Lecture 11 - GLM VII Interaction effects

Caspar J. van Lissa

2023-08-09

Recap regression

Regression model

$Y_i = a + b * X_i + e_i$

Symbol	Interpretation
Y_i	Individual i's score on dependent variable Y
a	Coefficient, intercept of the regression line
b	Coefficient, slope of the regression line
X_i	Individual i's score on independent variable X
e_i	Individual i's prediction error

Regression line

Predicted value (describes regression line)

$\hat{Y_i} = a + b * X_i$, and $Y_i = \hat{Y}_i + \epsilon_i$

Symbol	Interpretation
$\hat{Y_i}$	Individual i's **predicted** score on dependent variable Y
a	Coefficient, intercept of the regression line
b	Coefficient, slope of the regression line
X_i	Individual i's score on independent variable X

The road so far

- $Y_i = a + bX$: Bivariate linear regression
- $Y_i = a + bX$ where X is a dummy variable: comparing two groups, aka independent samples t-test
- $Y_i = a + b_1X_1 + ... + b_kX_k$ where $X_{1...k}$ are dummy variables: comparing multiple groups, aka ANOVA
- $Y_i = a + b_1X_1 + ... + b_kX_k$ where $X_{1...k}$ are continuous or dummy variables: multiple regression

Last thing we did is extend the linear model with building blocks that look like $+bX$

Introducing: interaction

> **Interaction:** The effect of one predictor depends on the level of another predictor.

To represent this, we add a *special* building block to our regression equation:
$Y = a + b_1 X_1 + b_2 X_2 + b_3(X_1 * X_2)$

When to use interaction?

In NL, women still take on the brunt of childrearing responsibilities (parental involvement). You hypothesize that progressive gender roles will predict greater involvement for men.

- Interaction between gender roles and sex
- Continuous and dummy

Personality dimension “agreeableness” positively predicts number of friends, but only when combined with a high level of “extraversion”.

- Include an interaction effect between agreeableness and extraversion
- Both are continuous variables

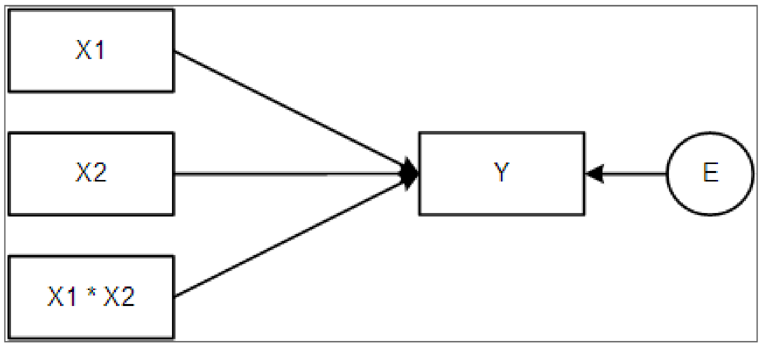
Treatment guidelines for heart failure are based mostly on research in men. There’s recent debates that commonly prescribed drugs affect recovery in men and women differently.

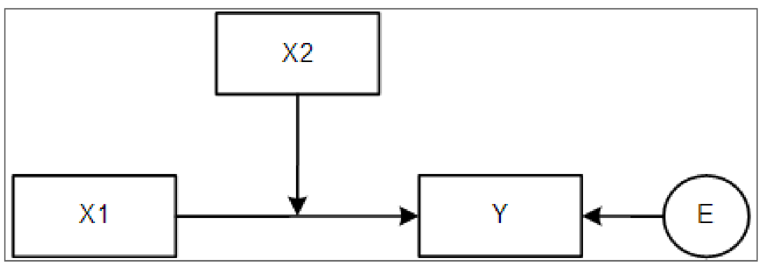
- Interaction between treatment (drug vs placebo) and sex (male vs female)
- Both are dummy variables

How to include interaction

$Y = a + b_1X_1 + b_2X_2 + b_3(X_1 * X_2)$

- Calculate a new variable that is the product of the two interacting variables
- Add it to the regression model, **along with the two original variables**





