Methods of Applied Statistics I

STA2101H F LEC9101

Week 8

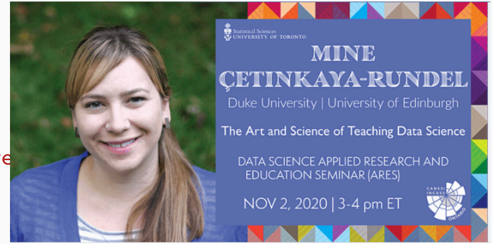
October 29 2020

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-eve`
- "The art and science of teaching data science"



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.

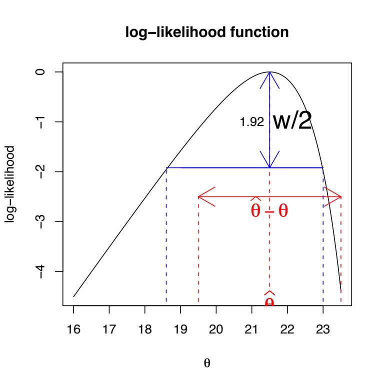
Recap

FLM-2 Ch. 5

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

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

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



Measures of risk

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

$$\mathrm{logit}(p_i) \sim \beta_0 + \beta_1 z_i, \quad i = 1, \ldots, n$$

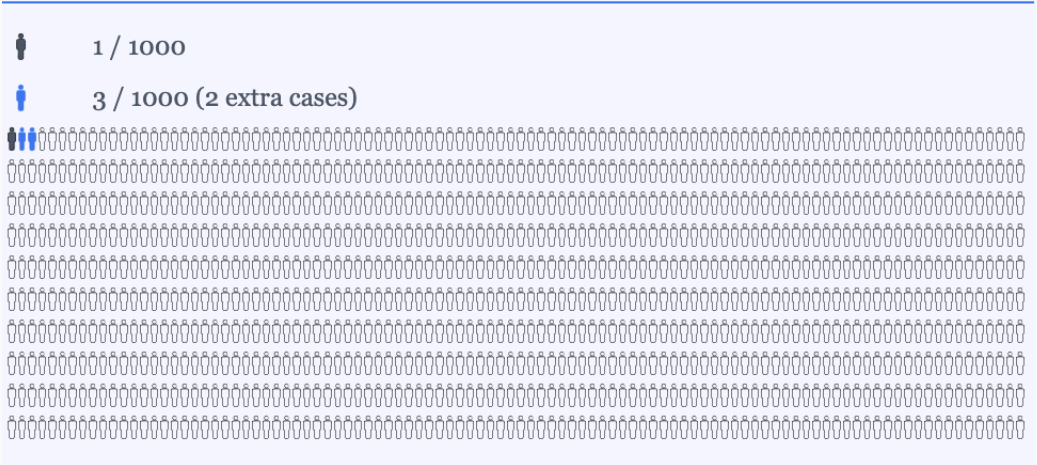
- no difference between groups $\iff$ odds-ratio $\equiv 1$

... 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
- 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

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

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

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’’ ...

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

$$z_i = x_i^T \gamma + \epsilon_i, \quad y_i = I(z_i > 0)$$

see also FELM Fig 2.3

Linear separability

FELM §2.8

ESTIMATION PROBLEMS 39

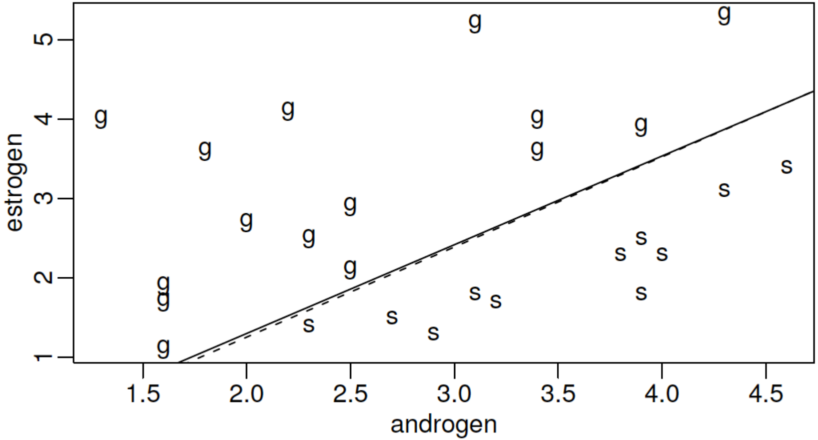


Figure 2.4 *Levels of androgen and estrogen for 15 homosexual (g) and 11 heterosexual (s)*

$\hat{p}(x^*) =$

ilogit

ED50 and delta method

- $Y_i \sim Bin(n_i, p_i) \Rightarrow \mathrm{E}(Y_i) = n_i p_i, \quad \mathrm{Var}(Y_i) = n_i p_i (1 - p_i)$
- variance is determined by the mean
- ```
  bmod <- glm(cbind(survive,total-survive) ~
                location + period,
              family = binomial,
              data = troutegg)

  summary(bmod)
   Null deviance: 1021.469  on 19  degrees of freedom
  ## Residual deviance:   64.495  on 12  degrees of freedom
  ## AIC: 157.03
  ```
- quasi-binomial: $\mathrm{E}(Y_i) = n_i p_i, \quad \mathrm{Var}(Y_i) = \phi n_i p_i (1 - p_i)$
- estimate ϕ?

over-dispersion parameter

- usually use $X^2/(n-p)$, where

$$X^2 = \sum \frac{(y_i - n_i \hat{p}_i)^2}{n \hat{p}_i (1 - \hat{p}_i)}$$
  ```

Quasibinomial

overdisp.Rmd; overdisp.html

Generalized linear models: theory

- $$f(y_i; \mu_i, \phi_i) = \exp\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i)\}$$
- $\mathrm{E}(y_i \mid x_i) = b'(\theta_i) = \mu_i$ 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

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

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

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

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

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

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

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

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) $\phi_i=\phi a_i$

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

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

Scale parameter ϕ_i

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

- 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

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

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

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

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

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

- does not involve ϕ_i

- iteratively re-weighted least squares — W, z both depend on β
- derived response $z = X\beta + W^{-1} u$ — linearized version of y