p_i^{r_i}(1-p_i)^{m_i - r_i};$ $\boxed{y_i = r_i/m_i}$

$\mu_i = E(\frac{r_i}{m_i})$

$$= \exp[m_i y_i \log\{p_i/(1-p_i)\} + m_i\log(1-p_i) + \underbrace{\log\binom{m_i}{m_i y_i}}_{c(y_i, \phi_i)}]$$

$\frac{1}{\phi_i}\theta_i y_i$ $\quad$ $\frac{1}{\phi_i} b(\theta_i)$

known

$$\boxed{\phi_i = 1/m_i,} \quad \underline{\theta_i = \log\{p_i/(1-p_i)\}}, \quad b(p_i) = -\log(1-p_i)$$

$$\boxed{\phi_i b''(\theta_i) = \text{var}(y_i)}$$

Note $p_i = \mu_i = E(y_i)$

$= E(y_i)$

$$f(y_i) = \binom{m_i}{m_i y_i} p_i^{m_i y_i} (1-p_i)^{m_i - m_i y_i}$$

$$= \exp[m_i(\log\frac{p_i}{1-p_i})y_i + m_i\log(1-p_i) + \quad]$$

Examples

- Normal: $f(y_i; \mu_i, \sigma^2) = \frac{1}{\sqrt(2\pi)\sigma} \exp\{-\frac{1}{2\sigma^2}(y_i - \mu_i^2)\}$

$$= \exp\{\frac{y_i\mu_i - (1/2)\mu_i^2}{\sigma^2} - (1/2)\log\sigma^2 - y_i^2/2\sigma^2 - (1/2)\log\sqrt(2\pi)\}$$

$$\phi_i = \sigma^2, \quad \theta_i = \mu_i, \quad b(\mu_i) = \mu_i^2/2\sigma^2$$

note $b''(\mu_i) = 1$

- Binomial: $f(r_i; p_i) = \binom{m_i}{r_i} p_i^{r_i}(1 - p_i)^{m_i - r_i}; \quad y_i = r_i/m_i$

$$= \exp[m_i y_i \log\{p_i/(1 - p_i)\} + m_i \log(1 - p_i) + \log\binom{m_i}{m_i y_i}]$$

$$\phi_i = 1/m_i, \quad \theta_i = \log\{p_i/(1 - p_i)\}, \quad b(p_i) = -\log(1 - p_i)$$

Note $p_i = \mu_i = \mathrm{E}(y_i)$

- ELM (p.115) uses $a_i(\phi)$ in place of ϕ_i, later (p.117) $a_i(\phi) = \phi/w_i$; later (p.118) w_i used for weights in IRWLS algorithm; SM uses ϕ_i, later (p. 483) $\phi_i = \phi a_i$

Inference

- $\ell(\beta; y) = \sum\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i)\}$

Inference

- $\ell(\beta; y) = \sum\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i)\}$
- $b'(\theta_i) = \mu_i;\quad g(\mu_i) = g(b'(\theta_i)) = \eta_i = x_i^{\mathrm{T}}\beta$

Inference

- $\ell(\beta; y) = \sum\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i)\}$
- $b'(\theta_i) = \mu_i; \quad g(\mu_i) = g(b'(\theta_i)) = \eta_i = x_i^{\mathrm{T}}\beta$
- $\dfrac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \dfrac{\partial \ell_i}{\partial \theta_i}\dfrac{\partial \theta_i}{\partial \beta_j} = \sum \dfrac{y_i - b'(\theta_i)}{\phi_i}\dfrac{\partial \theta_i}{\partial \beta_j}$

Inference

- $\ell(\beta; y) = \sum\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i)\}$
- $b'(\theta_i) = \mu_i; \quad g(\mu_i) = g(b'(\theta_i)) = \eta_i = x_i^{\mathrm{T}}\beta$
- $\dfrac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \dfrac{\partial \ell_i}{\partial \theta_i}\dfrac{\partial \theta_i}{\partial \beta_j} = \sum \dfrac{y_i - b'(\theta_i)}{\phi_i}\dfrac{\partial \theta_i}{\partial \beta_j}$
- $g'(b(\theta_i))b''(\theta_i)\dfrac{\partial \theta_i}{\partial \beta_j} = x_{ij} = g'(\mu_i)V(\mu_i)$ See Slide 2

