

Chapter 4

Logistic Regression = Regression in a GLM sense as the function is modeled by a linear combination of X -variables

↳ Appropriate for binary dependent variable. Computes associated probabilities with class assignments

In logistic regression, the probability of the event occurrence = π .

$$\rightarrow P(Y=1) = \bar{p} \text{ (Probability that event occurs)}$$

$$\rightarrow E(Y) = \bar{p}.$$

Since $p \in [0, 1]$, we use a function of p so it covers the real line $(-\infty, \infty)$.

$$\bar{p} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad \leftarrow p = \text{Number of } X\text{-variables here}$$

$$\logit(p) = \log \frac{\bar{p}}{1-\bar{p}} = \frac{\beta_0 + \beta_1 X}{\frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}} \quad \leftarrow \text{This is undefined if } p=0 \text{ and } p=1, \text{ but this relationship is deterministic}$$

$$\bar{p} = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad \text{and} \quad \bar{p} = \expit(\beta_0 + \beta_1 X)$$

- The shape of this function is sigmoidal

- Estimated coefficients follow a t -distribution ^{in linear regression}, whereas in logistic regression, estimated coefficients are normally distributed. Thus, it follows a Z -statistic (hypothesis testing based on normal distribution)

$$\logit\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

↳ There is no error term

- Log odds = When all x -variables are zero: β_0 . Then $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0$

$$\text{Odds} = \frac{\pi}{1-\pi} = \exp(\beta_0)$$

$$\exp(\beta_1) = \frac{\text{Odds}_i}{\text{Odds}_j}$$

- Given an example with one variable

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 (x+1) \quad \leftarrow x \text{ increased by 1 unit}$$

$$\log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_1 x$$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) - \log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_1 (x+1) - \beta_1 (x) = \beta_1$$

$$\text{Hence } \beta_1 = \log \frac{\frac{\pi_i}{1-\pi_i}}{\frac{\pi_j}{1-\pi_j}} = \log\left(\frac{\text{Odds}_i}{\text{Odds}_j}\right) \quad \text{Hence, a coefficient can be interpreted as a log odds ratio (log OR)}$$

Given $\exp(B_1) = \frac{\text{odds}_j}{\text{odds}_i}$, exponentiated coefficients give the OR corresponding to an increase of one unit in x

⊗ If $B_1 = 2$, increasing x_1 by 1 unit increases the odds of the event by a factor of $\exp(2)$ on average

Thus, $\text{odds}_i \times \exp(B_1) = \text{odds}_j$.

Increasing x_1 by 1 unit is associated with an increase in the odds of the event

⊗ $\text{logit}(P_{\text{admit}}) = -10 + 0.15 X_{\text{AV}}$

On average, each additional percentage point on the major average score increases the odds of

admission by a factor of $\exp(0.15) = 1.16$ (with all other variables held constant)

If $x = 80$ and we want to find the odds,

$$\text{odds} = \frac{P_{\text{admit}}}{1 - P_{\text{admit}}} = \exp(-10 + 0.15 \times 80) = 7.38$$

From here, if your major average increases by 1 percent point, the new odd is =

$$7.389 \cdot \exp(0.15) = 8.585$$

Estimation

Given $Y_i \sim \text{Bernoulli}(\pi_i) \rightarrow \text{From } L(\beta) = \prod_{i=1}^n p_i^{Y_i} (1-p_i)^{1-Y_i}$

$$L(\beta) = \prod_{i=1}^n (\pi(x_i; \beta))^{Y_i} \cdot [1 - \pi(x_i; \beta)]^{1-Y_i} \quad \text{where } \pi(x_i; \beta) = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}$$

- Estimating equations of the coefficients by maximum likelihood

Where $\beta = (\beta_0, \dots, \beta_p)^T$ is the vector of coefficients and $x_i = (x_{i1}, \dots, x_{ip})^T$

and Y_i is a scalar (Bernoulli, hence 1 or 0?)

$$\begin{aligned} \text{Take log} \rightarrow l(\beta) &= \sum_{i=1}^n \{Y_i \log p(x_i; \beta) + (1-Y_i) \log [1 - p(x_i; \beta)]\} \\ &= \sum_{i=1}^n \{Y_i \beta^T x_i - Y_i \log(1 + e^{\beta^T x_i}) + (1-Y_i) \log \left[\frac{1}{1 + e^{\beta^T x_i}} \right]\} \\ &= Y_i \log \left(\frac{e^{\beta^T x_i}}{1 + e^{\beta^T x_i}} \right) + (1-Y_i) \log \left(1 - \frac{e^{\beta^T x_i}}{1 + e^{\beta^T x_i}} \right) \\ &= Y_i \log(e^{\beta^T x_i}) - \log(1 + e^{\beta^T x_i}) \\ &= Y_i e^{\beta^T x_i} - Y_i \log(1 + e^{\beta^T x_i}) + (1-Y_i) [\log(1) - \log(1 + e^{\beta^T x_i})] \\ &= Y_i e^{\beta^T x_i} - Y_i \log(1 + e^{\beta^T x_i}) + [-\log(1 + e^{\beta^T x_i}) + Y_i \log(1 + e^{\beta^T x_i})] \\ &= Y_i e^{\beta^T x_i} - \log(1 + e^{\beta^T x_i}) \approx \sum_{i=1}^n \{Y_i \beta^T x_i - \log(1 + e^{\beta^T x_i})\} \end{aligned}$$

To find a maximum, we take the derivative w.r.t. β and set it to 0.

$$\begin{aligned} l(\beta) &= \sum_{i=1}^n \{Y_i \beta^T x_i - \log(1 + e^{\beta^T x_i})\} \\ \frac{dl}{d\beta} &= \sum_{i=1}^n x_i \left\{ Y_i - \frac{e^{\beta^T x_i}}{1 + e^{\beta^T x_i}} \right\} = \sum_{i=1}^n x_i \{Y_i - p(x_i; \beta)\} = 0 \end{aligned}$$

Non-linear equations in β

Use a Newton-Robson algorithm to solve for β in closed form.

$$\text{Newton-Robson step: } \beta^{\text{new}} = \beta^{\text{old}} - \left(\frac{d^2 l(\beta)}{d\beta d\beta^T} \right)^{-1} \frac{dl(\beta)}{d\beta}$$

$$\text{Need a second derivative} = \frac{d^2 l(\beta)}{d\beta d\beta^T} = - \sum_{i=1}^n x_i x_i^T p(x_i; \beta) (1 - p(x_i; \beta))$$

$$\text{or, } \frac{dl}{d\beta} = X'(y - p) \quad (\text{in Matrix notation})$$

$$\text{and } \frac{d^2 l}{d\beta d\beta^T} = -X'WX, \text{ where } X = (n, n) \text{ diagonal matrix with the } i^{\text{th}} \text{ diagonal element defined as } p(x_i; \beta_{old}) [1 - p(x_i; \beta_{old})]$$