Assumptions

Wednesday 1 May 2024 9:34 PM

NOTE : violating these assumptions doesn't mean we can't use logistic regression, but it may impact the validity of the results & we should proceed with caution.

1) Binary logistic regression requires the dependent variable to be binary

i.e. $y \rightarrow$ bernoulli distribution

2) Independence of observations

3) Absence of multicollinearity in independent variables

4) large sample size

5) linearity of independent variables and log odds

odds : the odds of an event is the ratio of the probability of the event happening to the probability of the event not happening. it's a way of expressing the likelihood of an event. if the odds are greater than 1, the event is more likely to happen than not.

$$\boxed{\frac{p}{1-p}} \quad [0 \rightarrow \infty)$$

$\log_e(\text{odds}) \rightarrow$ brings symmetry to scale

log(odds) connection to logistic regression?

$$\log(\text{odds}) = \log\left(\frac{P}{1-P}\right)$$

X_1	X_2	X_3	X_4	$\hat{y}$

$$\hat{y} = P(1) = P = \frac{1}{1 + e^{-\beta X}}$$

$$P = \frac{1}{1 + e^{-\beta X}}$$

$$1 + e^{-\beta X} = \frac{1}{P}$$

$$e^{-\beta X} = \frac{1}{P} - 1$$

$$e^{-\beta X} = \frac{1-P}{P}$$

$$\frac{1}{e^{\beta X}} = \frac{1-P}{P}$$

$$\frac{P}{1-P} = e^{\beta X}$$

$$\log\left(\frac{P}{1-P}\right) = \beta X$$

$$\log(\text{odds}) = \beta X$$

$$\boxed{P = \frac{1}{1 + e^{-\text{logodds}}}}$$