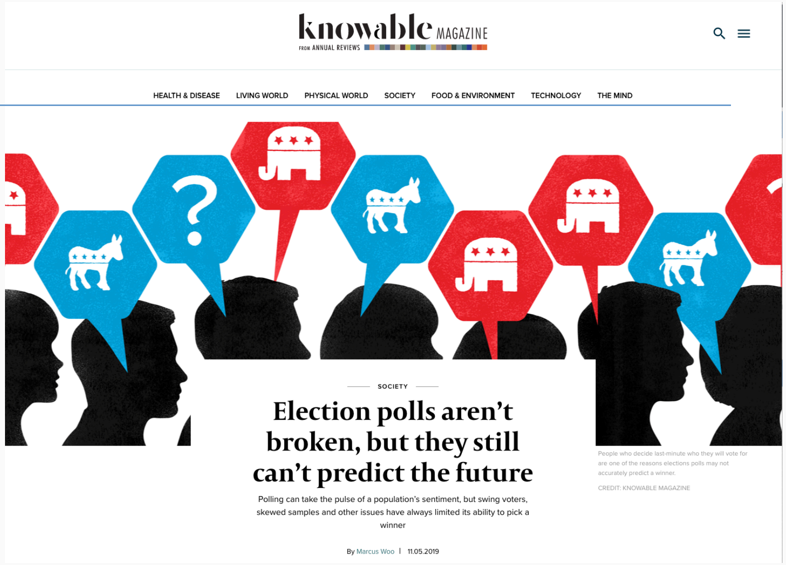
Methods of Applied Statistics I

STA2101H F LEC9101

Week 9

November 5 2020



Technology & Ideas

Polling Failed. It's Time to Kick the Addiction

Doubling down won't help Americans understand themselves.

By Cathy O'Neil
November 4, 2020, 2:13 PM EST

"It's no longer reasonable to assume that we can get better. For all the effort that a lot of smart people have put into it, polling is just hard. There's too much problematic, biased and missing data. People who don't trust the polls don't talk to pollsters. Sometimes they flat out lie."

link

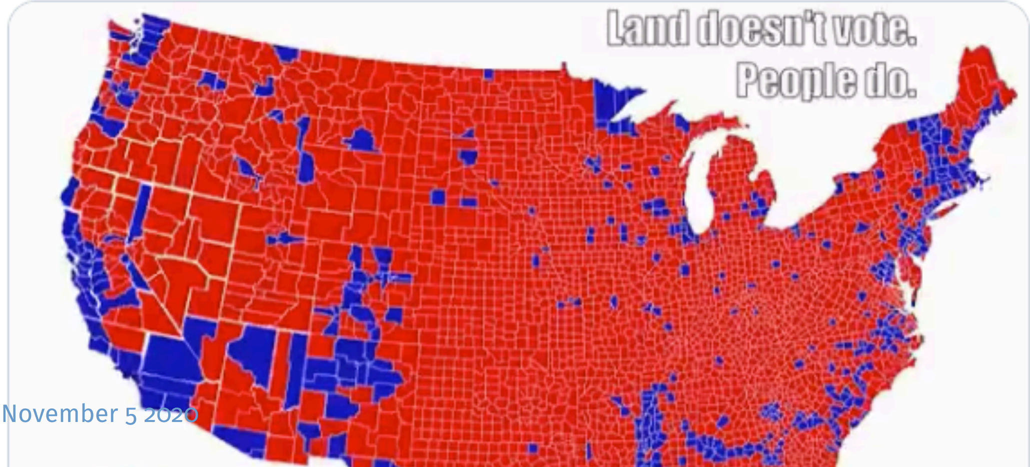
The Polls Underestimated Trump — Again. Nobody Agrees on Why.

No matter who ends up winning, the industry failed to fully account for the missteps that led it to miscalculate Donald J. Trump's support four years ago.

