HOW TO APPLY THE MULTIPHASE OPTIMIZATION STRATEGY (MOST) IN YOUR INTERVENTION DEVELOPMENT RESEARCH

Module 4
Some conceptual and technical aspects of the factorial experiment

Lesson 1: The regression model and coding

NYU | SCHOOL OF GLOBAL PUBLIC HEALTH

Intervention Optimization Initiative

This course was developed by

Linda M. Collins

School of Global Public Health

New York University

(narrator)

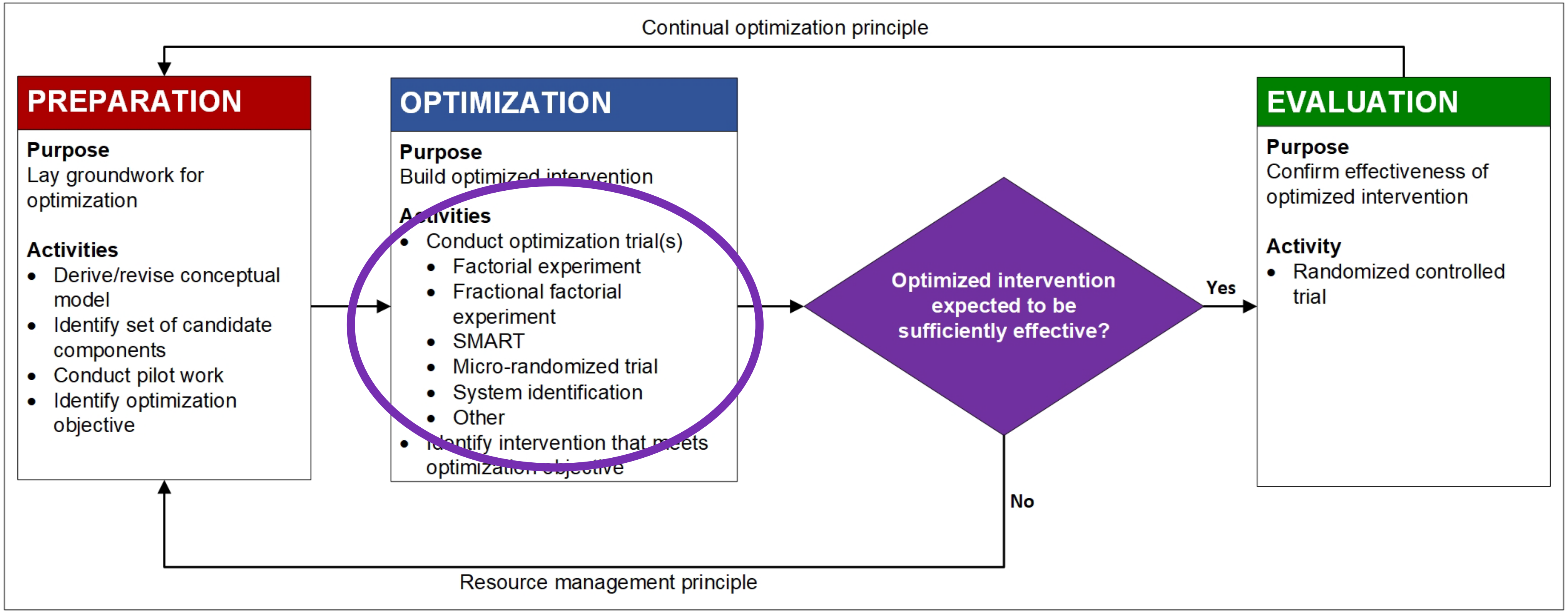
Kate Guastaferro

College of Health and Human Development

The Pennsylvania State University

In this lesson you will learn how to:

- Describe the regression model for a factorial experiment
- Explain the difference between dummy coding and effect coding



Flow chart of the three phases of the multiphase optimization strategy (MOST). Rectangle = action. Diamond = decision.

Figure adapted from Collins, L.M. (2018)

What are we trying to estimate with a factorial experiment? Main effects and interactions

- Main effect of each factor
  - DEFINITION OF MAIN EFFECT OF FACTOR *A*:
    - Effect of Factor *A* averaged across all levels of all other factors

What are we trying to estimate with a factorial experiment? Main effects and interactions

- Interactions
  - DEFINITION OF INTERACTION BETWEEN FACTOR *A* AND FACTOR *B* (assuming each factor has two levels):
    - ½ [(effect of Factor *A* at level 1 of Factor *B*) – (effect of Factor *A* at level 2 of Factor *B*)] averaged across all levels of all other factors

- Suppose we conducted this experiment and now want to analyze the data.

