# Methods of Applied Statistics I

STA2101H F LEC9101

Week 10

November 24 2021

# Today

Start Zoom

1. Upcoming events
2. In the News
3. Theory of GLMs
4. Reminder: HW10 ready Nov 25, due Dec 2, is the final HW for the term
5. Reminder: Final Project due Dec 17 23.59 — PhD Dec 20 09.00
6. Office Hour Nov 24 16.30 – 18.00 — this week only

# Upcoming

- Monday Nov 29 3.30 Data Science ARES series
  Policy Questions, Messy Data: Three approaches to turning messy data into information for public policy Link

Dr. Krista Gile, U MASS

# Upcoming

- Monday Nov 29 3.30 Data Science ARES series
  Policy Questions, Messy Data: Three approaches to turning messy data into information for public policy Link

Dr. Krista Gile, U MASS

- Friday Nov 26 Toronto Data Workshop

Kieran Campbell U of T

Dear friends,

Toronto Data Workshop this Friday, 26 November, at noon (Toronto time) hosts Professor Kieran Campbell on the intersection of biomedical data and data science.

Dr. Kieran Campbell is an investigator at the Lunenfeld-Tanenbaum Research Institute and an assistant professor at the Departments of Molecular Genetics and Statistical Sciences, University of Toronto. His research focusses on Bayesian models and machine learning for high dimensional biomedical data, including single-cell and cancer genomics. Recently, he has led efforts to develop statistical machine learning methodology to integrate single-cell RNA and DNA sequencing data to uncover the effects of tumour clonal identity on gene expression, as well as methods to automatically delineate the tumour microenvironment from single-cell RNA-sequencing data. Such findings can improve our understanding of cancer progression and of why certain tumours are resistant to therapies, leading to relapse. https://www.camlab.ca

# ... Upcoming

- Thursday Nov 4 3.30
  Diffusion Schrödinger Bridges with Applications to Score-Based Generative Modelling

  Arnaud Doucet, U Oxford

Zoom Link

# Project

- Part I 3–5 pages, non-technical
  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
  1. models and analysis
  2. summary for a statistician of the analysis and conclusions
- Part III Appendix

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

- 40 points total

- Part I:
  description of data and scientific problem 5
  suitability of plots and tables 5
  quality of the presentation 5

clear, non-technical, concise but thorough

- 40 points total
- Part I:
  description of data and scientific problem 5
  suitability of plots and tables 5
  quality of the presentation 5

  clear, non-technical, concise but thorough
- Part II:
  summary of the modelling and methods 5
  suitability and thoroughness of the analysis 10

  justification for choices
  model checks, data checks

- 40 points total
- Part I:
  description of data and scientific problem 5
  suitability of plots and tables 5
  quality of the presentation 5

  clear, non-technical, concise but thorough
- Part II:
  summary of the modelling and methods 5
  suitability and thoroughness of the analysis 10

  justification for choices
  model checks, data checks
- Part III:
  relevance of additional material 5
  complete and reproducible submission 5

# Recap

- binomial regression: deviance residuals, Pearson residuals, Pearson $X^2$, non-canonical link functions
- Poisson regression: deviance residuals, Pearson residuals, Pearson $X^2$, non-canonical link functions HW 8
- overdispersion, quasi-Poisson, quasi-Binomial
- measures of risk: odds ratio, risk ratio, risk difference, prospective/retrospective sampling
- glm theory Part I

# Residuals

thanks Adewale

I managed to compute the residuals following the formula on the slides and got results which agree exactly with R. I used the formula

$$r_{DI} = \text{sign}(y_i - \hat{p}_i)\sqrt{2[y_i \log\{y_i/n_i\hat{p}_i\} + (n_i - y_i)\log\{(n_i - y_i)/(n_i - n_i\hat{p}_i)\}]}.$$

The only difference just being substituting the "$\pm$" with "sign$(y_i - \hat{y}_i)$. Since the data is bernoulli, I used $n_i =$ and $\hat{p}_i = \hat{y}_i$.

