Poisson Regression

Linear regression is only ideal for positive output.

Poisson regression is a GLM used when the response variable is a count:

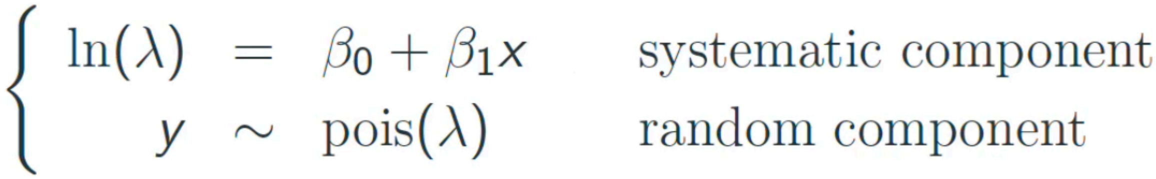
- 0, 1, 2,
- Number of events in a fixed time/space

Examples:

- Number of emails received per day
- Number of accidents at an intersection
- Number of customers entering a store per hour

Assumptions

1. The response (Y) follows a **Poisson distribution**



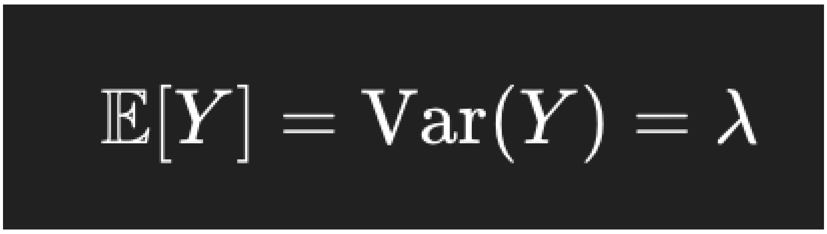
$$\begin{cases} \ln(\lambda) = \beta_0 + \beta_1 x & \text{systematic component} \\ y \sim \text{pois}(\lambda) & \text{random component} \end{cases}$$

Where lamdba is the expected number of occurrences when the value of the explanatory variable is x.
And yes-- the B and B1 are unknown coefficient.

2. Counts are:

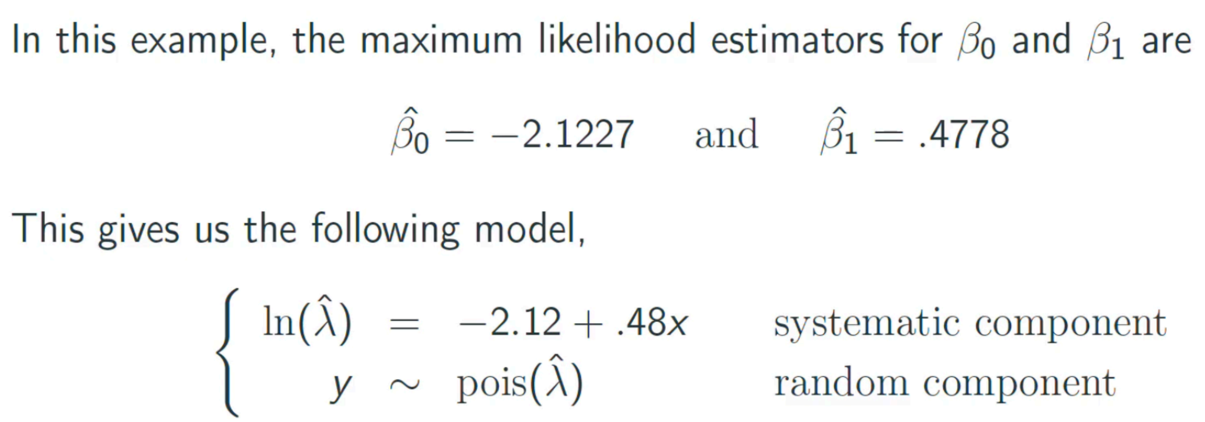
- Non-negative integers
- Independent

3. Mean = Variance



$$\mathbb{E}[Y] = \text{Var}(Y) = \lambda$$

The model is fit using the method of maximum likelihood estimation.
Example:

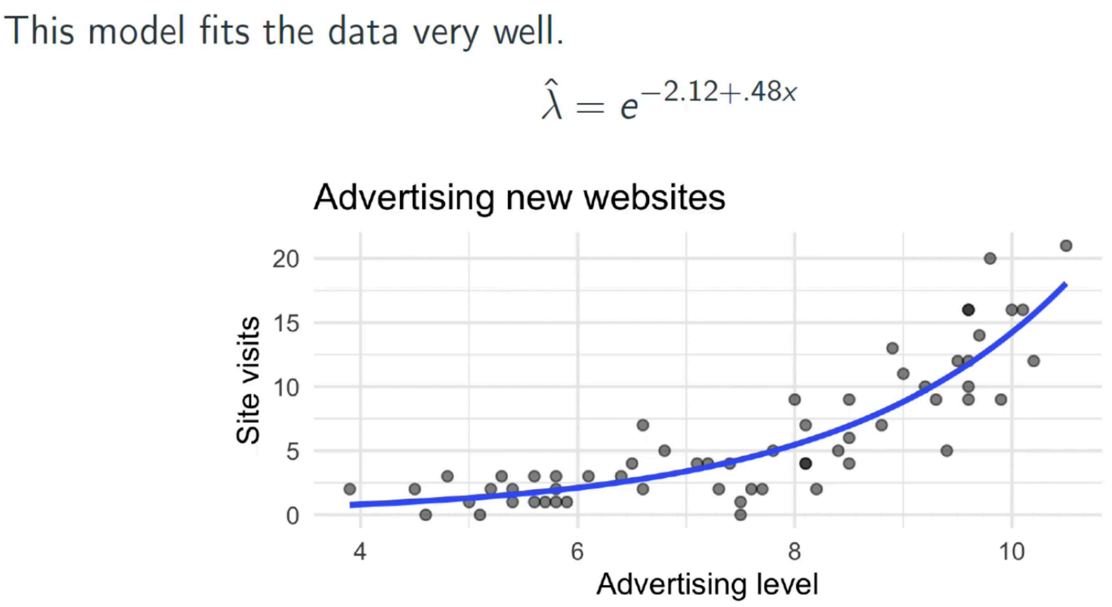


In this example, the maximum likelihood estimators for β_0 and β_1 are

$$\hat{\beta}_0 = -2.1227 \quad \text{and} \quad \hat{\beta}_1 = .4778$$

This gives us the following model,

$$\begin{cases} \ln(\hat{\lambda}) & = & -2.12 + .48x & \text{systematic component} \\ y & \sim & \text{pois}(\hat{\lambda}) & \text{random component} \end{cases}$$



This model fits the data very well.

$$\hat{\lambda} = e^{-2.12+.48x}$$

When Poisson regression is NOT ideal

Overdispersion
If: **Variance > Mean**
Then Poisson is too restrictive.
--> Solutions:

- **Quasi-Poisson**
- **Negative Binomial regression**

Difference between poisson regression and linear regression

-There is no normal distribution in a poisson regression model. At every x-value. the distribution of the response variable is poisson.
-The variance of the response variable is not constant in a poisson regression model. In fact, it is larger for larger value of lambda.
-Raw residuals (the difference between the observed and expected y-value) are not good

measure of model fit sine larger errors are expected for larger lambda. Deviance and Pearson (Standardized) residuals are typically used instead.