UNIVERSITY OF EDINBURGH
SCHOOL OF MATHEMATICS

Generalised Regression Models

GRM: Solutions 5 **Semester 1, 2022–2023**

---

1. (a) The problem can be analysed as a GLM since it has the following components.

   (i) **Model matrix and parameters**:

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad \beta^T = (\theta_1, \theta_2)$$

   (ii) **Link function:** $\lambda_i = g(\lambda_i) = \eta_i = \mathbf{x}_i^T\beta$. Since $\lambda_i \equiv \mu_i$ we have that g is the identity link function.

   (iii) **Exponential family**: the Poisson distribution is a member of the exponential family since

$$f(y;\lambda) = \frac{\lambda^y e^{-\lambda}}{y!} = \exp\{y\log\lambda - \lambda - \log y!\},$$

   i.e., $f(y;\lambda)$ is of form $\exp\{a(y)b(\lambda)+c(\lambda)+d(y)\}$

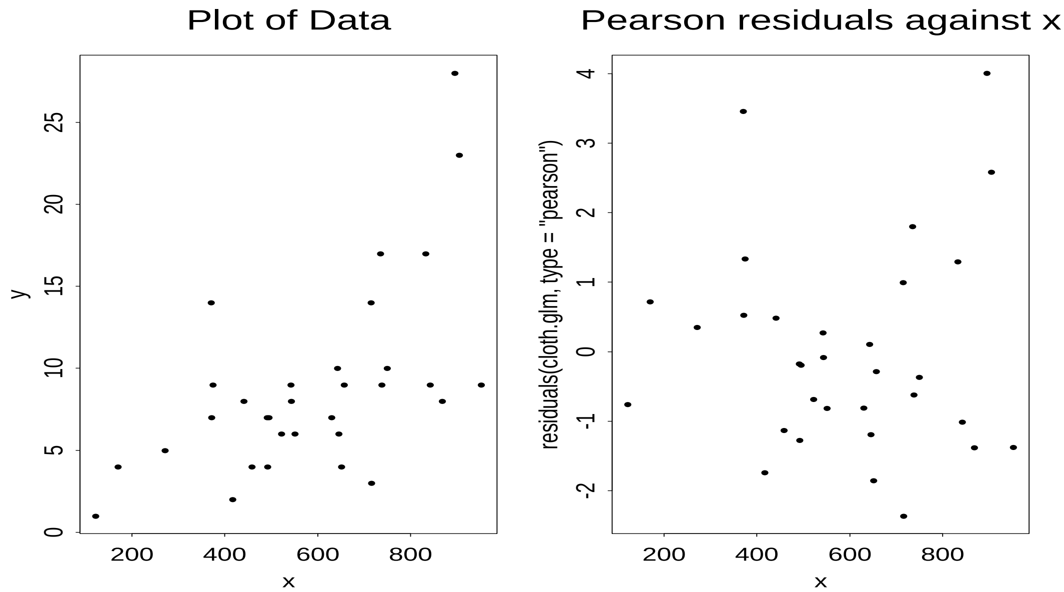
(b)
```
> x <- c(551,651,832,375,715,868,271,630,491,372,645,441,895,458,642,492,
    543,842,905,542,522,122,657,170,738,371,735,749,495,716,952,417)
> y <- c(6,4,17,9,14,8,5,7,7,7,6,8,28,4,10,4,
    8,9,23,9,6,1,9,4,9,14,17,10,7,3,9,2)
> cloth.df <- data.frame(x=x, y=y)
> attach(cloth.df)

> par(mfrow=c(1,2))
> plot(x,y,main='Plot of Data')

> cloth.glm <- glm(y~x,poisson(link=identity))
> summary(cloth.glm)
Call: glm(formula = y ~ x, family = poisson(link = identity))
Deviance Residuals:
       Min        1Q     Median       3Q      Max
 -2.798506 -1.104746 -0.2399216 0.550989 3.490582

Coefficients:
                Value  Std. Error   t value
(Intercept) 0.3234857 1.111843792 0.2909453
          x 0.0145519 0.002079579 6.9975214

(Dispersion Parameter for Poisson family taken to be 1 )

    Null Deviance: 103.7138 on 31 degrees of freedom

Residual Deviance: 64.45047 on 30 degrees of freedom

Number of Fisher Scoring Iterations: 3

Correlation of Coefficients:
  (Intercept)
x -0.9024138
```