# In the News

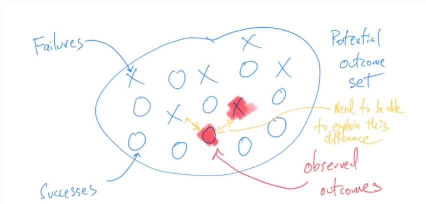
- Guardian, Nov 14 Spiegelhalter & Masters
  "On Covid we need to be careful when we talk about numbers"


On Covid, we need to be careful when we talk about numbers
*David Spiegelhalter and Anthony Masters*

A recent wave of mistakes shows how misinterpreting data risks misrepresenting the impact of the virus

| Cases | Deaths | Healthcare | Testing |
|---|---|---|---|
| **People tested positive** | **Deaths within 28 days of positive test** | **Patients admitted** | **Virus tests conducted** |
| Latest data provided on 25 October 2021 | Latest data provided on 25 October 2021 | Latest data provided on 19 October 2021 | Latest data provided on 24 October 2021 |
| Daily 36,567 | Daily 38 | Daily 1,082 | Daily 785,553 |
| Last 7 days 315,698 ↑6,685 (2.2%) | Last 7 days 942 ↑73 (8.4%) | Last 7 days 6,730 ↑1,125 (20.1%) | Last 7 days 6,307,015 ↓-147,832 (-2.3%) |
| Rate per 100,000 people: 486.9 | Rate per 100,000 people: 1.3 | | |
| All cases data | All deaths data | All healthcare data | All testing data |

- Simply Statistics, Nov 10 Peng
  Thinking about failure in data analysis



Data analysis potential outcome set.

# In the News

Nature Behaviour, Nov Wagenmakers et al.

Seven steps towards more transparency in statistical practice

nature human behaviour

PERSPECTIVE

https://doi.org/10.1038/s41562-021-01211-8

Check for updates

## Seven steps toward more transparency in statistical practice

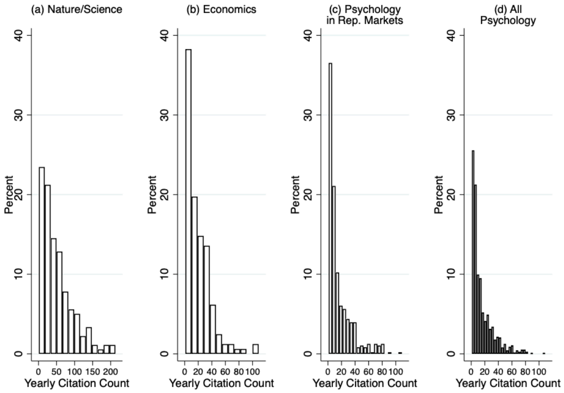
Eric-Jan Wagenmakers[1]✉, Alexandra Sarafoglou[1], Sil Aarts[2], Casper Albers[3], Johannes Algermissen[4], Štěpán Bahník[5], Noah van Dongen[1], Rink Hoekstra[6], David Moreau[7], Don van Ravenzwaaij[8], Aljaž Sluga[9], Franziska Stanke[10], Jorge Tendeiro[8,11] and Balazs Aczel[12]

**We argue that statistical practice in the social and behavioural sciences benefits from transparency, a fair acknowledgement of uncertainty and openness to alternative interpretations. Here, to promote such a practice, we recommend seven concrete statistical procedures: (1) visualizing data; (2) quantifying inferential uncertainty; (3) assessing data preprocessing choices; (4) reporting multiple models; (5) involving multiple analysts; (6) interpreting results modestly; and (7) sharing data and code. We discuss their benefits and limitations, and provide guidelines for adoption. Each of the seven procedures finds inspiration in Merton's ethos of science as reflected in the norms of communalism, universalism, disinterestedness and organized scepticism. We believe that these ethical considerations—as well as their statistical consequences—establish common ground among data analysts, despite continuing disagreements about the foundations of statistical inference.**

Science Advances May 2021 Serra-Garcia & Gneezy

Nonreplicable publications are cited more than replicable ones

**FIGURE S1. Distribution of yearly citation counts by replication project**




Notes: This figure shows distribution of yearly citation counts of each replicated study, across the three replication projects. Panel (a) for (*7*), Panel (b) for (*6*), and Panel (c) for (*5*), including those studies in replication markets. Panel (d) includes all papers in the psychology RP (*5*), that featured a significant result in the replicated paper.

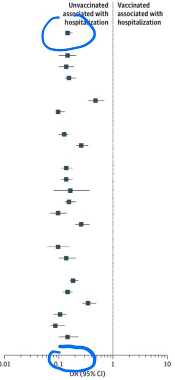
"the distribution of citation counts is highly right-skewed. We hence use Poisson regression models for the main specification in the paper."

J Am. Medical Assoc. Nov 4, 2021 Tenforde et al.