Inference

- $\ell(\beta; y) = \sum\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i)\}$
- $b'(\theta_i) = \mu_i; \quad g(\mu_i) = g(b'(\theta_i)) = \eta_i = x_i^{\mathrm{T}}\beta$
- $\dfrac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \dfrac{\partial \ell_i}{\partial \theta_i}\dfrac{\partial \theta_i}{\partial \beta_j} = \sum \dfrac{y_i - b'(\theta_i)}{\phi_i}\dfrac{\partial \theta_i}{\partial \beta_j}$
- $g'(b(\theta_i))b''(\theta_i)\dfrac{\partial \theta_i}{\partial \beta_j} = x_{ij} = g'(\mu_i)V(\mu_i)$

See Slide 2

- $\dfrac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \dfrac{y_i - \mu_i}{\phi_i g'(\mu_i)V(\mu_i)}x_{ij} = \sum \dfrac{y_i - \mu_i}{a_i g'(\mu_i)V(\mu_i)}x_{ij}$

when $\phi_i = a_i\phi$

Inference

- $\ell(\beta; y) = \sum\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i)\}$
- $b'(\theta_i) = \mu_i; \quad g(\mu_i) = g(b'(\theta_i)) = \eta_i = x_i^{\mathrm{T}}\beta$
- $\dfrac{\partial\ell(\beta; y)}{\partial\beta_j} = \sum \dfrac{\partial\ell_i}{\partial\theta_i}\dfrac{\partial\theta_i}{\partial\beta_j} = \sum \dfrac{y_i - b'(\theta_i)}{\phi_i}\dfrac{\partial\theta_i}{\partial\beta_j}$
- $g'(b(\theta_i))b''(\theta_i)\dfrac{\partial\theta_i}{\partial\beta_j} = x_{ij} = g'(\mu_i)V(\mu_i)$

See Slide 2

- $\dfrac{\partial\ell(\beta; y)}{\partial\beta_j} = \sum \dfrac{y_i - \mu_i}{\phi_i g'(\mu_i)V(\mu_i)}x_{ij} = \sum \dfrac{y_i - \mu_i}{a_i g'(\mu_i)V(\mu_i)}x_{ij}$

when $\phi_i = a_i\phi$

- matrix notation:

$$\frac{\partial\ell(\beta)}{\partial\beta} = X^{\mathrm{T}}u(\beta), \quad X = \frac{\partial\eta}{\partial\beta^{\mathrm{T}}}, \quad u = (u_1, \dots, u_n)$$

Scale parameter ϕ_i

- in most cases, either ϕ_i is known, or $\phi_i = \phi a_i$, where a_i is known

Scale parameter ϕ_i

- in most cases, either ϕ_i is known, or $\phi_i = \phi a_i$, where a_i is known
- Normal distribution, $\phi = \sigma^2$

Scale parameter ϕ_i

- in most cases, either ϕ_i is known, or $\phi_i = \phi a_i$, where a_i is known
- Normal distribution, $\phi = \sigma^2$
- Binomial distribution $\phi_i = m_i^{-1}$

Scale parameter ϕ_i

- in most cases, either ϕ_i is known, or $\phi_i = \phi a_i$, where a_i is known
- Normal distribution, $\phi = \sigma^2$
- Binomial distribution $\phi_i = m_i^{-1}$
- Gamma distribution, $\phi = 1/\nu$

Scale parameter ϕ_i

- in most cases, either ϕ_i is known, or $\phi_i = \phi a_i$, where a_i is known
- Normal distribution, $\phi = \sigma^2$
- Binomial distribution $\phi_i = m_i^{-1}$
- Gamma distribution, $\phi = 1/\nu$
- $\dfrac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \dfrac{y_i - \mu_i}{\phi_i g'(\mu_i) V(\mu_i)} x_{ij} = \sum \dfrac{y_i - \mu_i}{a_i g'(\mu_i) V(\mu_i)} x_{ij}$

