CSCU Cornell Statistical Consulting Unit

Interpreting Interactions in Logistic Regression

Hongyu Li and Jay Barry

1 Introduction

Logistic regression is useful when modeling a binary (i.e. two category) response variable. This newsletter focuses on how to interpret an interaction term between a continuous predictor and a categorical predictor in a logistic regression model. We suggest two techniques to aid in interpretation of such interactions: 1) numerical summaries of a series of odds ratios and 2) plotting predicted probabilities.

2 Example

To explore this topic we consider data from a study of birth weight in 189 infants and characteristics of their mothers. The response variable is binary, low birth weight status: lowbwt $= 1$ if the birth weight is less than 2500 grams and lowbwt $= 0$ otherwise. The continuous predictor is the age of the mother in years, and the categorical predictor ftv is whether or not the mother made frequent physician visits during the first trimester of pregnancy, i.e. ftv $=$ 0 if no and ftv $= 1$ if yes. To simplify the interpretation of the effect of age by ftv status on the outcome, the age variable was centered at the sample mean of 23 years (i.e., age_c in the model below is equal to age minus 23).

In our model, the log odds of a low birth weight infant is assumed to be a linear function of the two predictors and their interaction:

$$\text{logit}(\text{lowbwt}) = \ln\left(\frac{P(\text{lowbwt} = 1)}{1 - P(\text{lowbwt} = 1)}\right) = \beta_0 + \beta_1 \text{age}_c + \beta_2 \text{ftv} + \beta_3 \text{age}_c \times \text{ftv}$$

We estimate the coefficients of this logistic regression model using the method of maximum likelihood. Table 2.1 displays the coefficient estimates and their standard errors.

Table 2.1: Coefficient estimates, standard errors, z statistic and p-values for the logistic regression model of low birth weight. Note that dummy coding is used with ftv $= 0$ *as the reference category.*

Coefficient estimate	Standard Error	z	p-value
intercept	-0.52	0.21	-2.44
age_c	0.04	0.05	0.95
ftv	-0.47	0.33	-1.41
$\text{age}_c \times \text{ftv}$	-0.18	0.07	-2.59

3 Odds Ratios

Although Table 2.1 tells us we have a significant interaction, interpreting the effect of the interaction term may be challenging. One method to understand the interaction can be through exploring several odds ratios expressing the association between low birth weight and frequent physician visits, at different levels of mother's age. The odds ratios in Table 3.1 can be calculated using model coefficients reported in the previous table and the following formula:

$$\text{odds} = \frac{P(\text{lowbwt} = 1)}{1 - P(\text{lowbwt} = 1)} = e^{\beta_0 + \beta_1 \text{age}_c + \beta_2 \text{ftv} + \beta_3 \text{age}_c \times \text{ftv}}$$

Recall that an odds ratio of 1 means no association between predictor and outcome (holding other predictors fixed). Odds ratios from the low birth weight example can be summarized as in Table 3.1.

Table 3.1: Odds ratios comparing mothers who frequently visit the doctor to those who do not, given the mother's age

Mother's age	OR_{ftv}	p-value	95% confidence interval
17	1.868	0.209	(0.705,4.949)
23	0.625	0.158	(0.325,1.201)
24	0.521	0.063	(0.262,1.036)
25	0.434	0.028	(0.206,0.916)
30	0.174	0.006	(0.050,0.607)

For example, the last row shows that a mother at the age of 30 who visits the physician frequently has 0.174 times the odds of having a low birth weight baby as compared to those of the same age who don't visit the doctor frequently, and it is a statistically significant association. For women whose ages are between 17 and 24, the 95% confidence intervals of the odds ratios include the null value of 1, so we do not have strong evidence of an association between frequent doctor visits and low birth weight for that age range. For mothers aged 25 years and older, the odds of having a low birth weight baby significantly decrease if the mother frequently visits her physician.

4 Probabilities

Another approach to investigating the nature of this interaction is through calculating predicted probabilities of having a low birth weight infant across different levels of mother's age and frequent physician visits. In this logistic model, predicted probabilities are given by the following equation:

$$P(\text{lowbwt} = 1) = \frac{e^{\beta_0 + \beta_1 \text{age}_c + \beta_2 \text{ftv} + \beta_3 \text{age}_c \times \text{ftv}}}{1 + e^{\beta_0 + \beta_1 \text{age}_c + \beta_2 \text{ftv} + \beta_3 \text{age}_c \times \text{ftv}}}.$$

Differences in predicted probabilities of low birth weight between those who visit the physician and those who do not (along with p-value for the test if this difference is significantly different from zero) are summarized in Table 4.1 for five fixed values of the mother's age.

Table 4.1: Odds ratios comparing mothers who frequently visit the doctor to those who do not, given the mother's age

Mother's age	Difference in probability	p-value	95% confidence interval
17	0.157	0.192	(−0.788,0.393)
23	−0.092	0.191	(−0.197,0.088)
24	−0.130	0.072	(−0.232,0.046)
25	−0.165	0.030	(−0.315, −0.016)
30	−0.316	0.006	(−0.540, −0.092)

The results from this method are in agreement with the findings based on odds ratios, although it is noteworthy that the p-values do not have to match exactly between these two metrics. For young mothers (less than 24 years old), we do not have strong evidence of an association between low birth weight and frequent physician visits. For mothers aged 25 years and older, we reject the null hypothesis of no difference in probability between those with frequent physician visits and those without. Overall, the difference between the probability of low birth weight comparing those with frequent visits to those without increases as the mother's age increases.

These probabilities are also summarized in Figure 4.1, which displays the predicted probability of low birth weight, along with confidence intervals, as a function of mother's age for those with and without frequent physician visits.

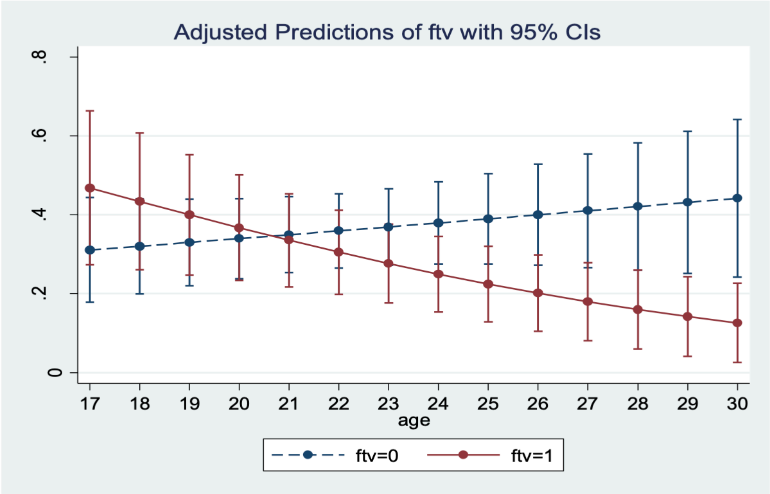

Figure 4.1: Predicted probability of low birth weight as a function of mother's age for those with frequent physician visits (solid line) and those without frequent visits (dashed line).

Importantly, the substantive conclusions that an interaction is present and the direction of the interaction will not be affected by the minor discrepancies that come about from using different metrics. They are equally valid techniques for exploring the nature of an interaction in a logistic regression model.

Created October 2012. Last updated April 2022.