Association Between mRNA Vaccination and COVID-19 Hospitalization and Disease Severity

Figure 2. Association Between Hospitalization for COVID-19 and Prior Vaccination With a 2-Dose mRNA Vaccine

| Subgroup | Vaccinated case patients/total case patients (%) | Vaccinated control patients/total control patients (%) | Absolute difference (95% CI), % | Adjusted odds ratio (95% CI)[a] |
|---|---|---|---|---|
| Overall | 314/1983 (15.8) | 1386/2530 (54.8) | -38.9 (-41.5 to -36.4) | 0.15 (0.13 to 0.18) |
| By age group, y | | | | |
| 18-49 | 54/758 (7.1) | 216/690 (31.3) | -24.2 (-28.1 to -20.3) | 0.15 (0.10 to 0.21) |
| 50-64 | 82/656 (12.5) | 384/756 (50.8) | -38.3 (-42.7 to -33.9) | 0.14 (0.10 to 0.19) |
| ≥65 | 178/569 (31.3) | 786/1084 (72.5) | -41.2 (-45.9 to -36.6) | 0.16 (0.13 to 0.21) |
| By immunocompromising condition[b] | | | | |
| Yes (immunocompromised) | 128/319 (40.1) | 344/585 (58.8) | -18.7 (-25.4 to -12.0) | 0.49 (0.35 to 0.69) |
| No (immunocompetent) | 186/1662 (11.2) | 1039/1942 (53.5) | -42.3 (-45.0 to -39.6) | 0.10 (0.09 to 0.13) |
| By time between vaccine dose 2 and illness onset | | | | |
| 14-120 Days since vaccination | 179/1848 (9.7) | 1134/2278 (49.8) | -40.1 (-42.6 to -37.6) | 0.13 (0.10 to 0.15) |
| >120 Days since vaccination | 135/1804 (7.5) | 252/1396 (18.1) | -10.6 (-12.9 to -8.2) | 0.27 (0.21 to 0.35) |
| By month of illness onset overall and time between vaccine dose 2 and illness onset | | | | |
| March-June 2021 overall (Alpha period) | 123/1126 (10.9) | 903/1748 (51.7) | -40.7 (-43.7 to -37.8) | 0.14 (0.11 to 0.18) |
| 14-120 Days since vaccination | 115/1118 (10.3) | 849/1694 (50.1) | -39.8 (-42.8 to -36.9) | 0.14 (0.11 to 0.18) |
| >120 Days since vaccination | 8/1011 (0.8) | 54/899 (6.0) | -5.2 (-6.9 to -3.6) | 0.17 (0.08 to 0.37) |
| July-August 2021 overall (Delta period) | 191/857 (22.3) | 483/782 (61.8) | -39.5 (-43.9 to -35.1) | 0.16 (0.13 to 0.21) |
| 14-120 Days since vaccination | 64/730 (8.8) | 285/584 (48.8) | -40.0 (-44.6 to -35.5) | 0.10 (0.07 to 0.14) |
| >120 Days since vaccination | 127/793 (16.0) | 198/497 (39.8) | -23.8 (-28.8 to -18.8) | 0.27 (0.20 to 0.37) |
| By SARS-CoV-2 lineage, if sequenced[c] | | | | |
| Alpha (B.1.1.7) | 21/242 (8.7) | 903/1748 (51.7) | -43.0 (-47.2 to -38.7) | 0.10 (0.06 to 0.16) |
| Delta (B.1.617.2 or AY) | 63/288 (21.9) | 483/782 (61.8) | -39.9 (-45.8 to -34.0) | 0.14 (0.10 to 0.21) |
| By vaccine product overall and by time between vaccine dose 2 and illness onset | | | | |
| BNT162b2 overall | 226/1895 (11.9) | 810/1954 (41.5) | -29.5 (-32.2 to -26.9) | 0.19 (0.16 to 0.23) |
| 14-120 Days since vaccination | 123/1792 (6.9) | 661/1805 (36.6) | -29.8 (-32.3 to -27.2) | 0.15 (0.12 to 0.18) |
| >120 Days since vaccination | 103/1772 (5.8) | 149/1293 (11.5) | -5.7 (-7.8 to -3.7) | 0.36 (0.27 to 0.49) |
| mRNA-1273 overall | 88/1757 (5.0) | 576/1720 (33.5) | -28.5 (-30.9 to -26.0) | 0.11 (0.08 to 0.14) |
| 14-120 Days since vaccination | 56/1725 (3.2) | 473/1617 (29.3) | -26.0 (-28.4 to -23.6) | 0.09 (0.07 to 0.13) |
| >120 Days since vaccination | 32/1701 (1.9) | 103/1247 (8.3) | -6.4 (-8.0 to -4.7) | 0.15 (0.09 to 0.23) |




"We used a test-negative case-control design to assess the association between hospitalization for COVID-19 and prior vaccination with an mRNA COVID-19 vaccine. In this analysis, case patients were those hospitalized with COVID-19 and control patients were those hospitalized for other reasons."