when $\phi_i = a_i \phi$

Scale parameter ϕ_i

- in most cases, either ϕ_i is known, or $\phi_i = \phi a_i$, where a_i is known
- Normal distribution, $\phi = \sigma^2$
- Binomial distribution $\phi_i = m_i^{-1}$
- Gamma distribution, $\phi = 1/\nu$
- $\frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \frac{y_i - \mu_i}{\phi_i g'(\mu_i) V(\mu_i)} x_{ij} = \sum \frac{y_i - \mu_i}{a_i g'(\mu_i) V(\mu_i)} x_{ij}$

when $\phi_i = a_i \phi$

- if $\theta_i = g(\mu_i)$ canonical link, then $g'(\mu_i) = 1/V(\mu_i)$, and

$$\sum \frac{y_i x_{ij}}{a_i} = \sum \frac{y_i \hat{\mu}_i x_{ij}}{a_i}$$

Solving maximum likelihood equation

- Newton-Raphson: $\ell'(\hat{\beta}) = 0 \approx \ell'(\beta) + (\hat{\beta} - \beta)\ell''(\beta)$

defines iterative scheme

Solving maximum likelihood equation

- Newton-Raphson: $\ell'(\hat{\beta}) = 0 \approx \ell'(\beta) + (\hat{\beta} - \beta)\ell''(\beta)$

defines iterative scheme

- $\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} - \{\ell''(\hat{\beta}^{(t)})\}^{-1}\ell'(\hat{\beta}^{(t)})$

Solving maximum likelihood equation

- Newton-Raphson: $\ell'(\hat\beta) = 0 \approx \ell'(\beta) + (\hat\beta - \beta)\ell''(\beta)$

defines iterative scheme

- $\hat\beta^{(t+1)} = \hat\beta^{(t)} - \{\ell''(\hat\beta^{(t)})\}^{-1}\ell'(\hat\beta^{(t)})$

- Fisher scoring: $-\ell''(\beta) \leftarrow \mathrm{E}\{-\ell''(\beta)\} = i(\beta)$

many books use $I(\beta)$

Solving maximum likelihood equation

- Newton-Raphson: $\ell'(\hat{\beta}) = 0 \approx \ell'(\beta) + (\hat{\beta} - \beta)\ell''(\beta)$

defines iterative scheme

- $\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} - \{\ell''(\hat{\beta}^{(t)})\}^{-1}\ell'(\hat{\beta}^{(t)})$
- Fisher scoring: $-\ell''(\beta) \leftarrow \mathrm{E}\{-\ell''(\beta)\} = i(\beta)$

many books use $I(\beta)$

- $\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + \{i(\hat{\beta}^{(t)})\}^{-1}\ell'(\hat{\beta}^{(t)})$

Solving maximum likelihood equation

- Newton-Raphson: $\ell'(\hat\beta) = 0 \approx \ell'(\beta) + (\hat\beta - \beta)\ell''(\beta)$

defines iterative scheme

- $\hat\beta^{(t+1)} = \hat\beta^{(t)} - \{\ell''(\hat\beta^{(t)})\}^{-1}\ell'(\hat\beta^{(t)})$

- Fisher scoring: $-\ell''(\beta) \leftarrow \mathrm{E}\{-\ell''(\beta)\} = i(\beta)$

many books use $I(\beta)$

- $\hat\beta^{(t+1)} = \hat\beta^{(t)} + \{i(\hat\beta^{(t)})\}^{-1}\ell'(\hat\beta^{(t)})$

- applied to matrix version: $X^{\mathrm{T}}u(\hat\beta) = 0 \doteq X^{\mathrm{T}}u(\beta) + (\hat\beta - \beta)X^{\mathrm{T}}\dfrac{\partial u(\beta)}{\partial\beta^{\mathrm{T}}}$

slide 5

Solving maximum likelihood equation

- Newton-Raphson: $\ell'(\hat{\beta}) = 0 \approx \ell'(\beta) + (\hat{\beta} - \beta)\ell''(\beta)$

defines iterative scheme

- $\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} - \{\ell''(\hat{\beta}^{(t)})\}^{-1}\ell'(\hat{\beta}^{(t)})$
- Fisher scoring: $-\ell''(\beta) \leftarrow \mathrm{E}\{-\ell''(\beta)\} = i(\beta)$

