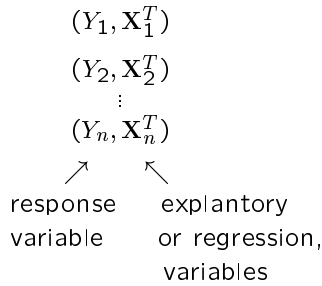


Generalized Linear Models:

A subclass of nonlinear models with a "linear" component

Data:



where

$$\mathbf{X}_j^T = (X_{1j}, X_{2j}, \dots, X_{pj})$$

describe the conditions under which the response Y_j was obtained.

3. (Systematic part of the model)

There is a link function $h()$ that links the conditional mean of Y_j given $\mathbf{X}_j = (X_{1j}, \dots, X_{pj})^T$ to a linear function of unknown parameters, e.g.

$$h(E(Y_j|\mathbf{X}_j)) = \mathbf{X}_j^T \boldsymbol{\beta} = \beta_1 X_{1j} + \dots + \beta_p X_{pj}$$

Members of the exponential family of distributions available in

S-PLUS: glm() function

SAS: PROC GENMOD

Basic Features:

1. Y_1, Y_2, \dots, Y_n are independent
2. Y_j has a distribution in the exponential family of distributions, $j = 1, 2, \dots, n$.

The joint likelihood has the form (canonical form):

$$L(\boldsymbol{\theta}, \phi; \mathbf{Y}) = \pi_{j=1}^n f(Y_j; \theta_j, \phi)$$

where

$$f(Y_j; \theta_j, \phi) = \exp \left\{ \frac{Y_j \theta_j - b(\theta_j)}{a(\phi)} \right\}$$

for some functions

$$a(), b(), \text{ and } c().$$

Here, θ_j is called a canonical parameter.

Normal distribution: $Y \sim N(\mu, \sigma^2)$

$$f(Y; \theta, \phi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-(Y - \mu)^2}{2\sigma^2} \right\}$$

$$= \exp \left\{ \frac{Y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{1}{2} \left[\frac{Y^2}{\sigma^2} + \log(2\pi\sigma^2) \right] \right\}$$

$\uparrow C(Y, \phi)$

Here

$$E(Y) = \mu \equiv \theta \quad \text{and} \quad b(\theta) = \frac{\theta^2}{2}$$

$$Var(Y) = \sigma^2 = \phi \quad \text{and} \quad a(\phi) = \phi$$

Binomial Distribution: $Y \sim Bin(n, \pi)$

$$\begin{aligned} f(Y; \theta, \phi) &= \binom{n}{Y} \pi^Y (1 - \pi)^{n-Y} \\ &= \binom{n}{Y} \left(\frac{\pi}{1 - \pi} \right)^Y (1 - \pi)^n \\ &= \exp \left\{ Y \log \left(\frac{\pi}{1 - \pi} \right) + n \log(1 - \pi) + \log \binom{n}{Y} \right\} \\ &\quad \uparrow C(Y, \phi). \end{aligned}$$

Here

$$\theta = \log \left(\frac{\pi}{1 - \pi} \right) \Rightarrow \pi = \frac{e^\theta}{1 + e^\theta}$$

and

$$b(\theta) = n \log(1 - \pi) = n \log(1 + e^\theta).$$

There is no dispersion parameter, so we will take $\phi \equiv 1$ and $a(\phi) = \phi \equiv 1$.

5

Two parameter gamma distribution: $Y \sim \text{gamma}(\alpha, \beta)$

$$\begin{aligned} f(Y; \theta, \phi) &= \frac{Y^{\alpha-1} \exp(-Y/\beta)}{\beta^\alpha \Gamma(\alpha)} \\ &= \exp \left\{ \frac{-Y}{\beta} + (\alpha - 1) \log(Y) - \alpha \log(\beta) - \log(\Gamma(\alpha)) \right\} \\ &= \exp \left\{ \frac{Y \left(\frac{-1}{\beta} \right) + \log \left(\frac{1}{\beta} \right)}{\alpha^{-1}} \right. \\ &\quad \left. + [(\alpha - 1) \log(Y) + \alpha^{-1} \log(\alpha^{-1}) - \log(\Gamma(\alpha))] \right\} \\ &\quad \uparrow C(Y, \phi) \end{aligned}$$

Here, $\theta = \frac{-1}{\beta}$

$$\phi = \alpha^{-1}, \quad b(\theta) = -\log(-\theta), \quad a(\phi) = \phi$$

7

Poisson distribution: $Y \sim \text{Poisson}(\mu)$

$$\begin{aligned} f(Y; \theta, \phi) &= \frac{\mu^Y}{Y!} e^{-\mu} \\ &= \exp \left\{ Y \log(\mu) - \mu - \log(Y!) \right\} \\ &\quad \uparrow C(Y, \phi). \end{aligned}$$

Here

$$\theta = \log(\mu) \Rightarrow \mu = \exp(\theta)$$

and

$$b(\theta) = \mu = \exp(\theta)$$

and there is no additional dispersion parameter. We will take $\phi \equiv 1$ and $a(\phi) = \phi \equiv 1$.