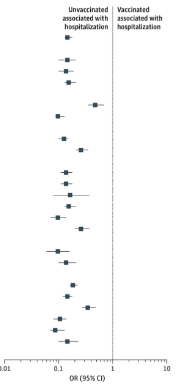
# … in the News

medRXiv preprint; Roessler et al.

Post COVID-19 in children, adolescents and adults: results of a matched cohort study

Figure 2. Association Between Hospitalization for COVID-19 and Prior Vaccination With a 2-Dose mRNA Vaccine

| Subgroup | Vaccinated case patients/total case patients (%) | Vaccinated control patients/total control patients (%) | Absolute difference (95% CI), % | Adjusted odds ratio (95% CI)[a] |
|---|---|---|---|---|
| Overall | 314/1983 (15.8) | 1386/2530 (54.8) | -38.9 (-41.5 to -36.4) | 0.15 (0.13 to 0.18) |
| By age group, y | | | | |
| 18-49 | 54/758 (7.1) | 216/690 (31.3) | -24.2 (-28.1 to -20.3) | 0.15 (0.10 to 0.21) |
| 50-64 | 82/656 (12.5) | 384/756 (50.8) | -38.3 (-42.7 to -33.9) | 0.14 (0.10 to 0.19) |
| ≥65 | 178/569 (31.3) | 786/1084 (72.5) | -41.2 (-45.9 to -36.6) | 0.16 (0.13 to 0.21) |
| By immunocompromising condition[b] | | | | |
| Yes (immunocompromised) | 128/319 (40.1) | 344/585 (58.8) | -18.7 (-25.4 to -12.0) | 0.49 (0.35 to 0.69) |
| No (immunocompetent) | 186/1662 (11.2) | 1039/1942 (53.5) | -42.3 (-45.0 to -39.6) | 0.10 (0.09 to 0.13) |
| By time between vaccine dose 2 and illness onset | | | | |
| 14-120 Days since vaccination | 179/1848 (9.7) | 1134/2278 (49.8) | -40.1 (-42.6 to -37.6) | 0.13 (0.10 to 0.15) |
| >120 Days since vaccination | 135/1804 (7.5) | 252/1396 (18.1) | -10.6 (-12.9 to -8.2) | 0.27 (0.21 to 0.35) |
| By month of illness onset overall and time between vaccine dose 2 and illness onset | | | | |
| March-June 2021 overall (Alpha period) | 123/1126 (10.9) | 903/1748 (51.7) | -40.7 (-43.7 to -37.8) | 0.14 (0.11 to 0.18) |
| 14-120 Days since vaccination | 115/1118 (10.3) | 849/1694 (50.1) | -39.8 (-42.8 to -36.9) | 0.14 (0.11 to 0.18) |
| >120 Days since vaccination | 8/1011 (0.8) | 54/899 (6.0) | -5.2 (-6.9 to -3.6) | 0.17 (0.08 to 0.37) |
| July-August 2021 overall (Delta period) | 191/857 (22.3) | 483/782 (61.8) | -39.5 (-43.9 to -35.1) | 0.16 (0.13 to 0.21) |
| 14-120 Days since vaccination | 64/730 (8.8) | 285/584 (48.8) | -40.0 (-44.6 to -35.5) | 0.10 (0.07 to 0.14) |
| >120 Days since vaccination | 127/793 (16.0) | 198/497 (39.8) | -23.8 (-28.8 to -18.8) | 0.27 (0.20 to 0.37) |
| By SARS-CoV-2 lineage, if sequenced[c] | | | | |
| Alpha (B.1.1.7) | 21/242 (8.7) | 903/1748 (51.7) | -43.0 (-47.2 to -38.7) | 0.10 (0.06 to 0.16) |
| Delta (B.1.617.2 or AY) | 63/288 (21.9) | 483/782 (61.8) | -39.9 (-45.8 to -34.0) | 0.14 (0.10 to 0.21) |
| By vaccine product overall and by time between vaccine dose 2 and illness onset | | | | |
| BNT162b2 overall | 226/1895 (11.9) | 810/1954 (41.5) | -29.5 (-32.2 to -26.9) | 0.19 (0.16 to 0.23) |
| 14-120 Days since vaccination | 123/1792 (6.9) | 661/1805 (36.6) | -29.8 (-32.3 to -27.2) | 0.15 (0.12 to 0.18) |
| >120 Days since vaccination | 103/1772 (5.8) | 149/1293 (11.5) | -5.7 (-7.8 to -3.7) | 0.36 (0.27 to 0.49) |
| mRNA-1273 overall | 88/1757 (5.0) | 576/1720 (33.5) | -28.5 (-30.9 to -26.0) | 0.11 (0.08 to 0.14) |
| 14-120 Days since vaccination | 56/1725 (3.2) | 473/1617 (29.3) | -26.0 (-28.4 to -23.6) | 0.09 (0.07 to 0.13) |
| >120 Days since vaccination | 32/1701 (1.9) | 103/1247 (8.3) | -6.4 (-8.0 to -4.7) | 0.15 (0.09 to 0.23) |




