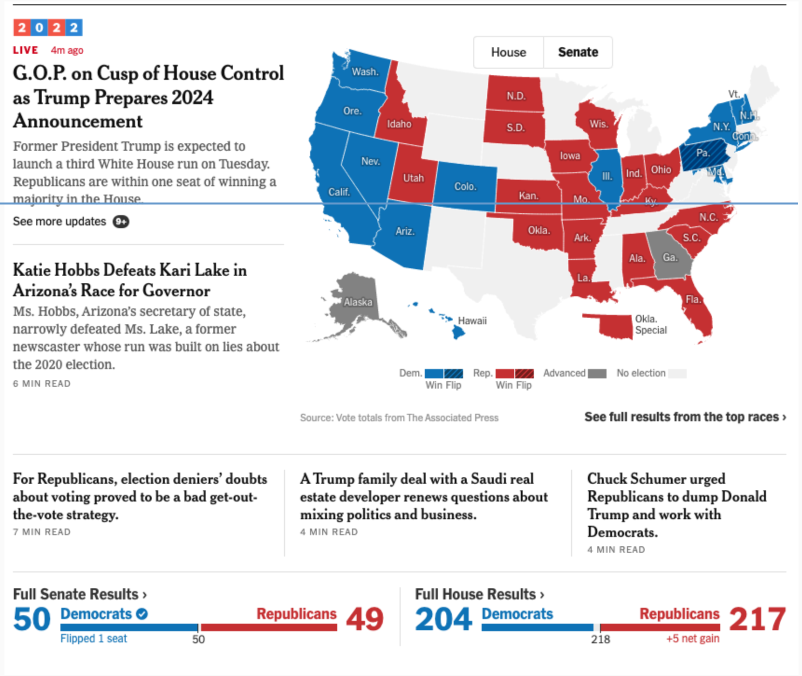
Methods of Applied Statistics I

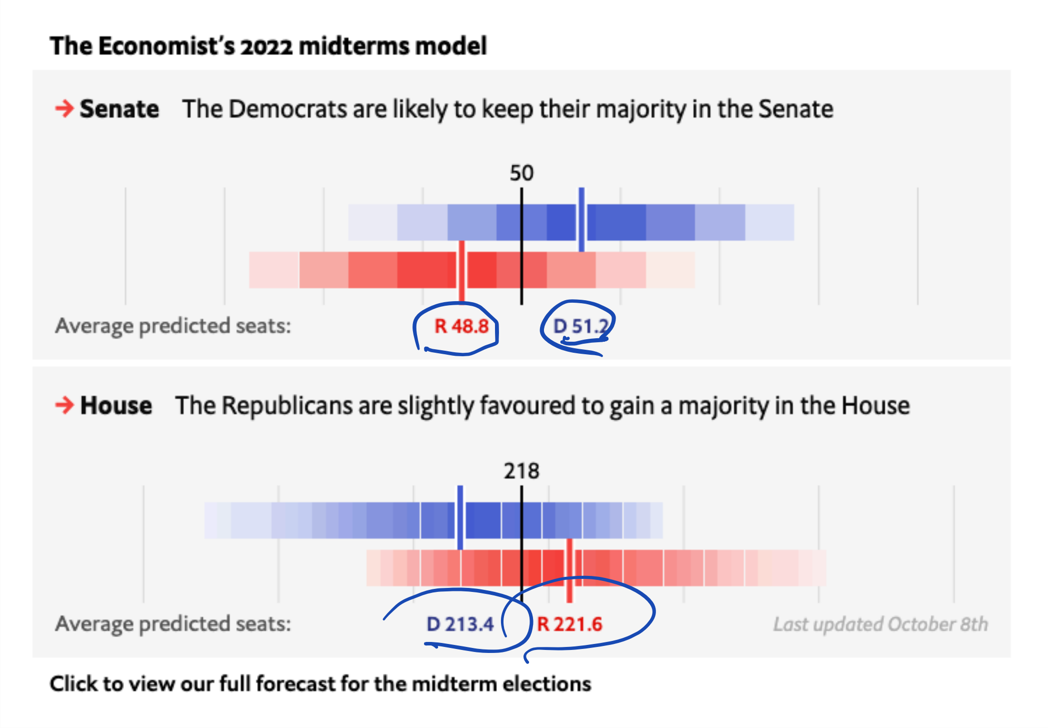
STA2101H F LEC9101

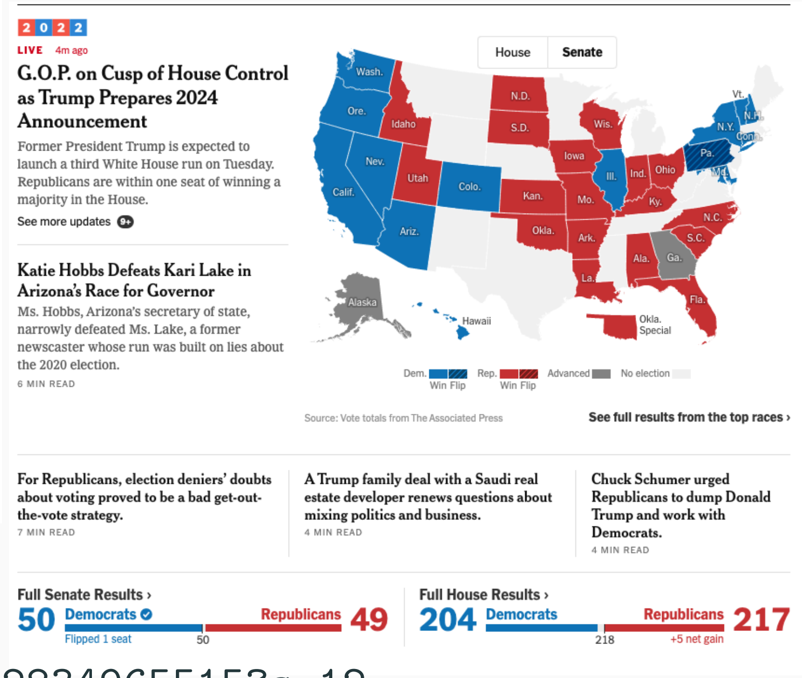
Week 9

November 16 2022



Predictions





And see "Land doesn't vote"

twitter.com/simongerman600/status/1591175192834965515?s=12

Today

1. Upcoming events
2. Reminder re Project
3. Recap
4. Finish generalized linear models
5. Survival data

random effects

nonpar. regression

Project Guidelines STA 2101F: Methods of Applied Statistics I 2022

Outline

- Part I 3–5 pages, non-technical 12 point type, 1.5 vertical spacing, thank you
  1. a description of the scientific problem of interest
  2. how (and why) the data being analyzed was collected
  3. preliminary description of the data (plots and tables)
  4. non-technical summary for a non-statistician of the analysis and conclusions
- Part II 3–5 pages, technical LaTeX or R markdown; submit .Rmd and .pdf files
  1. models and analysis
  2. summary for a statistician of the analysis and conclusions
- Part III Appendix submit .Rmd and .pdf or .html files

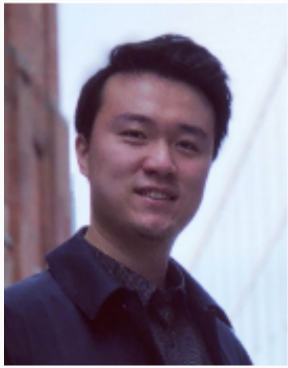
  R script or .Rmd file; additional plots; additional analysis; References

Project Marking

- 40 points total
- Part I:
  description of data and scientific problem 5
  suitability of plots and tables 5 clear, non-technical, concise but thorough
  quality of the presentation 5
- Part II:
  summary of the modelling and methods 5 justification for choices
  suitability and thoroughness of the analysis 10 model checks, data checks
- Part III:
  relevance of additional material 5
  complete and reproducible submission 5

- November 17 3.30-4.30 Statistical Sciences Seminar
  Room 9014, Hydro Building
  and online

  Qiongshi Li, U Wisconsin
  "New Advances in Genetic Risk Prediction"



- November 18 12.00-1.00 Toronto Data Workshop
  UY 9195 and zoom
  Lindsay Katz, U Toronto
  "... a new comprehensive database of all proceedings of Australian Parliamentary debates"

link

We are currently hiring a Statistical Programmer for our client, a global pharmaceutical company and one of the largest pharmaceutical companies in the world. The company is known for its success in researching developing and marketing innovative drugs.