Then,

$$f(Y; \theta, \phi) = \exp\{Y_\theta - \exp(\theta) - \log(\Gamma(Y + 1))\}$$

6

Inverse Gaussian Distribution: $Y \sim IG(\mu, \sigma^2)$

$$\begin{aligned} f(Y; \theta, \phi) &= \frac{1}{\sqrt{2\pi\sigma^2 Y^3}} \exp \left\{ \frac{-(Y - \mu)^2}{2\sigma^2 \mu^2 Y} \right\} \\ &\quad \text{for } Y > 0, \mu > 0, \sigma^2 > 0 \\ &\quad \text{also } E(Y) = \mu, \text{Var}(Y) = \sigma^2 \mu^3 \\ &= \exp \left\{ \frac{Y \left(\frac{-1}{2\mu^2} \right) - \frac{1}{\mu}}{\sigma^2} + \left[\frac{1}{2\sigma^2 Y} - \frac{1}{2} \log(2\pi\sigma^2 Y^3) \right] \right\} \\ &\quad \uparrow C(Y, \phi) \end{aligned}$$

Here,

$$\begin{aligned} \theta &= \frac{-1}{2\mu^2} \\ b(\theta) &= \frac{1}{\mu} = \sqrt{-2\theta} \\ \phi &= \sigma^2 \quad \text{"dispersion parameter"} \\ a(\phi) &= \phi = \sigma^2 \end{aligned}$$

8

Canonical form of the log-likelihood:

$$\begin{aligned}\ell(\theta, \phi; Y) &= \log(f(Y; \theta, \phi)) \\ &= \frac{Y\theta - b(\theta)}{a(\phi)} + C(Y, \phi)\end{aligned}$$

partial derivatives:

$$(i) \frac{\partial \ell(\theta, \phi; Y)}{\partial \theta} = \frac{Y - b'(\theta)}{a(\phi)}$$

$$(ii) \frac{\partial^2 \ell(\theta, \phi; Y)}{\partial \theta^2} = \frac{-b''(\theta)}{a(\phi)}$$

Apply the following results

$$E\left(\frac{\partial \ell(\theta, \phi; Y)}{\partial \theta}\right) = 0$$

$$E\left(\frac{\partial^2 \ell(\theta, \phi; Y)}{\partial \theta^2}\right) + E\left[\frac{\partial \ell(\theta, \phi; Y)}{\partial \theta}\right]^2 = 0$$

to obtain

$$0 = E\left(\frac{\partial \ell(\theta, \phi; Y)}{\partial \theta}\right) = \frac{E(Y) - b'(\theta)}{a(\phi)}$$

9

Normal distribution: $Y \sim N(\mu, \sigma^2)$

$$\begin{aligned}b(\theta) &= \frac{\theta^2}{2} \\ b'(\theta) &= \theta \Rightarrow \mu \equiv E(Y) = \theta\end{aligned}$$

$$\left. \begin{aligned}b''(\theta) &= 1 \\ a(\phi) &= \phi = \sigma^2\end{aligned} \right\} \Rightarrow \text{Var}(Y) = b''(\theta)a(\phi) \\ &= (1)\phi \\ &= \phi = \sigma^2$$

↑ the variance does not depend on the parameter that define the mean

11

which implies

$$E(Y) = b'(\theta)$$

Furthermore, from (i), (ii) and (iv) we have

$$\begin{aligned}0 &= E\left(\frac{\partial^2 \ell(\theta, \phi; Y)}{\partial \theta^2}\right) + E\left(\frac{\partial \ell(\theta, \phi; Y)}{\partial \theta}\right)^2 \\ &= \frac{-b''(\theta)}{a(\phi)} + \text{Var}\left(\frac{Y - b'(\theta)}{a(\theta)}\right) \\ &= \frac{-b''(\theta)}{a(\theta)} + \frac{\text{Var}(Y)}{[a(\phi)]^2}\end{aligned}$$

which implies

$$V(Y) = \frac{a(\phi)}{\uparrow} \frac{b''(\theta)}{\uparrow}$$

function of the dispersion parameter called the variance function

10

Binomial distribution: $Y \sim \text{Bin}(n, \pi)$

$$\begin{aligned}\theta &= \log\left(\frac{\pi}{1-\pi}\right) \\ b(\theta) &= (n)\log(1 + e^\theta) \\ b'(\theta) &= \frac{ne^\theta}{1+e^\theta} \Rightarrow E(Y) = \frac{ne^\theta}{1+e^\theta} = n\pi\end{aligned}$$

$$\left. \begin{aligned}b''(\theta) &= \frac{ne^\theta}{[1+e^\theta]^2} \\ a(\phi) &\approx 1\end{aligned} \right\} \text{Var}(Y) = \frac{ne^\theta}{[1+e^\theta]^2}(1) \\ &= n\left(\frac{e^\theta}{1+e^\theta}\right)\left(\frac{1}{1+e^\theta}\right) \\ &= n\pi(1 - \pi)$$

↑ $\text{Var}(Y)$

is completely determined by the parameter that defines the mean.

12

Poisson Distribution: $Y \sim \text{Poisson}(\mu)$

$$b(\theta) = \exp(\theta)$$

$$b'(\theta) = \exp(\theta) \Rightarrow \mu = E(Y) = \exp(\theta)$$

$$\left. \begin{array}{l} b''(\theta) = \exp(\theta) \\ a(\phi) = \theta \equiv 1 \end{array} \right\} \text{Var}(Y) = \begin{array}{l} \exp(\theta) \\ = E(Y) \\ = \mu \end{array}$$

Gamma Distribution:

$$b(\theta) = -\log(-\theta)$$

$$b'(\theta) = -\frac{1}{\theta} \Rightarrow E(Y) = -\frac{1}{\theta}$$

$$\left. \begin{array}{l} b''(\theta) = \frac{1}{\theta^2} \\ a(\phi) = \phi \end{array} \right\} \text{Var}(Y) = \begin{array}{l} a(\phi)b''(\theta) \\ = \frac{\phi}{\theta^2} \\ = \phi[E(Y)]^2 \end{array}$$

13

Binomial distribution:

$$\theta = \mathbf{X}^T \boldsymbol{\beta}$$

$$n\pi = E(Y) = b'(\theta) = \frac{ne^\theta}{1+e^\theta}$$

$$\Rightarrow \theta = \log\left(\frac{\pi}{1-\pi}\right) = \mathbf{X}^T \boldsymbol{\beta}$$

Hence, the canonical (default) link is the logit link function

$$h(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \mathbf{X}^T \boldsymbol{\beta}$$

Other available link functions:

Probit: $h(\pi) = \Phi^{-1}(\pi) = \mathbf{X}^T \boldsymbol{\beta}$

where

$$\pi = \Phi(\mathbf{X}^T \boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{X}^T \boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