"We used a test-negative case-control design to assess the association between hospitalization for COVID-19 and prior vaccination with an mRNA COVID-19 vaccine. In this analysis, case patients were those hospitalized with COVID-19 and control patients were those hospitalized for other reasons."

used $x_i$ as offset w/ log link

$\lambda_i = \beta x_i$ ("time")

y ~ x - 1, fam = poiss (link = "id")

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

# GLM Models: Examples

- Normal: $f(y_i; \mu_i, \sigma^2) = \frac{1}{\sqrt{(2\pi)}\sigma} \exp\{-\frac{1}{2\sigma^2}(y_i - \mu_i^2)\}$ $\longleftrightarrow$ $\exp\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i)\}$

$$= \exp\left\{-\frac{1}{2\sigma^2} y_i^2 + \frac{y_i}{\sigma^2}\mu_i - \frac{\mu_i^2}{2\sigma^2} - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\ln(2\pi)\right\}$$

$\theta_i = \mu_i \qquad \phi_i = \sigma^2$

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

$\phi_i = \sigma^2 \qquad c(y_i, \phi_i) = \left(-\frac{1}{2\sigma^2} y_i^2 - \frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2\right)$

$E(y_i) = \mu_i$

$\mathrm{var}(y_i) = \sigma^2 \qquad \phi_i V(\mu_i)$

$V(\mu_i) = 1$

$\phi_i = \sigma^2$

# GLM Models: Examples

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

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

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

$= \exp\{m_i \frac{r_i}{m_i} \log\left(\frac{p_i}{1-p_i}\right) - m_i \ln(1-p_i) + \ln\binom{m_i}{r_i}\}$

$y_i = r_i/m_i$

$p_i = \mu_i$

$y_i = \frac{r_i}{m_i}$ $\quad \theta_i = \log\left(\frac{p_i}{1-p_i}\right)$ $\quad \phi_i = \frac{1}{m_i}$

$\mathrm{var}(y_i) = V(p_i) \cdot \phi_i = p_i(1-p_i)/m_i$

glm(r/m ~ x, family=bin, weights = ...)

cbind(r, m-r) ~ x, ...

# GLM Models: Examples

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

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

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

- ELM (p.115) uses $a_i(\phi)$ in place of $\phi_i$, later (p.117) $a_i(\phi)=\phi/w_i$;
SM uses $\phi_i$, later (p. 483) $\phi_i=\phi a_i$

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

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

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

- ELM (p.115) uses $a_i(\phi)$ in place of $\phi_i$, later (p.117) $a_i(\phi) = \phi/w_i$; SM uses $\phi_i$, later (p. 483) $\phi_i = \phi a_i$

$-c(y_i, \phi_i)$

$+\ln\binom{m_i}{r_i}$

$w_i = m_i$

not in

$\phi a_i \longleftrightarrow \phi/w_i$ IRWLS

# ... Examples



| Family | Canonical link | Variance function | $\phi_i$ |
|---|---|---|---|
| Normal | $\eta = \mu$ | 1 | $\sigma^2$ |
| Binomial | $\eta = \log\{\mu/(1-\mu)\}$ | $\mu(1-\mu)$ | $1/m_i$ |
| Poisson | $\eta = \log(\mu)$ | $\mu$ | 1 |
| Gamma | $\eta = 1/\mu$ | $\mu^2$ | $1/\nu$ |
| Inverse Gaussian | $\eta = 1/\mu^2$ | $\mu^3$ | $\xi$ |

# … Examples



| Family | Canonical link | Variance function | $\phi_i$ |
|---|---|---|---|
| Normal | $\eta = \mu$ | 1 | $\sigma^2$ |
| Binomial | $\eta = \log\{\mu/(1-\mu)\}$ | $\mu(1-\mu)$ | $1/m_i$ |
| Poisson | $\eta = \log(\mu)$ | $\mu$ | 1 |
| Gamma | $\eta = 1/\mu$ | $\mu^2$ | $1/\nu$ |
| Inverse Gaussian | $\eta = 1/\mu^2$ | $\mu^3$ | $\xi$ |