(c) The ith Pearson residual is $r_i = \frac{y_i - \widehat{\mu}_i}{\sqrt{V(\widehat{\mu}_i)}}$. For Poisson responses the variance function is $V(\mu_i) = \mu_i$, and thus the 1st Pearson residual is $r_1 = \frac{6-8.342}{\sqrt{8.342}} = -0.811$

```
> fitted(cloth.glm)[1:8]                    #first 8 fitted values
       1        2        3        4        5        6       7        8
 8.341582 9.796772 12.43067 5.780448 10.72809 12.95453 4.26705 9.491182

> residuals(cloth.glm,type="pearson")[1:8]  #first 8 Pearson residuals
        1         2        3        4         5         6        7          8
 -0.810754 -1.851991 1.29595 1.339207 0.9989161 -1.376488 0.354876 -0.8086141
```

(d) R/S-PLUS plot of Pearson residuals shows some large values.

```
> plot(x,residuals(cloth.glm,type="pearson"))
> title('Pearson residuals against x')
```



2. (a) Model ω with $\lambda_i = \theta_1$ (constant) has fitted values $\widehat{\mu}_i = \widehat{\theta}_1 = \bar{y} = 8.875$, i.e., the usual MLE under an iid Poisson model.

(b) To test the null hypothesis that $\theta_2 = 0$, use analysis of deviance.

```
> anova(cloth.glm)
Analysis of Deviance Table

Poisson model

Response: y

Terms added sequentially (first to last)
     Df Deviance Resid. Df Resid. Dev
NULL                    31   103.7138
   x  1 39.26328        30    64.4505
```

From this analysis of deviance table, we see that the deviance of the model ω with $\eta = \theta_1$ is D_ω =103.71 (31 df) and the deviance of the model Ω with $\eta = \theta_1 + \theta_2 x$ is D_Ω = 64.45 (30 df).

Therefore the change in deviance is $D_\omega - D_\Omega = 39.26$. Under the null hypothesis that $\theta_2 = 0$, $D_\omega - D_\Omega$ has a χ^2 distribution with 1 degree of freedom. Since the 5% critical value is 3.84, there is strong evidence that $\theta_2 \neq 0$.

(Note that the model Ω has a residual deviance of $D_\Omega = 64.45$ with 30 df, and thus the full model does not provide a very good fit to the data.)

3. For the gamma distribution with pdf $f(y;\theta) = \frac{y^{\phi-1}\theta^{\phi}e^{-y\theta}}{\Gamma(\phi)}$ (ϕ known), we have $\mu_i = \mathrm{E}(Y_i) = \frac{\phi}{\theta_i}$, and thus $\theta_i = \frac{\phi}{\mu_i}$.

   (a) log likelihood

$$\begin{aligned} l(\beta) = \sum_{i=1}^{n} \log f(y_i;\theta_i) &= \sum_{i=1}^{n} \{-y_i\theta_i + \phi\log\theta_i - \log\Gamma(\phi) + (\phi-1)\log y_i\} \\ &= \sum_{i=1}^{n} \left\{-y_i\frac{\phi}{\mu_i} + \phi\log\frac{\phi}{\mu_i} - \log\Gamma(\phi) + (\phi-1)\log y_i\right\} \end{aligned}$$

   Thus, the deviance of a model with fitted values $\widehat{\mu}_i = g^{-1}(\mathbf{x}_i^T\widehat{\beta})$ is

$$\begin{aligned} D &= -2\{l(\widehat{\beta}) - l(\text{saturated or maximal model with fitted values } \widehat{\mu}_i^{saturated} = y_i)\} \\ &= -2\sum_{i=1}^{n}\left\{-\phi\frac{y_i}{\widehat{\mu}_i} + \phi\log\frac{\phi}{\widehat{\mu}_i} - (-\phi\frac{y_i}{y_i} + \phi\log\frac{\phi}{y_i})\right\} \\ &= 2\phi\sum_{i=1}^{n}\left\{\log\frac{\widehat{\mu}_i}{y_i} + \frac{y_i - \widehat{\mu}_i}{\widehat{\mu}_i}\right\} \end{aligned}$$

   (b) Since, for gamma data the variance function (as a function of μ_i) is $V(\mu_i) = \frac{\phi}{\theta_i^2} = \phi\left(\frac{\mu_i}{\phi}\right)^2 = \frac{\mu_i^2}{\phi}$, the Pearson residuals are given by $r_i = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mu}_i^2/\phi}}$.

