Logistic Regression

Ananda Biswas

Last Updated : October 25, 2025

Contents

1 Interpretation of Parameters

1.1 Single Explanatory Variable

Consider a logistic regression model with single explanatory variable X and a dichotomous dependent variable Y as follows.

$$\eta(X) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X \tag{1}$$

where $\pi = P(Y = 1|X)$ is the probability of the event of interest; β_0 and β_1 are parameters.

- β_0 : With $X = 0$, $\beta_0 = \log\left(\frac{\pi}{1-\pi}\right)$. So β_0 is log-odds of the event when $X = 0$, or e^{β_0} is the odds of the event of interest when $X = 0$.
- β_1 : Following (1), at $X = x$, $\eta(x) = \beta_0 + \beta_1 x$ and at $X = x+1$, $\eta(x+1) = \beta_0 + \beta_1(x+1)$.

  $\therefore \beta_1 = \eta(x+1) - \eta(x) \Rightarrow \beta_1$ is the change in log-odds of the event of interest for a one-unit increase in X.

  Again,

$$\begin{aligned}
\beta_1 &= \eta(x+1) - \eta(x) \\
&= \log[\text{odds}(x+1)] - \log[\text{odds}(x)] \\
&= \log\left[\frac{\text{odds}(x+1)}{\text{odds}(x)}\right] \\
\Rightarrow e^{\beta_1} &= \frac{\text{odds}(x+1)}{\text{odds}(x)}
\end{aligned}$$

In the set-up as in (1), e^{β_1} is called **Crude Odds Ratio** as it shows the relationship between the outcome and the predictor - without taking into account the effect of any other variable.

1.2 Multiple Explanatory Variables

Consider a logistic regression model with $p-$many explanatory variables $X_1, X_2, \ldots, X_p$ and a dichotomous dependent variable Y as follows.

$$\eta(X_1, X_2, \ldots, X_p) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \tag{2}$$

where $\pi = P(Y = 1|X_1, X_2, \ldots, X_p)$ is the probability of the event of interest; $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are parameters.

- β_0 : With $X_1 = X_2 = \ldots = X_p = 0$, $\beta_0 = \log\left(\frac{\pi}{1-\pi}\right)$. So β_0 is log-odds of the event when $X_1 = X_2 = \ldots = X_p = 0$, or e^{β_0} is the odds of the event of interest when $X_1 = X_2 = \ldots = X_p = 0$.
- β_1 : Following (2), at $X_1 = x_1, X_2 = x_2, \ldots X_j = x_j, \ldots, X_p = x_p$,

$$\eta(x_1, x_2, \ldots, x_j, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j + \ldots + \beta_p x_p$$

and at $X_1 = x_1, X_2 = x_2, \ldots X_j = x_j + 1, \ldots, X_p = x_p$,

$$\eta(x_1, x_2, \ldots, x_j + 1, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j(x_j + 1) + \ldots + \beta_p x_p.$$

$\therefore \beta_j = \eta(x_1, x_2, \ldots, x_j + 1, \ldots, x_p) - \eta(x_1, x_2, \ldots, x_j, \ldots, x_p)$

$\Rightarrow \beta_j$ is the change in log-odds of the event of interest for a one-unit increase in X_j.

Again,

$$\begin{aligned}
\beta_j &= \eta(x_1, x_2, \ldots, x_j + 1, \ldots, x_p) - \eta(x_1, x_2, \ldots, x_j, \ldots, x_p) \\
&= \log[\text{odds}(x_1, x_2, \ldots, x_j + 1, \ldots, x_p)] - \log[\text{odds}(x_1, x_2, \ldots, x_j, \ldots, x_p)] \\
&= \log\left[\frac{\text{odds}(x_1, x_2, \ldots, x_j + 1, \ldots, x_p)}{\text{odds}(x_1, x_2, \ldots, x_j, \ldots, x_p)}\right] \\
\Rightarrow e^{\beta_j} &= \frac{\text{odds}(x_1, x_2, \ldots, x_j + 1, \ldots, x_p)}{\text{odds}(x_1, x_2, \ldots, x_j, \ldots, x_p)}
\end{aligned}$$

In the set-up as in (2), e^{β_j} is called **Adjusted Odds Ratio** as it represents the effect of X_j on the outcome after controlling(*adjusting*) all other predictors in the model.

Thus, *Adjusted Odds Ratio* is very useful to assess *individual risk factors* for an outcome, as we get to see how that factor alone impacts the outcome when all other risk factors are fixed. This was not possible with *Crude Odds Ratio* as it shows the effect of a single risk factor without considering the latent effect of other risk factors.

- Is it the true impact of this risk factor ? 🤔
- Are you sure there is no synergy among the risk factors ? 🤨
- Are you sure this particular risk factor does not surrogate any other risk factor ? 😤 – These kind of questions are best answered by *Adjusted Odds Ratio*. 😼