Continuous and binary

Binary predictor

There is a difference in Parental Involvement between males (0) and females (1)
$Y_i = a + b * X_i + \epsilon_i$
This regression will give us:

- The mean level of involvement for males, a
- The difference in mean level of involvement between males and females, b
- We can calculate the mean involvement of females: $a + b$

Binary and Continuous Predictor

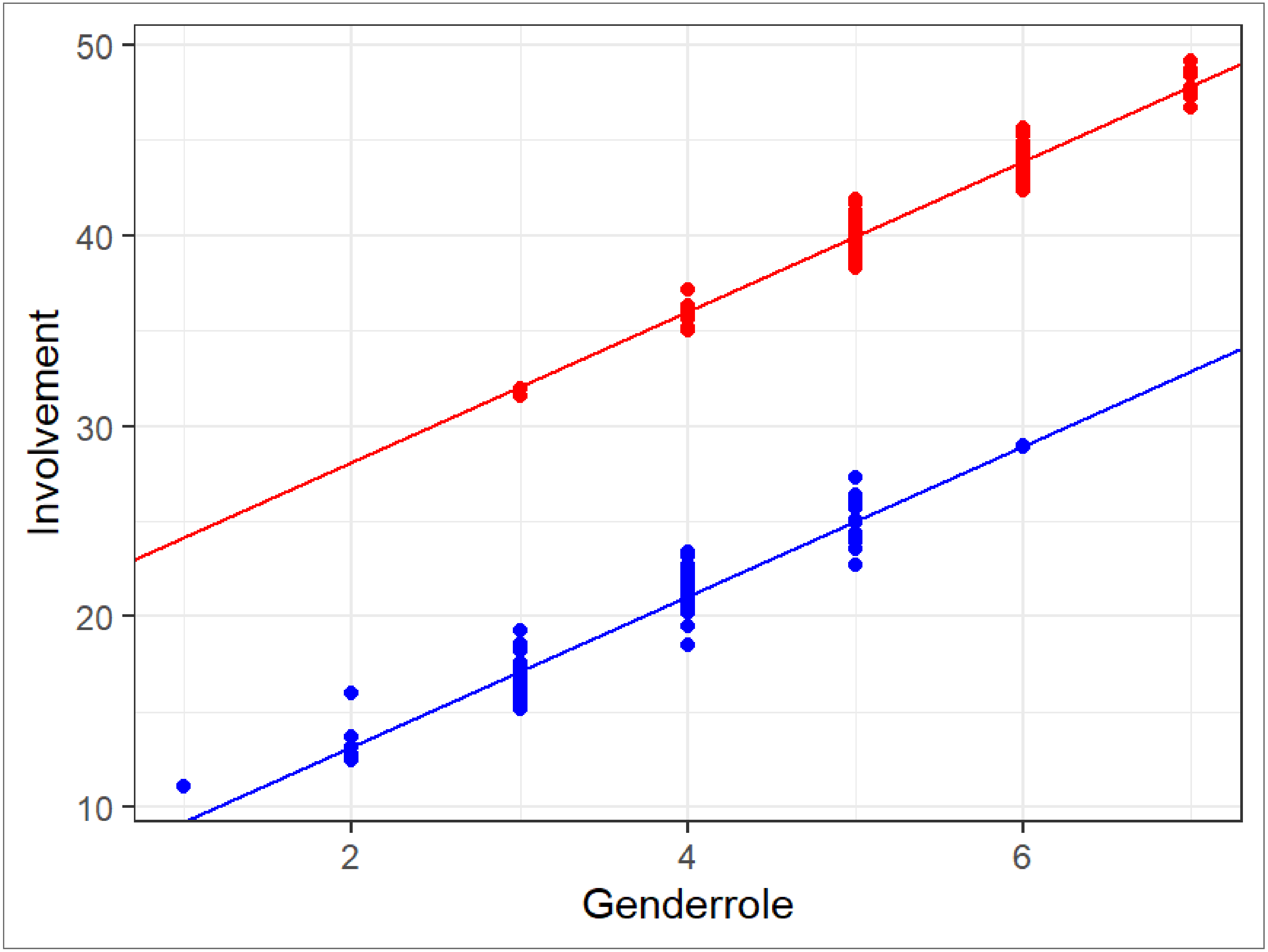
Aside from a sex difference X_1, there is an effect of gender role attitudes, X_2:
$Y_i = a + b_1 * X_{1i} + b_2 * X_{2i} + \epsilon_i$