Visualization

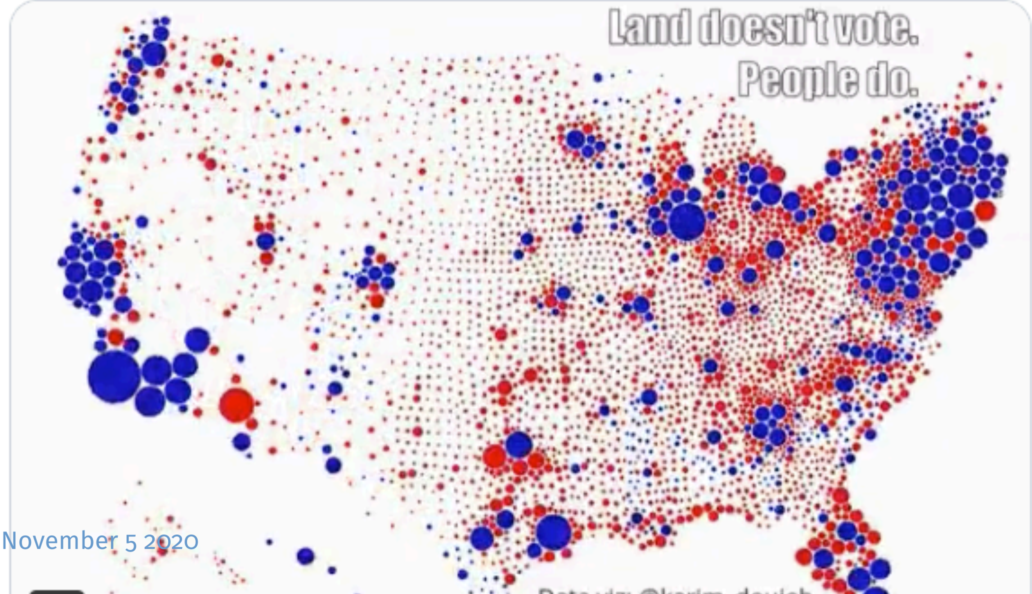
Bettina Forget
@BettinaForget

Data visualization insights:
Land doesn't vote. People do.
#USElection2020

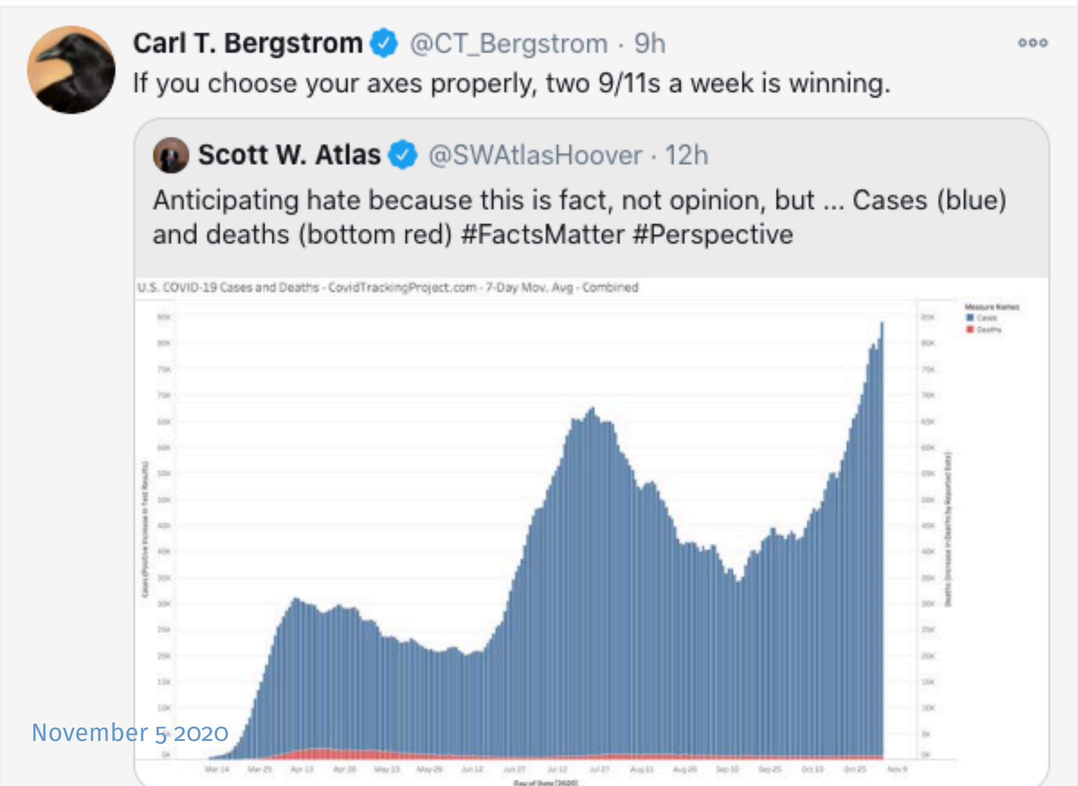


Visualization

Data visualization insights:
Land doesn't vote. People do.
#USElection2020



Visualization



The double *y*-axis chart

blog.datawrapper.de › dualaxis

Why not to use two axes, and what to use instead | Chartable

