

Assignment 2 MATH 523

YUNHEUM DAN SEOL

2018-02-19

A5

Given a Gamma GLM, we deal with response that has a conditional distribution

$$f(y|\alpha, \beta_i) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{(-\beta y)}; & y > 0 \\ 0 & y \leq 0 \end{cases}$$

then we know

$$\mathbb{E}[Y_i] = \frac{\alpha}{\beta} \quad \text{Var}[Y_i] = \frac{\alpha}{\beta^2}$$

and from class we know

$$I(\beta) = \mathbb{X}^\top W \mathbb{X} = [I_{jk}]$$

where

$$I_{jk} = \sum_{i=1}^n (x_{ij} \left(\frac{(\frac{\partial \mu_i}{\partial \eta_i})^2 x_{ik}}{\text{Var}[Y_i]} \right))$$

$$W = \text{diag}[W_1, \dots, W_n]; \quad W_i = \frac{(\frac{\partial \mu_i}{\partial \eta_i})^2}{\text{Var}[Y_i]} = \frac{\beta^2}{\alpha(g'(\mu_i))^2}$$

$$\begin{aligned} I(\beta) &= \begin{bmatrix} 1 & \dots & \dots & 1 \\ x_1 & \dots & \dots & x_n \end{bmatrix} \begin{bmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & W_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} & \frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i \\ \frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i & \frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2 \end{bmatrix} \end{aligned}$$

for an arbitrary link

when the canonical link is used, i.e.

$$g(\mu_i) = \frac{1}{\mu_i}; g'(\mu_i) = -\frac{1}{\mu_i^2} \implies \left(\frac{1}{g'(\mu_i)} \right)^2 = \mu_i^4 = \frac{\alpha^4}{\beta_i^4}$$

Our expected information matrix gets to be simplified as

$$\begin{bmatrix} \alpha^3 \sum_{i=1}^n \frac{1}{\beta_i^2} & \alpha^3 \sum_{i=1}^n \frac{x_i}{\beta_i^2} \\ \alpha^3 \sum_{i=1}^n \frac{x_i}{\beta_i^2} & \alpha^3 \sum_{i=1}^n \frac{x_i^2}{\beta_i^2} \end{bmatrix} = \begin{bmatrix} \alpha \sum_{i=1}^n \mu_i^2 & \alpha \sum_{i=1}^n \mu_i^2 x_i \\ \alpha \sum_{i=1}^n \mu_i^2 x_i & \alpha \sum_{i=1}^n \mu_i^2 x_i^2 \end{bmatrix}$$

##(b) From class, we know

$$\hat{\beta} \sim_{app} N(\beta, I(\beta)^{-1})$$

as

$$n \rightarrow \infty$$

Therefore once we find the variance-covariance matrix $I(\beta)^{-1}$ The asymptotic variances will be its diagonal entries.

$$\begin{aligned}
& \begin{bmatrix} \frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} & \frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i \\ \frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i & \frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2 \end{bmatrix}^{-1} \\
&= \frac{1}{\left(\frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2}\right) \left(\frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2\right) - \left(\frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i\right)^2} \begin{bmatrix} \frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2 & -\frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i \\ -\frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i & \frac{1}{\alpha} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} \end{bmatrix} \\
&= \frac{\alpha^2}{\left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2}\right) \left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2\right) - \left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i\right)^2} \frac{1}{\alpha} \begin{bmatrix} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2 & -\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i \\ -\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i & \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} \end{bmatrix} \\
&= \frac{\alpha}{\left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2}\right) \left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2\right) - \left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i\right)^2} \begin{bmatrix} \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2 & -\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i \\ -\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i & \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} \end{bmatrix}
\end{aligned}$$

Since we have acquired the variance-covariance matrix, thus

$$\begin{aligned}
\mathbb{V}_{\mathcal{D} \setminus [\widehat{\beta}_0]} &= \frac{\alpha \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2}{\left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2}\right) \left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2\right) - \left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i\right)^2} \\
\mathbb{V}_{\mathcal{D} \setminus [\widehat{\beta}_1]} &= \frac{\alpha \sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2}}{\left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2}\right) \left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i^2\right) - \left(\sum_{i=1}^n \frac{\beta_i^2}{g'(\mu_i)^2} x_i\right)^2}
\end{aligned}$$