The Statistical Programmer will be responsible for the development of SAS programs and statistical output for the management and reporting of clinical trial data managed by the Unit. Ensure quality of statistical output produced by external provider, programs tools to support data review activities and data visualization and collaborate on the interpretation and communication of trial results. ...

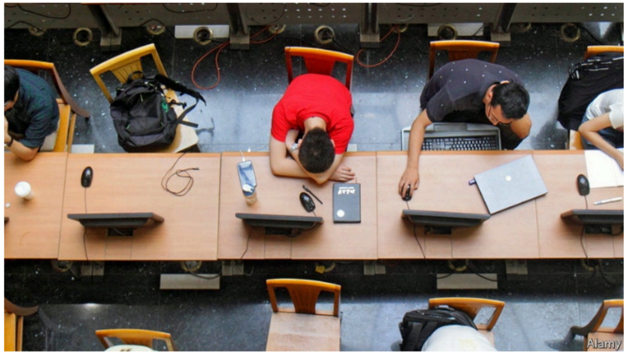
ssc.ca/employment

→ other job pages

Science & technology | Exam nerves' real effect

Anxiety does not cause bad results in exams

The problem is in the run-up, not the main event



Nov 2nd 2022

Save Share Give

"Anxiety on the day of the test did not predict exam performance at all. ... What actually hampered students, it turned out, were high levels of anxiety during the weeks before the exam took place. "

Economist

Source: Psychological Science, Theobold et al. 2022

@emitanaka@fosstodon.org @statsgen · Nov 12 ···

I highly recommend watching this talk by @monjalexander if you are interested in modern demography research or interesting use of facebook ad data. As someone who knows very little about demography, this was a great intro and an interesting application!

youtube.com

Toronto Data Workshop - Monica Alexander - Usin...

Talk on 3 December 2020 by Professor Monica Alexander, Statistical Sciences and Sociology, ...

YouTube Link

Recap

- brief review of likelihood theory
- overdispersion; `troutegg` example; Pearson's χ^2; beta-binomial `mgcv::betar`
- measures of risk: odds ratio, risk ratio, risk difference; prospective and retrospective sampling
- binary responses: non-convergence; latent variables
- generalized linear models: families, density, linear predictor, link function, variance function

Generalized linear models

`glm` has several options for `family`

```
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")
```

← write your own

Generalized linear models

`glm` has several options for `family` default link (canonical)

```
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")
```

Each of these is a member of the class of generalized linear models

Generalized: distribution of response is not assumed to be normal except G.

Linear: some transformation of $E(y_i)$ is of the form $x_i^T\beta$ link function

μ_i

Link functions

```
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" (such as generated by make.link)
provided it is not specified via one of the standard names given next.

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);
the Gamma family the links inverse, identity and log;
the poisson family the links log, identity, and sqrt;
and the inverse.gaussian family the links 1/mu^2, inverse, identity and log.

The quasi family accepts the links logit, probit, cloglog, identity, inverse, log
```

Generalized linear models: notation

- density: $f(y_i; \mu_i, \phi_i) = \exp\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i)\}$
- moments: $E(y_i \mid x_i) = b'(\theta_i) = \mu_i \quad Var(y_i \mid x_i) = \phi_i b''(\theta_i) = \phi_i V(\mu_i)$

μ_i is a function of θ_i

canonical par.

dispersion

exponential family

$V(\cdot)$ var fn

of the family

$$E(y_i) = \int y_i \, e^{\{y_i\theta_i - b(\theta_i)\}/\phi_i + c(y_i, \phi_i)} \, dy_i$$

$$\int e^{\{\ \}} dy_i = 1$$

$$\frac{d}{d\theta} \int e^{\{\ \}} = 0$$

Generalized linear models: notation

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

  μ_i is a function of θ_i
- link function: $g(\mu_i) = x_i^T\beta =$ links the n observations together via covariates
- linear predictor: $\eta_i = x_i^T\beta$

  residual plots

Generalized linear models: notation

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

  μ_i is a function of θ_i
- link function: $g(\mu_i) = x_i^T\beta =$ links the n observations together via covariates
- linear predictor: $\eta_i = x_i^T\beta$

  residual plots
- $\mathrm{Var}(y_i \mid x_i) = \phi_i b''(\theta_i) = \phi_i V(\mu_i)$
- $V(\cdot)$ is the variance function