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

# ... Examples

| Family | Canonical link | Variance function | $\phi_i$ |
|---|---|---|---|
| Normal | $\eta = \mu$ | 1 | $\sigma^2$ |
| Binomial | $\eta = \log\{\mu/(1-\mu)\}$ | $\mu(1-\mu)$ | $1/m_i$ |
| Poisson | $\eta = \log(\mu)$ | $\mu$ | 1 |
| Gamma | $\eta = 1/\mu$ | $\mu^2$ | $1/\nu$ |
| Inverse Gaussian | $\eta = 1/\mu^2$ | $\mu^3$ | $\xi$ |

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

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

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

$\phi_i = \frac{1}{\nu}$ $\quad \frac{1}{\Gamma(\nu)}\beta^{\nu} y^{\nu-1} e^{-\beta y}$ $\quad E(y_i) = \mu_i$ $\quad var(y_i) = \frac{\mu_i^2}{\nu}$

$\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i,\phi_i)\}$ $\quad$ rate $\beta_i = \frac{\nu}{\mu_i}$

# Inference

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

# Inference

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

messy

$$\theta_i = \theta_i(\mu_i) = \theta_i\{g^{-1}(x_i^T\beta)\}$$

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} \left\{ \frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i) \right\} \qquad j = 1, \dots, p$$

$$= \sum \frac{\partial}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta_j} \{ \quad \}$$

$$= \sum_{i=1}^{n} \left\{ \frac{y_i - b'(\theta_i)}{\phi_i} \right\} \frac{\partial \theta_i}{\partial \beta_j} = \sum \frac{(y_i - \mu_i)}{\phi_i} \left( \frac{\partial \theta_i}{\partial \beta_j} \right)$$

# Inference

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

# Inference

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

$$g'(b'(\theta_i))\, b''(\theta_i) \frac{\partial \theta_i}{\partial \beta_j} = \frac{\partial}{\partial \beta_j}(x_i^T \beta) = x_{ij}$$

$$\frac{\partial \theta_i}{\partial \beta_j} = \frac{x_{ij}}{g'(\mu_i)\, \phi_i\, V(\mu_i)}$$

# Inference

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

# Inference

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

# Inference

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

when $\phi_i = a_i\phi$

ELM has $\phi_i = \phi/w_i$

# Inference

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

when $\phi_i = a_i\phi$

ELM has $\phi_i = \phi/w_i$

# Inference

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

when $\phi_i = a_i\phi$

ELM has $\phi_i = \phi/w_i$

- matrix notation:

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

# ... Inference

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

when $\phi_i = a_i\phi$

ELM has $\phi_i = \phi/w_i$

- matrix notation:

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

# Scale parameter $\phi_i$

- in most cases, either $\phi_i$ 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 | $\phi_i$ |
|---|---|---|---|
| Normal | $\eta = \mu$ | 1 | $\sigma^2$ |
| Binomial | $\eta = \log\{\mu/(1-\mu)\}$ | $\mu(1-\mu)$ | $1/m_i$ |
| Poisson | $\eta = \log(\mu)$ | $\mu$ | 1 |
| Gamma | $\eta = 1/\mu$ | $\mu^2$ | $1/\nu$ |
| Inverse Gaussian | $\eta = 1/\mu^2$ | $\mu^3$ | $\xi$ |

# Scale parameter $\phi_i$

- in most cases, either $\phi_i$ 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 | $\phi_i$ |
|---|---|---|---|
| Normal | $\eta = \mu$ | 1 | $\sigma^2$ |
| Binomial | $\eta = \log\{\mu/(1-\mu)\}$ | $\mu(1-\mu)$ | $1/m_i$ |
| Poisson | $\eta = \log(\mu)$ | $\mu$ | 1 |
| Gamma | $\eta = 1/\mu$ | $\mu^2$ | $1/\nu$ |
| Inverse Gaussian | $\eta = 1/\mu^2$ | $\mu^3$ | $\xi$ |

- $\dfrac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \dfrac{y_i - \mu_i}{\phi_i g'(\mu_i) V(\mu_i)} x_{ij} = \sum \dfrac{y_i - \mu_i}{a_i \phi g'(\mu_i) V(\mu_i)} x_{ij}$ when $\phi_i = a_i \phi$
- if $\theta_i = g(\mu_i)$ canonical link, then $g'(\mu_i) = 1/V(\mu_i)$, and

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

$\sum \frac{y_i - \mu_i}{a_i} x_{ij} = 0$

$= 0$

$a_i$ known

Poiss.