If you use canonical link,

$$\begin{aligned}
& \begin{bmatrix} \alpha \sum_{i=1}^n \mu_i^2 & \alpha \sum_{i=1}^n \mu_i^2 x_i \\ \alpha \sum_{i=1}^n \mu_i^2 x_i & \alpha \sum_{i=1}^n \mu_i^2 x_i^2 \end{bmatrix}^{-1} = \frac{\alpha}{\left(\alpha \sum_{i=1}^n \mu_i^2\right) \left(\alpha \sum_{i=1}^n \mu_i^2 x_i^2\right) - \left(\alpha \sum_{i=1}^n \mu_i^2 x_i\right)^2} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 x_i^2 & -\sum_{i=1}^n \mu_i^2 x_i \\ -\sum_{i=1}^n \mu_i^2 x_i & \sum_{i=1}^n \mu_i^2 \end{bmatrix} \\
&= \frac{\alpha}{\alpha^2 \left[\left(\sum_{i=1}^n \mu_i^2\right) \left(\sum_{i=1}^n \mu_i^2 x_i^2\right) - \left(\sum_{i=1}^n \mu_i^2 x_i\right)^2\right]} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 x_i^2 & -\sum_{i=1}^n \mu_i^2 x_i \\ -\sum_{i=1}^n \mu_i^2 x_i & \sum_{i=1}^n \mu_i^2 \end{bmatrix} \\
&= \frac{1}{\alpha \left[\left(\sum_{i=1}^n \mu_i^2\right) \left(\sum_{i=1}^n \mu_i^2 x_i^2\right) - \left(\sum_{i=1}^n \mu_i^2 x_i\right)^2\right]} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 x_i^2 & -\sum_{i=1}^n \mu_i^2 x_i \\ -\sum_{i=1}^n \mu_i^2 x_i & \sum_{i=1}^n \mu_i^2 \end{bmatrix}
\end{aligned}$$

and

$$\begin{aligned}
\text{VAR}[\widehat{\beta}_0] &= \frac{\sum_{i=1}^n \mu_i^2 x_i^2}{\alpha \left[\left(\sum_{i=1}^n \mu_i^2\right) \left(\sum_{i=1}^n \mu_i^2 x_i^2\right) - \left(\sum_{i=1}^n \mu_i^2 x_i\right)^2\right]} \\
\text{VAR}[\widehat{\beta}_1] &= \frac{\sum_{i=1}^n \mu_i^2}{\alpha \left[\left(\sum_{i=1}^n \mu_i^2\right) \left(\sum_{i=1}^n \mu_i^2 x_i^2\right) - \left(\sum_{i=1}^n \mu_i^2 x_i\right)^2\right]}
\end{aligned}$$

(c)

It is analogous to the setting of A2, thus we can classify our data into two groups such that

$$x_i = \begin{cases} 1 & \text{if } i = 1, \dots, n_A \\ 0 & \text{if } i = n_A + 1, \dots, n_A + n_B = n \end{cases}$$

From class, we know

$$\widehat{\beta} = (\mathbb{X}^\top W \mathbb{X})^{-1} \mathbb{X}^\top W Y$$

where

$$\begin{aligned}
\mathbb{X}^\top W Y &= \begin{bmatrix} 1 & \dots & \dots & 1 \\ x_1 & \dots & \dots & x_n \end{bmatrix} \begin{bmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & W_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} \alpha \sum_{i=1}^n \mu_i^2 y_i \\ \alpha \sum_{i=1}^n \mu_i^2 x_i y_i \end{bmatrix} \\
&= \alpha \begin{bmatrix} \sum_{i=1}^n \mu_i^2 y_i \\ \sum_{i=1}^n \mu_i^2 x_i y_i \end{bmatrix} \\
&\implies (\mathbb{X}^\top W \mathbb{X})^{-1} \mathbb{X}^\top W Y = \\
&= \frac{1}{\alpha[(\sum_{i=1}^n \mu_i^2)(\sum_{i=1}^n \mu_i^2 x_i^2) - (\sum_{i=1}^n \mu_i^2 x_i)^2]} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 x_i^2 & -\sum_{i=1}^n \mu_i^2 x_i \\ -\sum_{i=1}^n \mu_i^2 x_i & \sum_{i=1}^n \mu_i^2 \end{bmatrix} \alpha \begin{bmatrix} \sum_{i=1}^n \mu_i^2 y_i \\ \sum_{i=1}^n \mu_i^2 x_i y_i \end{bmatrix} \\
&= \frac{\alpha}{\alpha[(\sum_{i=1}^n \mu_i^2)(\sum_{i=1}^n \mu_i^2 x_i^2) - (\sum_{i=1}^n \mu_i^2 x_i)^2]} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 x_i^2 & -\sum_{i=1}^n \mu_i^2 x_i \\ -\sum_{i=1}^n \mu_i^2 x_i & \sum_{i=1}^n \mu_i^2 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 y_i \\ \sum_{i=1}^n \mu_i^2 x_i y_i \end{bmatrix} \\
&= \frac{1}{[(\sum_{i=1}^n \mu_i^2)(\sum_{i=1}^n \mu_i^2 x_i^2) - (\sum_{i=1}^n \mu_i^2 x_i)^2]} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 x_i^2 & -\sum_{i=1}^n \mu_i^2 x_i \\ -\sum_{i=1}^n \mu_i^2 x_i & \sum_{i=1}^n \mu_i^2 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 y_i \\ \sum_{i=1}^n \mu_i^2 x_i y_i \end{bmatrix}
\end{aligned}$$