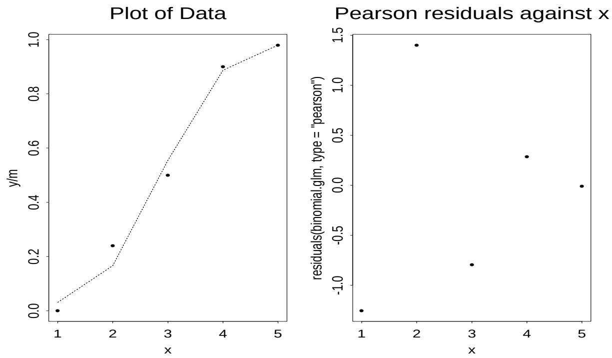
4. Let the linear component of the model be denoted by $\eta = \alpha + \beta x$, then

$$\theta = \frac{\exp(\eta)}{1+\exp(\eta)} \Rightarrow \theta + \theta\exp(\eta) = \exp(\eta) \Rightarrow \theta = \exp(\eta)(1-\theta) \Rightarrow \log\left(\frac{\theta}{1-\theta}\right) = \eta \Rightarrow \mathrm{logit}(\theta) = \alpha + \beta x$$

   (a) Ordinary least squares with x as the explanatory variable and $\log\left(\frac{(R+\frac{1}{2})/N}{1-(R+\frac{1}{2})/N}\right)$ as the response variable gives estimates of α and β as $\tilde{\alpha} = -6.287$ and $\tilde{\beta} = 2.179$ (using `lm` function). However, it is important to note that this is **not** the best use of the data as the variances of the responses are non-constant: the appropriate analysis is given below, i.e., the analysis of the model as a GLM.

   (b)
```
> x <- c(1,2,3,4,5)
> y <- c(0,12,25,45,49)
> m <- c(50,50,50,50,50)
> binomial.df <- data.frame(x=x, y=y, my = m-y)
> attach(binomial.df)
> binomial.glm <- glm( cbind(y,my) ~x,binomial(link=logit))
> summary(binomial.glm)

Call: glm(formula = cbind(y, my) ~ x, family = binomial(link = logit))
Deviance Residuals:
         1        2           3        4            5
 -1.766399 1.334817 -0.7922928 0.291387 -0.008383783

Coefficients:
                Value Std. Error   t value
(Intercept) -5.289629  0.6516660 -8.117086
          x  1.837986  0.2154278  8.531797

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 179.2624 on 4 degrees of freedom

Residual Deviance: 5.614606 on 3 degrees of freedom

Number of Fisher Scoring Iterations: 3
```

```
> par(mfrow=c(1,2))
> plot(x,y/m,main='Plot of Data')
> lines(x, fitted(binomial.glm), lty=2)
> plot(x,residuals(binomial.glm,
            type="pearson"))
> title('Pearson residuals against x')
```



(c) Fitted values for model with constant probability across groups, i.e., logit(θ_i) = α (constant): usual MLE estimate of $\theta_i = \theta$ (in iid binomial case) is

$$\widehat{\theta}_i = \widehat{\theta} = \frac{\sum_{k=1}^{n} R_i}{\sum_{k=1}^{n} N_i} = \frac{131}{250} = 0.524.$$

Thus, the fitted values (on y-scale) are $\widehat{\mu}_i = N_i \widehat{\theta}_i = 50(0.524) = 26.2$.