many books use $I(\beta)$

- $\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + \{i(\hat{\beta}^{(t)})\}^{-1}\ell'(\hat{\beta}^{(t)})$
- applied to matrix version: $X^{\mathrm{T}}u(\hat{\beta}) = 0 \doteq X^{\mathrm{T}}u(\beta) + (\hat{\beta} - \beta)X^{\mathrm{T}}\dfrac{\partial u(\beta)}{\partial \beta^{\mathrm{T}}}$ slide 5
- change to Fisher scoring: $\hat{\beta} = \beta + i(\beta)^{-1}X^{\mathrm{T}}u(\beta)$

… maximum likelihood equation

- $\hat{\beta} = \beta + i(\beta)^{-1} X^{\mathrm{T}} u(\beta)$

$$\frac{\partial^2 \ell(\beta; y)}{\partial \beta_j \partial \beta_k} = \sum \frac{-b''(\theta_i)}{\phi_i} \left(\frac{\partial \theta_i}{\partial \beta_j} \right) \left(\frac{\partial \theta_i}{\partial \beta_k} \right) + \sum \frac{y_i - b'(\theta_i)}{\phi_i} \frac{\partial^2 \theta_i}{\partial \beta_j \partial \beta_k}$$

… maximum likelihood equation

- $\hat{\beta} = \beta + i(\beta)^{-1} X^{\mathrm{T}} u(\beta)$

$$\frac{\partial^2 \ell(\beta; y)}{\partial \beta_j \partial \beta_k} = \sum \frac{-b''(\theta_i)}{\phi_i} \left(\frac{\partial \theta_i}{\partial \beta_j} \right) \left(\frac{\partial \theta_i}{\partial \beta_k} \right) + \sum \frac{y_i - b'(\theta_i)}{\phi_i} \frac{\partial^2 \theta_i}{\partial \beta_j \partial \beta_k}$$

- $\mathrm{E} \left(- \frac{\partial^2 \ell(\beta; y)}{\partial \beta_j \partial \beta_k} \right) = \sum \frac{V(\mu_i)}{\phi_i} \frac{x_{ij}}{g'(\mu_i) V(\mu_i)} \frac{x_{ik}}{g'(\mu_i) V(\mu_i)} = \sum \frac{x_{ij} x_{ik}}{\phi_i \{g'(\mu_i)\}^2 V(\mu_i)}$

… maximum likelihood equation

- $\hat{\beta} = \beta + i(\beta)^{-1} X^{\mathrm{T}} u(\beta)$

$$\frac{\partial^2 \ell(\beta; y)}{\partial \beta_j \partial \beta_k} = \sum \frac{-b''(\theta_i)}{\phi_i} \left(\frac{\partial \theta_i}{\partial \beta_j} \right) \left(\frac{\partial \theta_i}{\partial \beta_k} \right) + \sum \frac{y_i - b'(\theta_i)}{\phi_i} \frac{\partial^2 \theta_i}{\partial \beta_j \partial \beta_k}$$

- $\mathrm{E} \left(-\frac{\partial^2 \ell(\beta; y)}{\partial \beta_j \partial \beta_k} \right) = \sum \frac{V(\mu_i)}{\phi_i} \frac{x_{ij}}{g'(\mu_i) V(\mu_i)} \frac{x_{ik}}{g'(\mu_i) V(\mu_i)} = \sum \frac{x_{ij} x_{ik}}{\phi_i \{g'(\mu_i)\}^2 V(\mu_i)}$
-

$$\begin{aligned} \hat{\beta} &= \beta + (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} u(\beta) = (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W X \beta + X^{\mathrm{T}} u(\beta)\} \\ &= (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W (X\beta + W^{-1} u(\beta)\} \\ &= (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \end{aligned}$$