- a: Mean level of involvement for males who score 0 on gender role
- b_1: Difference in mean level of involvement between males and females, b
- b_2: Increase in involvement associated with a 1-point increase in gender roles

Distinct intercepts

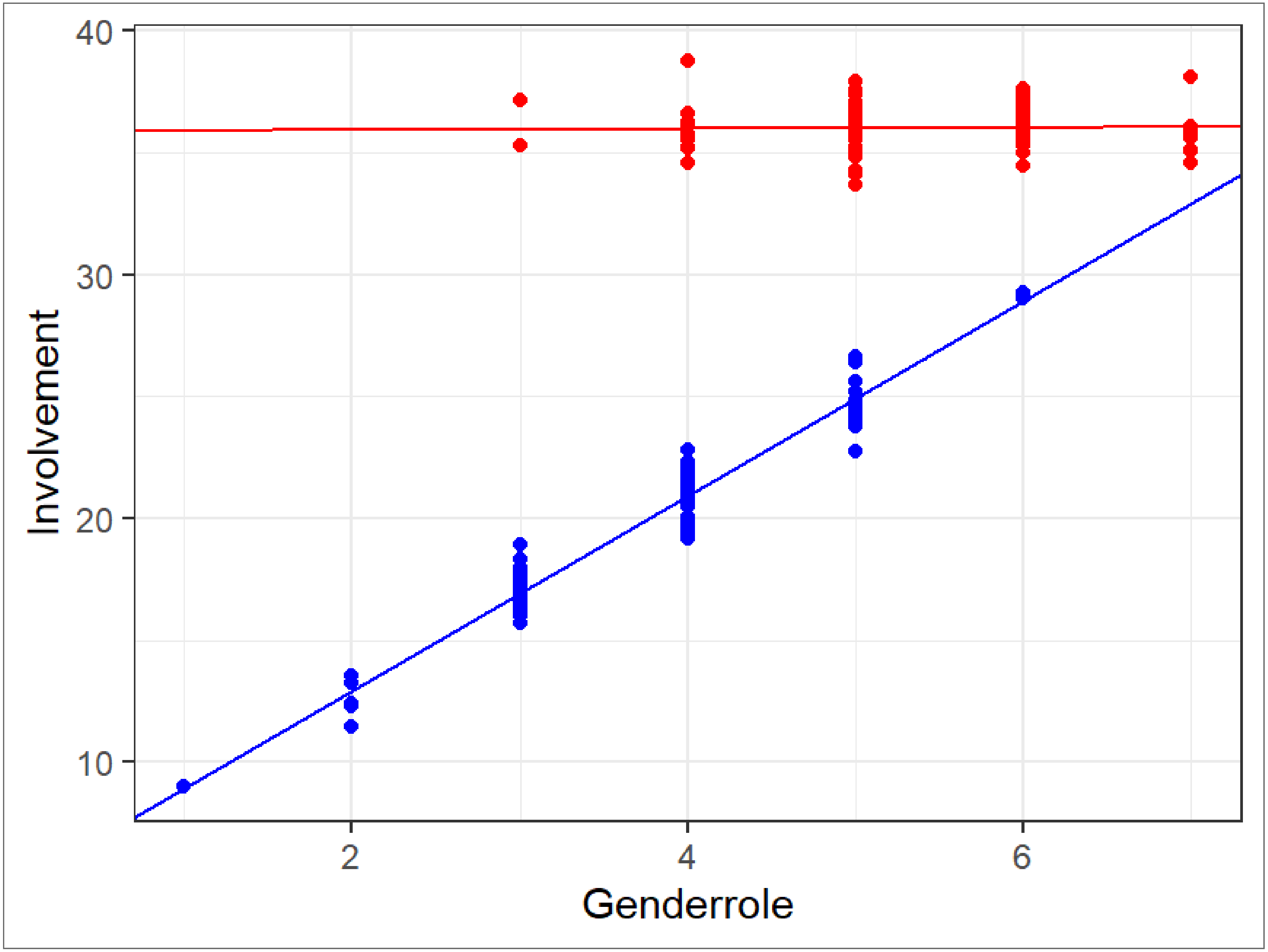
$Y_i = a + b_1 * X_{1i} + b_2 * X_{2i} + \epsilon_i$

Here, males and females have separate intercepts:


50
40
30
20
10
Involvement
2
4
6
Genderrole

Distinct regression lines

But what if we not only want to estimate distinct intercepts, but also distinct slopes for men and women?


