

The Generalized Linear Model GzLM- Two examples (Poisson and Binomial)

The General linear model (GLM) applies to a ratio scale response variable Y with normal error ε_{Normal} .
Count data usually violate the assumption of homogeneous residuals.

Ratio scale counts (counts in defined units, ranging from zero upward)

The appropriate error model is Poisson or Negative Binomial $\varepsilon_{Poisson}$ or $\varepsilon_{NegBinomial}$.
So we use the Generalized Linear Model, which allows us to use a better error model.
The GzLM (which includes the GLM as a special case) has three components

1. The structural model consisting of linear predictors.

For the GLM, the linear predictor is the sum of fixed factors and covariates.

The linear predictor ANCOVA example (Brussard) was $\eta = \beta_0 + \beta_{SP}SP + \beta_{Hsl}Hsl + \beta_{SP \cdot Hsl}SP \cdot Hsl$

2. A linkfunction, that links the linear predictor to the response variable.

3. The error model.

For the GLM the model is: $Y = \beta_0 + \beta_x X + \varepsilon_{Normal}$

The error model is: $Y \sim Normal(\beta_0 + \beta_x X, \sigma^2)$

This is read as: Y is normally distributed, given the parameters β_0 , β_x , and σ^2 (the fixed variance)

The distributional assumption, given the parameters, can only be checked after estimating the parameters

Count data usually violate this assumption. The result-- > heterogeneous residuals

So we use a better error model (Generalized Linear Model)

Ratio scale counts (counts in defined units, ranging from zero upward)

We use an error model such as Poisson model $\varepsilon_{Poisson}$ or Negative Binomial error $\varepsilon_{NegBinomial}$

For example

$$\begin{aligned} \text{Count} &= e^\eta + \varepsilon_{Poisson} \\ \eta &= \beta_0 + \beta_{V1}V1 + \dots \end{aligned}$$

In this example the log link is used – the structural model appears in the exponent

1. Death by horsekick. The classic example of Poisson data is the number of deaths by horse kick for each of 16 corps in the Prussian army, from 1875 to 1894. Bortkiewicz (1898 *The Law of Small Numbers*) showed that the horsekick data fit a Poisson distribution.

Corps	Deaths
Guard	16
First	16
Second	12
Third	12

Symbol for response variable _____ and for explanatory variable _____

Write the model

$$Odds = e^{\eta} + \varepsilon_{Poisson}$$

(Fit to 1:1:1:1 assumes Poisson error)

$$\eta = \underline{\hspace{10em}}$$

Likelihood Ratio Test: $\Delta G = 1.147$ $df = 3$ $p = 0.7658$ Therefore, 1:1:1:1 is a good fit to the data.

http://www.mun.ca/biology/schneider/b4605/LNotes/Pt5/Ch17_2.pdf

GzLM with fixed explanatory variables. - 2nd example

Nominal scale counts (units scored Y or N) Use binomial error model $\epsilon_{binomial}$

$$\text{Yes/No} = \text{Odds} \quad \text{Odds} = e^{\eta} + \epsilon_{binomial}$$

$$\eta = \beta_o + \beta_{V1}V1 + \dots$$

2. Example – Cancer in cigarette smokers. Data from Cornfield (1951) who established the mathematical basis for using case-control samples to estimate risk in a population.

Odds of tumor for
Heavy smokers _____

	Lung Tumors		Total
	Present	Absent	
heavy smokers	27	99	126
light smokers	8	72	80

Light smokers _____

Symbol for response variable _____ and for explanatory variable _____

Write the model $\text{Odds} = e^{\eta} + \epsilon_{binomial}$
(contingency test not correct. it assumes Poisson error)

$$\eta = \underline{\hspace{10em}}$$

The odds are 2.455 times higher for heavy than light smokers. The 95% confidence limits are 1.05 to 5.1. The lower limit excludes the null hypotheses OR = 1 (Odds the same for light and heavy smokers)

http://www.mun.ca/biology/schneider/b4605/LNotes/Pt5/Ch18_3.pdf