Experimental Condition (Cell)	Factorial Experiment				
	MI	*PEER*	*TEXT*	*outcome*	*n*
1	No	No	No	$\bar{Y}_1$	40
2	No	No	Yes	$\bar{Y}_2$	40
3	No	Yes	No	$\bar{Y}_3$	40
4	No	Yes	Yes	$\bar{Y}_4$	40
5	Yes	No	No	$\bar{Y}_5$	40
6	Yes	No	Yes	$\bar{Y}_6$	40
7	Yes	Yes	No	$\bar{Y}_7$	40
8	Yes	Yes	Yes	$\bar{Y}_8$	40

Suppose we have conducted the 2^3 experiment and want to analyze the data

- We can perform Analysis of Variance (ANOVA) via regression
- We need to specify a regression model to predict Y

Suppose we have conducted the 2^3 experiment and want to analyze the data

- In a full ANOVA model for this 2^3 experiment there are:
  - Three main effects (*MI*, *PEER*, *TEXT*)
  - Three 2-way interactions (*MI*×*PEER*, *MI*×*TEXT*, *PEER*×*TEXT*)
  - One 3-way interaction (*MI*×*PEER*×*TEXT*)
- These effects can be estimated via regression

There is a $\hat{Y}$ corresponding to each experimental condition, e.g.:

$\hat{Y}_3$ is the predicted Y when

- *MI*=No,
- *PEER*=Yes, and
- *TEXT*=No.

Experimental Condition (Cell)	Factorial Experiment				
	MI	*PEER*	*TEXT*	*predicted outcome*	*n*
1	No	No	No	$\hat{Y}_1$	40
2	No	No	Yes	$\hat{Y}_2$	40
3	No	Yes	No	$\hat{Y}_3$	40
4	No	Yes	Yes	$\hat{Y}_4$	40
5	Yes	No	No	$\hat{Y}_5$	40
6	Yes	No	Yes	$\hat{Y}_6$	40
7	Yes	Yes	No	$\hat{Y}_7$	40
8	Yes	Yes	Yes	$\hat{Y}_8$	40

The regression model

$$\hat{Y} = \underbrace{\beta_0}_{\text{Intercept}} + \underbrace{\beta_1 X_{MI} + \beta_2 X_{PEER} + \beta_3 X_{TEXT}}_{\text{Main effects}} + \underbrace{\beta_4 X_{MI \times PEER} + \beta_5 X_{MI \times TEXT} + \beta_6 X_{PEER \times TEXT} + \beta_7 X_{MI \times PEER \times TEXT}}_{\text{Interactions}}$$

The βs correspond to effect estimates.

Where do the X's come from?

Coding main effects: Dummy coding

Dummy Codes							
Experimental Condition (cell)	Main effects			Interactions			
	MI	*PEER*	*TEXT*				
	X_1	X_2	X_3				
1	0	0	0				
2	0	0	1				
3	0	1	0				
4	0	1	1				
5	1	0	0				
6	1	0	1				
7	1	1	0				
8	1	1	1				

Experimental Condition (Cell)	Factorial Experiment		
	MI	*PEER*	*TEXT*
1	No	No	No
2	No	No	Yes
3	No	Yes	No
4	No	Yes	Yes
5	Yes	No	No
6	Yes	No	Yes
7	Yes	Yes	No
8	Yes	Yes	Yes

Coding main effects: Effect coding

Effect Codes							
Experimental Condition (cell)	Main effects			Interactions			
	MI	*PEER*	*TEXT*				
	X_1	X_2	X_3				
1	-1	-1	-1				
2	-1	-1	+1				
3	-1	+1	-1				
4	-1	+1	+1				
5	+1	-1	-1				
6	+1	-1	+1				
7	+1	+1	-1				
8	+1	+1	+1				

Experimental Condition (Cell)	Factorial Experiment		
	MI	*PEER*	*TEXT*
1	No	No	No
2	No	No	Yes
3	No	Yes	No
4	No	Yes	Yes
5	Yes	No	No
6	Yes	No	Yes
7	Yes	Yes	No
8	Yes	Yes	Yes

Coding main effects: Effect coding

Effect Codes							
Experimental Condition (cell)	Main effects			Interactions			
	MI	*PEER*	*TEXT*				
	X_1	X_2	X_3				
1	-1	-1	-1				
2	-1	-1	+1				
3	-1	+1	-1				
4	-1	+1	+1				
5	+1	-1	-1				
6	+1	-1	+1				
7	+1	+1	-1				
8	+1	+1	+1				

Experimental Condition (Cell)	Factorial Experiment		
	MI	*PEER*	*TEXT*
1	No	No	No
2	No	No	Yes
3	No	Yes	No
4	No	Yes	Yes
5	Yes	No	No
6	Yes	No	Yes
7	Yes	Yes	No
8	Yes	Yes	Yes