40
30
20
10
Involvement
2
4
6
Genderrole

Interaction effect

For one binary predictor (male = 0, female = 1) and gender roles:

$\hat{Y}_i = a + b_1 * X_{1i} + b_2 * X_{2i} + b_3 * (X_{1i} * X_{2i})$

Symbol	Interpretation
$\hat{Y}_i$	Individual predicted value for Y (involvement)
a	Expected value for men who score 0 on gender role
b_1	Mean difference between men and women who score 0 on gender role
b_2	Effect of gender role for men
b_3	Difference in the effect of gender role between men and women

Complete the formula

$\hat{Y}_i = a + b_1 * X_{1i} + b_2 * X_{2i} + b_3 * (X_{1i} * X_{2i})$

Complete for men:

- $$\begin{aligned}\hat{Y}_i = & a + b_1 * 0 + b_2 * X_{2i} + b_3 * (0 * X_{2i}) = \\ & a + b_2 * X_{2i}\end{aligned}$$
- So the regression line for men is $a + b_2 * X_{2i}$

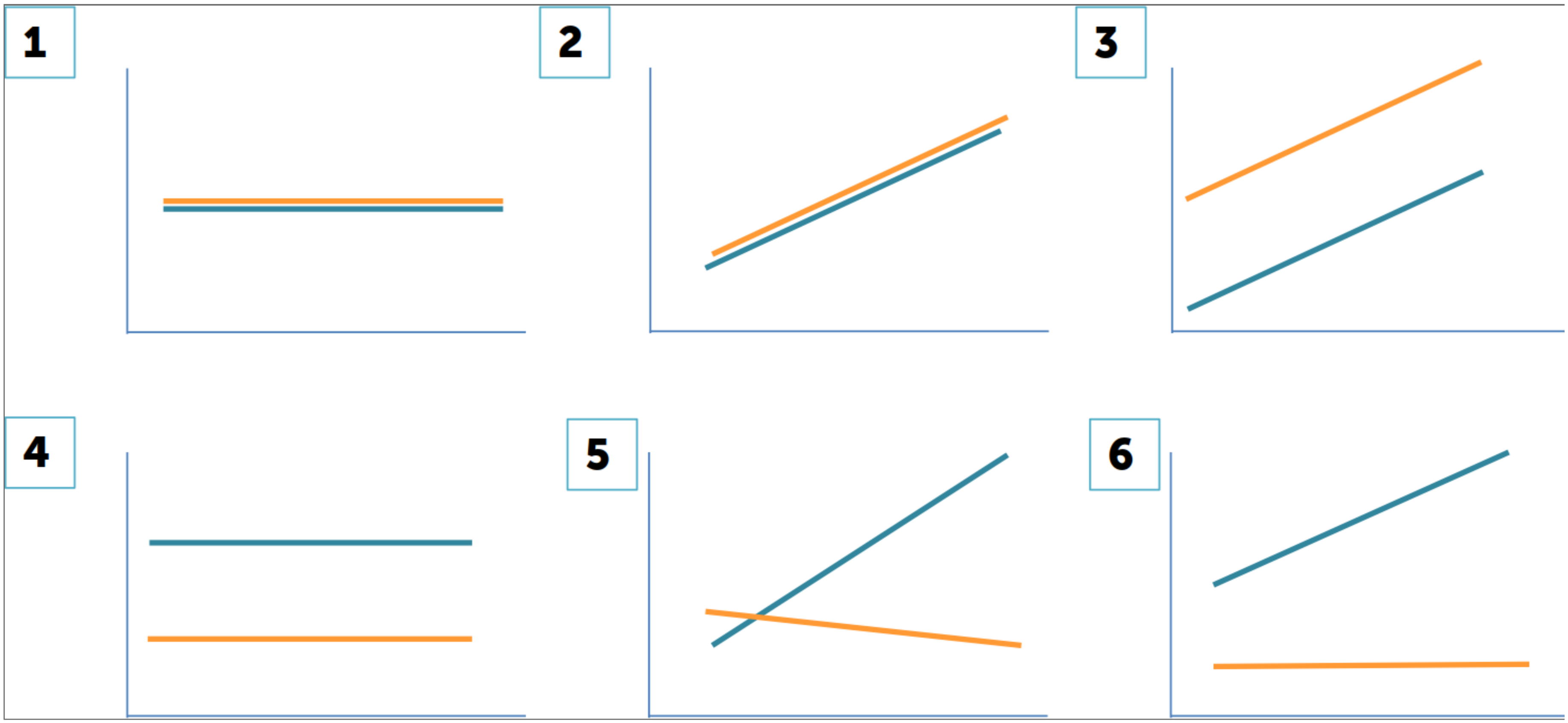
Complete for women:

- $$\begin{aligned}\hat{Y}_i = & a + b_1 * 1 + b_2 * X_{2i} + b_3 * (1 * X_{2i}) = \\ & a + b_1 + b_2 * X_{2i} + b_3 * X_{2i}\end{aligned}$$
- So the regression line for women is $(a + b_1) + (b_2 + b_3) * X_{2i}$
- An extra "bump" on top of the intercept and slope

Examples

Which parameters are non-zero?

$$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i} + b3(X_{1i} X_{2i})$$



Simple Effects

Simple effects

If the interaction is significant, we might ask:

> What is the effect of X1 on Y for each level of G?

For categorical moderator

Easy trick to obtain the effect for each category:

- Create k dummies for a categorical variable with k categories (instead of the usual k-1)
- Compute interaction term with each dummy
- Specify regression with all these interaction effects, and without the main effect of the continuous variable

Formulas

Standard model:

$$\hat{Y}' = b_0 + b_1 X1 + b_2 D_2 + b_3 D_3 + b_4(D_2 X_1) + b_5(D_3 X_1)$$

- Main effect of the continuous predictor and k-1 dummies
- Interaction effect with k-1 dummies

$$\hat{Y}' = b_0 + b_1 D_2 + b_2 D_3 + b_3(D_1 X_1) + b_4(D_2 X_1) + b_5(D_3 X_1)$$

- Main effect of k-1 dummies
- NO main effect of X_1
- k interaction terms (so the effect of X_1 is split across all k categories)

Filling in Formulas

Effect of X_1 for group 1 ($D_2 = 0, D_3 = 0$):

$\hat{Y}' = b_0 + b_1 X1 + b_2 * 0 + b_3 * 0 + b_4(0 * X_1) + b_5(0 * X_1) = b_0 + b_1 * X1$

Effect of X_1 for group 2 ($D_2 = 1, D_3 = 0$):

$\hat{Y}' = b_0 + b_1 X1 + b_2 * 1 + b_3 * 0 + b_4(1 * X_1) + b_5(0 * X_1) = (b_0 + b_2) + (b_1 + b_4) X_1$

Using the model with all interaction effects:

Effect of X_1 for group 1 ($D_2 = 0, D_3 = 0$):

$\hat{Y}' = b_0 + b_1 * 0 + b_2 * 0 + b_3(1 * X_1) + b_4(0 * X_1) + b_5(0 * X_1) = b_0 + b_3 * X_1$

Effect of X_1 for group 2 ($D_2 = 1, D_3 = 0$):

$\hat{Y}' = b_0 + b_1 * 1 + b_2 * 0 + b_3(0 * X_1) + b_4(1 * X_1) + b_5(0 * X_1) = (b_0 + b_1) + b_4 X_1$

Two continuous predictors

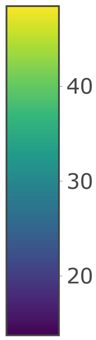
Difference with previous example

- An interaction between one binary and one continuous predictor results in **two regression lines**
  - One for each unique value of the binary predictor
- An interaction betweeb two continuous predictors also gives us a unique regression line for every value of each predictor
  - But both predictors can take on infinite unique values

