⇒ Generalized Linear Models for Binary Data:

Many categorical response variable have only two categories. for example whether you take public transport today (Yes, No) - Denote a binary response variable by Y and its two possible outcomes by 1(success) and 0(failure).

The distribution of Y is specified by probabilities $P(Y=1) = \pi$ of success and $P(Y=0) = (1-\pi)$ of failure. for each binary observation its mean is $E(Y) = \pi$.

The following section introduces GLMs for binary responses. Although GLMs can have multiple explanatory variables. for simplicity we introduce them using single x. The value of π can vary as the value of x changes. and we replace π by $\pi(x)$ when we want to describe its independence on that value

(1) Linear Probability model:

In ordinary regression, $\mu = E(Y)$ is a linear function of x. for a binary response, an analogous model is

$$\pi(x) = \alpha + \beta x$$

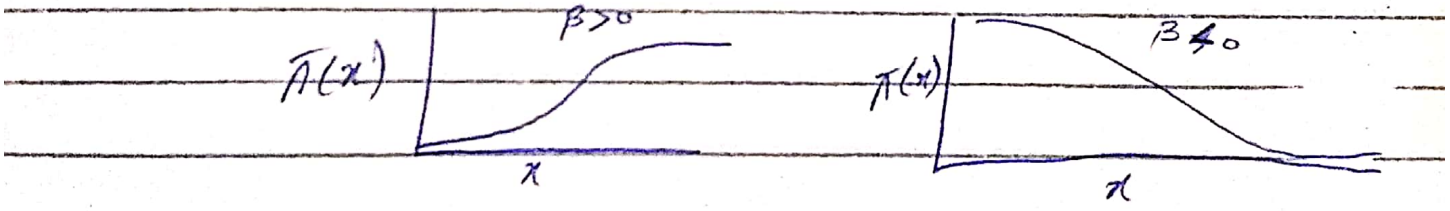
This is called a linear probability model because the probability of success changes linearly in x - The parameter β represent the change in the probability per unit change in x. This model is a GLM with binomial random component and identity link function.

This model is simple, but unfortunately it has a structural defect. Probability fall between 0 and 1, whereas linear functions take value over the entire real line. This model predicts $\pi(x) < 0$ and $\pi(x) > 1$ for sufficiently large or small x values. The model can fit adequate over a restricted range of x values.

② Logistic Regression Model:

Relationship between $\pi(x)$ and x are usually non linear rather than linear. (Prob. of buying a new car increase when income increase in an)

In practice, $\pi(x)$ often either increases continuously or decreases continuously as x increase. The S-shaped curves are often realistic shapes for the relationship



The most important matematical function with this shape has formula.

$$\pi(x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)} = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}.$$

This is called the logistic regression function.

The corresponding logistic regression model form is

$$\boxed{\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x}$$

$$\frac{\pi(x)}{1-\pi(x)} = \frac{\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}}{1-\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}}$$

$$= \frac{e^{\alpha+\beta x}}{\frac{1+e^{\alpha+\beta x}-e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}}$$

$$\frac{\pi(x)}{1-\pi(x)} = e^{\alpha+\beta x}$$

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha+\beta x \log(e)$$

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha+\beta x \Rightarrow \text{logit}(\pi(x)) = \alpha+\beta x$$

↳ logit function of $\pi(x)$

This is a GLM with binomial random component and logit link function. The logistic regression models are often called logit models.

Whereas π is restricted to the 0-1 range, the logit can be any real numbers. The real numbers are also the potential range for linear predictors (such as $\alpha+\beta x$) that form the systematic component of a GLM, so this model does not have the structural problem that the linear

probability model has.

The parameter β determines the rate of increase or decrease of the curve.

when $\beta > 0$, $\pi(x)$ increase as x increase.

when $\beta < 0$, $\pi(x)$ decrease as x increase

The magnitude of β determines how fast the curve increases or decreases.

when $\beta = 0$, the curve flattens to a horizontal straight line.

③ Probit Regression Model:

Another model that has S-shaped curves is called probit model. The link function for the model, called the probit link, transform probabilities to z-scores from the standard normal distribution. The probit model has expression

$$\text{Probit}[\pi(x)] = \alpha + \beta x$$

The probit link function applied to $\pi(x)$ gives the standard normal z-score at which the left-tail probability equals $\pi(x)$. For instant probit$(0.05) = -1.645$ because 5% of the standard normal distribution falls betaw -1.645.

Like wise

$$\text{Probit}(0.50) = 0$$
$$\text{probit}(0.95) = 1.645$$
$$\text{Probit}(0.975) = 1.96$$