Chap 3 Logistic regression

1. Logistic regression model

$Y_i \sim$ Bernoulli with $P(Y_i=1) = p(x_i) \in (0,1)$, $x_i = (1, x_1, \cdots, x_p)^T$

Generalized linear model:

$$\underbrace{g}_{\text{link function}}(P(x_i)) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} = x_i^T \beta$$

linear in parameters.

"logit" $g(p) = \log \frac{p}{1-p} : (0,1) \to \mathbb{R}$. $\quad g^{-1}(l) = \frac{e^l}{1+e^l}$ "sigmoid."

<u>Ungrouped data</u> $\quad y_i \in \{0,1\}$, $x_i \in \mathbb{R}^{p+1}$, $i=1,\cdots,n$

<u>Grouped data</u> $\quad n_i \in \mathbb{N}_+$, $r_i \in \{1,\cdots,n_i\}$, $x_i \in \mathbb{R}^p$, $i=1,\cdots,s$ $(s > p+1)$

Rewritten model: $\quad y_i = r_i \sim B(n_i, p(x_i))$

with $\quad \hat{P}(Y_i=1) = \hat{p}(x_i) = \frac{r_i}{n_i}$.

$\Rightarrow$ Regression model:

$$\log\left[\frac{\hat{P}_i}{1-\hat{P}_i}\right] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + e_i \qquad i = 1, 2, \ldots, s$$

with $\hat{w}_i \coloneqq \frac{1}{\hat{\sigma}_i^2} \approx n_i \hat{P}_i (1-\hat{P}_i)$, where $\hat{P}_i = r_i/n_i$.

2. Weighted least squares

$$Y = X\beta + \varepsilon$$

where $Var(\varepsilon) = V = diag(\sigma_1^2, \cdots, \sigma_n^2)$.

Weighted LSE: $\quad \hat{\beta}^{WLSE} = \arg\min_\beta \{ SS_{Res(weighted)} := \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2 \}$ $\quad (w_i = \frac{1}{\sigma_i^2})$

$$= (X^T V^{-1} X)^{-1} X^T V^{-1} Y.$$

3. Estimation (MLE)

<u>Ungrouped data</u>:

n observations with n_1 successes.

Likelihood: $\quad L(\beta) = \prod_{i=1}^{n_1} \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}} \prod_{i=n_1+1}^{n} \frac{1}{1+e^{x_i^T\beta}}$

MLE: $\hat{\beta} = \underset{\beta}{\operatorname{argmax}} L(\beta)$

$$= \underset{\beta}{\operatorname{argmax}} \log L(\beta)$$

$$= \underset{\beta}{\operatorname{argmax}} -\sum_{i=1}^{n_1} \log(1+e^{-x_i^T\beta}) - \sum_{i=n_1+1}^{n} \log(1+e^{x_i^T\beta})$$

Numerical algorithm: Newton-Raphson.

<u>Grouped data</u>

$$L(\beta) = \prod_{i=1}^{2} \left(\frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right)^{r_i} \left(\frac{1}{1+e^{x_i^T\beta}}\right)^{n_i - r_i}.$$

<u>Interpretation of regression coefficients.</u>

$$odds_i = \frac{P_i}{1-P_i} = \exp(\beta^T x_i)$$

$$\text{odds ratio} = \frac{odd_i}{odd_j}. \qquad i \neq j \in \{1, \cdots, n\}.$$

4. CI (asymptotic normality of MLE).

CI of β_j: $\hat{\beta}_j \mp z_{\frac{\alpha}{2}} \sqrt{C^{jj}}$ $\quad j = 0, 1, \cdots, p.$

where $C^{jj} = (C^{-1})_{j+1, j+1}$, $C^{-1} = \widehat{Cov}(\hat{\beta})$, $C_{ij} = -\frac{\partial^2 \log L(\hat{\beta})}{\partial \beta_i \partial \beta_j}$.

5. HT.

1) Deviance goodness of fit test

For testing

$$H_o: P_i = \frac{e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}},$$

the test statistic is

$$\lambda(\underset{\sim}{\beta}) = -2\log\left[\frac{L(\hat{\underset{\sim}{\beta}})}{L(\hat{\underset{\sim}{P}})}\right]$$

(a) Ungrouped data:

$$\begin{aligned}
L(\underset{\sim}{P}) &= \prod_{i=1}^{n} P_i^{y_i}(1-P_i)^{1-y_i} \\
\text{As} \quad \hat{P}_i &= y_i \text{ if all } \boldsymbol{x}_i \text{ are distinct} \\
\Rightarrow \quad L(\hat{\underset{\sim}{P}}) &= \prod_{i=1}^{n} y_i^{y_i}(1-y_i)^{1-y_i} \\
L(\hat{\underset{\sim}{\beta}}) &= \prod_{i=1}^{n_1}\left(\frac{e^{\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}}}{1+e^{\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}}}\right)\prod_{i=n_1+1}^{n}\left(\frac{1}{1+e^{\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}}}\right) \\
\Rightarrow \quad \lambda(\underset{\sim}{\beta}) &\sim \chi^2_{n-(p+1)}
\end{aligned}$$

(b) Grouped data:

$$
\begin{aligned}
L(\underset{\sim}{P}) &= \prod_{i=1}^{s} P_i^{r_i}(1-P_i)^{n_i-r_i} \\
\text{As} \quad \hat{P}_i &= \frac{r_i}{n_i} \\
\Rightarrow \quad L(\hat{\underset{\sim}{P}}) &= \prod_{i=1}^{s} \left(\frac{r_i}{n_i}\right)^{r_i} \left(\frac{n_i-r_i}{n_i}\right)^{n_i-r_i} \\
L(\hat{\underset{\sim}{\beta}}) &= \prod_{i=1}^{s} \left(\frac{e^{\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}}}{1+e^{\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}}}\right)^{r_i} \left(\frac{1}{1+e^{\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}}}\right)^{n_i-r_i} \\
\Rightarrow \quad \lambda(\underset{\sim}{\beta}) &\sim \chi^2_{s-(p+1)}
\end{aligned}
$$

2) Wald test

3) LRT

4) Score test

See summary_20201117.

6. Measure of performance

Pseudo R^2

Chap 4 Model Selection

1. Sequential Variable selection
   - Forward selection
   - Backward elimination
   - Stepwise selection
2. Best subset selection
   - Cp stat. $= 2p' - n + \frac{RSS_{p'}}{\hat{\sigma}^2_{full}}$
   - Cross validation, PRESS stat.
   - R^2, adjusted $R^2 = 1 - \frac{n-1}{n-p'}(1-R^2)$
   - AIC, BIC