under our setting, this is equivalent to

$$\begin{aligned}
&\frac{1}{[(\sum_{i=1}^n \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2) - (\sum_{i=1}^{n_A} \mu_i^2)^2]} \begin{bmatrix} \sum_{i=1}^{n_A} \mu_i^2 & -\sum_{i=1}^{n_A} \mu_i^2 \\ -\sum_{i=1}^{n_A} \mu_i^2 & \sum_{i=1}^{n_A} \mu_i^2 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 y_i \\ \sum_{i=1}^n \mu_i^2 x_i y_i \end{bmatrix} \\
&= \frac{1}{[(\sum_{i=1}^{n_A} \mu_i^2 + \sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2) - (\sum_{i=1}^{n_A} \mu_i^2)^2]} \begin{bmatrix} \sum_{i=1}^{n_A} \mu_i^2 & -\sum_{i=1}^{n_A} \mu_i^2 \\ -\sum_{i=1}^{n_A} \mu_i^2 & \sum_{i=1}^{n_A} \mu_i^2 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 y_i \\ \sum_{i=1}^{n_A} \mu_i^2 y_i \end{bmatrix} \\
&= \frac{1}{[(\sum_{i=1}^{n_A} \mu_i^2 + \sum_{i=n_A+1}^{n_A+n_B} \mu_i^2 - \sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2)]} \begin{bmatrix} \sum_{i=1}^{n_A} \mu_i^2 & -\sum_{i=1}^{n_A} \mu_i^2 \\ -\sum_{i=1}^{n_A} \mu_i^2 & \sum_{i=1}^{n_A} \mu_i^2 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 y_i \\ \sum_{i=1}^{n_A} \mu_i^2 y_i \end{bmatrix} \\
&= \frac{1}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2)]} \begin{bmatrix} \sum_{i=1}^{n_A} \mu_i^2 & -\sum_{i=1}^{n_A} \mu_i^2 \\ -\sum_{i=1}^{n_A} \mu_i^2 & \sum_{i=1}^{n_A} \mu_i^2 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n \mu_i^2 y_i \\ \sum_{i=1}^{n_A} \mu_i^2 y_i \end{bmatrix} \\
&= \frac{1}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2)]} \begin{bmatrix} (\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^n \mu_i^2 y_i) - (\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2 y_i) \\ (-\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^n \mu_i^2 y_i) + (\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2 y_i) \end{bmatrix} \\
&= \frac{1}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2)]} \begin{bmatrix} (\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^n \mu_i^2 y_i) - (\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2 y_i) \\ (-\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2 y_i) + (\sum_{i=1}^{n_A} \mu_i^2 + \sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2 y_i) \end{bmatrix} \\
&= \frac{1}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2)]} \begin{bmatrix} (\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^n \mu_i^2 y_i) - (\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2 y_i) \\ (-\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2 y_i) + (\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2 y_i) \end{bmatrix} \\
&= \frac{1}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2)]} \begin{bmatrix} (\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^n \mu_i^2 y_i) - (\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2 y_i) \\ (\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2 y_i) - (\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2 y_i) \end{bmatrix} \\
&= \frac{1}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2)]} \begin{bmatrix} \frac{(\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2 y_i)}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2)]} \\ \frac{(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2 y_i)}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2)]} - \frac{(\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2 y_i)}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2)]} \end{bmatrix} \\
&= \begin{bmatrix} \frac{(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2 y_i)}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)]} \\ \frac{(\sum_{i=1}^{n_A} \mu_i^2 y_i)}{[(\sum_{i=1}^{n_A} \mu_i^2)]} - \frac{(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2 y_i)}{[(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)]} \end{bmatrix}
\end{aligned}$$

from A2 we know that we can estimate group means as below

$$\widehat{\mu}_A = \bar{y}_A = \frac{1}{n_A} \sum_{i=1}^{n_A} y_i$$

$$\widehat{\mu}_B = \bar{y}_B = \frac{1}{n_B} \sum_{i=n_A+1}^{n_A+n_B} y_i$$

so our β can be estimated as

$$\left[\begin{array}{c} \frac{(\sum_{i=n_A+1}^{n_A+n_B} \bar{y}_B^2 y_i)}{[(\sum_{i=n_A+1}^{n_A+n_B} \bar{y}_B^2)]} \\ \frac{(\sum_{i=1}^{n_A} \bar{y}_A^2 y_i)}{[(\sum_{i=1}^{n_A} \bar{y}_A^2)]} - \frac{(\sum_{i=n_A+1}^{n_A+n_B} \bar{y}_B^2 y_i)}{[(\sum_{i=n_A+1}^{n_A+n_B} \bar{y}_B^2)]} \end{array} \right] = \left[\begin{array}{c} (\sum_{i=n_A+1}^{n_A+n_B} y_i) \\ (\sum_{i=1}^{n_A} y_i) - (\sum_{i=n_A+1}^{n_A+n_B} y_i) \end{array} \right]$$

