Computing in Epidemiology and Biostatistics

Find the maximum likelihood estimates for a Poisson regression

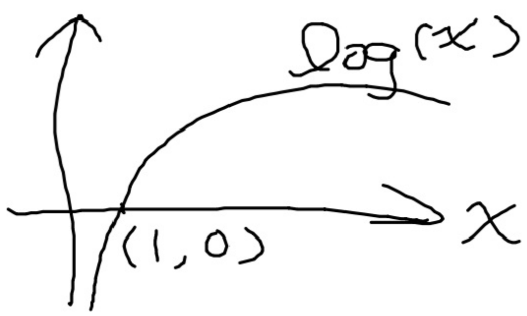
Wan-Yu Lin

Ex 20-2：Data: seizure.csv please read it into R and call it “seizure” object：

(1) Please use the Newton-Raphson method to find the maximum likelihood estimate (MLE) of the regression coefficients of Poisson regression ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$), and compare your result with the R built-in function glm(y~trt+age, data=seizure, offset=lweek, family=poisson)

(2) Please find the variance-covariance matrix for $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, and compare your result with the R built-in function vcov(model)

(3) Please find the log likelihood at $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, and compare your result with the R built-in function logLik(model)



[Hint]

the average seizure count per week >= 0

$$\log\left(\frac{E(Y_i)}{week_i}\right) = \beta_0 + \beta_1 trt_i + \beta_2 age_i = \begin{bmatrix}1 & trt_i & age_i\end{bmatrix}\begin{bmatrix}\beta_0 \\ \beta_1 \\ \beta_2\end{bmatrix} = \boldsymbol{x_i}'\boldsymbol{\beta}$$

$$\log(E(Y_i)) - \log(week_i) = \boldsymbol{x_i}'\boldsymbol{\beta}$$

$$\log(E(Y_i)) = \boldsymbol{x_i}'\boldsymbol{\beta} + \log(week_i)$$

$$\log(\mu_i) = \boldsymbol{x_i}'\boldsymbol{\beta} + \log(week_i) \quad => \quad \mu_i = e^{\boldsymbol{x_i}'\boldsymbol{\beta}} week_i = e^{\boldsymbol{x_i}'\boldsymbol{\beta}} e^{\log(week_i)} = e^{\boldsymbol{x_i}'\boldsymbol{\beta} + \log(week_i)}$$

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ -\mu_i + y_i \log \mu_i - \log(y_i!) \right\}$$

$$= \sum_{i=1}^{n} \left\{ -e^{x_i'\beta} week_i + y_i \left(\boldsymbol{x_i}'\boldsymbol{\beta} + \log(week_i) \right) - \log(y_i!) \right\}$$

$$= \sum_{i=1}^{n} \left\{ -e^{x_i'\beta} week_i + y_i \boldsymbol{x_i}'\boldsymbol{\beta} + y_i \log(week_i) - \log(y_i!) \right\}$$

$$\frac{d \log L(\boldsymbol{\beta})}{d\boldsymbol{\beta}}_{3\times1} = \sum_{i=1}^{n} \left\{ \boldsymbol{x}_{\boldsymbol{i}_{3\times1}} y_{i_{1\times1}} - \boldsymbol{x}_{\boldsymbol{i}_{3\times1}} e^{x_i'\beta}{}_{1\times1} week_{i_{1\times1}} \right\} = \sum_{i=1}^{n} \boldsymbol{x_i} \left\{ y_i - e^{x_i'\beta} week_i \right\} = \sum_{i=1}^{n} \boldsymbol{x_i} \left\{ y_i - \mu_i \right\}$$

$$= \boldsymbol{x}_1 \{y_1 - \mu_1\} + \boldsymbol{x}_2 \{y_2 - \mu_2\} + \cdots + \boldsymbol{x}_n \{y_n - \mu_n\}$$

$$= \begin{bmatrix} 1 \\ trt_1 \\ age_1 \end{bmatrix} \{y_1 - \mu_1\} + \begin{bmatrix} 1 \\ trt_2 \\ age_2 \end{bmatrix} \{y_2 - \mu_2\} + \cdots + \begin{bmatrix} 1 \\ trt_n \\ age_n \end{bmatrix} \{y_n - \mu_n\}$$