ELM has $a_i(\phi)$ instead of ϕ_i, later $a_i(\phi) = \phi/w_i$

SM has ϕ_i, later $\phi_i = a_i\phi$

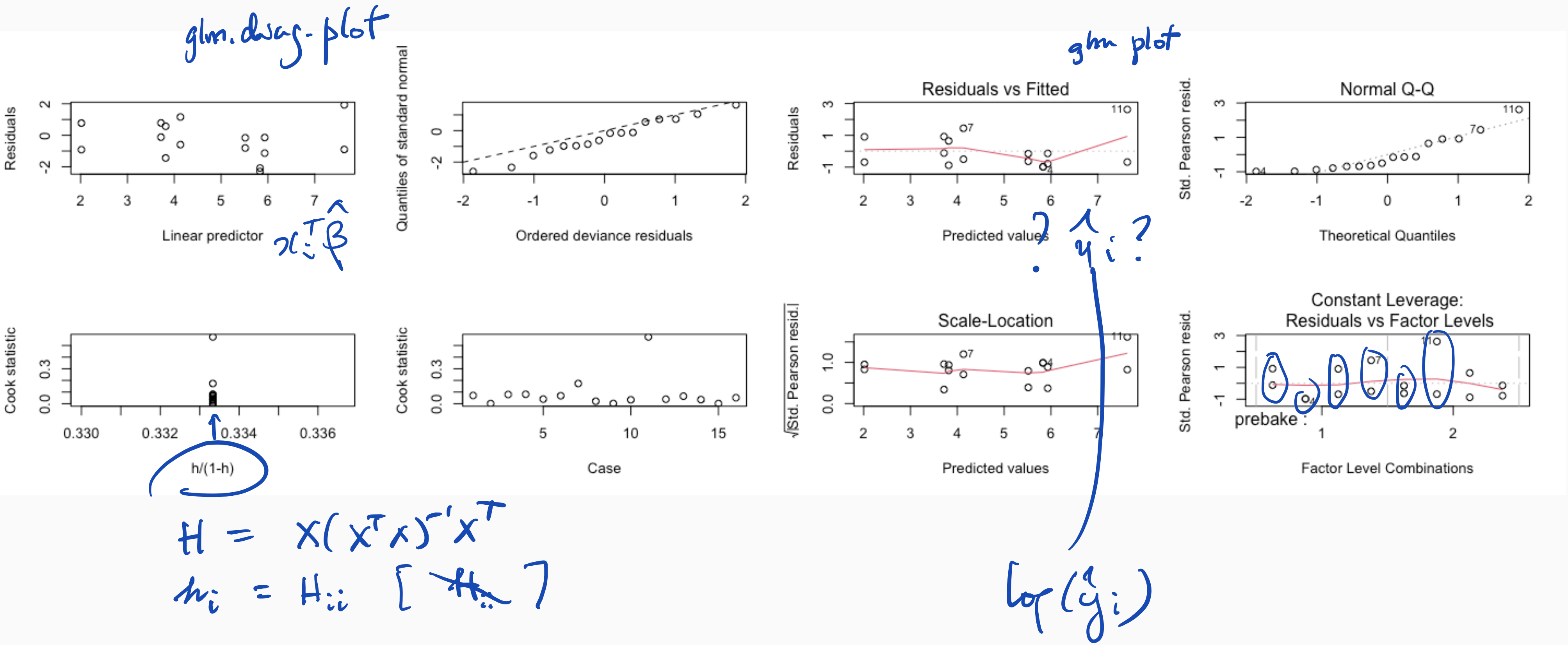
ELM $\phi_i \to \phi/w_i$

w_i known

SM $\phi_i \to a_i\phi$

a_i known

Diagnostics



$$H = X(X^T X)^{-1} X^T$$

$$h_i = H_{ii}$$

	y1	y2	y3	prebake	flux	speed	preheat	cooling	agitator	temp	mean	var
1	13	30	26	1	1	1	1	1	1	1	23.00000	79.000000
2	4	16	11	1	1	1	2	2	2	2	10.33333	36.333333
3	20	15	20	1	1	2	1	1	2	2	18.33333	8.333333
4	42	43	46	1	1	2	2	2	1	1	43.66667	4.333333
5	14	15	17	1	2	1	1	2	1	2	15.33333	2.333333
6	10	17	16	1	2	1	2	1	2	1	14.33333	14.333333
7	36	29	53	1	2	2	1	2	2	1	39.33333	152.333333
8	5	9	16	1	2	2	2	1	1	2	10.00000	31.000000
9	29	0	14	2	1	1	1	2	2	1	14.33333	210.333333
10	10	26	9	2	1	1	2	1	1	2	15.00000	91.000000
11	28	173	19	2	1	2	1	2	1	2	73.33333	7470.333333
12	100	129	151	2	1	2	2	1	2	1	126.66667	654.333333
13	11	15	11	2	2	1	1	1	2	2	12.33333	5.333333
14	17	2	17	2	2	1	2	2	1	1	12.00000	75.000000
15	53	70	89	2	2	2	1	1	1	1	70.66667	324.333333
16	23	22	7	2	2	2	2	2	2	2	17.33333	80.333333

Examples

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

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

$$\phi_i = \sigma^2,\quad \theta_i = \mu_i,\quad b(\mu_i) = \mu_i^2/2, b'(\mu_i) = \mu_i, b''(\mu_i) = 1$$

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

$$\frac{y_i\theta_i - b(\theta_i)}{\sigma^2 \Leftarrow \phi_i}$$

$$var(Y_i|x_i) = \sigma^2 \cdot 1 = \sigma^2$$

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

Examples

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

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

$$\phi_i = \sigma^2, \quad \theta_i = \mu_i, \quad b(\mu_i) = \mu_i^2/2, b'(\mu_i) = \mu_i, b''(\mu_i) = 1$$

Examples

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

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

$\phi_i = \sigma^2, \quad \theta_i = \mu_i, \quad b(\mu_i) = \mu_i^2/2, b'(\mu_i) = \mu_i, b''(\mu_i) = 1$

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

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

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

$\phi_i = \frac{1}{m_i}$

$= a_i \phi$

$\frac{1}{m_i}$ 1

$\frac{y_i\theta_i}{\phi_i}$

$b'(\theta_i) = E(y_i)$

(n_i)

Examples

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

$\phi_i=\sigma^2, \quad \theta_i=\mu_i, \quad b(\mu_i)=\mu_i^2/2, b'(\mu_i)=\mu_i, b''(\mu_i)=1$

- Binomial: $f(r_i;p_i)=\binom{m_i}{r_i}p_i^{r_i}(1-p_i)^{m_i-r_i}; \quad y_i=r_i/m_i$
$=\exp[m_iy_i\log\{p_i/(1-p_i)\}+m_i\log(1-p_i)+\log\binom{m_i}{m_iy_i}]$

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

- ELM (§8.1/6.1) uses $a_i(\phi)$ in place of ϕ_i, later $a_i(\phi)=\phi/w_i$;
SM uses ϕ_i, later (p. 483) $\phi_i=\phi a_i$

... Examples

Family	Canonical link	Variance function	ϕ_i
Normal	$\eta = \mu$	1	σ^2
Binomial	$\eta = \log\{\mu/(1-\mu)\}$	$\mu(1-\mu)$	$1/m_i$
Poisson	$\eta = \log(\mu)$	μ	1
Gamma	$\eta = 1/\mu$	μ^2	$1/\nu$
Inverse Gaussian	$\eta = 1/\mu^2$	μ^3	ξ

ϕa_i $a_i \equiv 1$

$a_i = \frac{1}{m_i}$ $\phi = 1$ $\phi = \sigma^2$

shape ν

sat'd model

μ_i not a regr. but $E(Y_i)$ $i = 1, \dots, n$

$2\{\ell(\hat{p}_i) - \ell(p_i(\hat{\beta}))\}$ = resid. dev. binom

... Examples

Family	Canonical link	Variance function	ϕ_i
Normal	$\eta = \mu$	1	σ^2
Binomial	$\eta = \log\{\mu/(1-\mu)\}$	$\mu(1-\mu)$	$1/m_i$
Poisson	$\eta = \log(\mu)$	μ	1
Gamma	$\eta = 1/\mu$	μ^2	$1/\nu$
Inverse Gaussian	$\eta = 1/\mu^2$	μ^3	ξ

Gamma: $f(y_i; \mu_i, \nu) = \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu_i}\right)^{\nu} y_i^{\nu-1}\exp(-\frac{\nu}{\mu_i})y_i$

$$= \exp[-\frac{\nu}{\mu_i}y_i - \nu\log(\frac{1}{\mu_i}) + (\nu-1)\log(y_i) + \nu\log(\nu) - \log\{\Gamma(\nu)\}]$$

$$= \exp[\nu\{\frac{y_i}{-\mu_i} - \log(\frac{1}{\mu_i}) + \log(y_i) - \log\Gamma(\nu) + \log(\nu) - \frac{1}{\nu}\log(y_i)\}]$$

canonical link conventionally ignores the $-$sign

Inference

$y_i \quad i = 1, \ldots, n \qquad f(y_i; \mu_i, \phi_i)$

log lik

- $\ell(\beta; y) = \sum_{i=1}^{n} \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i) \right\} \qquad b'(\theta_i) = \mu_i; \quad b''(\theta_i) = V(\mu_i)$
- $g(\mu_i) = g\{b'(\theta_i)\} = \eta_i = x_i^{\mathrm{T}} \beta$
- $\frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta_j} = \sum \frac{y_i - b'(\theta_i)}{\phi_i} \frac{\partial \theta_i}{\partial \beta_j}$

$g'(b'(\theta_i)) b''(\theta_i) \frac{\partial \theta_i}{\partial \beta_j} = x_{ij}$

$\sum \left(\frac{y_i - \mu_i}{\phi_i} \right) \frac{x_{ij}}{g'(\mu_i) V(\mu_i)} \qquad \frac{\partial \theta_i}{\partial \beta_j} = \frac{x_{ij}}{g'(\mu_i) V(\mu_i)}$

Inference

- $\ell(\beta; y) = \sum_{i=1}^{n} \left\{ \frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i) \right\}$ $\quad b'(\theta_i) = \mu_i; \quad b''(\theta_i) = V(\mu_i)$
- $g(\mu_i) = g\{b'(\theta_i)\} = \eta_i = x_i^{\mathrm{T}}\beta$
- $\frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta_j} = \sum \frac{y_i - b'(\theta_i)}{\phi_i} \frac{\partial \theta_i}{\partial \beta_j}$
- $g'\{b'(\theta_i)\}b''(\theta_i)\frac{\partial \theta_i}{\partial \beta_j} = x_{ij} = g'(\mu_i)V(\mu_i)$