May 8, 2018 — We believe that **charts** with **two** different **y-axes** make it hard for most people to intuitively make right statements about **two** data series.

People also search for

- how to read dual y-axis graph
- dual axis bar chart maker
- when to use dual axis
- dual line graph
- bar charts different scales
- what is the purpose of a linear trendline?

infogram.com › create › dual-axis-chart

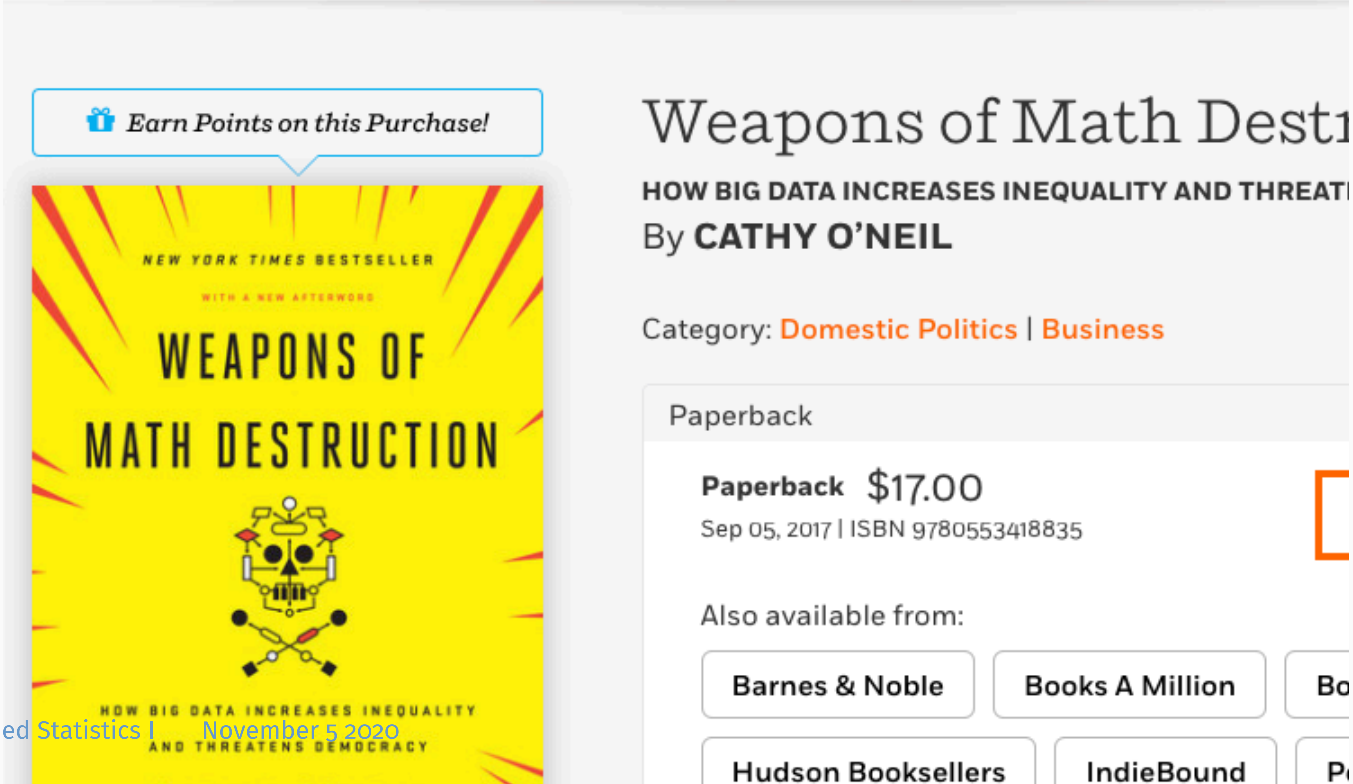
Create a stunning dual axis chart and engage your viewers

