GLMs – Poisson regression revisited

Suppose that we have count data $y_1, \dots, y_n$ and adopt a model of the form $Y_i \sim Po(\lambda_i)$, $i = 1, \dots, n$. [We may also have covariate info of the form $(x_{i1}, x_{i2}, \dots, x_{ip})^T$, $i = 1, \dots, n$.]

- The probability mass f^ is $f(y) = \dfrac{\lambda^y e^{-\lambda}}{y!} = \underbrace{\exp\{y \log\lambda - \lambda - \log y!\}}_{\text{Exp. family form}}$.

- Identify the natural link as $g(\lambda) = g(\underset{E(Y)}{\underset{\uparrow}{\mu}}) = \log\lambda$.

- Hence our GLM is $Y_i \mid \underline{x}_i \sim Po(\lambda_i)$, $\lambda_i = \underbrace{e^{\underline{x}_i^T \underline{\beta}}}_{\text{Response f}^\wedge\ h(\cdot)}$

where $\underline{x}_i^T \underline{\beta} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$.

- Code: glm($\underset{\text{Response}}{\_\_}$ ~ $\underset{\text{covariate info}}{\_\_}$, family=poisson(link=log), data=__)

Dealing with rate data (offsets)

Example

Sppse we want to model no of violent crimes in schools or colleges. Given:

y_i : no of violent crimes in a given school or college

x_i : 0 for school, 1 for college

n_i : no students enrolled $\qquad i = 1, \dots, n$.

- To allow / account for the differing no's of students enrolled, we are interested in $\frac{\lambda_i}{n_i}$ as the number per students enrolled, to allow comparison across schools / colleges of different sizes.

- Hence use

$$\log \frac{\lambda_i}{n_i} = \beta_0 + \beta_1 x_i$$

$$\Rightarrow \log \lambda_i - \log n_i = \beta_0 + \beta_1 x_i$$

$$\Rightarrow \log \lambda_i = \beta_0 + \beta_1 x_i + \log n_i$$

We are still modelling λ_i, but adding an <u>offset</u> to adjust for the different enrollments.

- Code: glm(___ ~ ___ , family = poisson(link = log),
  offset = log(n), data = ___)

  column in data frame.

<u>A note on residuals</u>

- Linear regression: $\hat{\varepsilon}_i = \frac{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{\hat{\sigma}}$ [standardised]

- Adopting a similar principle for Poisson regn, leads to

$$\hat{\varepsilon}_i = \frac{y_i - \hat{\lambda}_i}{\sqrt{\hat{\lambda}_i}}$$ since $E(Y_i) = Var(Y_i) = \lambda_i$