What:

- **Modeling:** Add GLM models in your toolbox, and how to use R for these models.
- **Inference:** Estimation of parameters.
- **Evaluation:** Able to evaluate and compare your models.

GLM

- **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:** Quantitative x (continuous or ordinal).

Exponential family

Definition

$f(y;\theta)$ belongs to the exponential family if

$$f(y;\theta) = \exp[a(y)b(\theta)) + c(\theta) + d(y)]$$

Examples:

- Normal
- Binomial
- Poisson
- Chi-square
- Gamma
- Beta

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

Score statistics

Let $l(\theta; y_i)$ be log-likelihood function. Then the score statistic is:

$$U(\theta; y) = \frac{\partial l(\theta; y_i)}{\partial \theta}$$

Information

Let $U = U(\theta; Y)$ be the score statistic. Then the information is

$$\Im = Var(U)$$

If Y_i has pdf from exponential family:

- $\Im = E(U^2) = -E(\frac{\partial U}{\partial \theta}) = -E(\frac{\partial^2 l(\theta;y)}{\partial \theta^2})$

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.