Bin.

ntbc

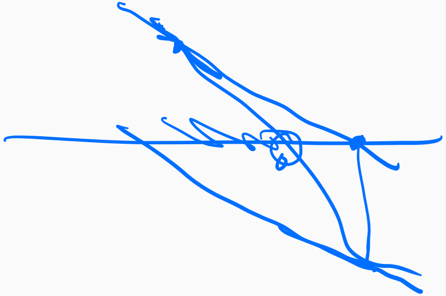
# Solving maximum likelihood equation

$X^T u(\hat\beta) = 0$

- 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

$+ \{-\ell''(\hat{\beta}^{(t)})\}^{-1}$ obs'd information

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

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

many books use $I(\beta)$

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

$$\hat{\beta} - \beta = \frac{+\ell'(\beta)}{-\ell''(\beta)}$$

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

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

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

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

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

$$E\left(-\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k}\right) \quad \sum_{i=1}^{n} \frac{x_{ij} x_{ik}}{\phi a_i \{g'(\mu_i)\}^2 V(\mu_i)}$$

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

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

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

$$\hat\beta = \beta + \dots = (X^T W_\beta X)^{-1} X^T W_\beta z_\beta$$

$$X^T W_\beta X$$

$$W_\beta = \mathrm{diag}\left(\frac{1}{g'(\mu_i)^2 V(\mu_i)}\right)$$

# … maximum likelihood equation

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

$X^T u(\hat\beta) = 0$

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

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

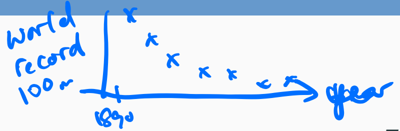
$X\beta + W^{-1}u(\beta)$ derived response

- does not involve $\phi_i$ — iteratively re-weighted least squares $W, z$ both depend on $\beta$
- derived response $z = X\beta + W^{-1}u$ — linearized version of $y$

$\hat\beta - (X^T\hat W X)^{-1} X^T \hat W \hat z = 0$ at conv.

# Summary

Model:




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

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:

$$\mathrm{Var}(\hat\beta) = (X^TWX)^{-1}$$

$W$ is diagonal

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

# Summary 2

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

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

$$W_{ii} =$$

$$u_i =$$

$\phi_i = a_i \phi$ $a_i$ known

$\hat{\beta}$ free of $\phi$ →

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

Warning: in ELM $W$ is defined slightly differently (no $\phi$), so he has $\text{Var}(\hat{\beta}) = (X^T W X)^{-1}\hat{\phi}$

# Summary 2

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

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

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

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

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

Warning: in ELM $W$ is defined slightly differently (no $\phi$), so he has Var$(\hat\beta) = (X^TWX)^{-1}\hat\phi$
Further, the $w_i$ on p.117 is not the same as the $w_i$ on p. 118; SM uses $a_i$ instead which would have been better for ELM

# Analysis of data using GLMs: overview

- choose a model, often based on type of response    or on mean/variance relationship

# 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

# 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`

exception

binary data that is separable

perfect prediction is possible

$$\frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}$$

$\beta \to \infty$

$p_i \to 1$

$\beta \to -\infty$

$p_i \to 0$

$\hat{p}_i = 1, 0$

# 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

as in lin reg.

# 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 $\phi$ based on Pearson's Chi-square

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

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

gamma
$\phi = \frac{1}{\nu}$
Inv G
$\phi = \lambda$

not m.l.e.
$s^2 = \frac{RSS}{n-p}$

# 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 $\phi$ based on Pearson's Chi-square

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

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

- analysis of deviance: see p. 121 (near bottom) — likelihood ratio tests

# 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 $\phi$ based on Pearson's Chi-square

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

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

- analysis of deviance: see p. 121 (near bottom) — likelihood ratio tests
- diagnostics: same as for `lm` — ELM p.124; SM p.477

# 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 $\phi$ based on Pearson's Chi-square

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

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

- 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

# 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 $\phi$ based on Pearson's Chi-square

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

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

- 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`

$y = X\beta + \varepsilon \quad \text{var}(\varepsilon) \sim \sigma^2 W \leftarrow$ known

# The last slide about GLM theory

- special to `glm`

$$W^{1/2}y = W^{1/2}X\beta + W^{1/2}\varepsilon \quad \text{WLS}$$

$$y_1 = X_1\beta + \varepsilon_1 \quad \text{OLS}$$

# The last slide about GLM theory

- special to `glm`
- two models, Poisson and Binomial, have no $\phi$ parameter

# The last slide about GLM theory

- special to `glm`
- two models, Poisson and Binomial, have no $\phi$ parameter
- this has two consequences

# The last slide about GLM theory

- special to `glm`
- two models, Poisson and Binomial, have no $\phi$ parameter
- this has two consequences
- the residual deviance can be used as a test of fit of the model

