The Problem with Interactions

Max R. P. Grossmann

December 25, 2025

1 Setup

Suppose an outcome is binomial. Consider the following hypothetical 'success' proportions in a 2×2 design:

Table 1: Success proportions per treatment

	$h = 0$	$h = 1$
$g = 0$	0.6	0.3
$g = 1$	0.8	π_{11}

Now, pray tell, to what value of π_{11} could we assign a meaning of "no effect on the interaction?" Troublingly, it depends on the model!

2 A true fact

If π_{11} differs between models, then, with sufficient sample size, what is considered a null effect in one model is a *true* effect in another model!

3 Linear model

With $y_i = \beta_0 + \beta_1 g_i + \beta_2 h_i + \beta_3 g_i h_i + \varepsilon_i$,

$$\pi_{11} = 0.5. \tag{1}$$

The same holds for the "correct" model—in terms of main effects—where $g_i' = g_i - \frac{1}{2}$, $h_i' = h_i - \frac{1}{2}$, and $y_i = \beta_0 + \beta_1 g_i' + \beta_2 h_i' + \beta_3 g_i' h_i' + \varepsilon_i$.

4 The exponential model

With $y_i \sim \text{Bernoulli}(p_i)$ and $p_i = \exp\left(\beta_0 + \beta_1 g_i + \beta_2 h_i + \beta_3 g_i h_i\right)$,

$$\pi_{11} = 0.4. \tag{2}$$

Proof. Taking logarithms: $\log p_i = \beta_0 + \beta_1 g_i + \beta_2 h_i + \beta_3 g_i h_i$. From the known cells, $\log(0.6) = \beta_0$, $\log(0.3) = \beta_0 + \beta_2$, and $\log(0.8) = \beta_0 + \beta_1$. Thus $\beta_0 = \log(0.6)$, $\beta_1 = \log(0.8) - \log(0.6) = \log(4/3)$, and $\beta_2 = \log(0.3) - \log(0.6) = \log(1/2)$. For no interaction ($\beta_3 = 0$),

$$\begin{aligned}\pi_{11} &= \exp(\beta_0 + \beta_1 + \beta_2) = \exp(\log(0.6) + \log(4/3) + \log(1/2)) \\ &= 0.6 \times \frac{4}{3} \times \frac{1}{2} = 0.4.\end{aligned}$$ □

5 The logit model

With $y_i \sim \text{Bernoulli}(p_i)$ and $p_i = L\left(\beta_0 + \beta_1 g_i + \beta_2 h_i + \beta_3 g_i h_i\right)$, where $L(x) = [1 + \exp(-x)]^{-1}$,

$$\pi_{11} = \frac{8}{15}. \tag{3}$$

Proof. The logit link gives $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 g_i + \beta_2 h_i + \beta_3 g_i h_i$. From the known cells, $\beta_0 = \log(0.6/0.4) = \log(3/2)$, $\beta_0 + \beta_2 = \log(0.3/0.7) = \log(3/7)$, and $\beta_0 + \beta_1 = \log(0.8/0.2) = \log(4)$. Thus $\beta_1 = \log(4) - \log(3/2) = \log(8/3)$ and $\beta_2 = \log(3/7) - \log(3/2) = \log(2/7)$. For no interaction ($\beta_3 = 0$),

$$\beta_0 + \beta_1 + \beta_2 = \log(3/2) + \log(8/3) + \log(2/7) = \log\left(\frac{3}{2} \cdot \frac{8}{3} \cdot \frac{2}{7}\right) = \log(8/7).$$

Therefore, $\pi_{11} = L(\log(8/7)) = \frac{\exp(\log(8/7))}{1+\exp(\log(8/7))} = \frac{8/7}{1+8/7} = \frac{8}{15}$. □

6 The probit model

With $y_i \sim \text{Bernoulli}(p_i)$ and $p_i = \Phi\left(\beta_0 + \beta_1 g_i + \beta_2 h_i + \beta_3 g_i h_i\right)$, where Φ is the standard-normal cdf,

$$\pi_{11} \approx 0.525. \tag{4}$$

Proof. The probit link gives $\Phi^{-1}(p_i) = \beta_0 + \beta_1 g_i + \beta_2 h_i + \beta_3 g_i h_i$. From the known cells, $\beta_0 = \Phi^{-1}(0.6)$, $\beta_0 + \beta_2 = \Phi^{-1}(0.3)$, and $\beta_0 + \beta_1 = \Phi^{-1}(0.8)$. Thus $\beta_1 = \Phi^{-1}(0.8) - \Phi^{-1}(0.6)$ and $\beta_2 = \Phi^{-1}(0.3) - \Phi^{-1}(0.6)$. For no interaction ($\beta_3 = 0$),

$$\begin{aligned}\pi_{11} &= \Phi(\beta_0 + \beta_1 + \beta_2) \\ &= \Phi\left(\Phi^{-1}(0.6) + \Phi^{-1}(0.8) - \Phi^{-1}(0.6) + \Phi^{-1}(0.3) - \Phi^{-1}(0.6)\right) \\ &= \Phi\left(\Phi^{-1}(0.8) + \Phi^{-1}(0.3) - \Phi^{-1}(0.6)\right) \\ &\approx \Phi(0.8416 - 0.5244 - 0.2533) = \Phi(0.0639) \approx 0.525.\end{aligned}$$ □

7 A note

This problem is neither novel (e.g., Ai & Norton, 2003; McCabe, Halvorson, King, Cao, & Kim, 2022; Mize, 2019) nor universal. For some choices of Table 1, different conclusions emerge. In that sense, this example is pedagogical.

Mize (2019) provided practical guidance for interpreting interaction effects. Ai and Norton (2003) showed that the probability-scale interaction effect in nonlinear models is not simply the marginal effect of the interaction term.[1] McCabe et al. (2022) extended this finding to count outcomes and documented its widespread misinterpretation in psychological research. These contributions (see also `https://datacolada.org/57`) assume the probability scale is the quantity of interest—precisely the question at issue here.

References

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123–129.

McCabe, C. J., Halvorson, M. A., King, K. M., Cao, X., & Kim, D. S. (2022). Interpreting interaction effects in generalized linear models of nonlinear probabilities and counts. *Multivariate Behavioral Research*, *57*(2-3), 243–263.

Mize, T. D. (2019). Best practices for estimating, interpreting, and presenting nonlinear interaction effects. *Sociological Science*, *6*, 81–117.

[1]Their definition of "no interaction effect" as additivity on the probability scale yields $\pi_{11} = 0.5$, coinciding with the linear model.