BIOSTAT 651
Notes #6: GLM: Inference

- Lecture Topics:
  - Nested models
  - Hypothesis testing
- Text (Dobson & Barnett, 3rd Ed.): Chapter 5

GLM: Basic Set-Up

- Canonical exponential family,

$$f(y_i; \theta_i, \phi) \quad = \quad \exp\left\{ \frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c(Y_i, \phi) \right\}$$

- Key quantities:

$$\begin{aligned} E[Y_i] &= \mu_i \\ V(Y_i) &= v(\mu_i)a(\phi) \\ g(\mu_i) &= \eta_i \\ \eta_i &= \mathbf{x}_i^T \boldsymbol{\beta} \end{aligned}$$

- Connecting parameters:

$$\begin{aligned} \mu_i &= b'(\theta_i) \\ V(Y_i) &= b''(\theta_i)a(\phi) \\ v(\mu_i) &= b''(\theta_i) = \frac{\partial \mu_i}{\partial \theta_i} \end{aligned}$$

- Under canonical link ($\eta_i = \theta_i$):

$$v(\mu_i) \quad = \quad \frac{1}{g'(\mu_i)}$$

Comparison of Nested Models

- Suppose we wish to compare the fit of two models
  - *Full model*:

$$\eta_i = \mathbf{x}_{i1}^T \boldsymbol{\beta}_1 + \mathbf{x}_{i2}^T \boldsymbol{\beta}_2$$

  - *Reduced model*:

$$\eta_i = \mathbf{x}_{i2}^T \boldsymbol{\beta}_2$$

  where $\mathbf{x}_i^T = (\mathbf{x}_{i1}^T, \mathbf{x}_{i2}^T)$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$ are $q \times 1$ vectors, with $q = q_1 + q_2$

- The above-listed models are *nested*
  - reduced model contains proper subset of full model's covariates

- Testing the null hypothesis

  $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ (vs $H_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$)

Hypothesis Testing: Example 1

- Example: Suppose that the impact of a treatment (Z_i) on Y_i is of interest, with age (A_i) and bodymass index (B_i) serving as adjustment covariates. The n patients in the study are randomized to receive treatment ($Z_i = 1$), or not ($Z_i = 0$).
  - Since Z_i was randomly generated, the investigators may be interested in whether A_i and B_i are actually worth including in the model.
    - Full model:

$$\eta_i = \beta_0 + \beta_1 Z_i + \beta_2 A_i + \beta_3 B_i$$

    - Reduced model:

$$\eta_i = \beta_0 + \beta_1 Z_i$$

  - Implies the hypothesis test,

    $H_0 : \beta_2 = \beta_3 = 0$ versus

    $H_1 : \beta_2 \neq 0 \cup \beta_3 \neq 0$

Hypothesis Testing: Example 2

- Example: As a continuation of the previously-described study, a follow-up study is carried out in which all patients received treatment, although the dose (D_i) differed substantially across patients. The dose/response relationship is of chief interest and, although parsimony is a goal, the investigators are willing to work with more complicated models.
  - Two models that the investigator might compare are

$$\begin{aligned} \eta_i &= \beta_0 + \beta_1 D_i + \beta_2 D_i^2 + \beta_3 D_i^3 \\ \eta_i &= \beta_0 + \beta_1 D_i \end{aligned}$$

  - Hypothesis test,

    $H_0 : \beta_2 = \beta_3 = 0$ versus

    $H_1 : \beta_2 \neq 0 \cup \beta_3 \neq 0$

Hypothesis Testing: Nested Models

- Note that the hypothesis testing methods we will study in this lecture related to nested models
  - full model contains each of the reduced model's covariates, plus some additional factors
- e.g., Referring back to the previous examples, the models

$$\begin{aligned} \eta_i &= \beta_0 + \beta_1 D_i + \beta_2 A_i \\ \eta_i &= \beta_0 + \beta_1 Z_i + \beta_2 A_i + \beta_3 B_i \end{aligned}$$

  are *not nested*

Hypothesis Testing: General Set-Up

- The model (GLM) is given by,

$$
\begin{aligned}
\eta_i &= \mathbf{x}_{i1}^T \boldsymbol{\beta}_1 + \mathbf{x}_{i2}^T \boldsymbol{\beta}_2 \\
\boldsymbol{\beta} &= \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} \\
\mathbf{x}_i^T &= [\mathbf{x}_{i1}^T, \mathbf{x}_{i2}^T]
\end{aligned}
$$

  where $\boldsymbol{\beta}_j$ is a $q_j \times 1$ vector $(j = 1, 2)$, with $q = q_1 + q_2$