15

Canonical link function:

is the link function $h(\mu) = \mathbf{X}^T \boldsymbol{\beta}$ that corresponds to

$$\theta = \mathbf{X}^T \boldsymbol{\beta}$$

$$\mu = E(Y) = \frac{\partial b(\theta)}{\partial \theta} = b'(\theta)$$

Normal (Gaussian) distribution:

$$\theta = \mathbf{X}^T \boldsymbol{\beta}$$

$$\mu = b'(\theta) = \theta$$

then $E(Y) = \mu = \theta = \mathbf{X}^T \boldsymbol{\beta}$

and the canonical link function is the identity function

$$h(\mu) = \mu = \mathbf{X}^T \boldsymbol{\beta}$$

14

Complimentary log-log link (clog log):

$$h(\pi) = \log[-\log(1-\pi)] = \mathbf{X}^T \boldsymbol{\beta}$$

The inverse of this link function is

$$\pi = 1 - \exp[-\exp(\mathbf{X}^T \boldsymbol{\beta})]$$

The glm() function

glm(model, family, data, weights, control)

family = binomial

family = binomial(link=logit)

family = binomial(link=probit)

family = binomial(link=cloglog)

16

Poisson regression (log-linear models)

$$Y \sim \text{Poisson}(\mu)$$

$$\theta = \mathbf{X}^T \boldsymbol{\beta}$$

$$\mu = E(Y) = b'(\theta) = \exp(\theta)$$

and the canonical link function is

$$\theta = \log(\mu) = \mathbf{X}^T \boldsymbol{\beta}$$

This is called the log link function. Other available link functions:

identity link: $h(\mu) = \mu = \mathbf{X}^T \boldsymbol{\beta}$

square root link: $h(\mu) = \sqrt{\mu} = \mathbf{X}^T \boldsymbol{\beta}$

`glm(model, family, data, weights, control)`

`family = poisson`

`family = poisson(link=log)`

`family = poisson(link=identity)`

`family = poisson(link=sqrt)`

17

Inverse Gaussian Distribution

$$\theta = \mathbf{X}^T \boldsymbol{\beta}$$

$$b(\theta) = -\sqrt{-2\theta}$$

$$b'(\theta) = \frac{1}{\sqrt{-2\theta}} = E(Y) = \mu$$

The canonical link is

$$\theta = h(\mu) = \frac{-1}{2\mu^2} = \mathbf{X}^T \boldsymbol{\beta}$$

This is the only built-in link function for the inverse gaussian distribution.

`glm(model, family, data, weights, controls)`

`family = inverse.gaussian`

`family = inverse.gaussian(link=1/\mu^2)`

19

Gamma distribution:

$$\theta = \mathbf{X}^T \boldsymbol{\beta}$$

$$E(Y) = \mu = \frac{-1}{\theta}$$

the canonical link function is

$$h(\mu) = \frac{-1}{\mu} = \mathbf{X}^T \boldsymbol{\beta}$$

This is called the "inverse" link.

`glm(model, family, data, weights, control)`

`family = gamma`

`family = gamma(link=inverse)`

`family = gamma(link=identity)`

`family = gamma(link=log)`

18

Normal-theory Gauss-Markov model:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

or Y_1, Y_2, \dots, Y_n are independent with $\mathbf{Y}_j \sim N(\mathbf{X}_j^T \boldsymbol{\beta}, \sigma^2)$. Then

$$f(\mathbf{Y}_j; \boldsymbol{\beta}_j, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-(Y_j - \mathbf{X}_j^T \boldsymbol{\beta})^2}{2\sigma^2} \right\}$$

$$= \exp \left\{ \frac{(Y_j)(\mathbf{X}_j^T \boldsymbol{\beta}) - \frac{(\mathbf{X}_j^T \boldsymbol{\beta})^2}{2}}{\sigma^2} - \frac{1}{2} \left[\frac{Y_j^2}{\sigma^2} + \log(2\pi\sigma^2) \right] \right\}$$

↑ $C(Y_j, \sigma^2)$

Here,

$$b(\boldsymbol{\beta}) = \frac{(\mathbf{X}_j^T \boldsymbol{\beta})^2}{2}$$

$$\phi = \sigma^2 \text{ and } a(\phi) = \phi$$

$$E(\mathbf{Y}_j) = \boldsymbol{\mu}_j = \mathbf{X}_j^T \boldsymbol{\beta}$$

20

The link function is the identity function

$$\begin{aligned} h(\mu_j) &= \mu_j \\ &= \mathbf{X}_j^T \boldsymbol{\beta} \end{aligned}$$

The class of generalized linear models includes the class of linear models:

(i) Systematic part of the model (identity link function)

$$\begin{aligned} \mu_j = E(Y_j) &= \mathbf{X}_j^T \boldsymbol{\beta} \\ &= \beta_1 X_{1j} + \dots + \beta_p X_{pj} \end{aligned}$$

(ii) Stochastic part of the model

$$Y_j \sim N(\mu_j, \sigma^2)$$

and $Y_1 Y_2 \dots Y_n$ are independent.

21

Logistic Regression:

- Y_1, Y_2, \dots, Y_8 are independent random counts with

$$Y_j \sim \text{Bin}(n_j, \pi_j)$$

- $\log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_1 X_j$

Comments:

- * Here the logit link function is the canonical link function

- * Use of the binomial distribution follows from

(i) Each beetle exposed to log-dose X_j responds independently

(ii) Each beetle exposed to log-dose X_j has probability π_j of dying

23

Example 13.1 Mortality of a certain species of beetle after 5 hours exposure to gaseous carbon disulfide (Bliss, 1935).