# The last slide about GLM theory

- special to `glm`
- two models, Poisson and Binomial, have no $\phi$ parameter
- this has two consequences
- the residual deviance can be used as a test of fit of the model
- two pseudo-models are available called `quasibinomial`, `quasipoisson`

# The last slide about GLM theory

- special to `glm`
- two models, Poisson and Binomial, have no $\phi$ parameter
- this has two consequences
- the residual deviance can be used as a test of fit of the model
- two pseudo-models are available called `quasibinomial, quasipoisson`
- quasi-binomial: $\text{var}(y_i) = \phi p_i(1 - p_i)$

# The last slide about GLM theory

- special to `glm`
- two models, Poisson and Binomial, have no $\phi$ parameter
- this has two consequences
- the residual deviance can be used as a test of fit of the model
- two pseudo-models are available called `quasibinomial`, `quasipoisson`
- quasi-binomial: $\text{var}(y_i) = \phi p_i(1 - p_i)$
- quasi-Poisson: $\text{var}(y_i) = \phi \mu_i$

# The last slide about GLM theory

- special to `glm`
- two models, Poisson and Binomial, have no $\phi$ parameter
- this has two consequences
- the residual deviance can be used as a test of fit of the model
- two pseudo-models are available called `quasibinomial`, `quasipoisson`
- quasi-binomial: $\text{var}(y_i) = \phi p_i(1 - p_i)$
- quasi-Poisson: $\text{var}(y_i) = \phi\mu_i$
- quasi- is a quick way to fit proportion or count responses, but allow the variance to be bigger (or rarely, smaller) than it would be under the binomial or Poisson model

# The last slide about GLM theory

- special to `glm`
- two models, Poisson and Binomial, have no $\phi$ parameter
- this has two consequences
- the residual deviance can be used as a test of fit of the model
- two pseudo-models are available called `quasibinomial, quasipoisson`
- quasi-binomial: $\text{var}(y_i) = \phi p_i(1 - p_i)$
- quasi-Poisson: $\text{var}(y_i) = \phi\mu_i$
- quasi- is a quick way to fit proportion or count responses, but allow the variance to be bigger (or rarely, smaller) than it would be under the binomial or Poisson model
- caveat – none of this works for binary data, only binomial

$n_i \geq 5$, approx

Link

Save

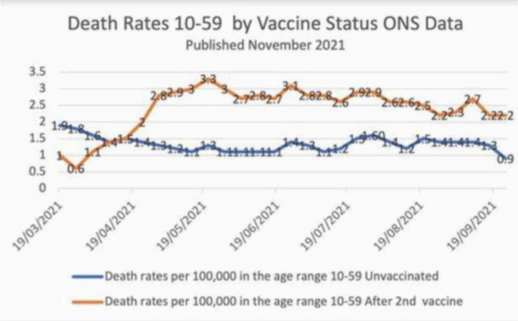
## Vaccinated English adults under 60 are dying at twice the rate of unvaccinated people the same age

And have been for six months. This chart may seem unbelievable or impossible, but it's correct, based on weekly data from the British government.

Alex Berenson
Nov 20 ♡1,137 ◯627

The brown line represents weekly deaths from all causes of vaccinated people aged 10-59, per 100,000 people.

The blue line represents weekly deaths from all causes of unvaccinated people per 100,000 in the same age range.




I have checked the underlying dataset myself and this graph is correct. Vaccinated people under 60 are twice as likely to die as unvaccinated people. And overall deaths in Britain are running well above normal.

I don't know how to explain this other than vaccine-caused mortality.

# Blog post Jeffrey Morris

Link

Is watching the 1984 Ghostbusters movie killing people? A Statistician's Perspective

Is watching the 1984 Ghostbusters movie killing people?

English adults under 60 who have watched the 1984 Ghostbusters movie are dying at twice the rate of people who have watched the 2021 Ghostbusters move the same age.

And have been for six months. The chart may seem unbelievable or impossible but it is correct, based on weekly data from the British government.

The brown line represents weekly deaths from call causes of people aged 10-59 who have watched the 1984 Ghostbusters movie but not the 2021 Ghostbusters movie, per 100,000 people.

The blue line represents weekly deaths from all causes of unvaccinated people per 100,000 in the same age range.

 Link

## Is watching the 1984 Ghostbusters movie killing people?


I have checked the underlying dataset myself and the graph plotted above is correct. People under 60 who watched the 1984 Ghostbusters movie are twice as likely to die as people who watched the 2021 Ghostbusters movie. The overall deaths in Britain are running well above normal.

I don't know how to explain this other than movie-caused mortality.

# Still time?

$\longrightarrow$ casestudies.pdf