$$= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ trt_1 & trt_2 & \cdots & trt_n \\ age_1 & age_2 & \cdots & age_n \end{bmatrix}_{3\times n} \begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \\ \vdots \\ y_n - \mu_n \end{bmatrix}_{n\times1}$$

$$\frac{d^2 \log L(\boldsymbol{\beta})}{d\boldsymbol{\beta}^2}_{3\times3} = \sum_{i=1}^{n} \left\{ -\boldsymbol{x}_{\boldsymbol{i}_{3\times1}} e^{x_i'\beta}{}_{1\times1} week_{i_{1\times1}} \boldsymbol{x_i}'_{1\times3} \right\} = \sum_{i=1}^{n} \left\{ -\boldsymbol{x_i} \mu_i \boldsymbol{x_i}' \right\}$$

$$= -\boldsymbol{x}_1 \mu_1 \boldsymbol{x}_1' - \boldsymbol{x}_2 \mu_2 \boldsymbol{x}_2' - \cdots - \boldsymbol{x}_n \mu_n \boldsymbol{x}_n'$$

$$= -\begin{bmatrix} 1 \\ trt_1 \\ age_1 \end{bmatrix} \mu_1 \begin{bmatrix} 1 & trt_1 & age_1 \end{bmatrix} - \begin{bmatrix} 1 \\ trt_2 \\ age_2 \end{bmatrix} \mu_2 \begin{bmatrix} 1 & trt_2 & age_2 \end{bmatrix} - \cdots$$

$$- \begin{bmatrix} 1 \\ trt_n \\ age_n \end{bmatrix} \mu_n \begin{bmatrix} 1 & trt_n & age_n \end{bmatrix}$$

$$= -\begin{bmatrix} 1 & 1 & \cdots & 1 \\ trt_1 & trt_2 & \cdots & trt_n \\ age_1 & age_2 & \cdots & age_n \end{bmatrix}_{3\times n} \begin{bmatrix} \mu_1 & 0 & 0 & 0 \\ 0 & \mu_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mu_n \end{bmatrix}_{n\times n} \begin{bmatrix} 1 & trt_1 & age_1 \\ 1 & trt_2 & age_2 \\ \vdots & \vdots & \vdots \\ 1 & trt_n & age_n \end{bmatrix}_{n\times3}$$

Ex 20-3：Data: rate.csv please read it into R and call it “rate” object：

(1) Please use the Newton-Raphson method to find the maximum likelihood estimate (MLE) of the regression coefficients of Poisson regression ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$), and compare your result with the R built-in function “glm(Death~Age+sex, offset=log(PY/100000), data=rate, family=poisson)”.

(2) Please find the variance-covariance matrix for $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, and compare your result with the R built-in function vcov(model).

(3) Please find the log likelihood at $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, and compare your result with the R built-in function logLik(model).

