1 Poisson Regression

1.1 The Poisson Regression Model

- When the response variable Y is **Poisson**, ordinary linear regression models (that assume normality of the errors) aren't appropriate.

  Instead, we use the *Poisson regression model*.

- Recall that a ***generalized linear model*** is a statistical model for which:

  1. Y doesn't necessarily have to be normally distributed.
  2. Some function $g(\mu)$ of $\mu = \mathrm{E}(Y)$ (called the ***link function***) is a linear function of X, i.e.

  $$g(\mu) \;=\; \beta_0 + \beta_1 X\,.$$

- Recall also that a ***Poisson(μ)*** random variable takes values $\boldsymbol{Y = 0, 1, 2, \ldots}$ with probabilities given by

  $$p(Y) \;=\; \frac{\mu^Y e^{-\mu}}{Y!}\,.$$

  They're used to model **random counts**.

- Suppose $\boldsymbol{Y_1, Y_2, \ldots, Y_n}$ are independent **Poisson($\boldsymbol{\mu_i}$)** random variables. Note that $\boldsymbol{\mu_i}$ is allowed to **differ** from **one individual** to the **next**.

  The **mean response** (which we'll call the ***mean count***) for the ith individual is

  $$\mathrm{E}(Y_i) \;=\; \mu_i$$

  and the variance is also

  $$\mathrm{V}(Y_i) \;=\; \mu_i$$

  Thus the **variance** and **mean** are **equal**.

- In *Poisson regression*, we'll model the mean count as a function of a predictor variable X.

- The ***Poisson regression model*** is defined as follows.

  Poisson Regression Model: Suppose $Y_1, Y_2, \ldots, Y_n$ are independent

Poisson(μ_i) random variables, so

$$\text{E}(Y_i) \;=\; \mu_i\,.$$

The ***Poisson regression model*** is

$$\log(\mu_i) \;=\; \beta_0 + \beta_1 X_i\,, \tag{1}$$

which (by exponentiating both sides) can be written as

$$\mu_i \;=\; e^{\beta_0+\beta_1 X_i}\,, \tag{2}$$

where for $i = 1, 2, \ldots, n$,

- ▷ X_i is the value of the predictor variable X for the ith individual
- ▷ μ_i is the mean count for the ith individual

and β_0 and β_1 are parameters of the model.

- The function

  $$g(\mu_i) \;=\; \log(\mu_i)$$

  is the ***link function*** which "links" the mean count $\text{E}(Y) = \mu$ to the predictor X via the linear function $\beta_0 + \beta_1 X$ in (1).

  More precisely, it's called the ***log link***, and is the most widely use link function in Poisson regression.

  The Poisson regression model is an example of a so-called ***loglinear model*** (because $\log(\mu)$ is a linear model).

- As a function of X, the **mean count function**

  $$\mu(X) \;=\; e^{\beta_0+\beta_1 X}$$

  of the **Poisson regression model** (2) is graphed in Fig. 1. It has the following properties:

  1. μ is constrained to be greater than zero.
  2. If $\beta_1 > 0$, then μ is an increasing function of X.

3. If $\beta_1 < 0$, then μ is decreasing function of X.
4. The parameter β_1 determines how "fast" the graph of μ rises (or falls) as a function of X. A larger value of β_1 results in a "faster" rise in the graph.
5. The parameter β_0 determines the value of μ when $X = 0$.

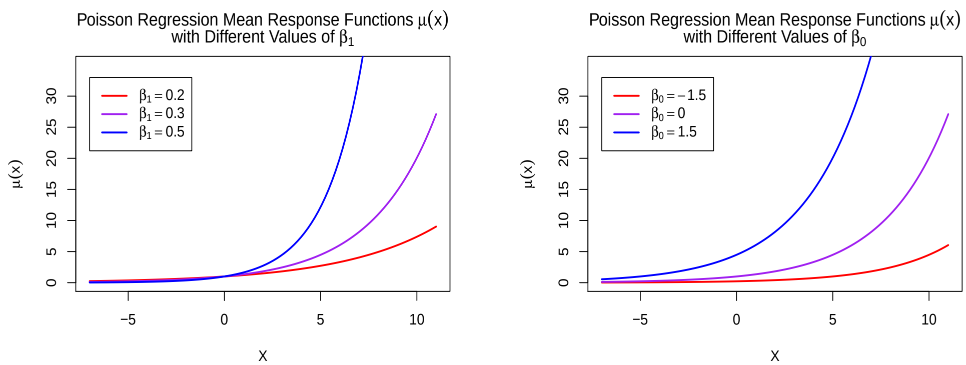

Figure 1

- The parameters β_0 and β_1 are estimated *not* by least squares, but by the ***maximum likelihood*** method.
- Once the **maximum likelihood** estimates $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$ of β_0 and β_1 are obtained, the ***fitted values*** (or ***predicted values***), denoted $\boldsymbol{\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_n}$, are defined as

Fitted (Predicted) Values:

$$\hat{\mu}_i \;=\; e^{b_0 + b_1 X_i} \qquad \text{for } i = 1, 2, \ldots, n$$

The ith **fitted value** $\hat{\mu}_i$ is the **estimated mean count** for individuals whose predictor value is X_i.

1.2 Interpretation of the Estimate b_1 of β_1

- We can examine how the **mean count changes** as $\boldsymbol{X}$ **increases** by **one unit**. The fitted value at some (generic) value of X is

$$\hat{\mu}_1 \;=\; e^{b_0 + b_1 X}$$

and after increasing X by one unit, the fitted value changes to

$$\hat{\mu}_2 = e^{b_0 + b_1(X+1)} .$$

It's easy to see that

$$\hat{\mu}_2 = \hat{\mu}_1 e^{b_1} .$$

so a one-unit increase in X results in an change in the mean count by the **multiplicative factor** e^{b_1}.

For example, if $e^{b_1} = 1.1$ (which would be the case if $b_1 = \log(1.1) = 0.095$), the **mean count** μ would increase by **10%** for each one-unit increase in X.