… maximum likelihood equation

- $\hat{\beta} = \beta + i(\beta)^{-1}X^{\mathrm{T}}u(\beta)$

$$\frac{\partial^2 \ell(\beta;y)}{\partial\beta_j\partial\beta_k} = \sum \frac{-b''(\theta_i)}{\phi_i}\left(\frac{\partial\theta_i}{\partial\beta_j}\right)\left(\frac{\partial\theta_i}{\partial\beta_k}\right) + \sum \frac{y_i - b'(\theta_i)}{\phi_i}\frac{\partial^2\theta_i}{\partial\beta_j\partial\beta_k}$$

- $\mathrm{E}\left(-\frac{\partial^2 \ell(\beta;y)}{\partial\beta_j\partial\beta_k}\right) = \sum \frac{V(\mu_i)}{\phi_i}\frac{x_{ij}}{g'(\mu_i)V(\mu_i)}\frac{x_{ik}}{g'(\mu_i)V(\mu_i)} = \sum \frac{x_{ij}x_{ik}}{\phi_i\{g'(\mu_i)\}^2V(\mu_i)}$
-

$$\begin{aligned}\hat{\beta} &= \beta + (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}u(\beta) = (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}WX\beta + X^{\mathrm{T}}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}W(X\beta + W^{-1}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Wz\end{aligned}$$

- does not involve ϕ_i

… maximum likelihood equation

- $\hat{\beta} = \beta + i(\beta)^{-1} X^{\mathrm{T}} u(\beta)$

$$\frac{\partial^2 \ell(\beta; y)}{\partial \beta_j \partial \beta_k} = \sum \frac{-b''(\theta_i)}{\phi_i} \left(\frac{\partial \theta_i}{\partial \beta_j}\right) \left(\frac{\partial \theta_i}{\partial \beta_k}\right) + \sum \frac{y_i - b'(\theta_i)}{\phi_i} \frac{\partial^2 \theta_i}{\partial \beta_j \partial \beta_k}$$

- $\mathrm{E}\left(-\frac{\partial^2 \ell(\beta; y)}{\partial \beta_j \partial \beta_k}\right) = \sum \frac{V(\mu_i)}{\phi_i} \frac{x_{ij}}{g'(\mu_i)V(\mu_i)} \frac{x_{ik}}{g'(\mu_i)V(\mu_i)} = \sum \frac{x_{ij}x_{ik}}{\phi_i \{g'(\mu_i)\}^2 V(\mu_i)}$
-

$$\begin{aligned} \hat{\beta} &= \beta + (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} u(\beta) = (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W X \beta + X^{\mathrm{T}} u(\beta)\} \\ &= (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W (X\beta + W^{-1} u(\beta)\} \\ &= (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \end{aligned}$$

- does not involve ϕ_i
- iteratively re-weighted least squares

W, z both depend on β

… maximum likelihood equation

- $\hat{\beta} = \beta + i(\beta)^{-1} X^{\mathrm{T}} u(\beta)$

$$\frac{\partial^2 \ell(\beta; y)}{\partial \beta_j \partial \beta_k} = \sum \frac{-b''(\theta_i)}{\phi_i} \left(\frac{\partial \theta_i}{\partial \beta_j}\right) \left(\frac{\partial \theta_i}{\partial \beta_k}\right) + \sum \frac{y_i - b'(\theta_i)}{\phi_i} \frac{\partial^2 \theta_i}{\partial \beta_j \partial \beta_k}$$

- $\mathrm{E}\left(-\frac{\partial^2 \ell(\beta; y)}{\partial \beta_j \partial \beta_k}\right) = \sum \frac{V(\mu_i)}{\phi_i} \frac{x_{ij}}{g'(\mu_i) V(\mu_i)} \frac{x_{ik}}{g'(\mu_i) V(\mu_i)} = \sum \frac{x_{ij} x_{ik}}{\phi_i \{g'(\mu_i)\}^2 V(\mu_i)}$
-

$$\begin{aligned} \hat{\beta} &= \beta + (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} u(\beta) = (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W X \beta + X^{\mathrm{T}} u(\beta)\} \\ &= (X^{\mathrm{T}} W X)^{-1} \{X^{\mathrm{T}} W (X\beta + W^{-1} u(\beta)\} \\ &= (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \end{aligned}$$

- does not involve ϕ_i

- iteratively re-weighted least squares
- derived response $z = X\beta + W^{-1} u$

W, z both depend on β

linearized version of y