- Hypothesis of interest:
  - $H_0 : \boldsymbol{\beta}_1 = \mathbf{d}$ versus
    $H_1 : \boldsymbol{\beta}_1 \neq \mathbf{d}$
  - often, we will set $\mathbf{d} = \mathbf{0}$

- We use the above general framework for the three tests we will now describe:
  - Wald test
  - Score test
  - Likelihood ratio test

Wald Test

- For the Wald test, fit the full model

  i.e., *unrestricted* MLE of $(\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$

- Reference distribution: under H_0,

$$X_W^2 \quad \sim \quad \chi^2_{q_1}$$

- Special case (β_1: scalar):

$$X_W^2 = \left\{ \frac{\widehat{\beta}_1}{\widehat{SE}(\widehat{\beta}_1)} \right\}^2 \quad \sim \quad \chi^2_1$$

Wald Test (continued)

- More generally, can use the Wald test for any null hypothesis of the form $H_0 : \mathbf{C}\boldsymbol{\beta} - \mathbf{d} = \mathbf{0}$
  - then, test statistic given by:

$$\begin{aligned} X_W^2 &= (\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d})^T \left\{ \widehat{V}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d}) \right\}^{-1} (\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d}) \\ &\sim \chi_r^2 \end{aligned}$$

    where C is of rank r.

  - The variance estimator of $\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d}$ is

$$\begin{aligned} V(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d}) &= V(\mathbf{C}\widehat{\boldsymbol{\beta}}) = \mathbf{C}V(\widehat{\boldsymbol{\beta}})\mathbf{C}^T \\ &= \mathbf{C}I(\boldsymbol{\beta})^{-1}\mathbf{C}^T \end{aligned}$$

  - Combining the results

$$\begin{aligned} X_W^2 &= (\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d})^T \left\{ \mathbf{C}I(\widehat{\boldsymbol{\beta}})^{-1}\mathbf{C}^T \right\}^{-1} (\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d}) \\ &\sim \chi_r^2 \end{aligned}$$

- Recall the linear regression analog:
  - $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$
  - General Linear Hypothesis (GLH) test,

$$
\begin{aligned}
F &= \frac{(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d})^T \left\{ \widehat{V}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d}) \right\}^{-1} (\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d})}{r} \\
&= \frac{(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d})^T \left\{ \mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T \right\}^{-1} (\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{d})}{\widehat{\sigma}^2 \; r} \\
&\sim F_{r,n-q}
\end{aligned}
$$

Score Test

- To carry out the score test, compute the *restricted* MLE

  i.e., maximize the likelihood under the constraints imposed by H_0

- The test is then based on the original (*unrestricted*) score function

- Score test statistic for $H_0 : \boldsymbol{\beta}_1 = \mathbf{d}$,

$$\begin{aligned} X_S^2 &= U(\widehat{\boldsymbol{\beta}}_H)^T I(\widehat{\boldsymbol{\beta}}_H)^{-1} U(\widehat{\boldsymbol{\beta}}_H) \\ \widehat{\boldsymbol{\beta}}_H &= \begin{bmatrix} \mathbf{d} \\ \widehat{\boldsymbol{\beta}}_{2H} \end{bmatrix} \end{aligned}$$

  where $\widehat{\boldsymbol{\beta}}_{2H}$ is computed under H_0

- Null distribution:

$$X_S^2 \sim \chi^2_{q_1}$$

Likelihood Ratio Test

- Likelihood Ratio Test (LRT),
  - fit the full model, hence maximizing the unrestricted likelihood
  - fit the reduced model, computing the H_0-restricted MLE
  - the LRT statistic is then given by:

$$\begin{aligned} X_L^2 &= 2\{\ell(\widehat{\boldsymbol{\beta}}) - \ell(\widehat{\boldsymbol{\beta}}_H)\} \\ &\sim \chi^2_{q_1} \end{aligned}$$

Comparison of Tests

- The Wald, Score and Likelihood Ratio tests are *asymptotically* equivalent
  - not (numerically) equivalent

    need not be equal in finite samples
- The LRT tends to perform better than Wald or Score in smaller samples
- Note: the Wald test is not invariant under reparametrization
  - e.g., test of $H_0 : \beta_1 = 0$ could yield a different result from a test of $H_0 : e^{\beta_1} = 1$
- The Score and LR tests are invariant