Inference

- $\ell(\beta; y) = \sum_{i=1}^{n} \left\{ \frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i) \right\}$ $\quad b'(\theta_i) = \mu_i; \quad b''(\theta_i) = V(\mu_i)$
- $g(\mu_i) = g\{b'(\theta_i)\} = \eta_i = x_i^{\mathrm{T}}\beta$
- $\frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \frac{\partial \ell_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta_j} = \sum \frac{y_i - b'(\theta_i)}{\phi_i}\frac{\partial \theta_i}{\partial \beta_j}$
- $g'\{b'(\theta_i)\}b''(\theta_i)\frac{\partial \theta_i}{\partial \beta_j} = x_{ij} = g'(\mu_i)V(\mu_i)$
- $\frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \frac{y_i - \mu_i}{\phi_i g'(\mu_i)V(\mu_i)}x_{ij} = \sum \frac{y_i - \mu_i}{a_i\phi g'(\mu_i)V(\mu_i)}x_{ij}$

when $\phi_i = a_i\phi$

$\phi_i = a_i\phi$

Inference

- $\ell(\beta; y) = \sum_{i=1}^{n} \left\{ \frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i) \right\}$ $\quad b'(\theta_i) = \mu_i; \quad b''(\theta_i) = V(\mu_i)$
- $g(\mu_i) = g\{b'(\theta_i)\} = \eta_i = x_i^{\mathrm{T}}\beta$
- $\frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \frac{\partial \ell_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta_j} = \sum \frac{y_i - b'(\theta_i)}{\phi_i}\frac{\partial \theta_i}{\partial \beta_j}$
- $g'\{b'(\theta_i)\}b''(\theta_i)\frac{\partial \theta_i}{\partial \beta_j} = x_{ij} = g'(\mu_i)V(\mu_i)$
- $\frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \frac{y_i - \mu_i}{\phi_i g'(\mu_i)V(\mu_i)} x_{ij} = \sum \frac{y_i - \mu_i}{a_i \phi g'(\mu_i)V(\mu_i)} x_{ij} = 0$

when $\phi_i = a_i\phi$

- matrix notation:

$$\frac{\partial \ell(\beta)}{\partial \beta} = X^{\mathrm{T}}u(\beta), \quad X = \frac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad u = (u_1, \ldots, u_n), \quad u_i = \frac{y_i - \mu_i}{\phi_i g'(\mu_i)V(\mu_i)}$$

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, $\phi = 1/\nu$

Family	Canonical link	Variance function	ϕ_i
Normal	$\eta = \mu$	1	σ^2
Binomial	$\eta = \log\{\mu/(1-\mu)\}$	$\mu(1-\mu)$	$1/m_i$
Poisson	$\eta = \log(\mu)$	μ	1
Gamma	$\eta = 1/\mu$	μ^2	$1/\nu$
Inverse Gaussian	$\eta = 1/\mu^2$	μ^3	ξ

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, $\phi = 1/\nu$

Family	Canonical link	Variance function	ϕ_i
Normal	$\eta = \mu$	1	σ^2
Binomial	$\eta = \log\{\mu/(1-\mu)\}$	$\mu(1-\mu)$	$1/m_i$
Poisson	$\eta = \log(\mu)$	μ	1
Gamma	$\eta = 1/\mu$	μ^2	$1/\nu$
Inverse Gaussian	$\eta = 1/\mu^2$	μ^3	ξ

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

when $\phi_i = a_i \phi$

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

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

Solving maximum likelihood equation

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

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

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

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: $X^{\mathrm{T}}u(\hat\beta) = 0 \doteq X^{\mathrm{T}}u(\beta) + (\hat\beta - \beta)X^{\mathrm{T}}\dfrac{\partial u(\beta)}{\partial\beta^{\mathrm{T}}}$ $u_i = \frac{y_i - \mu_i}{\phi_i g'(\mu_i)V(\mu_i)}$
- change to Fisher scoring: $X^{\mathrm{T}}u(\hat\beta) = 0 \doteq X^{\mathrm{T}}u(\beta) + (\hat\beta - \beta)X^{\mathrm{T}}\mathrm{E}\left\{\dfrac{\partial u(\beta)}{\partial\beta^{\mathrm{T}}}\right\}$

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

… maximum likelihood equation

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

- $\dfrac{\partial^2 \ell(\beta; y)}{\partial \beta_j \partial \beta_k} =$

... maximum likelihood equation

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

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

... maximum likelihood equation

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

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

... maximum likelihood equation

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

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

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

$W = W(\beta)$ $\quad z = z(\beta) = X\beta + W(\beta)^{-1} u(\beta)$

... maximum likelihood equation

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

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

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

- 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

Summary

Model:

$$\mathbb{E}(y_i) = \mu_i; \qquad g(\mu_i) = x_i^T\beta; \qquad \text{Var}(y_i) = \phi_i V(\mu_i) \qquad \phi_i = a_i\phi$$

Estimation:

$$\hat{\beta} = (X^TWX)^{-1}X^TWz; \quad z = X\beta + W^{-1}u; \qquad z(\beta) = X\beta + W^{-1}(\beta)u(\beta)$$

Variance:

$$\widehat{\text{Var}}(\hat{\beta}) \doteq (X^T\widehat{W}X)^{-1}$$

W is diagonal

On pp. 118-119 of ELM, this iteration is carried out in `R` on the `bliss` data

Summary 2

$$\hat{\beta} = (X^T WX)^{-1} X^T Wz; \quad z = X\beta + W^{-1}u; \qquad z(\beta) = X\beta + W^{-1}(\beta)u(\beta)$$

$$\text{Var}(\hat{\beta}) \doteq (X^T WX)^{-1} \qquad W \text{ is diagonal}$$

$$W_{ii} = \frac{1}{\phi a_i \{g'(\mu_i)\}^2 V(\mu_i)}$$

$$u_i = \frac{y_i - \mu_i}{\phi a_i g'(\mu_i) V(\mu_i)}$$

Note $\hat{\beta}$ is free of ϕ because of W and W^{-1}, but $\text{Var}(\hat{\beta})$ depends on ϕ

Warnings

1. in ELM W is defined slightly differently (no ϕ), so he writes $\widehat{\text{Var}}(\hat{\beta}) = (X^T WX)^{-1}\hat{\phi}$
2. ELM uses w_i where SM uses $1/a_i$

Analysis of data using GLMs: overview

- choose a model, often based on type of response — or on mean/variance relationship
- fit a model, using maximum likelihood estimation — convergence (almost) guaranteed
- inference for individual coefficients $\hat{\beta}_j$ from `summary`
- inference for groups of coefficients by analysis of deviance

Analysis of data using GLMs: overview

- choose a model, often based on type of response — or on mean/variance relationship
- fit a model, using maximum likelihood estimation — convergence (almost) guaranteed
- inference for individual coefficients $\hat{\beta}_j$ from `summary`
- inference for groups of coefficients by analysis of deviance

- estimation of ϕ based on Pearson's Chi-square

typo in ELM p.121: cross out $=$ var($\hat{\mu}$)

$\hat{\beta} = (X^T W X)^{-1} X^T W z$

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

$\phi = \frac{1}{\nu}$

- analysis of deviance: see p. 121 (near bottom) — likelihood ratio tests
- diagnostics: same as for `lm` — ELM p.124; SM p.477
  - residuals: deviance or Pearson; can be standardized — ELM likes 1/2 normal plots
  - influential observations: uses hat matrix — `SMPracticals` has very good GLM diagnostics: `glm.diag`, `plot.glm.diag`