And they will have variance

$$\text{VAR}[\widehat{\beta}_0] = \frac{\sum_{i=1}^{n_A} \mu_i^2}{\alpha[(\sum_{i=1}^n \mu_i^2)(\sum_{i=1}^{n_A} \mu_i^2) - (\sum_{i=1}^{n_A+n_B} \mu_i^2)]} = \frac{\sum_{i=1}^{n_A} \mu_i^2}{\alpha(\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)} = \frac{1}{\alpha(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)}$$

$$\text{VAR}[\widehat{\beta}_1] = \frac{\sum_{i=1}^n \mu_i^2}{\alpha[(\sum_{i=1}^n \mu_i^2)(\sum_{i=1}^n \mu_i^2 x_i^2) - (\sum_{i=1}^n \mu_i^2 x_i)^2]} = \frac{\sum_{i=1}^n \mu_i^2}{\alpha(\sum_{i=1}^{n_A} \mu_i^2)(\sum_{i=n_A+1}^{n_A+n_B} \mu_i^2)}$$

##(d) if you use the log link

$$g(\mu_i) = \ln(\mu_i) \implies g'(\mu_i) = \frac{1}{\mu_i} \implies \frac{1}{g'(\mu_i)} = \mu_i$$

$$W_i = \frac{\beta^2}{\alpha(g'(\mu_i))^2} = \frac{\beta^2}{\alpha} * \mu_i^2 = \frac{\beta^2}{\alpha} * \frac{\alpha^2}{\beta^2} = \alpha$$

Thus

$$W = \alpha I$$

where I is the $n \times n$ identity matrix. so our $\widehat{\beta}$ would be

$$\widehat{\beta} = (\mathbb{X}^\top W \mathbb{X})^{-1} \mathbb{X}^\top W Y$$

where

$$(\mathbb{X}^\top W \mathbb{X})^{-1} = [\alpha(\mathbb{X}^\top I \mathbb{X})]^{-1} = [\alpha(\mathbb{X}^\top \mathbb{X})]^{-1} = \frac{1}{\alpha} \begin{bmatrix} n & n_A \\ n_A & n_A \end{bmatrix}^{-1} = \frac{1}{\alpha} \begin{bmatrix} n_A & -n_A \\ -n_A & n \end{bmatrix}$$

and

$$\mathbb{X}^\top W Y = \alpha \mathbb{X}^\top Y = \alpha \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^{n_A} y_i \end{bmatrix}$$

then

$$\widehat{\beta} = \begin{bmatrix} n_A & -n_A \\ -n_A & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^{n_A} y_i \end{bmatrix} = \begin{bmatrix} n_A \sum_{i=1}^n y_i - n_A \sum_{i=1}^{n_A} y_i \\ -n_A \sum_{i=1}^n y_i + n \sum_{i=1}^{n_A} y_i \end{bmatrix} = \begin{bmatrix} n_A \sum_{i=n_A+1}^n y_i \\ n_B \sum_{i=1}^{n_A} y_i - n_A \sum_{i=n_A+1}^n y_i \end{bmatrix}$$

##(e) Deviance for

$$Y_i \sim \Gamma(\alpha, \beta)$$

with Y_i independent We know the formula for deviance for GLM of data distributed according to a pdf in exponential dispersion family(without weight) is

$$D(y, \bar{\mu}) = 2 \sum_{i=1}^n \{y_i \tilde{\theta}_i - y_i \widehat{\theta}_i - b(\tilde{\theta}_i) + b(\widehat{\theta}_i)\}$$

where

$$\theta = -\frac{1}{\mu}, \quad b(\theta) = \ln(\mu)$$

At saturated model $\tilde{\theta}_i = -\frac{1}{y_i}$ whereas in the model with p parameters we have $\hat{\theta}_i = -\frac{1}{\hat{\mu}_i}$ so the deviance would be

$$D(y, \hat{\mu}) = 2 \sum_{i=1}^n \left\{ -\left(y_i\right) \frac{1}{y_i} + \frac{y_i}{\hat{\mu}_i} - \ln\left(\frac{1}{y_i}\right) + \ln\left(\frac{1}{\hat{\mu}_i}\right) \right\} = 2 \sum_{i=1}^n \left\{ \frac{y_i}{\hat{\mu}_i} - 1 + \ln\left(\frac{y_i}{\hat{\mu}_i}\right) \right\}$$

##(f) Pearson residual:

$$r_{pi} = \frac{r_i}{\sqrt{v(\mu_i)}} = \frac{y_i - \hat{\mu}_i}{\sqrt{v(\mu_i)}}$$

. For a Gamma GLM, we have

$$\mu_i = \frac{\alpha}{\beta_i}, \text{VAR}[Y_i] = \frac{\alpha}{\beta_i^2}, a(\phi) = \frac{1}{\alpha}$$