Binary vs continuous interaction

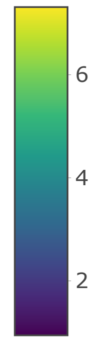
Multiple regression

Multiple regression demo



Multiple regression with interaction

Multiple regression demo



Complete the formula

$$\hat{Y}_i = a + b_1 * X_{1i} + b_2 * X_{2i} + b_3 * (X_{1i} * X_{2i})$$

- Y is involvement, X1 is gender roles, X2 is work
- Imagine we have found these coefficients:

\hat{Y}_i &= 12.50 + 1.50*X_{1i} - .20 * X_{2i} + 0.07 * (X_{1i} * X_{2i})

What's the effect of gender roles for someone who works 40 hours?

$$\hat{Y}_i = 12.50 + 1.50 * X_{1i} - .20 * 40 + 0.07 * (40 * X_{1i})$$

$$\hat{Y}_i = (12.50 - .20 * 40) + (1.50 + 0.07 * 40) * X_{1i} = 4.5 + 4.3 * X_{1i}$$

Complete the formula

$\hat{Y}_i = a + b_1 * X_{1i} + b_2 * X_{2i} + b_3 * (X_{1i} * X_{2i})$

Imagine we have found these coefficients:

$\hat{Y}_i = 12.50 + 1.50 * X_{1i} - .20 * X_{2i} + 0.07 * (X_{1i} * X_{2i})$

What’s the effect of work hours for someone who scores 0 on gender roles?

$\hat{Y}_i = 12.50 + 1.50 * 0 - .20 * X_{2i} + 0.07 * (0 * X_{2i})$

$\hat{Y}_i = (12.50 + 1.50 * 0) - (.20 + 0.07 * 0) * X_{2i} = 12.50 - .20 * X_{2i}$

Centering

As the effect of X1 is now dependent on the value of X2, and vice versa, it's essential to **center the variables**
Not centering…

- Makes coefficients hard to interpret
- Introduces (artificial) multicolinearity

When you center, the interpretation is:

- a: Expected value of Y for people who score average on all predictors
- b1: Slope of predictor 1 for people who score average on predictor 2
- b2: Slope of predictor 2 for people who score average on predictor 1

Simple Slopes

Simple Slopes

If the interaction is significant, we might ask:

> What is the effect of X1 on Y for different levels of X2?

With a categorical moderator, we obtained the effect of X1 for each discrete level of X2
With a continuous moderator, we need to pick specific values of X2

Which one is the moderator?

Mathematically, there is no difference between saying:

- The effect of X1 depends on the value of X2
- The effect of X2 depends on the value of X1
- Because $X1 * X2$ is the same as $X2 * X1$

Theory makes the difference! You decide which variable moderates the effect of the other

Use centering

Remember that multiple regression gives you the effect of each predictor while controlling all other predictors at zero

- You can change the zero-value
- You've done this before: by centering variables you make the zero value equal to "the mean" of that variable
- With centered predictors centered at the mean, regression with interaction gives us the effect of X1 for people with an average score on X2

Center to high and low values

To get the effect of X1 for people who score high on X2, just center X2 to a high value!

- Center X2 at +1 SD
- Center X2 at -1 SD

Recompute the regression for each of these re-centered variables

Simple slopes steps

1. Center the interacting predictors at their mean value
2. Compute the interaction term
3. Determine whether the interaction is significant
4. Center X2 at +1SD
5. Re-compute the interaction term
6. Note the effect of X1 for this level of X2
7. Center X2 at -1SD
8. Re-compute the interaction term
9. Note the effect of X1 for this level of X2

Important

- To center at +1SD you have to **subtract** 1 SD from the centered variable
- To center at -1SD you have to **add** 1 SD from the centered variable

Important 2

After changing how a variable is centered, you have to re-compute the interaction term

- Make sure you always use the correctly centered variable and its corresponding interaction term
- Using syntax helps prevent mistakes

Error

×