GLM and vocabulary

- **Response**, y
- **Explanatory variables:** $x_1, x_2, \ldots x_p$ or X.

GLM

$$Y_i \sim f()$$

$$\mu_i = E(Y_i)$$

$$g(\mu_i) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

- **Factor**, Categorical/ qualitative x, with I levels / classes
- **Covariate:** Qualtatative x (continiuous or ordinal).

House sparrows

Response modeled with **explanatory variables**.

1. Tarsus length continuous
   - Hatch island factor, 5 levels
   - Sex factor, 2 levels
   - NAO covariate
2. Dispersal, categorical, binary
   - Hatch island factor, 5 levels
   - Sex factor, 2 levels
   - NAO covariate
   - Wing length covariate
3. Number of off-spring categorical, counts
   - Dispersal factor, 2 levels
   - Hatch island factor, 5 levels
   - Body mass covariate

Maximum Likelihood Estimation (MLE)

Likelihood function: Joint probability function for all data seen as function of parameter(s).

MLE: Optimum for parameter(s).

1. Find likelihood function $L(\theta, y)$
2. Find optimum:
   - Find log-likelihood: $l(\theta, y) = \log(L(\theta, y))$
   - Solve $\frac{\partial}{\partial \theta} l(\theta, y) = 0$ for θ

Standardized residuals

Normal model: $E(Y_i) = \mu_i,\ Y_i = N(\mu_i, \sigma^2)$

Poisson model: $E(Y_i) = \theta_i,\ Y_i \sim Po(\theta_i)$

- $r_i = \frac{y_i - \hat{\mu}_i}{\hat{\sigma}}$
- $r_i = \frac{y_i - \hat{\theta}}{\sqrt{\hat{\theta}}}$

If model is correct: Approximately: $r_i \sim N(0, 1)$

Plots for r_i

- qq-plot
- against each explanatory variable
- other potential explanatory variables
- plot r_i vs $\hat{y}_i$ (check assumption of constant variance / homoscedasity)
- plot r_i in order y_i was measured.

Example: Chronically medical conditions

- Women in rural area see GP less then women in urban area.
- Why? Less sick or less accessible?

Example: Chronically medical conditions

- Women in rural area see GP less then women in urban area.
- Why? Less sick or less accessible?

Data

Group 1: No. of chronically conditions for 26 town women with ≤ 3 GP visits.

Group 2: No. of chronically conditions for 23 country women with ≤ 3 GP visits.

Do women in the two groups with the same number of visits have the same need?

Model and hypothesis

y_{jk}: Woman j from group k.

H_0 Same need: $Y_{jk} \sim Po(\theta)$

H_1 Different needs: $Y_{jk} \sim Po(\theta_k)$

Assignment 5

- We recommend binary data with at least one continuous and one nominal/ordinal explanatory variable.
- But any data / model that fit the course is OK.
- Discuss with Ingelin and Xiangping!

Find a friend and start today!