Since

$$\text{VAR}[Y_i] = a(\phi)v(\mu_i) = \frac{1}{\alpha}v(\mu_i) \implies v(\mu_i) = \alpha \text{VAR}[Y_i] = \frac{\alpha^2}{\beta_i^2} = \mu_i^2$$

So our pearson residual would be of the form

$$\frac{y_i - \hat{\mu}_i}{\sqrt{v(\mu_i)}} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i^2}} = \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}$$

Anscombe residual:

$$r_{Ai} = \frac{A(y_i) - A(\hat{\mu}_i)}{\sqrt{\{A'(\mu_i)\}^2 v(\mu_i)}}$$

where

$$A(y) = \int_0^y \frac{1}{(v(\mu))^{\frac{1}{3}}} d\mu = \int_0^y \frac{1}{(\mu^2)^{\frac{1}{3}}} d\mu = \int_0^y \mu^{-\frac{2}{3}} d\mu = 3\mu^{\frac{1}{3}}|_0^y = 3y^{\frac{1}{3}}$$

So our residual would be

$$\begin{aligned} \frac{A(y_i) - A(\hat{\mu}_i)}{\sqrt{\{A'(\mu_i)\}^2 v(\mu_i)}} &= \frac{3y_i^{\frac{1}{3}} - 3\hat{\mu}_i^{\frac{1}{3}}}{\sqrt{\left(\frac{d}{d\mu_i} 3\mu_i^{\frac{1}{3}}\right)^2 \hat{\mu}_i^2}} = \frac{3y_i^{\frac{1}{3}} - 3\hat{\mu}_i^{\frac{1}{3}}}{\sqrt{(\mu_i^{-\frac{2}{3}})^2 \hat{\mu}_i^2}} = \frac{3y_i^{\frac{1}{3}} - 3\hat{\mu}_i^{\frac{1}{3}}}{\sqrt{\hat{\mu}_i^{\frac{2}{3}}}} \\ &= \frac{3y_i^{\frac{1}{3}} - 3\hat{\mu}_i^{\frac{1}{3}}}{\hat{\mu}_i^{\frac{1}{3}}} \end{aligned}$$

Deviance Residual

$$r_{di} = \sqrt{2\omega_i(y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i))} = \sqrt{2\left(\frac{y_i}{\hat{\mu}_i} - 1 + \ln\left(\frac{y_i}{\hat{\mu}_i}\right)\right) \text{sgn}(y_i - \hat{\mu}_i)}$$

#A6 (canonical link assumed) Since for normal GLM,

$$\theta = \mu; \quad b(\theta) = \frac{\mu^2}{2}$$

The deviance for the normal GLM would be

$$\begin{aligned} D(y, \hat{\mu}) &= 2 \sum_{i=1}^n \left(y_i(y_i - \hat{\mu}_i) - \frac{y_i^2}{2} + \frac{\hat{\mu}_i^2}{2} \right) = \sum_{i=1}^n (y_i^2 - 2y_i\hat{\mu}_i + \hat{\mu}_i^2) \\ &= \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 \end{aligned}$$

So if we take the difference of deviances, it would be of the form

$$D(y, \tilde{\mu}) - D(y, \hat{\mu}) = \sum_{i=1}^n (y_i - \tilde{\mu}_i)^2 - \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 = SS_{res}(\beta_0) - SS_{res}(\beta_1)$$

This is analogous to the difference between residual sum-of-squares used for model selection that we used in Classical Linear Regression. It helps us to choose a better model by using F test, using the statistics

$$\frac{[SS_{res}(\beta_0) - SS_{res}(\beta_1)]/(p_1 - p_0)}{SS_{res}/(n - p_1)} \sim Fisher - F_{p_1 - p_0, n - p_1}$$

#A7

```
library(glmbb)
```

```
## Warning: package 'glmmbb' was built under R version 3.3.2
```

```
data(crabs)
```

```
##the response is satell
```

```
head(crabs)
```

```
##   color spine width satell weight y
## 1 medium  bad  28.3     8  3050  1
## 2  dark  bad  22.5     0  1550  0
## 3 light good  26.0     9  2300  1
## 4  dark  bad  24.8     0  2100  0
## 5  dark  bad  26.0     4  2600  1
## 6 medium  bad  23.8     0  2100  0
```

```
summary(crabs)
```

```
##      color      spine      width      satell      weight
## dark   :44   bad   :121   Min.   :21.0   Min.    : 0.000   Min.   :1200
## darker:22   good   : 37   1st Qu.:24.9   1st Qu.: 0.000   1st Qu.:2000
## light  :12   middle: 15   Median :26.1   Median : 2.000   Median :2350
## medium:95                      Mean   :26.3   Mean    : 2.919   Mean    :2437
##                      3rd Qu.:27.7   3rd Qu.: 5.000   3rd Qu.:2850
##                      Max.    :33.5   Max.    :15.000   Max.    :5200
##
##      y
## Min.   :0.0000
## 1st Qu.:0.0000
## Median :1.0000
## Mean    :0.6416
## 3rd Qu.:1.0000
## Max.    :1.0000
```

