Simple Logistic Regression

$x \in \mathbb{R}$, $y \in \{0, 1\}$

We want a simple specification for

$$P\left[Y = 1 \mid x \right]$$

- probability that Y turns out to be 1, given $X = x$

Two Steps

i Linearly transform x

$$x \to \beta_0 + \beta_1 x \equiv \eta$$

ii map η to $[0, 1]$

$$P[Y = 1 | x] = F(\eta)$$

So

$$P[Y = 1 | x] = F(\beta_0 + \beta_1 x)$$

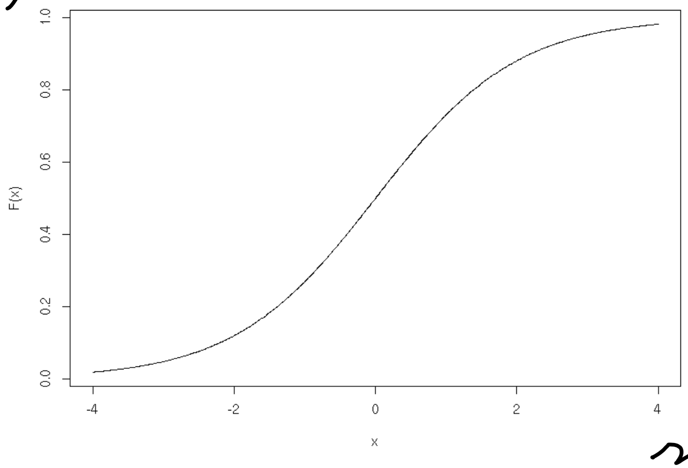
The parameters of the model are just (β_0, β_1)

Map to $(0,1)$

$$F(\eta) = \frac{e^{\eta}}{1+e^{\eta}}$$

$$= \frac{1}{1+e^{-\eta}}$$

$F(\eta)$



η

Likelihood:

Given data (x_i, y_i), $i = 1, 2, \ldots n$, we can estimate (β_0, β_1) using maximum likelihood.

$y = (y_1, y_2, \ldots \ldots y_n)$

$x = (x_1, x_2, \ldots \ldots x_n)$

$$L(\beta_0, \beta_1; x, y) \propto p(y \mid x, \beta_0, \beta_1)$$

$$L(\beta_0, \beta_1; x, y) \propto p(y \mid x, \beta_0, \beta_1)$$

Maximum Likelihood

$$\max_{(\beta_0, \beta_1)} L(\beta_0, \beta_1; x, y).$$

- choose the parameter values that make the observed data most likely!

Conditional Independence

Assume the Y_i are conditionally independent given x.

$$p(y \mid x, \beta_0, \beta_1) = \prod_{i=1}^{n} p(y_i \mid x_i, \beta_0, \beta_1)$$

$$p(y_i \mid x_i, \beta_0, \beta_1) = \begin{cases} F(\beta_0 + \beta_1 x_i) & y_i = 1 \\ 1 - F(\beta_0 + \beta_1 x_i) & y_i = 0 \end{cases}$$