Dose (mg/l)	Log(Dose) (X_j)	Number killed (Y_j)	Number of survivors ($n_j - Y_j$)	Number exposed (n_j)
49.057	3.893	6	53	59
52.991	3.970	13	47	60
56.911	4.041	18	44	62
60.842	4.108	28	28	56
64.759	4.171	52	11	63
68.691	4.230	53	6	59
72.611	4.258	61	1	62
76.542	4.338	60	0	60

22

Maximum Likelihood Estimation: Joint likelihood for Y_1, Y_2, \dots, Y_k

$$\begin{aligned} L(\boldsymbol{\beta}; \mathbf{Y}, X) &= \prod_{j=1}^k f(Y_j; \theta_j, \phi) \\ &= \prod_{j=1}^k \exp\left\{ \frac{Y_j \theta_j - b(\theta_j)}{a(\phi)} + C(Y_j, \phi) \right\} \\ &= \prod_{j=1}^k \exp\left\{ Y_j (\beta_0 + \beta_1 X_j) - n_j \log(1 + e^{\beta_0 + \beta_1 X_j}) + \log\binom{n_j}{Y_j} \right\} \end{aligned}$$

Since $\phi = 1$

$$a(\phi) = \phi = 1$$

$$\theta_j = \log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_1 X_j$$

$$b(\theta_j) = n_j \log(1 + e^{\beta_0 + \beta_1 X_j})$$

$$C(Y_j, \phi) = \log\binom{n_j}{Y_j}$$

24

log-likelihood function:

$$\begin{aligned} \ell(\beta) &= \sum_{j=1}^k \left[\frac{Y_j \theta_j - b(\theta_j)}{a(\phi)} + C(Y_j, \phi) \right] \\ &= \sum_{j=1}^k \left[Y_j (\beta_0 + \beta_1 X_j) - n_j \log(1 + e^{\beta_0 + \beta_1 X_j}) \right. \\ &\quad \left. + \log \binom{n_j}{Y_j} \right] \\ &= \beta_0 \sum_{j=1}^k Y_j + \beta_1 \sum_{j=1}^k Y_j X_j \\ &\quad - \sum_{j=1}^k n_j \log(1 + e^{\beta_0 + \beta_1 X_j}) + \sum_{j=1}^k \log \binom{n_j}{Y_j} \end{aligned}$$

To maximize the log-likelihood, solve the system of equations obtained by setting first partial derivatives equal to zero.

25

The likelihood equations (or score function) can be expressed as

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = X^T(\mathbf{Y} - \mathbf{m}) = Q$$

↑ score function where

$$X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_k \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{bmatrix} \quad \mathbf{m} = \begin{bmatrix} n_1 \pi_1 \\ n_2 \pi_2 \\ \vdots \\ n_8 \pi_8 \end{bmatrix}$$

and from the inverse link function

$$\pi_j = \frac{\exp(\beta_0 + \beta_1 X_j)}{1 + \exp(\beta_0 + \beta_1 X_j)}$$

We will use \hat{Q} , $\hat{\mathbf{m}}$, and $\hat{\pi}_j$ to indicate when these quantities are evaluated at $\beta = \hat{\beta}$.

27

Estimating Equations (likelihood equations):

$$\begin{aligned} 0 &= \frac{\partial \ell(\hat{\beta})}{\partial \hat{\beta}_0} = \sum_{j=1}^8 Y_j - \sum_{j=1}^8 n_j \frac{\exp(\hat{\beta} + \hat{\beta}_1 X_1)}{1 + \exp(\hat{\beta} + \hat{\beta}_1 X_1)} \\ &= \sum_{j=1}^8 (Y_j - n_j \hat{\pi}_j) \\ 0 &= \frac{\partial \ell(\hat{\beta})}{\partial \hat{\beta}_1} = \sum_{j=1}^8 X_j Y_j - \sum_{j=1}^8 n_j X_j \frac{\exp(\hat{\beta} + \hat{\beta}_1 X_1)}{1 + \exp(\hat{\beta} + \hat{\beta}_1 X_1)} \\ &\quad - \sum_{j=1}^8 X_j (Y_j - n_j \hat{\pi}_j) \end{aligned}$$

- Generally no closed form expression for the solution
- If a solution exists in the interior of the parameter space ($0 < \pi_j < 1$ for all $j = 1, \dots, k$) then it is unique.

26

Second partial derivatives of the log-likelihood function:

$$\begin{aligned} H &= \begin{bmatrix} \frac{-\partial^2 \ell(\beta)}{\partial \beta_0^2} & \frac{-\partial^2 \ell(\beta)}{\partial \beta_0 \partial \beta_1} \\ \frac{-\partial^2 \ell(\beta)}{\partial \beta_0 \partial \beta_1} & \frac{-\partial^2 \ell(\beta)}{\partial \beta_1^2} \end{bmatrix} \\ &= X^T V X \end{aligned}$$

↙ called the Hessian matrix

where

$$\begin{aligned} V &= \begin{bmatrix} n_1 \pi_1 (1 - \pi_1) & & \\ & \dots & \\ & & n_k \pi_k (1 - \pi_k) \end{bmatrix} \\ &= \text{Var}(\mathbf{Y}) \end{aligned}$$

We will use $\hat{H} = X^T \hat{V} X$ and \hat{V} to denote H and V evaluated at $\beta = \hat{\beta}$.

28

Fisher Information matrix:

$$H = \begin{bmatrix} -E\left(\frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0^2}\right) & -E\left(\frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1}\right) \\ -E\left(\frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1}\right) & -E\left(\frac{-\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_1^2}\right) \end{bmatrix}$$

$$= E(X^T V X)$$

$$= X^T V X = H$$

Since the second partial derivatives do not depend on $\mathbf{Y} = (Y_1, \dots, Y_k)^T$.

29

- Fisher scoring algorithm

$$\hat{\boldsymbol{\beta}}^{(S+1)} = \hat{\boldsymbol{\beta}}^{(S)} + \hat{I}^{-1} \hat{\mathbf{Q}}$$

↑ for the logistic regression model with binomial counts
 $\hat{I} = \hat{H}$

Modification:

- Check if the log-likelihood is larger at $\hat{\boldsymbol{\beta}}^{(S+1)}$ than at $\hat{\boldsymbol{\beta}}^{(S)}$.