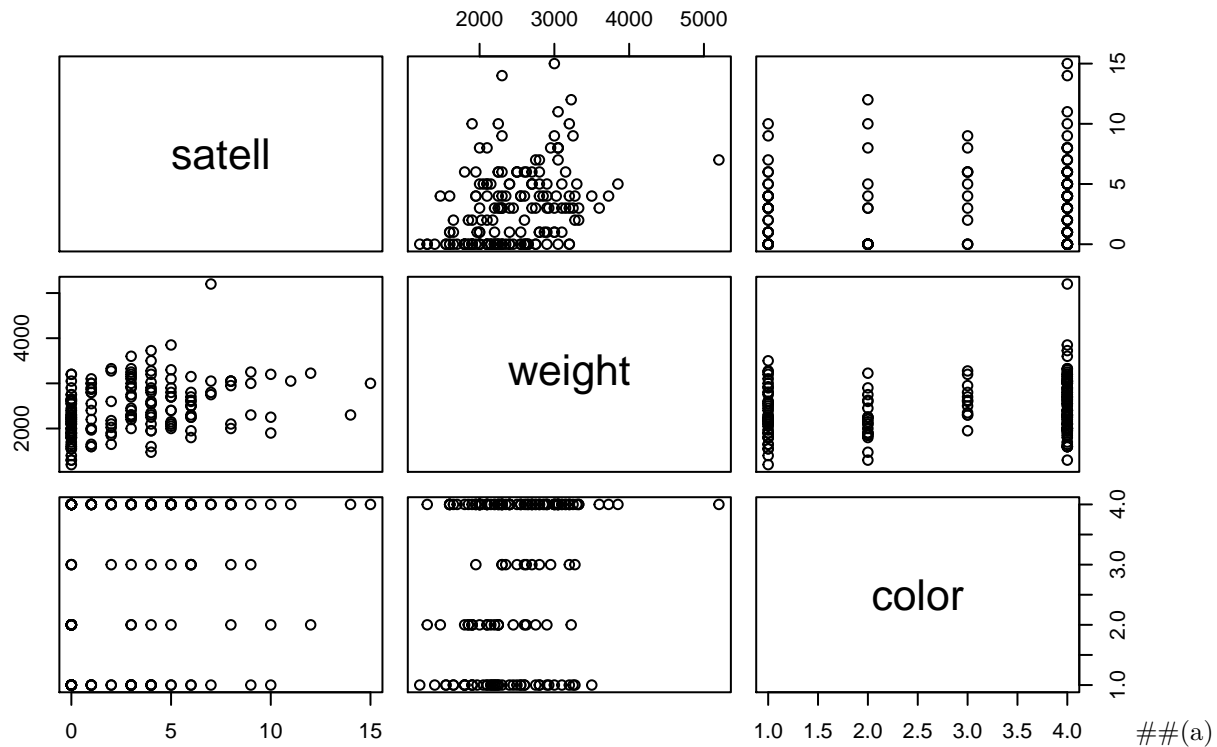
```
dim(crabs)
```

```
## [1] 173   6
```

```
#figure out what the data consists of
```

```
attach(crabs)
```

```
pairs(cbind(satell,weight,color))
```



```
#linear model for comparison
```

```
lmfit_1 <- lm(satell~(weight+color)^2)
```

```
summary(lmfit_1)
```

```
##
```

```
## Call:
```

```
## lm(formula = satell ~ (weight + color)^2)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -4.5248 -2.1020 -0.6264  1.4040 11.1470
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)    -2.685e+00  1.943e+00  -1.382  0.1688
```

```
## weight          2.137e-03  8.225e-04   2.598  0.0102 *
```

```
## colordarker     2.881e-02  3.704e+00   0.008  0.9938
```

```
## colorlight      8.961e+00  6.397e+00   1.401  0.1631
```

```
## colormedium     1.266e+00  2.342e+00   0.540  0.5897
```

```
## weight:colordarker 2.634e-05  1.642e-03   0.016  0.9872
```

```
## weight:colorlight -2.971e-03  2.438e-03  -1.218  0.2249
```

```
## weight:colormedium -2.789e-04  9.633e-04  -0.290  0.7725
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 2.952 on 165 degrees of freedom
```

```
## Multiple R-squared:  0.1567, Adjusted R-squared:  0.121
```

```
## F-statistic: 4.382 on 7 and 165 DF,  p-value: 0.0001748
```

```
#The responses are in discrete integers, so let's try poisson
```

```
mod1 <- glm(satell~(weight+factor(color))^2, family = poisson(link="log"))
```

```
summary(mod1)
```

```
##
## Call:
## glm(formula = satell ~ (weight + factor(color))^2, family = poisson(link = "log"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.9412  -1.8457  -0.6278   0.8346   4.7689
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.463e+00  4.851e-01  -3.016  0.00256 **
## weight           9.283e-04  1.839e-04   5.048 4.47e-07 ***
## factor(color)darker -5.214e-02  9.050e-01  -0.058  0.95406
## factor(color)light  3.408e+00  1.127e+00   3.023  0.00250 **
## factor(color)medium  1.386e+00  5.345e-01   2.594  0.00949 **
## weight:factor(color)darker  5.232e-05  3.645e-04   0.144  0.88587
## weight:factor(color)light -1.134e-03  4.287e-04  -2.645  0.00818 **
## weight:factor(color)medium -4.466e-04  2.002e-04  -2.231  0.02570 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 541.08  on 165  degrees of freedom
## AIC: 912.38
##
## Number of Fisher Scoring iterations: 6
```

