Generalized Linear Models Study Guide

By using Generalized Linear Models, we can include more types of response variables such as binary and count.

Generalized Linear Model Components

1. **Random** component which are response variable and its associated probability distribution
2. **Systematic** component which are explanatory variables and their relationships
3. **Link function** which describes the relationship between mean of response and systematic components.

Generalized Linear Model Expression

For observation i,

$$Y_i \sim G(\mu_i, \theta)$$
$$h(\mu_i) = X_i^T \beta$$

- G is the distribution of outcome/response
- μ_i is location parameter for i
- θ are extra parameters for density of G
- h is a link function
- X_i are covariates for i
- β is a vector of regression coefficients

Generalized Linear Model Assumptions

1. Data $Y_1, \cdots, Y_n$ are independently distributed.
   - Errors are independent but may not be normally distributed.
2. Dependent variable Y_i assumes a distribution, but the distribution may or may not be normal.
3. There is a linear relationship between transformed response and explanatory variables
   - There may or may not be a linear relationship between dependent and independent variables.
4. Homogeneity of variance may or may not be satisfied
5. Uses maximum likelihood estimation, so it relies on large-sample approximations

Binomial(Logistic) regression

$$Y_i \sim \text{Binomial}(N_i, p_i)$$
$$\log\left(\frac{p_i}{1-p_i}\right) = X_i\beta$$