1. Make sure your **Y-axes** are related · **2**. Place primary **Y-axis** on the left. · 3. Use contrasting colors. · 4. Give diversity to your data. · 5. Avoid clutter. · 6.

Recommended Holiday Reading

Now available! *Calling Bullshit: The Art of Skepticism in a Data-Driven World*, by Carl Bergstrom and Jevin West. Available here

Recommended Holiday Reading



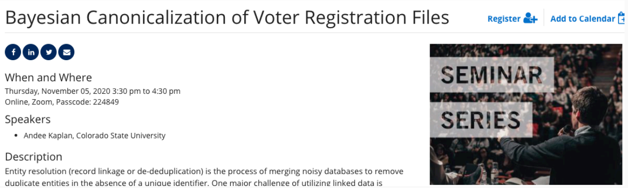
Today

Start Recording

1. correction re log link in gaussian glm
2. HW2 due November 5
3. HW 3 and Final HW
4. Generalized linear models theory
5. Generalized linear models examples

- November 6 12.00 – 13.00 Genevera Allen
- `https://canssiontario.utoronto.ca/?mec-events=stage_isss_genevera_allen`
- "Data Integration: Data-Driven Discovery from Diverse Data Sources"

1. correction re log link in gaussian glm
2. HW2 due November 5
3. HW 3 and Final HW
4. Generalized linear models theory
5. Generalized linear models examples



- November 6 12.00 – 13.00 Genevera Allen
- `https://cansiontario.utoronto.ca/?mec-events=stage_iss`
- "Data Integration: Data-Driven Discovery from Diverse Data Sources"

STAGE ISSS: Genevera I. Allen

Correction

```
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")
```

Arguments

`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

Recap

- odds ratio, risk ratio, risk difference
- case-control studies; prospective and retrospective sampling

Recap

- odds ratio, risk ratio, risk difference
- case-control studies; prospective and retrospective sampling
- binomial/binary response
  - link function
  - tolerance distribution
  - prediction intervals
  - ED50
  - overdispersion

Recap

- odds ratio, risk ratio, risk difference
- case-control studies; prospective and retrospective sampling
- binomial/binary response
  - link function
  - tolerance distribution
  - prediction intervals
  - ED50
  - overdispersion
- Generalized linear models theory
  - form of density
  - link function and linear predictor
  - variance function
  - normal and binomial examples

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

Generalized linear models: theory

- $f(y_i; \mu_i, \phi_i) = \exp\{\dfrac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i)\}$
- $E(y_i \mid x_i) = b'(\theta_i) = \mu_i$ defines μ_i as a function of θ_i

Generalized linear models: theory

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

- $\mathrm{E}(y_i \mid x_i) = b'(\theta_i) = \mu_i$ 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

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
- $g(\mu_i) = x_i^T\beta = \eta_i$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_i is the linear predictor