(d) Deviance (binomial; fitted values $\{\tilde{\mu}_i\}$): $D = 2\sum_{i=1}^{k} \left\{ y_i \log_e \left(\frac{y_i}{\tilde{\mu}_i} \right) + (m_i - y_i) \log_e \left(\frac{m_i - y_i}{m_i - \tilde{\mu}_i} \right) \right\}$

Model ω, logit(θ) $= \alpha$:

x_i	y_i	$m_i - y_i$	$\widehat{\theta}_i$	$\widehat{\mu}_i = m_i\widehat{\theta}_i$	$m_i - \widehat{\mu}_i$
1	0	50	0.524	26.2	23.8
2	12	38	0.524	26.2	23.8
3	25	25	0.524	26.2	23.8
4	45	5	0.524	26.2	23.8
5	49	1	0.524	26.2	23.8

Thus the deviance is

$$D_\omega = 2\{0\log(0/26.2) + 50\log(50/23.8) + \cdots + 49\log(49/26.2) + 1\log(1/23.8)\} = 179.26 \quad (4\text{ df}).$$

Model Ω, logit(θ) $= \alpha + \beta x$ (R/S-PLUS gives MLEs $\widehat{\alpha} = -5.290$ and $\widehat{\beta} = 1.838$):

x_i	y_i	$m_i - y_i$	$\widehat{\eta}_i = \widehat{\alpha} + \widehat{\beta}x_i = -5.290 + 1.838x_i$	$\widehat{\theta}_i = e^{\widehat{\eta}_i}/(1+e^{\widehat{\eta}_i})$	$\widehat{\mu}_i = m_i\widehat{\theta}_i$	$m_i - \widehat{\mu}_i$
1	0	50	−3.452	0.031	1.54	48.46
2	12	38	−1.614	0.166	8.30	41.70
3	25	25	0.224	0.556	27.79	22.21
4	45	5	2.062	0.887	44.36	5.64
5	49	1	3.900	0.980	49.01	0.99

Thus the deviance is

$$D_\Omega = 2\{0\log(0/1.54) + 50\log(50/48.46) + \cdots + 49\log(49/49.01) + 1\log(1/0.99)\} = 5.6 \quad (3\text{ df}).$$

(i) The deviance D_Ω should be compared with a χ^2 distribution with 3 degrees of freedom. Since the 5% critical value is 7.815, we conclude that there is no evidence of nonlinearity in the linear component η.

(ii) The change in deviance $D_\omega - D_\Omega$ has a χ^2_1 distribution under the null hypothesis that $\beta = 0$. Thus, since $D_\omega - D_\Omega = 173.6$, and the 5% critical value is 3.84, we conclude that $\beta \neq 0$.

```
> anova(binomial.glm)      # analysis of deviance in R/S-PLUS
Analysis of Deviance Table

Binomial model

Response: cbind(y, my)

Terms added sequentially (first to last)
     Df Deviance Resid. Df Resid. Dev
NULL                     4   179.2624
   x  1 173.6478         3     5.6146
```

5. (a) If $x_{\bullet\bullet}$ is fixed this gives a multinomial distribution for $\{x_{ij}\}$ with associated probabilities

$$p_{ij} = \Pr(\text{Observation falls in row } i \text{ and column } j) \quad (i = 1, \ldots, r; j = 1, \ldots, c).$$

Likelihood: $L = \prod_{i=1}^{r}\prod_{j=1}^{c} p_{ij}^{x_{ij}}$ Log likelihood: $l = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}\log p_{ij}$

(i) Under independence of rows and columns $p_{ij} = p_{i\bullet}p_{\bullet j}$ where

$$\begin{aligned} p_{i\bullet} &= \Pr(\text{Observation falls in row } i) \quad (i = 1, \ldots, r) \\ p_{\bullet j} &= \Pr(\text{Observation falls in column } j) \quad (j = 1, \ldots, c) \end{aligned}$$

with constraints $\sum_{i=1}^{r} p_{i\bullet} = 1$ and $\sum_{j=1}^{c} p_{\bullet j} = 1$.
To obtain MLEs for $p_{i\bullet}$ and $p_{\bullet j}$ maximize

$$G = \sum_{i=1}^{r} x_{i\bullet}\log p_{i\bullet} + \sum_{j=1}^{c} x_{\bullet j}\log p_{\bullet j} + \lambda(1 - \sum_{i=1}^{r} p_{i\bullet}) + \gamma(1 - \sum_{j=1}^{c} p_{\bullet j}).$$

