

Interactions in Linear Models

BIOE 210

What is an interaction?

Imagine we're modeling the response (y) from two input variables, x_1 and x_2 . The simplest model is

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

What is an interaction?

Imagine we're modeling the response (y) from two input variables, x_1 and x_2 . The simplest model is

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

The coefficient β_1 measures the effect of x_1 and β_2 measures the effect of x_2 . These effects are **independent**.

What is an interaction?

Imagine we're modeling the response (y) from two input variables, x_1 and x_2 . The simplest model is

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

The coefficient β_1 measures the effect of x_1 and β_2 measures the effect of x_2 . These effects are **independent**.

What if there is another effect that depends on both x_1 and x_2 ?
This is an **interaction** between x_1 and x_2 .

How do we model interactions?

We model the interaction of x_1 and x_2 using the product of these variables.

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$$

The coefficient β_{12} is the effect size of the interaction.

How do we model interactions?

We model the interaction of x_1 and x_2 using the product of these variables.

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$$

The coefficient β_{12} is the effect size of the interaction.

Why do we multiply x_1 and x_2 ? There are at least two ways to interpret this term.

The coded factor interpretation

Often we set up design matrices using **coded variables**. If we're testing the variable at two levels, we code the variable as "on/off" ($\{0, 1\}$) or "low/high" ($\{-1, +1\}$).

The coded factor interpretation

Often we set up design matrices using **coded variables**. If we're testing the variable at two levels, we code the variable as "on/off" ($\{0, 1\}$) or "low/high" ($\{-1, +1\}$).

on/off \rightarrow interaction when both "on"

x_1	x_2	x_1x_2
0	0	0
0	1	0
1	0	0
1	1	1

The coded factor interpretation

Often we set up design matrices using **coded variables**. If we're testing the variable at two levels, we code the variable as "on/off" ($\{0, 1\}$) or "low/high" ($\{-1, +1\}$).

on/off \rightarrow interaction when both "on"

x_1	x_2	x_1x_2
0	0	0
0	1	0
1	0	0
1	1	1

high/low \rightarrow interaction when both "high" or both "low"

x_1	x_2	x_1x_2
-1	-1	+1
-1	+1	-1
+1	-1	-1
+1	+1	+1

The augmented slope interpretation

We can also interpret the interaction as one variable changing the effect of the other variable.

$$\begin{aligned}y &= \beta_1 x_1 + \beta_2 (x_1) x_2 + \epsilon \\&= \beta_1 x_1 + (\beta_2 + \beta_{12} x_1) x_2 + \epsilon \\&= \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon\end{aligned}$$

Things to remember about interactions

- ▶ Interaction are modeled as the product of variables.
- ▶ The interaction effect is “above and beyond” the independent effects (synergy/super-additivity, antagonism/sub-additivity).
- ▶ Higher-order interactions are possible (e.g. $x_1x_2x_3$), but these are rare.