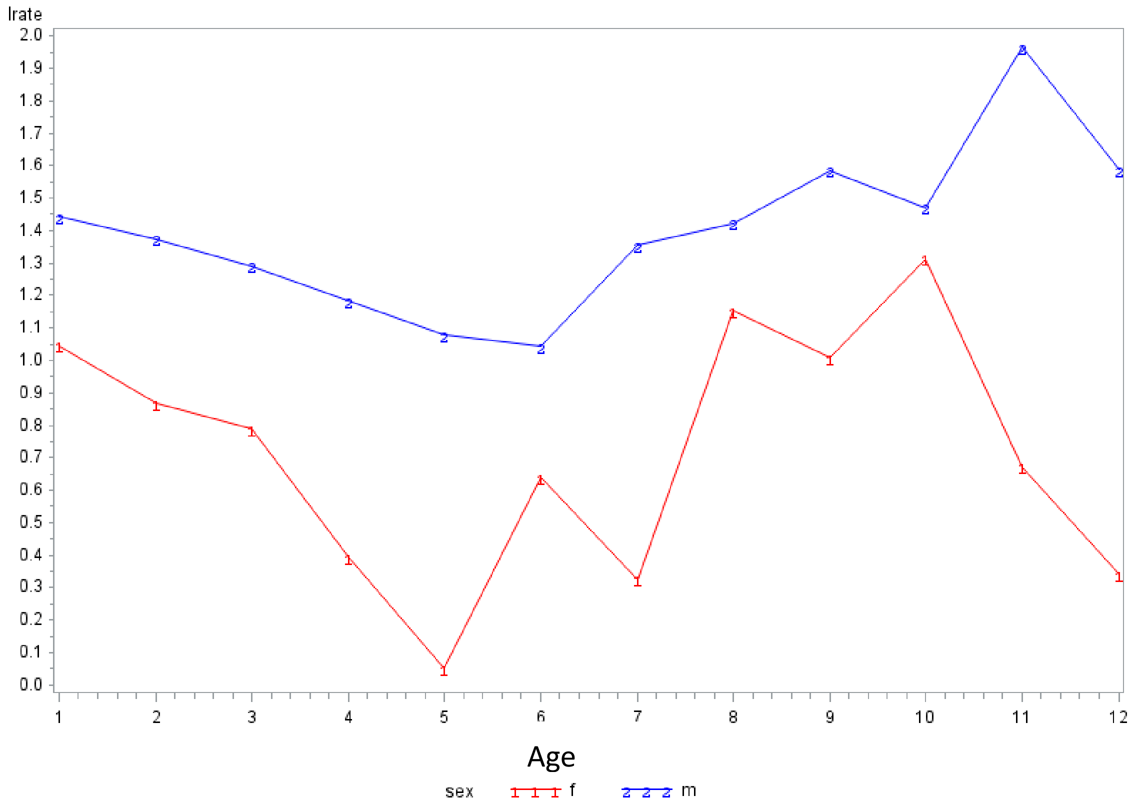
[Hint] Death rate per 100,000 person years

$$\log\left(\frac{E(Y_i)}{PY_i/100000}\right) = \beta_0 + \beta_1 Age_i + \beta_2 sex_i = \begin{bmatrix}1 & Age_i & sex_i\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix} = \boldsymbol{x_i}'\boldsymbol{\beta}$$

$$\log(E(Y_i)) - \log(PY_i/100000) = \boldsymbol{x_i}'\boldsymbol{\beta}$$

$$\log(E(Y_i)) = \boldsymbol{x_i}'\boldsymbol{\beta} + \log(PY_i/100000)$$

Standardized person years



1	30-34
2	35-39
3	40-44
4	45-49
5	50-54
6	55-59
7	60-64
8	65-69
9	70-74
10	75-79
11	80-84
12	85+

Homework: (8 points, please pay attention to all the words in this orange box):

Ex 20-4 ：Continue Ex 20-3, because log(Death rate per 100,000 person years) is not linear in Age, it will be more reasonable to recode Age as 11 dummy variables.

rate$Age.f <- factor(rate$Age)

(1) Please use the Newton-Raphson method to find the maximum likelihood estimate (MLE) of the regression coefficients of Poisson regression ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots$) (4 points)

Please note: No need to make a plot for the initial value. Using the R built-in function "glm(Death~Age.f+sex, offset=log(PY/100000), data=rate, family=poisson)" to answer this homework will be scored as 0, although you may use it to check your own answer.

(2) Please find the variance-covariance matrix for $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots$ (2 points)

Please note: Using the R built-in function "vcov(model)" to answer this homework will be scored as 0, although you may use it to check your own answer.

(3) Please find the log likelihood at $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots$ (2 points)

Please note: Using the R built-in function "logLik(model)" to answer this homework will be scored as 0, although you may use it to check your own answer.

[Hint]   Recode Age as 11 dummy variables

	Age.f2	Age.f3	Age.f4	Age.f5	Age.f6	Age.f7	Age.f8	Age.f9	Age.f10	Age.f11	Age.f12
1	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0	0
5	0	0	0	1	0	0	0	0	0	0	0
6	0	0	0	0	1	0	0	0	0	0	0
7	0	0	0	0	0	1	0	0	0	0	0
8	0	0	0	0	0	0	1	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	0
10	0	0	0	0	0	0	0	0	1	0	0
11	0	0	0	0	0	0	0	0	0	1	0
12	0	0	0	0	0	0	0	0	0	0	1