Differentiating l and setting equal to zero gives

$$\begin{aligned} (1)\quad 0 &= \frac{\partial G}{\partial p_{i\bullet}} = \frac{x_{i\bullet}}{p_{i\bullet}} - \lambda \quad \Rightarrow p_{i\bullet} = \frac{x_{i\bullet}}{\lambda} \\ (2)\quad 0 &= \frac{\partial G}{\partial p_{\bullet j}} = \frac{x_{\bullet j}}{p_{\bullet j}} - \gamma \quad \Rightarrow p_{\bullet j} = \frac{x_{\bullet j}}{\gamma} \\ 0 &= \frac{\partial G}{\partial \lambda} = 1 - \sum_{i=1}^{r} p_{i\bullet} \quad \text{and } (1) \Rightarrow \lambda = x_{\bullet\bullet} \\ 0 &= \frac{\partial G}{\partial \gamma} = 1 - \sum_{j=1}^{c} p_{\bullet j} \quad \text{and } (2) \Rightarrow \gamma = x_{\bullet\bullet} \end{aligned}$$

Thus $\widehat{p}_{i\bullet} = \frac{x_{i\bullet}}{x_{\bullet\bullet}}$ and $\widehat{p}_{\bullet j} = \frac{x_{\bullet j}}{x_{\bullet\bullet}}$.
Hence $\widehat{p}_{ij} = \frac{x_{i\bullet}x_{\bullet j}}{x_{\bullet\bullet}^2}$ and the expected values are $e_{ij} = x_{\bullet\bullet}\widehat{p}_{ij} = \frac{x_{i\bullet}x_{\bullet j}}{x_{\bullet\bullet}}$.

(ii) The corresponding estimates of p_{ij} for the unrestricted case are $\widehat{\widehat{p}}_{ij} = \frac{x_{ij}}{x_{\bullet\bullet}}$ which are obtained by maximizing

$$H = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}\log p_{ij} + \lambda_1(1 - \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij}).$$

(b) Log likelihood ratio

$$\begin{aligned} G^2 &= -2\log(LR) \\ &= -2\sum_{ij} x_{ij}\log\frac{\widehat{p}_{ij}}{\widehat{\widehat{p}}_{ij}} \\ &= 2\sum_{ij} x_{ij}\log\frac{x_{ij}/x_{\bullet\bullet}}{x_{i\bullet}x_{\bullet j}/x_{\bullet\bullet}^2} \\ &= 2\sum_{ij} x_{ij}\log\frac{x_{ij}}{e_{ij}} \end{aligned}$$

That is, $G^2 = -2\log LR$ is of the form $2\sum o\log\frac{o}{e}$. Under H_0 $G^2 \sim \chi^2_{(r-1)(c-1)}$

6. Fitted values under row/column independence from Question 5:

26.58	30.33	15.09	72
489.42	558.67	277.91	1326
516	589	293	1398

$G^2 = 2\sum o \log \frac{o}{e} = 2(19 \log \frac{19}{26.58} + \cdots + 269 \log \frac{269}{277.91}) = 2(3.66) = 7.32.$

Degrees of freedom for $-2\log LR$ test: $(6-1) - \{(3-1)+(2-1)\} = 2$.

Since $\chi^2_2(5\%) = 5.99$ there is evidence of an association between the rows and columns, i.e., between carries/non-carriers of *Streptococcus pyogenes* and size of tonsils.

7. The (log) likelihood ratio statistic is obtained by summing $o_s \log \frac{o_s}{e_s}$ over all subscripts s:

$$\begin{aligned} G^2 &= 2\sum_s o_s \log \frac{o_s}{e_s} \\ &= 2\sum_s e_s \left(1+\frac{d_s}{e_s}\right) \log\left(1+\frac{d_s}{e_s}\right) \qquad \text{where } d_s = o_s - e_s \\ &\approx 2\sum_s \left(d_s + \frac{d_s^2}{2e_s}\right) \end{aligned}$$

since d_s is small relative to e_s (for reasonable models). But $\sum_s d_s = \sum_s o_s - \sum_s e_s = 0$, thus we have

$$G^2 \approx 2\sum_s \frac{d_s^2}{2e_s} = \sum_s \frac{(o_s-e_s)^2}{e_s} = X^2.$$

$X^2 = \sum \frac{(o-e)^2}{e} = \frac{(19-26.58)^2}{26.58} + \cdots + \frac{(269-277.91)^2}{277.91} = 7.88$

8. The likelihood is given by $L(\theta_1,\theta_2) = \prod_{i=1}^n \left(\frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}\right)$, where $\lambda_i = \theta_1 + \theta_2 x_i$. Thus, the log likelihood is given by

$$l(\theta_1,\theta_2) = -\sum_{i=1}^n (\theta_1+\theta_2 x_i) + \sum_{i=1}^n Y_i \log(\theta_1+\theta_2 x_i)$$