- If it is, go to the next iteration

- If not, invoke a halving step

$$\hat{\boldsymbol{\beta}}_{new}^{(S+1)} = \frac{1}{2}(\hat{\boldsymbol{\beta}}^{(S)} + \hat{\boldsymbol{\beta}}_{old}^{(S+1)})$$

31

Iterative procedures for solving the likelihood equations:

- Start with some initial values

$$\hat{\boldsymbol{\beta}}^{(0)} = \begin{bmatrix} \hat{\beta}_0^{(0)} \\ \hat{\beta}_1^{(0)} \end{bmatrix} = \begin{bmatrix} \log\left(\frac{P}{1-P}\right) \\ 0 \end{bmatrix}$$

where

$$P = \frac{\sum_{j=1}^k Y_j}{\sum_{j=1}^k n_j}$$

- Newton-Raphson algorithm

$$\hat{\boldsymbol{\beta}}^{(S+1)} = \hat{\boldsymbol{\beta}}^{(S)} + \hat{H}^{-1} \hat{\mathbf{Q}}$$

↑ these are evaluated at $\hat{\boldsymbol{\beta}}^{(S)}$

30

Large sample inference: (see results 9.1-9.3)

From result 9.2,

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (X^T V^{-1} X)^{-1})$$

when n_1, n_2, \dots, n_k are all large

$$\left[\begin{array}{l} \text{"large" is } n_j \pi_j > 5 \\ \phantom{\text{"large" is }} n_j (1 - \pi_j) > 5 \end{array} \right]$$

Estimate the covariance matrix as

$$\widehat{Var}(\hat{\boldsymbol{\beta}}) = (X^T \hat{V}^{-1} X)^{-1}$$

where \hat{V} is V evaluated at $\hat{\boldsymbol{\beta}}$

32

For the Bliss beetle data:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} -60.717 \\ 14.883 \end{bmatrix}$$

with

$$\widehat{Var}(\hat{\beta}) = \begin{bmatrix} S_{\hat{\beta}_0}^2 & S_{\hat{\beta}_0, \hat{\beta}_1} \\ S_{\hat{\beta}_0, \hat{\beta}_1} & S_{\hat{\beta}_1}^2 \end{bmatrix} = \begin{bmatrix} 26.84 & -6.55 \\ -6.55 & 1.60 \end{bmatrix}$$

33

Approximate $(1 - \alpha) \times 100\%$ confidence intervals

$$\hat{\beta}_0 \pm Z_{\alpha/2} S_{\hat{\beta}_0}$$

$$\hat{\beta}_1 \pm Z_{\alpha/2} S_{\hat{\beta}_1}$$

↑ from the standard normal distribution

$$Z_{.025} = 1.96$$

35

Wald tests:

Test $H_0 : \beta_0 = 0$ vs. $H_A : \beta_0 \neq 0$

$$X^2 = \left[\frac{\hat{\beta}_0 - 0}{S_{\hat{\beta}_0}} \right]^2 = \left[\frac{-60.717}{5.18} \right]^2 = 137.36$$

$$\text{p-value} = Pr\{\chi_{(1)}^2 > 197.36\} < .0001$$

Test $H_0 : \beta_1 = 0$ vs. $H_A : \beta_1 \neq 0$

$$X^2 = \left[\frac{\hat{\beta}_1 - 0}{S_{\hat{\beta}_1}} \right]^2 = \left[\frac{14.883}{1.265} \right]^2 = 138.49$$

$$\text{p-value} = Pr\{\chi_{(1)}^2 > 138.49\} < .0001$$

34

Likelihood ratio (deviance) tests:

Model A: $\log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0$ and $Y_j \sim Bin(n_j, \pi_j)$

Model B: $\log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_1 X_j$ and $Y_j \sim Bin(n_j, \pi_j)$

Model C: $\log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \alpha_j$ and $Y_j \sim Bin(n_j, \pi_j)$

Model C simply says that $0 < \pi_j < 1$, $j = 1, \dots, k$

36

Null: Deviance residual

$$\begin{aligned} \text{Deviance}_{\text{Null}} &= -2 \left[\log\text{-likelihood}_{\text{Model A}} \right. \\ &\quad \left. - \log\text{-likelihood}_{\text{Model C}} \right] \\ &= 2 \left[\sum_{j=1}^k Y_j \log(P_j / \bar{P}) \right. \\ &\quad \left. + \sum_{j=1}^k (n_j - Y_j) \log\left(\frac{1 - P_j}{1 - \bar{P}}\right) \right] \\ &\quad \text{with } (k - 1) \text{ d.f.} \end{aligned}$$

where

$$P_j = \frac{Y_j}{n_j} \quad j = 1, \dots, k$$

$$\bar{P} = \frac{\sum Y_j}{\sum n_j} = \frac{\sum_{j=1}^k n_j P_j}{\sum_{j=1}^k n_j}$$

are the m.l.e.'s of π_j under models C and A, respectively.

37

LOOSE: DEV (Lack-of-fit test)

$$H_0 : \log\left(\frac{\pi_j}{1 - \pi_j}\right) = \beta_0 + \beta_1 X_j \quad (\text{model B})$$

$$H_A : \log\left(\frac{\pi_j}{1 - \pi_j}\right) = \beta_0 + \alpha_j \quad (\text{model C})$$

$$\begin{aligned} G^2 &= \text{deviance}_{\text{lack-of-fit}} = \text{deviance}_{\text{NULL}} \\ &\quad - \text{deviance}_{\text{LOOSE}} \\ &= 2 \left[\sum_{j=1}^k Y_j \log\left(\frac{P_j}{\hat{\pi}_j}\right) \right. \\ &\quad \left. + \sum_{j=1}^k (n_j - Y_j) \log\left(\frac{1 - P_j}{1 - \hat{\pi}_j}\right) \right] \\ &= 11.42 \end{aligned}$$

on

$$\begin{aligned} [(k - 1) - 1] &= k - 2 \text{ d.f.} \\ &= 8 - 2 = 6 \text{ d.f.} \\ &\quad (\text{p-value} = 0.762) \end{aligned}$$

