

Multiple Logistic Regression

The simple logistic regression model, discussed in the last chapter, uses only one predictor variable, while multiple logistic regression, as the name implies, allows for more than one predictor variable.

1.1 Multiple Logistic Regression

In multiple logistic regression, we want to model the relationship between a binary response variable and multiple predictor variables. Let y be the binary response variable and x_1, x_2, \dots, x_p be p predictor variables. The multiple logistic regression model has the form:

$$\ln \left(\frac{P(Y = 1|X)}{1 - P(Y = 1|X)} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

where $P(Y = 1|X)$ is the probability of the event $Y = 1$ given the predictor variables, and $\beta_0, \beta_1, \dots, \beta_p$ are the parameters of the model. This equation can also be rewritten in terms of the probability $P(Y = 1|X)$:

$$P(Y = 1|X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}$$

In this model, each one-unit increase in X_i multiplies the odds of $Y = 1$ by e^{β_i} , holding all other predictors constant.

1.2 Divide by 4 Rule

The "Divide by 4" rule is a rule of thumb for interpreting the coefficients in logistic regression. It says that for small values of β_i , a one-unit increase in X_i will change the probability $P(Y = 1|X)$ by approximately $\beta_i/4$ at the average value of X_i .

The rule arises from the derivative of the logistic function at its midpoint, and provides a useful and simple way to get an approximate sense of the effect size when interpreting the coefficients.

This is a draft chapter from the Kontinua Project. Please see our website (<https://kontinua.org/>) for more details.

Answers to Exercises