Method of score iterative step is given by

$$\theta_{r+1} = \theta_r - [-I_{\theta_r}]^{-1} U(\theta_r) \quad (r = 0,1,\ldots)$$

where

$$U(\theta) = \begin{pmatrix} \dfrac{\partial l}{\partial \theta_1} \\ \\ \dfrac{\partial l}{\partial \theta_2} \end{pmatrix} = \begin{pmatrix} -n + \displaystyle\sum_{i=1}^n \frac{Y_i}{\theta_1+\theta_2 x_i} \\ \\ -\displaystyle\sum_{i=1}^n x_i + \sum_{i=1}^n \frac{x_i Y_i}{\theta_1+\theta_2 x_i} \end{pmatrix}$$

and

$$I_\theta = -\mathrm{E}\begin{pmatrix} \dfrac{\partial^2 l}{\partial \theta_1^2} & \dfrac{\partial^2 l}{\partial \theta_1 \partial \theta_2} \\ \\ \dfrac{\partial^2 l}{\partial \theta_1 \partial \theta_2} & \dfrac{\partial^2 l}{\partial \theta_2^2} \end{pmatrix} = -\begin{pmatrix} -\displaystyle\sum_{i=1}^n \frac{\mathrm{E}(Y_i)}{(\theta_1+\theta_2 x_i)^2} & -\displaystyle\sum_{i=1}^n \frac{x_i \mathrm{E}(Y_i)}{(\theta_1+\theta_2 x_i)^2} \\ \\ -\displaystyle\sum_{i=1}^n \frac{x_i \mathrm{E}(Y_i)}{(\theta_1+\theta_2 x_i)^2} & -\displaystyle\sum_{i=1}^n \frac{x_i^2 \mathrm{E}(Y_i)}{(\theta_1+\theta_2 x_i)^2} \end{pmatrix}.$$

Note that $\mathrm{E}(Y_i) = \lambda_i = \theta_1 + \theta_2 x_i$.

Thus,

$$I_\theta = \begin{pmatrix} \displaystyle\sum_{i=1}^n \frac{1}{\theta_1+\theta_2 x_i} & \displaystyle\sum_{i=1}^n \frac{x_i}{\theta_1+\theta_2 x_i} \\ \\ \displaystyle\sum_{i=1}^n \frac{x_i}{\theta_1+\theta_2 x_i} & \displaystyle\sum_{i=1}^n \frac{x_i^2}{\theta_1+\theta_2 x_i} \end{pmatrix}.$$

Inverting I_θ is straightforward, and for large samples we have that $\mathrm{var}(\widehat{\theta}) \approx I_\theta^{-1}$, which may be estimated by $I_{\widehat{\theta}}^{-1}$.

9. (a) By hand with $\theta_0 = (0,1)$ ($\sum y = 284$, $\sum y/x = 0.487419375$, $\sum 1/x = 0.067588807$, and $\sum x = 18805$)

$$U(0,1) = \begin{pmatrix} -32 + \sum_{i=1}^{32} \frac{y_i}{x_i} \\ \\ -\sum_{i=1}^{32} x_i + \sum_{i=1}^{32} y_i \end{pmatrix} = \begin{pmatrix} -31.512580625 \\ -18521 \end{pmatrix}$$

and $$I_{\theta_0}^{-1} = \begin{pmatrix} \sum_{i=1}^{32} \frac{1}{x_i} & \sum_{i=1}^{32} 1 \\ \\ \sum_{i=1}^{32} 1 & \sum_{i=1}^{32} x_i \end{pmatrix}^{-1} = \begin{pmatrix} 0.067588807 & 32 \\ 32 & 18805 \end{pmatrix}^{-1} = \begin{pmatrix} 76.1312810 & -0.1295507 \\ -0.1295507 & 0.0002736306 \end{pmatrix}.$$

This gives θ_1 as (0.315, 0.0146).