39

LOOSE: deviance residuals

$$\begin{aligned} \text{Deviance}_{\text{LOOSE}} &= -2 \left[\log\text{-likelihood}_{\text{Model A}} \right. \\ &\quad \left. - \log\text{-likelihood}_{\text{Model B}} \right] \\ &= 2 \left[\sum_{j=1}^k Y_j \log\left(\frac{\hat{\pi}_j}{\bar{P}}\right) \right. \\ &\quad \left. + \sum_{j=1}^k (n_j - Y_j) \log\left(\frac{1 - \hat{\pi}_j}{1 - \bar{P}}\right) \right] \\ &= 272.7803 \text{ with } (2 - 1) = 1 \text{ d.f.} \end{aligned}$$

where

$$\hat{\pi}_j = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 X_j)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_j)}$$

$$\bar{P} = \frac{\exp(\hat{\beta}_0^*)}{1 + \exp(\hat{\beta}_0^*)} = \frac{\sum_{j=1}^k Y_j}{\sum_{j=1}^k n_j}$$

38

Pearson chi-square test:

$$\begin{aligned} X^2 &= \sum_{j=1}^k \frac{(Y_j - n_j \hat{\pi}_j)^2}{n_j \hat{\pi}_j} \\ &\quad + \sum_{j=1}^k \frac{(n_j - Y_j - n_j(1 - \hat{\pi}_j))^2}{n_j(1 - \hat{\pi}_j)} \\ &= \sum_{j=1}^k n_j \frac{(P_j - \hat{\pi}_j)^2}{\hat{\pi}_j} + \sum_{j=1}^k n_j \frac{(1 - P_j - [1 - \hat{\pi}_j])^2}{1 - \hat{\pi}_j} \\ &= 10.223 \end{aligned}$$

with d.f. = $k - 2 = 6$

p-value = .1155

40

Estimates of expected counts under Model B.

j	$n_j \hat{\pi}_j$	$n_j(1 - \hat{\pi}_j)$
1	3.43	55.57
2	9.78	50.22
3	22.66	39.34
4	33.83	22.17
5	50.05	12.95
6	53.27	5.73
6	59.22	2.78
8	58.74	1.26

Cochran's rule:

- All expected counts should be larger than one.
- At least 80% should be larger than 5.

41

Define

$$G_j = \begin{bmatrix} \frac{\partial \hat{\pi}_j}{\partial \beta_0} & \frac{\partial \hat{\pi}_j}{\partial \beta_1} \end{bmatrix} = [\hat{\pi}_j(1 - \hat{\pi}_j)X_j \hat{\pi}_j(1 - \hat{\pi}_j)]$$

Then, $S_{\hat{\pi}_j}^2 = G_j[\widehat{Var}(\hat{\beta})]G_j^T$ and an approximate standard error is

$$S_{\hat{\pi}_j} = \sqrt{G_j[\widehat{Var}(\hat{\beta})]G_j^T}$$

An approximate $(1 - \alpha) \times 100\%$ confidence interval for π_j is

$$\hat{\pi}_j \pm Z_{\alpha/2} S_{\hat{\pi}_j}$$

43

Estimate mortality rates:

$$\hat{\pi}_j = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 X_j)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_j)}$$

Use the delta method to compute an approximate standard error

$$\begin{aligned} \frac{\partial \hat{\pi}_j}{\partial \hat{\beta}_0} &= \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 X_j)}{[1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_j)]^2} \\ &= \hat{\pi}_j(1 - \hat{\pi}_j) \\ \frac{\partial \hat{\pi}_j}{\partial \hat{\beta}_1} &= \frac{X_j \exp(\hat{\beta}_0 + \hat{\beta}_1 X_j)}{[1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_j)]^2} \\ &= X_j \hat{\pi}_j(1 - \hat{\pi}_j) \end{aligned}$$

42

Estimate the log-dose that provides a certain mortality rate, say π_0 .

$$\log\left(\frac{\pi_0}{1 - \pi_0}\right) = \beta_0 + \beta_1 X_0 \Rightarrow$$

$$X_0 = \frac{\log\left(\frac{\pi_0}{1 - \pi_0}\right) - \beta_0}{\beta_1}$$

The m.l.e. for X_0 is

$$\hat{X}_0 = \frac{\log\left(\frac{\pi_0}{1 - \pi_0}\right) - \hat{\beta}_0}{\hat{\beta}_1}$$

44

Use the delta method to obtain an approximate standard error:

$$\frac{\partial \hat{X}_0}{\partial \hat{\beta}_0} = \frac{-1}{\hat{\beta}_1}$$

$$\frac{\partial \hat{X}_0}{\partial \hat{\beta}_1} = \frac{\left[\log\left(\frac{\pi_0}{1-\pi_0}\right) - \hat{\beta}_0 \right]}{\hat{\beta}_1^2}$$

Define $\hat{G} = \begin{bmatrix} \frac{\partial \hat{X}_0}{\partial \hat{\beta}_0} & \frac{\partial \hat{X}_0}{\partial \hat{\beta}_1} \end{bmatrix}$ and compute

$$S_{\hat{X}_0}^2 = \hat{G} [\widehat{Var}(\hat{\beta})] \hat{G}^T$$

45

The m.l.e. for the dose is

$$\widehat{dose}_0 = \exp(\hat{X}_0)$$

Apply the delta method

$$\frac{\partial \exp(\hat{X}_0)}{\partial \hat{X}_0} = \exp(\hat{X}_0)$$

Then

$$S_{\widehat{dose}_0} = [\exp(\hat{X}_0)]^2 \widehat{Var}(\hat{X}_0)$$

$$S_{\widehat{dose}} = [\exp(\hat{X}_0)] S_{\hat{X}_0} = (\widehat{dose}_0) S_{\hat{X}_0}$$

and an approximate $(1 - \alpha) \times 100\%$ confidence interval for $dose_0$ is

$$\exp(\hat{X}_0) \pm Z_{\alpha/2} [\exp(\hat{X}_0)] S_{\hat{X}_0}$$

47

The standard error is

$$S_{\hat{X}_0} = \sqrt{\hat{G} [\widehat{Var}(\hat{\beta})] \hat{G}^T}$$

and an approximate $(1 - \alpha) \times 100\%$ confidence interval for X_0 is

$$\hat{X}_0 \pm Z_{\alpha/2} S_{\hat{X}_0}$$

Suppose you want a 95% confidence interval for the dose that provides a mortality rate of π_0 .