(d) $\longrightarrow$ $S_i^2 \sim \Gamma\text{amma}(\nu, \mu_i)$ $\qquad \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu_i}\right)^{\nu} {S_i^2}^{\nu-1} e^{-S_i^2 \frac{\nu}{\mu_i}}$

$$L(\beta; S_i^2) = \prod_i e^{-\nu S_i^2/\mu_i - \nu \log \mu_i + \nu \log \nu - \log\Gamma(\nu) + (\nu-1)\log S_i^2}$$

$E(S_i^2) = \mu_i \qquad \log \mu_i = x_i^T \beta \qquad \mu_i = e^{x_i^T\beta}$

$\hat{\mu}_i = \hat{S}_i^2 = e^{x_i^T\hat{\beta}}$

$\left(-\frac{1}{\mu_i} = x_i^T \beta\right)$

(c) $\widehat{\log S_i^2} \sim N\left(\mu_i', \frac{2}{n-1}\right)$

$\log\left\{\frac{y_i}{m_i} \Big/ \left(1 - \frac{y_i}{m_i}\right)\right\}$

$\downarrow$

$\tilde{y}_i$

lm

$\mu_i' = E(\log S_i^2)$

$1 = \text{var}(\log S_i^2) \qquad n = 3$

$\mu_i' = x_i^T\beta$

$\hat{\mu}_i' = x_i^T\hat{\beta}$

Analysis of survival data

SM 5.4

- response is time to event – failure of a unit, death of a patient, length of unemployment
- response takes values in $[0, \infty)$
- responses may be censored

we know $y > c$ only

Analysis of survival data

- response is time to event – failure of a unit, death of a patient, length of unemployment
- response takes values in $[0, \infty)$
- responses may be censored — we know $y > c$ only

- density function $f(y)$; cumulative distribution function $F(y)$
- survivor function $S(y) = 1 - F(y) = \mathrm{pr}(Y \geq y)$ — assume distribution of y is continuous
- hazard function $\lambda(y) = \mathrm{pr}(\text{failure "at" } y \mid \text{survival to } y)$ — force of mortality

instantaneous failure rate

- response is time to event – failure of a unit, death of a patient, length of unemployment
- response takes values in $[0, \infty)$
- responses may be censored — we know $y > c$ only

- density function $f(y)$; cumulative distribution function $F(y)$
- survivor function $S(y) = 1 - F(y) = \text{pr}(Y \geq y)$ — assume distribution of y is continuous
- hazard function $\lambda(y) = \text{pr}(\text{failure “at” } y \mid \text{survival to } y)$ — force of mortality
- $$\lambda(y) = \lim_{h \to 0} \frac{1}{h} \text{pr}(y \leq Y < y + h \mid Y \geq y) = \frac{f(y)}{S(y)}$$
- cumulative hazard function
$$\Lambda(y) = \int_0^y \lambda(u) du$$

SM uses $h(\cdot)$ for $\lambda(\cdot)$

- exponential: density $f(y) = \lambda e^{-\lambda y}$
- Weibull: survivor fn $S(y) = \exp\{-(y/\theta)^{\alpha}\}$; hazard function $\lambda(y) = \alpha\theta^{-\alpha}y^{\alpha-1}$
- Gamma: hazard function

$$\lambda(y) = \frac{\lambda^{\alpha}y^{\alpha-1}e^{-\lambda y}}{\int_y^{\infty}\lambda^{\alpha}u^{\alpha-1}e^{-\lambda u}du}$$

- log-logistic: survivor function

$$S(y) = \frac{1}{1+(\lambda y)^{\alpha}}$$

- sometimes it's more convenient to specify the density, sometimes the hazard, sometimes the survivor function

- sample of size n, we observe $(y_1, d_1), \dots, (y_n, d_n)$
- $y_i = \min(y_i^o, c_i)$ where $Y_i^o \sim f(\cdot)$ and $C_i \sim g(\cdot)$ ← indep't censoring distⁿ independent
- C_i is an associated censoring time for unit i
- $d_i = 1$ if y_i is uncensored; $d_i = 0$ if y_i is censored

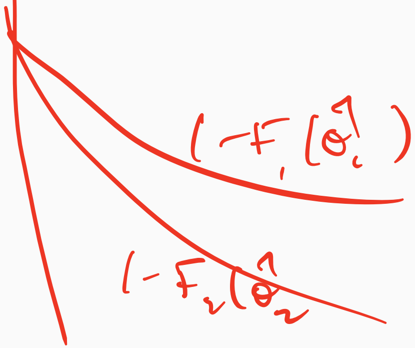
192 5 · *Models*

0+	1+	1+	3+	3+	7	10+	11+	12+	12+	15+	18+
20+	22+	22+	24+	25+	26+	31+	36+	36+	36	38	40
47+	47+	49+	53+	53+	55+	56+	57+	61+	67+	67+	70
73	75+	77+	83+	84+	88+	89+	99	121+	122+	123+	141+
0+	0+	2+	2+	2+	2+	3	3+	4+	5+	9+	10+
11	12+	13	13+	18+	22+	22+	24+	24+	24+	25+	26+
27	28	32+	35+	36	40+	43+	50+	54			

Table 5.3
Blalock–Taussig shunt data (Oakes, 1991). The table gives survival time of shunt (months after operation) for 48 infants aged over one month at time of operation, followed by times for 33 infants aged 30 or fewer days at operation. Infants whose shunt has not yet failed are marked +.

- data $(y_1, d_1), \ldots, (y_n, d_n)$
- likelihood function

$$
\begin{aligned}
L(\theta; y, d) &= \prod_{i=1}^{n} f(y_i; \theta)^{d_i} \{1 - F(y_i; \theta)\}^{1-d_i} \\
&= \prod_{i=1}^{n} \left[\frac{f(y_i; \theta)}{1 - F(y_i; \theta)} \right]^{d_i} \{1 - F(y_i; \theta)\} \\
&= \prod_{i=1}^{n} \lambda(y_i; \theta)^{d_i} \{1 - F(y_i; \theta)\}
\end{aligned}
$$



- log-likelihood function

cumulative hazard function

$$\ell(\theta; y, d) = \sum_{i=1}^{n} d_i \log\{\lambda(y_i; \theta)\} - \Lambda(y_i; \theta)$$

- any of the models above could be used for inference

SM Example 5.8: exponential, Weibull

- order the observed times $y_1 < y_2 < \cdots < y_n$
- Kaplan-Meier estimator of survivor function:

life-table

$$\widehat{S}(y) = \prod_{y_i \le y} \left(1 - \frac{1}{r_i}\right)^{d_i}$$

d_i = 1 failure
0 censored

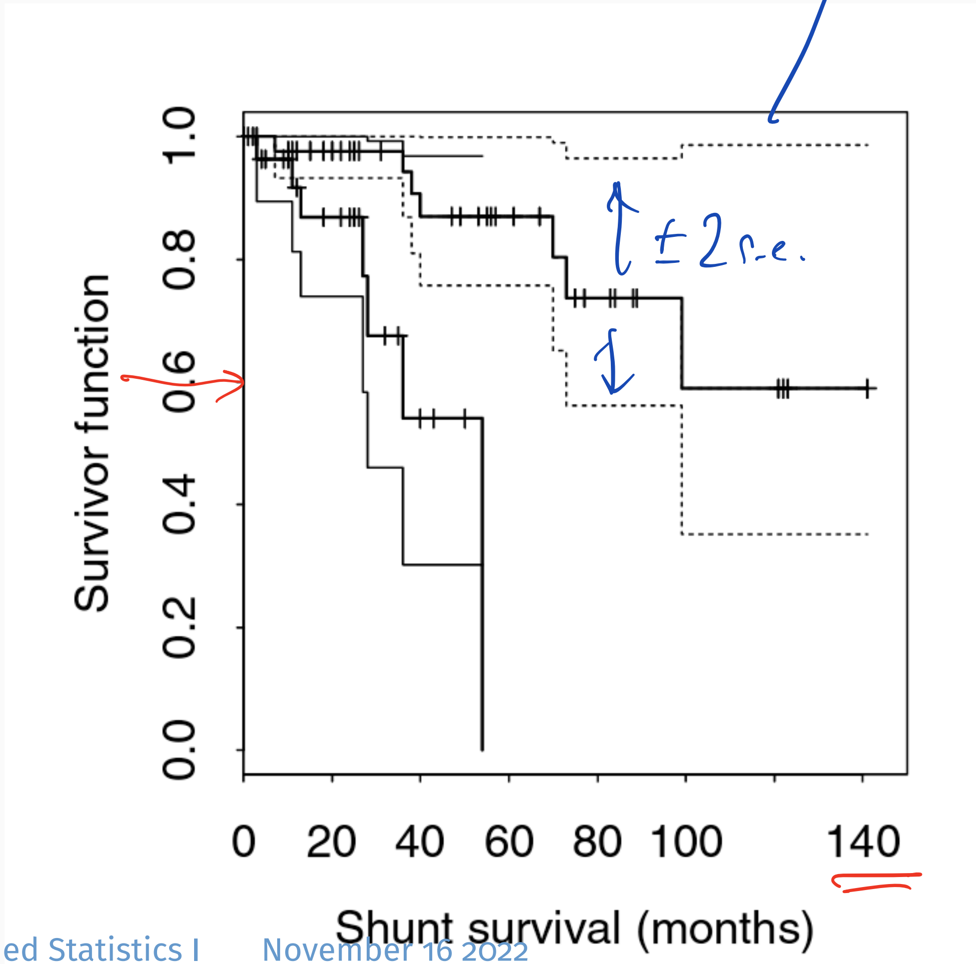
-

$$\text{var}\{\widehat{S}(y)\} \doteq \{\widehat{S}(y)\}^2 \sum_{y_i \le y} \frac{d_i}{r_i(r_i - d_i)}$$

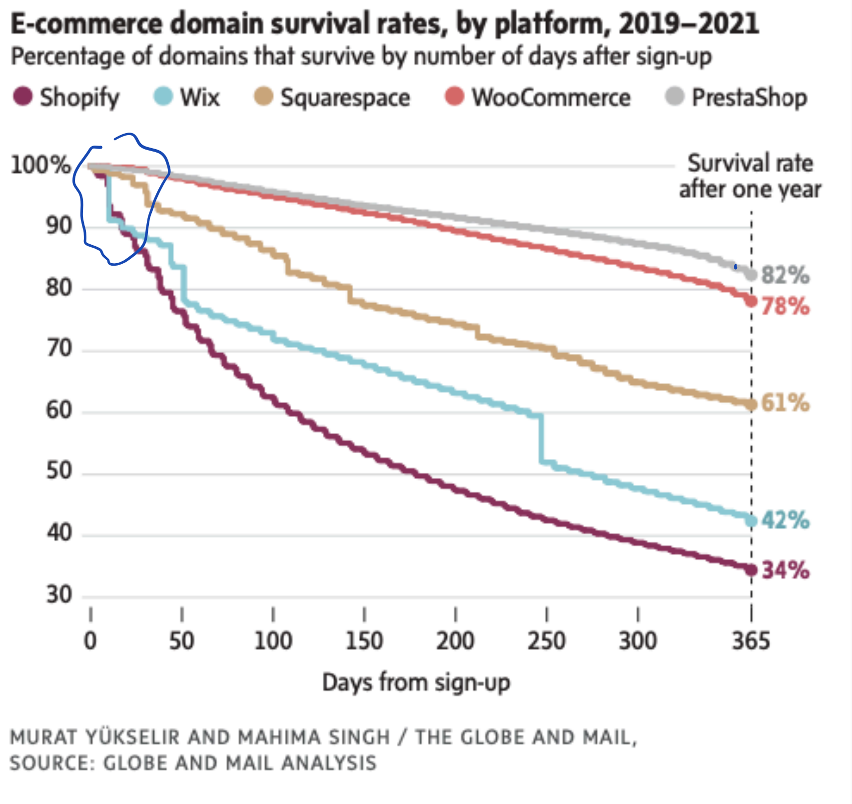
r_i =

- censoring indicator $d_i = 1$ if y_i is a failure time
- $r_i =$ number of items available to fail at time y_i
  $= \#\{j; y_j \ge y_i\}$
  $= n - i + 1$ if there are no ties in the data

… Non-parametric inference: single sample



S(t)
1−F(t) vs t



$E(y_i) = \frac{1}{\lambda_i} = \mu_i = e^{x_i^T\beta}$

- data $(y_i, d_i, x_i), i = 1, \ldots, n$ independent

  y_i still ordered
- parametric inference: model hazard function or density function in terms of $y_i \mid x_i$
- log-likelihood function $\ell(\beta; y, d, x) = \sum_{i=1}^n d_i \log\{\lambda(y_i; x_i, \beta)\} - \Lambda(y_i; x_i, \beta)$

SM Example 10.36

- proportional hazards model

$$\lambda(y_i; x_i, \beta) = \lambda_0(y_i) \exp\{x_i^T\beta\}$$

non par. parametric

cond'g on history & on "an event happened"

- Cox ~~partial~~ likelihood

$$L_{\text{part}}(\beta; t, x) = \prod_{i:\ \textit{failures}} \frac{\exp(x_i^T\beta)}{\sum_{j \in \mathcal{R}_i} \exp(x_j^T\beta)}$$

$\lambda_0(\cdot)$ gone

$d_i = 1$

- risk set $\mathcal{R}_i$ set of individuals still alive at the time the ith item fails

- partial log likelihood

$$\ell_{\text{part}}(\beta; t, x) = \sum_{failures} [x_i^T \beta - \log\{\sum_{j \in \mathcal{R}_i} \exp(x_j^T \beta)\}]$$

- inference

$$\ell'_{\text{part}}(\hat{\beta}) = 0; \qquad -\ell''_{\text{part}}(\hat{\beta}) \doteq \{\widehat{var}(\hat{\beta})\}^{-1}$$

- partial log likelihood

$$\ell_{\text{part}}(\beta; t, x) = \sum_{\text{failures}} [x_i^T \beta - \log\{\sum_{j \in \mathcal{R}_i} \exp(x_j^T \beta)\}]$$

- inference

$$\ell'_{\text{part}}(\hat\beta) = 0; \qquad -\ell''_{\text{part}}(\hat\beta) \doteq \{\widehat{var}(\hat\beta)\}^{-1}$$

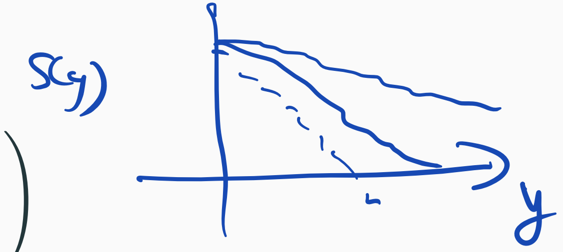
$$\hat\beta - \beta \dot\sim N(0, \widehat{var}(\hat\beta))$$

$$2\{\ell_{\text{part}}(\hat\beta) - \ell_{\text{part}}(\beta_0)\} \dot\sim \chi^2_p$$

- can be proved (but it's hard) that the usual likelihood theory applies to ℓ_{part}