(b)

```
# Fisher information at the last iteration
```

```
I <- (t(model.matrix(mod1))%*%diag(mod1$weights))%*%model.matrix(mod1)
```

```
# Inverse Fisher information matrix (i.e. as. covariance matrix of the MLEs)
```

```
I.inv <- solve(I)
dim(I.inv)
```

```
## [1] 8 8
```

```
# standard errors of the MLEs
```

```
sd <- sqrt(diag(I.inv))
beta <- coefficients(mod1)
```



```
p.val <- pchisq((beta/sd)^2,df=1,lower.tail=FALSE)
round(p.val,7)
```

```
##              (Intercept)              weight
##              0.0025591              0.0000004
## factor(color)darker      factor(color)light
##              0.9540605              0.0025003
## factor(color)medium weight:factor(color)darker
##              0.0094905              0.8858701
## weight:factor(color)light weight:factor(color)medium
##              0.0081774              0.0256999
```

Against the null hypothesis that the interaction of colorlvl and weight is not significant, we reject the null at $\alpha = 0.05$ and conclude that we have sufficient evidence to conclude that the interaction factor is significant, except for the color level “darker.” It seems that the interaction between weight and color is significant.

(c)

```
# Checking the fit of the model using the deviance statistic
# Fitting other models
modb <- glm(satell~weight+factor(color), family=poisson(link=log))
modc <- glm(satell~weight, family=poisson(link=log))
modd <- glm(satell~factor(color), family=poisson(link=log))
anova(mod1)
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: satell
##
## Terms added sequentially (first to last)
##
##
##              Df Deviance Resid. Df Resid. Dev
## NULL              172      632.79
## weight             1   71.925      171      560.87
## factor(color)       3    9.061      168      551.80
## weight:factor(color) 3   10.724      165      541.08
```

```
deviance(modb)
```

```
## [1] 551.8049
```

```
deviance(modc)
```

```
## [1] 560.8664
```

```
# The saturated model v. our model
pchisq(deviance(mod1), df=169, lower.tail=FALSE)
```

```
## [1] 1.580598e-40
```

```
# The model with interaction v. the model without interaction
pchisq((deviance(modb)-deviance(mod1)),df=1,lower.tail=FALSE)
```

```
## [1] 0.00105761
```

```
# The saturated model v. the model without interaction
pchisq(deviance(modb), df=170, lower.tail=FALSE)
```

```
## [1] 6.868553e-42
```

```
# The model with one predictor only
```

```
## weight only v. proposed model
```

```
pchisq((deviance(modc)-deviance(mod1)), df=2, lower.tail = FALSE)
```

```
## [1] 5.054307e-05
```

```
## color level only v. proposed model
```

```
pchisq((deviance(modd)-deviance(mod1)), df=2, lower.tail = FALSE)
```

```
## [1] 1.664875e-15
```

Under our null hypothesis that the nested model is an adequate simplification of the full model with more predictors, we get different results:

Model without the interaction v. model with interaction

We fail to reject the null hypothesis at $\alpha = 0.05$, and we conclude that we have failed to have sufficient evidence that the model with interaction explains the variation in response better. ### Other cases We can reject our null hypothesis at $\alpha = 0.05$, and we conclude that the saturated model might be a better fit.