$$dose_0 = \exp(X_0)$$

46

Fit a probit model to the Bliss beetle data:

$$\Phi^{-1}(\pi_j) = \beta_0 + \beta_1 X_j$$

where

$$X_j = \log(\text{dose})$$

$$Y_j = \text{Bin}(n_j, \pi_j)$$

$$j = 1, 2, \dots, k$$

Note that

$$\pi_j = \int_{-\infty}^{\beta_0 + \beta_1 X_j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

48

Fit the complimentary log-log model to the Bliss beetle data

$$\log[-\log(1 - \pi_j)] = \beta_0 + \beta_1 X_j$$

where

$$\begin{aligned} X_j &= \log(\text{dose}) \\ Y_j &= \text{Bin}(n_j, \pi_j) \end{aligned}$$

Note that

$$\pi_j = 1 - \exp[-e^{\beta_0 + \beta_1 X_j}] \quad j = 1, 2, \dots, k$$

49

Independent Poisson counts:

$$Y_{jk} \sim \text{Poisson}(m_j) \quad j = 1, \dots, 5 \quad k = 1, \dots, 3$$

where

$$\Pr\{Y_{jk} = y\} = \frac{m_j^y}{y!} e^{-m_j} \quad \text{for } y = 0, 1, 2, \dots$$

$$E(Y_{jk}) = m_j$$

$$\text{Var}(Y_{jk}) = m_j$$

Log-link function:

$$\log(m_j) = \beta_0 + \beta_1 X_j$$

51

Poisson Regression (Log-linear models)[.2in]

Example 13.2:

Y_{jk} = number of TA98 salmonella colonies on the k-th plate with the j-th dosage level of quinoline

X_j = $\log_{10}(\text{dose of quinoline}(\mu\text{g}))$

$$X_1 = 0$$

$$X_2 = 1.0$$

$$X_3 = 1.5$$

$$X_4 = 2.0$$

$$X_5 = 2.5$$

and $k = 1, 2, 3$ plates at each ??

50

Joint likelihood function

$$L(\beta) = \pi_{j=1}^5 \pi_{k=1}^3 \frac{m_j^{Y_{jk}}}{Y_{jk}!} e^{-m_j}$$

Log-likelihood function

$$\begin{aligned} \ell(\beta) &= \sum_{j=1}^5 \sum_{k=1}^3 [Y_{jk} \log(m_j) - m_j - \log(Y_{jk}!)] \\ &= \sum_{j=1}^5 \sum_{k=1}^3 [Y_{jk} [\beta_0 + \beta_1 X_j] - e^{\beta_0 + \beta_1 X_j} - \log(Y_{jk}!)] \\ &= \beta_0 Y_{00} + \beta_1 \sum_{j=1}^5 X_j Y_{j0} - 3 \sum_{j=1}^5 e^{\beta_0 + \beta_1 X_j} \\ &\quad - \sum_{j=1}^5 \sum_{k=1}^3 \log(Y_{jk}!) \end{aligned}$$

52

Likelihood equations:

$$0 = \frac{\partial \ell(\beta)}{\partial \beta_0} = Y_{..} - 3 \sum_{j=1}^5 e^{\beta_0 + \beta_1 X_j}$$

$$0 = \frac{\partial \ell(\beta)}{\partial \beta_1} = \sum_{j=1}^5 X_j Y_j - 3 \sum_{j=1}^5 X_j e^{\beta_0 + \beta_1 X_j}$$

Use the Fisher scoring algorithm:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{53} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_1 \\ \vdots & \vdots \\ 1 & X_5 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} \frac{\partial \ell(\beta)}{\partial \beta_0} \\ \frac{\partial \ell(\beta)}{\partial \beta_1} \end{bmatrix} = \mathbf{X}^T (\mathbf{Y} - \mathbf{m})$$

where

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_1 \\ \vdots \\ m_5 \end{bmatrix} = \exp(\mathbf{X}\beta)$$

53

Fisher information matrix

$$\begin{aligned} \mathbf{H} &= E \left[\begin{array}{cc} \frac{-\partial^2 \ell(\beta)}{\partial \beta_0^2} & \frac{-\partial^2 \ell(\beta)}{\partial \beta_0 \partial \beta_1} \\ \frac{-\partial^2 \ell(\beta)}{\partial \beta_0 \partial \beta_1} & \frac{-\partial^2 \ell(\beta)}{\partial \beta_1^2} \end{array} \right] \\ &= \begin{bmatrix} \sum_j \sum_k m_j & \sum_j \sum_k X_j m_j \\ \sum_j \sum_k X_j m_j & \sum_j \sum_k X_j^2 m_j \end{bmatrix} \\ &= \mathbf{X}^T \Delta_m \mathbf{X} \end{aligned}$$

where

$$\text{Var}(\mathbf{Y}) = \Delta_m = \begin{bmatrix} m_1 & & & & \\ & m_1 & & & \\ & & m_1 & & \\ & & & m_2 & \dots \\ & & & & \dots & m_5 \end{bmatrix}$$

55

Second partial derivatives of $\ell(\beta)$

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_0^2} = -3 \sum_{j=1}^5 e^{\beta_0 + \beta_1 X_j} = - \sum_j \sum_k m_j$$