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

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

Examples

family {stats} R Docur

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")
```

centering

Examples

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

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

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

Examples

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

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

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

Examples

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

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

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

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

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

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

Examples

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

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

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

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

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

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

- ELM (p.115) uses $a_i(\phi)$ in place of ϕ_i, later (p.117) $a_i(\phi) = \phi/w_i$; 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

Gamma: $\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu} y^{\nu-1}\exp(-\frac{\nu}{\mu})y$

Inference

- $\ell(\beta; y) = \sum\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i)\}$ $b'(\theta_i) = \mu_i$; $b''(\theta_i) = V(\mu_i)$
- $g(\mu_i) =$ $g\{b'(\theta_i)\} = x_i^T\beta = \eta_i$
- $\frac{\partial \ell(\beta; y)}{\partial \beta_j} =$ $\sum \frac{1}{\phi_i}\{y_i \frac{\partial\theta_i}{\partial\beta_j} - b'(\theta_i)\frac{\partial\theta_i}{\partial\beta_j}\} = \sum \frac{1}{\phi_i}(y_i - \mu_i)\frac{\partial\theta_i}{\partial\beta_j}$
- $g'\{b'(\theta_i)\}b''(\theta_i)\frac{\partial\theta_i}{\partial\beta_j} =$ x_{ij} $\frac{\partial\theta_i}{\partial\beta_j} = \frac{x_{ij}}{g'(\mu_i)V(\mu_i)}$
- $\frac{\partial \ell(\beta; y)}{\partial \beta_j} =$ $\sum \frac{1}{\phi_i}\frac{(y_i - \mu_i)x_{ij}}{g'(\mu_i)V(\mu_i)} = 0$ defines $\hat\beta_j$, $j = 1, \ldots, p$ when $\phi_i = a_i\phi$
- matrix notation:

$$\frac{\partial \ell}{\partial \beta} = X^T_{p\times n} \cdot \underset{\sim}{u}_{n\times 1}$$

$u_i = \frac{1}{\phi_i}(y_i - \mu_i) / g'(\mu_i)V(\mu_i)$

Inference

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

$1 \times \sigma^2$

m_i

1

- 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

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, $\phi = 1/\nu$
- $\frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \frac{y_i - \mu_i}{\phi_i g'(\mu_i) V(\mu_i)} x_{ij} = \sum \frac{y_i - \mu_i}{a_i \phi g'(\mu_i) V(\mu_i)} x_{ij}$

when $\phi_i = a_i \phi$

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, $\phi = 1/\nu$
- $$\frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \frac{y_i - \mu_i}{\phi_i g'(\mu_i) V(\mu_i)} x_{ij} = \sum \frac{y_i - \mu_i}{a_i \phi g'(\mu_i) V(\mu_i)} x_{ij}$$

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

defines iterative scheme

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

many books use $I(\beta)$

- $\hat\beta^{(t+1)} =$ $\hat\beta^{(t)} + i^{-1}(\hat\beta^{(t)}) \cdot \ell'(\hat\beta^{(t)})$ ($p \times p$, $p \times 1$)
- applied to matrix version: $X^{\mathrm{T}} u(\hat\beta) =$ $0 = X^T u(\beta) + (\hat\beta - \beta) X^T \frac{\partial u(\beta)}{\partial \beta^T}$ (ℓ'')
- change to Fisher scoring: $\hat\beta^{(t+1)} =$ $\beta^{(t)} + i^{-1}(\hat\beta^{(t)}) X^T u(\beta)$

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

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 $I(\beta)$

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

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

- change to Fisher scoring: $\hat{\beta} = \beta + i(\beta)^{-1}X^{\mathrm{T}}u(\beta)$

… maximum likelihood equation

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

$\ell(\beta) = \sum \frac{1}{\phi_i} y_i \theta_i \ldots$

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

- $\mathrm{E}\left(-\frac{\partial^2 \ell(\beta; y)}{\partial\beta_j \partial\beta_k}\right) =$
-

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

- 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

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

W, z both depend on β

linearized version of y

And still to come...

- estimation of ϕ
- analysis of deviance

likelihood ratio tests

- diagnostics
- overdispersion and quasi-likelihood

$\implies$ GLM Examples