(d)

We can try the “darker” factor level to the “dark” level (the base model) –

```
color_mod <- color
for(i in 1:length(color_mod)){
  if(color_mod[i] == "darker")
    color_mod[i] = "dark"
}
mod2 <- glm(satell~(weight+factor(color_mod))^2, family=poisson(link=log))
summary(mod2)
```

```
##
```

```
## Call:
```

```
## glm(formula = satell ~ (weight + factor(color_mod))^2, family = poisson(link = log))
```

```
##
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.9412  -1.8470  -0.6278   0.8346   4.7574
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.4451144   0.4013956  -3.600 0.000318 ***
## weight           0.0009299   0.0001559   5.963 2.48e-09 ***
## factor(color_mod)light    3.3893600   1.0936607   3.099 0.001941 **
## factor(color_mod)medium   1.3682995   0.4598657   2.975 0.002926 **
## weight:factor(color_mod)light -0.0011354   0.0004175  -2.720 0.006536 **
## weight:factor(color_mod)medium -0.0004482   0.0001749  -2.563 0.010380 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 632.79 on 172 degrees of freedom
## Residual deviance: 541.27 on 167 degrees of freedom
## AIC: 908.57
##
## Number of Fisher Scoring iterations: 6
```

From the visual inspection of the table we can conclude that all of the predictors are significant. ##(e)

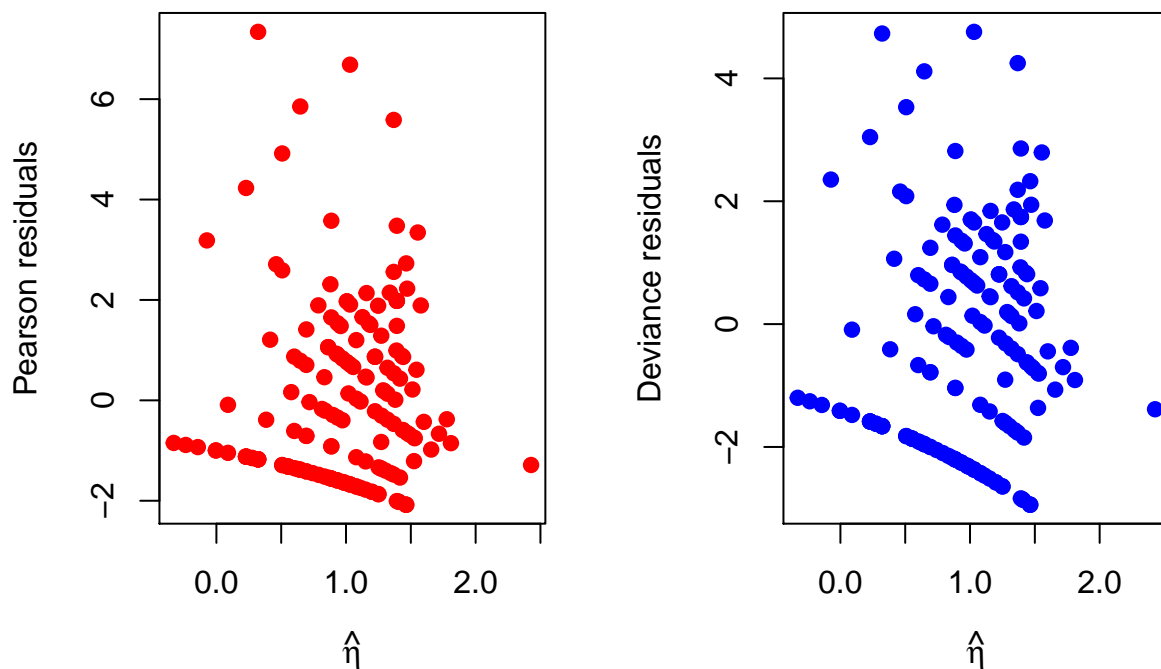
Next, compute the Pearson residuals (these are NOT the standardized residuals)

```
r.pearson <- residuals(mod2,"pearson")
```

Next, compute the deviance residuals

```
r.deviance <- residuals(mod2,"deviance")
par(mfrow=c(1,2))
```

```
plot(r.pearson~predict(mod2,type='link'), xlab=expression(hat(eta)), ylab="Pearson residuals", pch=19, col="red", main="Pearson residuals")
plot(r.deviance~predict(mod2,type='link'),xlab=expression(hat(eta)), ylab="Deviance residuals",pch=19, col="blue", main="Deviance residuals")
```



The saturated model v. the model without interaction
deviance(mod2)

```
## [1] 541.2675
```

```
deviance(mod1)
```

```
## [1] 541.081
```

```
pchisq(deviance(mod2), df=167, lower.tail=FALSE)
```

```
## [1] 4.547482e-41
```

```
pchisq(deviance(mod2)-deviance(mod1), df=1, lower.tail=FALSE)
```

```
## [1] 0.6658662
```

The residual plot gets closer to a null plot if the approximate normality is assumed for the response. Even if the other pieces of evidence suggest that the model is a good fit, it is really hard to make a correct judgment based on the residual plot and deviance only since Poisson-distributed random variables do not necessarily behave normally when the mean is not large. That explains some evident “pattern” in our residual plot.

For the deviance test, against the null hypothesis that the simpler model is a better model compared to the saturated model, we reject the null at $\alpha = 0.05$ and claim to have gathered sufficient evidence against the null and say the saturated model might fit the model better. However, among the test between models mod1 and mod2, mod2 can explain the variation in response certainly better.