(b) In MAPLE

```
x:=[551.,651,832,375,715,868,271,630,491,372,645,441,895,458,642,492,
     543,842,905,542,522,122,657,170,738,371,735,749,495,716,952,417]:
y:=[6.,4,17,9,14,8,5,7,7,7,6,8,28,4,10,4,
     8,9,23,9,6,1,9,4,9,14,17,10,7,3,9,2]:
U  := proc(theta)
     [-32 + sum(y[i]/(theta[1]+theta[2]*x[i]),i=1..32),
     -sum(x[i],i=1..32) + sum(y[i]*x[i]/(theta[1]+theta[2]*x[i]),i=1..32)];
     end proc:
InfMat := proc(theta)
     [[sum(1/(theta[1]+theta[2]*x[i]),i=1..32),
     sum(x[i]/(theta[1]+theta[2]*x[i]),i=1..32)],
     [sum(x[i]/(theta[1]+theta[2]*x[i]),i=1..32),
     sum(x[i]^2/(theta[1]+theta[2]*x[i]),i=1..32)]];
     end proc:
with(LinearAlgebra):with(linalg):
evalm([0.,1.]+inverse(InfMat([0.,1.])) &* U([0.,1.]));
                        [0.315461, 0.014565551]
evalm(%+inverse(InfMat(%))&*U(%));
                     [0.3234279155, 0.01455199716]
evalm(%+inverse(InfMat(%))&*U(%));
                     [0.3235092354, 0.01455185878]
evalm(%+inverse(InfMat(%))&*U(%));
                     [0.3235100231, 0.01455185744]
evalm(%+inverse(InfMat(%))&*U(%));
                     [0.3235100478, 0.01455185740]
evalm(%+inverse(InfMat(%))&*U(%));
                     [0.3235100478, 0.01455185740]
```

Thus the ML estimates of θ_1 and θ_2 are 0.32351 and 0.014552.

(c)

```
Inv := inverse(InfMat(%));
                     [    1.236830519    -0.002087480274]
           Inv :=    [                                  ]
                     [-0.002087480274  0.000004326042336]
sqrt(Inv[1,1]);sqrt(Inv[2,2]);
                            1.112128823
                          0.002079914021
```

Thus estimates of standard errors of the ML estimators are 1.11 and .00208.

10. Since $z = g(y)$

$$\begin{aligned} Z &= g(Y) \simeq g(\mu) + (Y-\mu)g'(\mu) + \frac{1}{2}(Y-\mu)^2 g''(\mu) \\ E(Z) &\simeq g(\mu) + \frac{1}{2}\mathrm{var}(Y)g''(\mu) \\ &\simeq g(\mu) \ \text{ if var}(Y) \text{ is small} \end{aligned}$$

(and higher moments of Y are also 'small'). Then

$$\begin{aligned} \mathrm{var}(Z) &\simeq E\left[(Z - g(\mu))^2\right] \\ &\simeq E\left[(Y-\mu)^2 g'(\mu)^2\right] \\ &= \mathrm{var}(Y)g'(\mu)^2 \end{aligned}$$

[Note that if $\mathrm{var}(Y) \sim O(\frac{1}{n})$, the error in this approximation will be $o(\frac{1}{n})$.]

Now for the 1-parameter exponential family, we have (see Question 5 on Problem Sheet 2)

$$\mathrm{var}(Y) = c''(\theta) = \frac{d\mu}{d\theta}$$

Since $g'(\mu) = \frac{d\theta}{d\mu}$,

$$\mathrm{var}(Z) \simeq \mathrm{var}(Y)g'(\mu)^2$$

becomes

$$\begin{aligned} \mathrm{var}(Z) &\simeq \frac{d\mu}{d\theta}\left(\frac{d\theta}{d\mu}\right)^2 \\ &= \frac{d\theta}{d\mu} \\ &= \frac{1}{\mathrm{var}(Y)} \end{aligned}$$

This is a generalization of our result for the logistic transformation where we had

$$\mathrm{var}(\mathrm{logit}p) = \frac{1}{n\pi(1-\pi)} = \frac{1}{\mathrm{var}(R)}.$$