- estimation of survivor function $S(y;x) = \mathrm{pr}(Y \geq y \mid x)$
- under PH model $S(y;x) = \{S_0(y)\}^{\exp(x^T\beta)}$

$$\lambda_0(t;x) = \lambda_0(t)e^{x^T\beta}$$



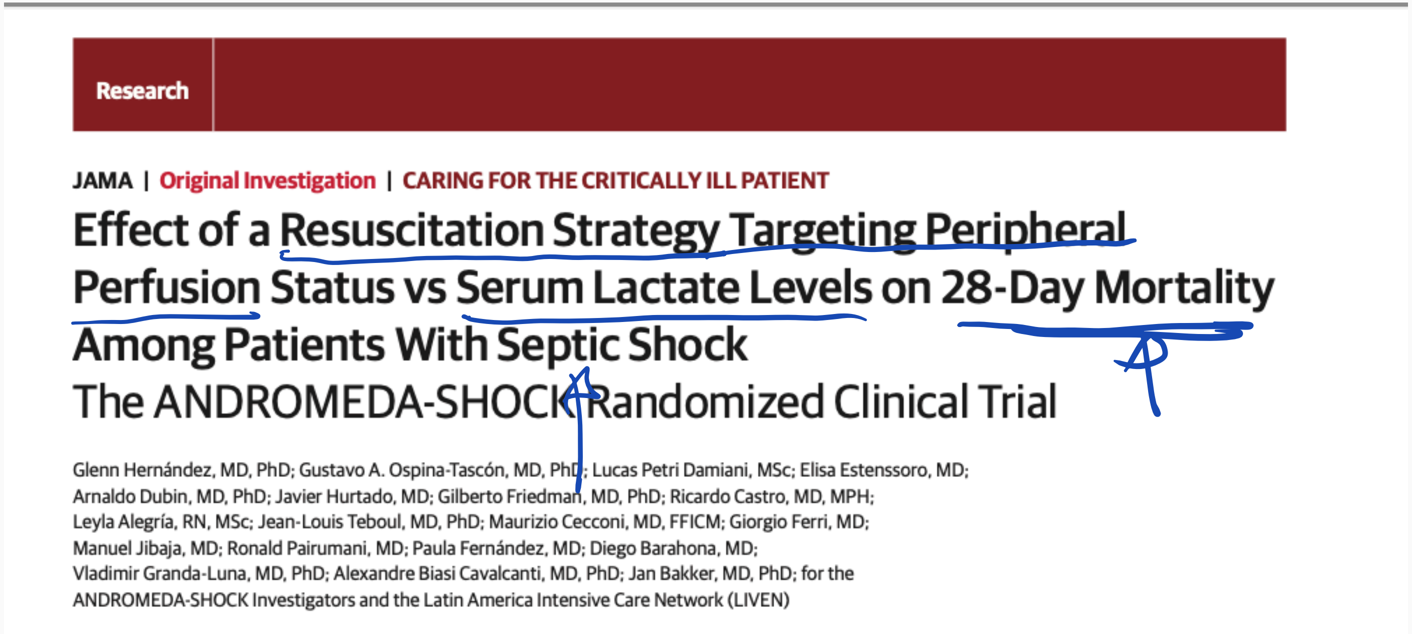
- estimate baseline survivor function

$$\widehat{S}_0(y) = \prod_{i:y_i \leq y}\left(1 - \frac{d_i}{\sum_{j\in\mathcal{R}_i}\exp(x_j^T\hat{\beta})}\right)$$

- estimate survivor function for individual with covariates x_+:

$$\widehat{S}(y;x_+) = \{\widehat{S}_0(y)\}^{exp(x_+^T\hat{\beta})}$$

Example



Research

JAMA | Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

Effect of a Resuscitation Strategy Targeting Peripheral Perfusion Status vs Serum Lactate Levels on 28-Day Mortality Among Patients With Septic Shock

The ANDROMEDA-SHOCK Randomized Clinical Trial

Glenn Hernández, MD, PhD; Gustavo A. Ospina-Tascón, MD, PhD; Lucas Petri Damiani, MSc; Elisa Estenssoro, MD; Arnaldo Dubin, MD, PhD; Javier Hurtado, MD; Gilberto Friedman, MD, PhD; Ricardo Castro, MD, MPH; Leyla Alegría, RN, MSc; Jean-Louis Teboul, MD, PhD; Maurizio Cecconi, MD, FFICM; Giorgio Ferri, MD; Manuel Jibaja, MD; Ronald Pairumani, MD; Paula Fernández, MD; Diego Barahona, MD; Vladimir Granda-Luna, MD, PhD; Alexandre Biasi Cavalcanti, MD, PhD; Jan Bakker, MD, PhD; for the ANDROMEDA-SHOCK Investigators and the Latin America Intensive Care Network (LIVEN)

link

"the treatment effect on the primary outcome was calculated with Cox proportional hazards, with adjustment for 5 pre-specified baseline covariates"

"results are reported as hazard ratios with 95% confidence intervals and Kaplan-Meier curves"

$$\hat{S}(t;\hat{\beta}) = \hat{S}_0(t)^{e^{x^T\hat{\beta}}}$$

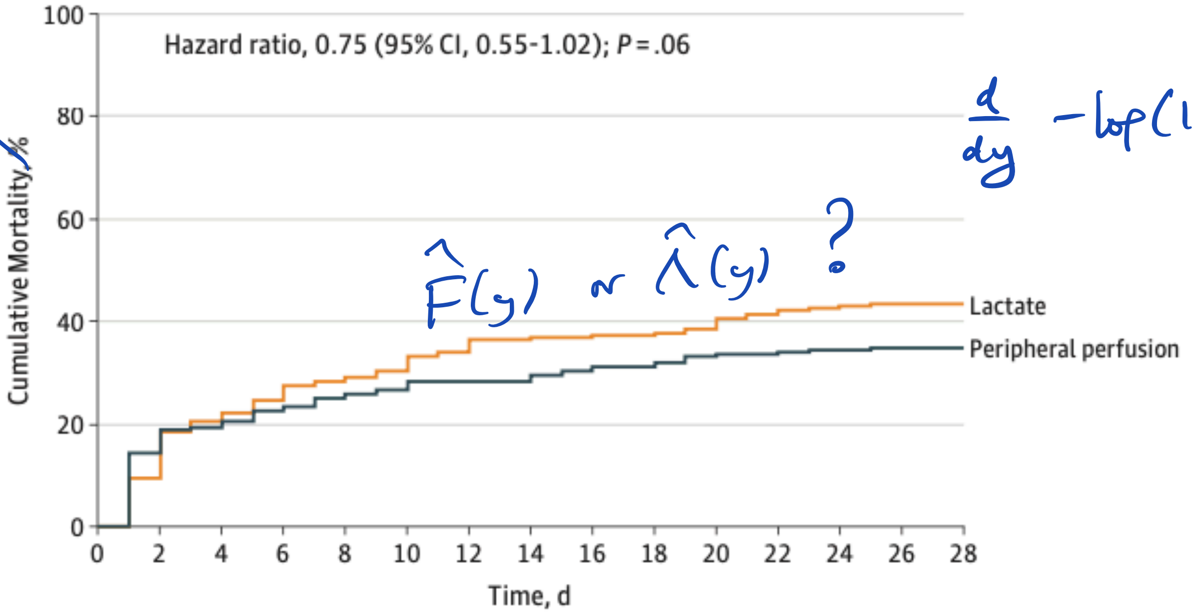
… Example

Research **Original Investigation** Effect on Septic Shock Mortality of Resuscitation Targeting Peripheral Perfusion vs Serum Lactate Levels