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_0 \partial \beta_1} = -3 \sum_{j=1}^5 X_j e^{\beta_0 + \beta_1 X_j} = - \sum_j \sum_k X_j m_j$$

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_1^2} = -3 \sum_{j=1}^5 X_j^2 e^{\beta_0 + \beta_1 X_j} = - \sum_j \sum_k X_j^2 m_j$$

54

Iterations:

$$\beta^{(s+1)} = \beta^{(s)} + \mathbf{H}^{-1} \mathbf{Q}$$

- Use halving steps as needed
- Starting values: Use the m.l.e.s for the model with $\beta_1 = 0$, i.e.,

$$Y_{ij} \sim \text{Poisson}(m)$$

are i.e.d. with $m = e^{\beta_0}$. Then,

$$\hat{m}^{(0)} = \bar{Y}_{..} \quad \text{and}$$

$$\hat{\beta}_0^{(0)} = \log(\bar{Y}_{..})$$

$$\hat{\beta}_1^{(0)} = 0$$

56

Large sample normal approximation for the distribution of $\hat{\beta}$

$$\hat{\beta} \sim N(\beta, H^{-1})$$

↗ estimate this by substituting $\hat{\beta}$ for β in H to obtain $\hat{\beta}$

57

Define

$$\hat{G} = \begin{bmatrix} \frac{\partial \hat{m}_*}{\partial \beta_0} & \frac{\partial \hat{m}_*}{\partial \beta_1} \end{bmatrix} = [\hat{m}_* X_* \hat{m}_*]$$

Then the estimated variance of

$$\hat{m}_* = e^{\hat{\beta}_0 + \hat{\beta}_1 X_*}$$

is

$$\begin{aligned} S_{\hat{m}_*}^2 &= \hat{G} \hat{H}^{-1} \hat{G}^T \\ &= \hat{G} [X_*^T \Delta_{\hat{m}} X_*]^{-1} \hat{G}^T \end{aligned}$$

The standard error is $S_{\hat{m}_*} = \sqrt{S_{\hat{m}_*}^2}$ and an approximate $(1 - \alpha) \times 100\%$ confidence interval is

$$\hat{m}_* \pm Z_{\alpha/2} S_{\hat{m}_*}$$

59

Estimate the mean number of colonies for a specific \log_{10} (dose) of quinoline, say X_*

The m.l.e. is

$$\hat{m}_* = e^{\hat{\beta}_0 + \hat{\beta}_1 X_*}$$

Apply the delta method:

$$\begin{aligned} \frac{\partial \hat{m}_*}{\partial \hat{\beta}_0} &= e^{\hat{\beta}_0 + \hat{\beta}_1 X_*} = \hat{m}_* \\ \frac{\partial \hat{m}_*}{\partial \hat{\beta}_1} &= X_* e^{\hat{\beta}_0 + \hat{\beta}_1 X_*} = X_* \hat{m}_* \end{aligned}$$

58

Four kinds of residuals:

Response residuals:

$$Y_i - \hat{m}u_i = \left[\frac{Y_i}{n_i} - \hat{\pi}_i \right]$$

Pearson residuals:

$$\begin{aligned} \frac{Y_i - \hat{\mu}_i}{\sqrt{\widehat{Var}(Y_i)}} &= \frac{Y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}} \quad (\text{binomial}) \\ &= \frac{Y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}} \quad (\text{Poisson}) \end{aligned}$$

60

Working residuals:

$$\begin{aligned} \frac{Y_i - \hat{\mu}_i}{\left(\frac{\partial \mu_i}{\partial \eta_i}\right)} &= \frac{Y_i - n\hat{\pi}_i}{n_i\hat{\pi}_i(1 - \hat{\pi}_i)} \text{ (binomial)} \\ &= \frac{Y_i - \hat{\mu}_i}{\hat{\mu}_i} \text{ (Poisson)} \end{aligned}$$

For the binomial family with logit link

$$\begin{aligned} \eta_i &= \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \log\left(\frac{n_i\pi_i}{n_i(1 - \pi_i)}\right) = \mathbf{X}_i^T \boldsymbol{\beta} \\ \mu_i &= n_i\pi_i = n_i \frac{\exp(\mathbf{X}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i^T \boldsymbol{\beta})} = n_i \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \\ \frac{\partial \mu_i}{\partial \eta_i} &= n_i \frac{\exp(\eta_i)}{[1 + \exp(\eta_i)]^2} = n_i\pi_i(1 - \pi_i) \end{aligned}$$

61

Poisson family

$$\begin{aligned} G^2 &= \text{(deviance)} \\ &= 2 \sum_{i=1}^k Y_i \log\left(\frac{Y_i}{\hat{\mu}_i}\right) \end{aligned}$$

where

$$\begin{aligned} \hat{\mu}_i &= \exp(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots) \\ \log(\mu_i) &= \beta_0 + \beta_1 X_{1i} + \dots \end{aligned}$$

then

$$e_i = \text{sign}(Y_i - \hat{\mu}_i) \sqrt{2Y_i \log\left(\frac{Y_i}{\hat{\mu}_i}\right)}$$

63

Deviance residual:

$$e_i = \text{sign}(Y_i - \hat{\mu}_i) \sqrt{|d_i|}$$

where d_i is the contribution of the i -th case to the deviance

binomial family

$$G^2 = 2 \sum_{i=1}^k \left[Y_i \log\left(\frac{Y_i}{n_i\hat{\pi}_i}\right) + (n_i - Y_i) \log\left(\frac{n_i - Y_i}{n_i(1 - \hat{\pi}_i)}\right) \right]$$

and

$$\begin{aligned} e_i &= \text{sign}(Y_i - n\hat{\pi}_i) \\ &\times \sqrt{\left| 2 \left[Y_i \log\left(\frac{Y_i}{n_i\hat{\pi}_i}\right) + (n_i - Y_i) \log\left(\frac{n_i - Y_i}{n_i(1 - \hat{\pi}_i)}\right) \right] \right|} \end{aligned}$$

62