11. As $f(y;\theta)=\exp[-\frac{y}{\mu}-\log\mu]$, it belongs to the family and $\theta=-\frac{1}{\mu}$ with $c(\theta)=\log\mu=\log(-\frac{1}{\theta})=-\log(-\theta)$.
[Note $c'(\theta)=-\frac{-1}{-\theta}=-\frac{1}{\theta}$ and $c''(\theta)=\frac{1}{\theta^2}$, which gives $\mathrm{var}(Y)=\frac{1}{\theta^2}=\mu^2$.]
The variance of Y is μ^2 and $l=-\frac{y}{\mu}-\log(\mu)$.
$\eta=\log(\mu)$ implies $\frac{\partial\eta}{\partial\mu}=\frac{1}{\mu}$ and $\frac{\partial\mu}{\partial\eta}=\mu$. And

$$\frac{\partial\theta}{\partial\eta}=\frac{\partial\theta}{\partial\mu}\frac{\partial\mu}{\partial\eta}=\frac{1}{\mu^2}\mu=\frac{1}{\mu}$$

The elements of the score vector may be determined by

$$\begin{aligned}\frac{\partial l}{\partial\beta_j} &= \frac{\partial l}{\partial\mu}\frac{\partial\mu}{\partial\eta}\frac{\partial\eta}{\partial\beta_j}\\ &= \left(\frac{y}{\mu^2}-\frac{1}{\mu}\right)\mu x_j\\ &= (y-\mu)\frac{1}{\mu}x_j\end{aligned}$$

The elements of the information matrix may be determined by

$$\begin{aligned}-\mathrm{E}\left[\frac{\partial^2 l}{\partial\beta_k\partial\beta_j}\right] &= \mathrm{E}\left[\frac{\partial l}{\partial\beta_k}\frac{\partial l}{\partial\beta_j}\right]\\ &= \mathrm{E}\left[(Y-\mu)\frac{1}{\mu}x_k\cdot(Y-\mu)\frac{1}{\mu}x_j\right]\\ &= \mathrm{E}\left[(Y-\mu)^2\frac{1}{\mu^2}x_kx_j\right]\\ &= \mathrm{E}(Y-\mu)^2\frac{1}{\mu^2}x_kx_j\\ &= \mathrm{var}(Y)\frac{1}{\mu^2}x_kx_j\\ &= x_kx_j\mathrm{var}(Y)\left(\frac{\partial\theta}{\partial\eta}\right)^2\end{aligned}$$

The variance of Y is μ^2 and $\frac{\partial\theta}{\partial\eta}=\frac{1}{\mu}$. Thus the expression for the information matrix simplifies to:

$$-\mathrm{E}\left[\frac{\partial^2 l}{\partial\beta_k\partial\beta_j}\right]=x_kx_j$$

[Note

$$\begin{aligned}\frac{\partial^2 l}{\partial\beta_k\partial\beta_j} &= -\frac{y}{\mu^2}\frac{\partial\mu}{\partial\beta_k}x_j\\ &= -\frac{y}{\mu^2}\frac{\partial\mu}{\partial\eta}\frac{\partial\eta}{\partial\beta_k}x_j\\ &= -\frac{y}{\mu^2}\mu x_kx_j\\ &= -\frac{y}{\mu}x_kx_j\\ \mathrm{E}\left[\frac{\partial^2 l}{\partial\beta_k\partial\beta_j}\right] &= -x_kx_j\\ -\mathrm{E}\left[\frac{\partial^2 l}{\partial\beta_k\partial\beta_j}\right] &= x_kx_j.]\end{aligned}$$

The method of scoring algorithm (using the expected information, I) yields

$$I\delta\beta = \mathbf{U}(\beta) = X'\mathbf{e}$$

where $e_i = (y_i - \mu_i)\frac{\partial\eta_i}{\partial\mu_i}$
i.e.

$$\begin{aligned}(X'X)\beta^{(2)} &= (X'X)\beta^{(1)} + X'\mathbf{e} \\ &= X'\left(X\beta^{(1)} + \mathbf{e}\right) \\ &= X'\mathbf{z}^{(1)}\end{aligned}$$

where $z_i = \eta_i^{(1)} + (y_i - \mu_i)\frac{\partial\eta_i}{\partial\mu_i}$.
So an iterated sequence of **unweighted** regressions of the adjusted variable $\mathbf{z}$ on X will converge to the solution of $\mathbf{U}(\beta) = 0$.

12. See page 4 of the document ‘Generalized Linear Mixed Models (GLMM)’ for the R analysis.

13. See pages 5-6 of the document ‘Generalized Linear Mixed Models (GLMM)’ for the R analysis.