Table 2. Main Outcomes of the Study of Resuscitation Strategies in Septic Shock

Outcome	Peripheral Perfusion-Targeted Resuscitation (n = 212)	Lactate Level-Targeted Resuscitation (n = 212)	Unadjusted Absolute Difference (95% CI)	Adjusted Relative Measure (95% CI)	P Value
Primary Outcome					
Death within 28 d, No. (%)	74 (34.9)	92 (43.4)	−8.5 (−18.2 to 1.2)[b]	HR, 0.75 (0.55 to 1.02)[a]	.06[a]
Secondary Outcomes					
Death within 90 d, No. (%)	87 (41.0)	99 (46.7)	−5.7 (−15.6 to 4.2)[b]	HR, 0.82 (0.61 to 1.09)[a]	.17[a]

Figure 2. Kaplan-Meier Estimates of Cumulative Mortality Within 28 Days Among Patients Treated With Peripheral Perfusion–Targeted Resuscitation vs Lactate Level–Targeted Resuscitation



No. at risk															
Lactate	212	192	168	160	152	148	140	135	134	133	130	124	122	120	120
Peripheral perfusion	212	182	171	164	159	155	152	152	148	146	142	141	139	138	138

$\hat{\lambda}(t)$

$\hat{\Lambda}(t)$

$\lambda_\theta(y) = \frac{f(y)}{S(y)}$

$\Lambda(y) = -\log S(y)$

$\frac{d}{dy} -\log(1-F(y)) = \frac{f(y)}{1-F(y)}$

Hazard ratio, 95% confidence interval, and *P* value were calculated with a Cox proportional hazards model that included as covariates baseline Acute Physiology and Chronic Health Evaluation (APACHE) II score,[23] Sequential Organ Failure Assessment (SOFA) score,[24] lactate level, capillary refill time, and source of infection. Median follow-up for peripheral perfusion–targeted resuscitation was 28 days (interquartile range, 8-28 days) and for lactate level–targeted resuscitation was 28 days (interquartile range, 6-28 days).

Example

Fox and Weisberg, 2011

See Appendix to *An R Companion to Applied Regression*

```
> library(car)
> data(Rossi)
> Rossi[1:5, 1:10]
  week arrest fin age  race wexp         mar paro prio educ
1   20      1  no  27 black   no not married  yes    3    3
2   17      1  no  18 black   no not married  yes    8    4
3   25      1  no  19 other  yes not married  yes   13    3
4   52      0 yes  23 black  yes     married  yes    1    5
5   52      0  no  19 other  yes not married  yes    3    3

> mod.allison <- coxph(Surv(week,arrest) ~ fin + age + race + wexp + mar + paro
    + prio, data = Rossi)
```

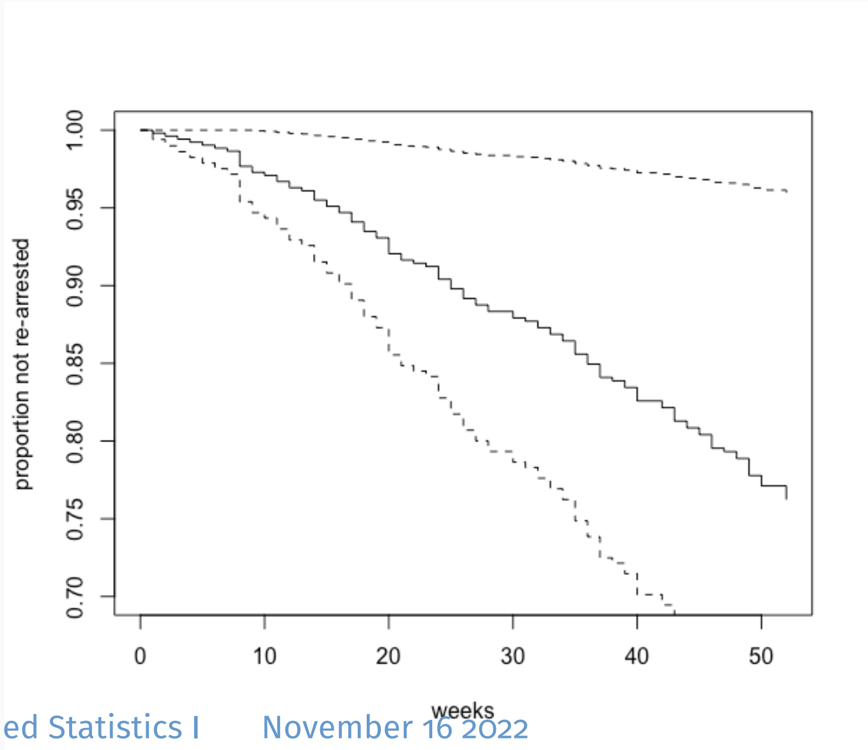
```
> summary(mod.allison)
Call:
coxph(formula = Surv(week, arrest) ~ fin + age + race + wexp +
    mar + paro + prio, data = Rossi)

  n= 432, number of events= 114

                  coef  exp(coef) se(coef)      z  Pr(>|z|)
finyes        -0.37942    0.68426  0.19138 -1.983  0.04742 *
age           -0.05744    0.94418  0.02200 -2.611  0.00903 **
raceother     -0.31390    0.73059  0.30799 -1.019  0.30812
wexpyes       -0.14980    0.86088  0.21222 -0.706  0.48029
marnot married 0.43370    1.54296  0.38187  1.136  0.25606
paroyes       -0.08487    0.91863  0.19576 -0.434  0.66461
prio           0.09150    1.09581  0.02865  3.194  0.00140 **
```

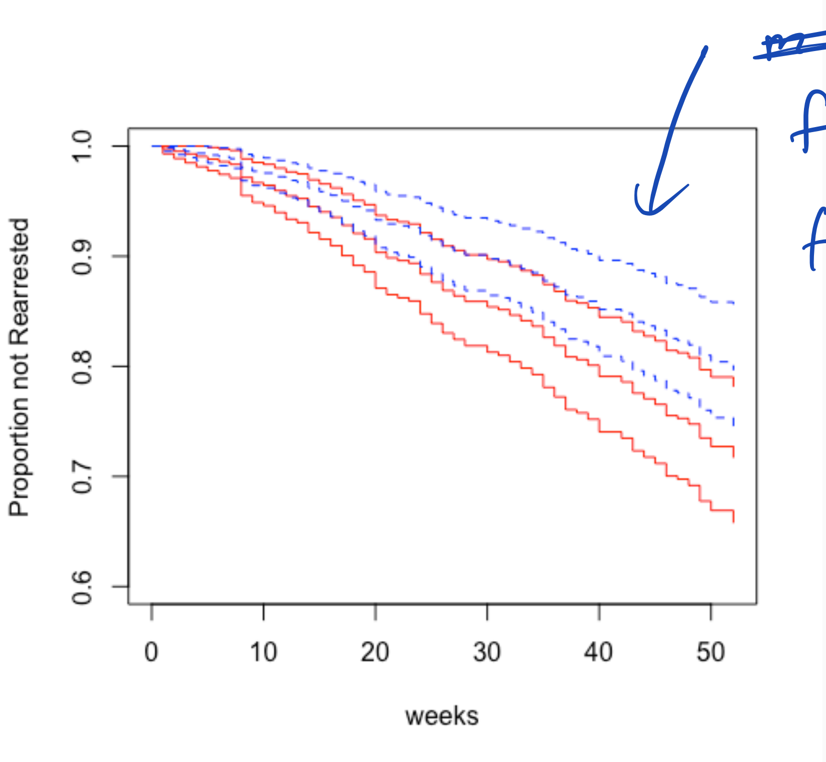
$\hat{\beta}$

… Example

Fox and Weisberg, 2011

```
> plot(survfit(mod.allison), ylim=c(.7,1), xlab="weeks",
    ylab = "proportion not re-arrested")
```





fin = 1

fin = 0