Coding main effects: Effect coding

Effect Codes							
Experimental Condition (cell)	Main effects			Interactions			
	MI	*PEER*	*TEXT*				
	X_1	X_2	X_3				
1	-1	-1	-1				
2	-1	-1	+1				
3	-1	+1	-1				
4	-1	+1	+1				
5	+1	-1	-1				
6	+1	-1	+1				
7	+1	+1	-1				
8	+1	+1	+1				

Experimental Condition (Cell)	Factorial Experiment		
	MI	*PEER*	*TEXT*
1	No	No	No
2	No	No	Yes
3	No	Yes	No
4	No	Yes	Yes
5	Yes	No	No
6	Yes	No	Yes
7	Yes	Yes	No
8	Yes	Yes	Yes

Coding main effects: Effect coding

Effect Codes							
Experimental Condition (cell)	Main effects			Interactions			
	MI	*PEER*	*TEXT*				
	X_1	X_2	X_3				
1	-1	-1	-1				
2	-1	-1	+1				
3	-1	+1	-1				
4	-1	+1	+1				
5	+1	-1	-1				
6	+1	-1	+1				
7	+1	+1	-1				
8	+1	+1	+1				

Experimental Condition (Cell)	Factorial Experiment		
	MI	*PEER*	*TEXT*
1	No	No	No
2	No	No	Yes
3	No	Yes	No
4	No	Yes	Yes
5	Yes	No	No
6	Yes	No	Yes
7	Yes	Yes	No
8	Yes	Yes	Yes

Coding interaction effects: Effect coding

Effect Codes							
Experimental Condition (cell)	Main effects			Interactions			
	MI	*PEER*	*TEXT*	*MI × PEER*	*MI × TEXT*	*PEER × TEXT*	*MI ×PEER × TEXT*
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	-1	-1	-1	+1	+1	+1	-1
2	-1	-1	+1	+1	-1	-1	+1
3	-1	+1	-1	-1	+1	-1	+1
4	-1	+1	+1	-1	-1	+1	-1
5	+1	-1	-1	-1	-1	+1	+1
6	+1	-1	+1	-1	+1	-1	-1
7	+1	+1	-1	+1	-1	-1	-1
8	+1	+1	+1	+1	+1	+1	+1

Coding interaction effects: Effect coding

Effect Codes							
Experimental Condition (cell)	Main effects			Interactions			
	MI	*PEER*	*TEXT*	*MI × PEER*	*MI × TEXT*	*PEER × TEXT*	*MI ×PEER × TEXT*
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	-1	-1	-1	+1	+1	+1	-1
2	-1	-1	+1	+1	-1	-1	+1
3	-1	+1	-1	-1	+1	-1	+1
4	-1	+1	+1	-1	-1	+1	-1
5	+1	-1	-1	-1	-1	+1	+1
6	+1	-1	+1	-1	+1	-1	-1
7	+1	+1	-1	+1	-1	-1	-1
8	+1	+1	+1	+1	+1	+1	+1

Coding interaction effects: Effect coding

Effect Codes							
Experimental Condition (cell)	Main effects			Interactions			
	MI	*PEER*	*TEXT*	*MI × PEER*	*MI × TEXT*	*PEER × TEXT*	*MI ×PEER × TEXT*
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	-1	-1	-1	+1	+1	+1	-1
2	-1	-1	+1	+1	-1	-1	+1
3	-1	+1	-1	-1	+1	-1	+1
4	-1	+1	+1	-1	-1	+1	-1
5	+1	-1	-1	-1	-1	+1	+1
6	+1	-1	+1	-1	+1	-1	-1
7	+1	+1	-1	+1	-1	-1	-1
8	+1	+1	+1	+1	+1	+1	+1

Coding interaction effects: Effect coding

Effect Codes							
Experimental Condition (cell)	Main effects			Interactions			
	MI	*PEER*	*TEXT*	*MI × PEER*	*MI × TEXT*	*PEER × TEXT*	*MI ×PEER × TEXT*
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	-1	-1	-1	+1	+1	+1	-1
2	-1	-1	+1	+1	-1	-1	+1
3	-1	+1	-1	-1	+1	-1	+1
4	-1	+1	+1	-1	-1	+1	-1
5	+1	-1	-1	-1	-1	+1	+1
6	+1	-1	+1	-1	+1	-1	-1
7	+1	+1	-1	+1	-1	-1	-1
8	+1	+1	+1	+1	+1	+1	+1

To sum up:

- To analyze the data from a factorial experiment using a regression model, it's necessary to use numeric codes to represent each main effect and interaction
- This can be done using dummy codes (0,1) or effect codes (for a 2^k, -1,1)

To sum up:

- All else being equal, effect coding and dummy coding produce the same omnibus F
- However, these two approaches define the main effects and interactions differently
- This means hypothesis tests for individual main and interaction effects will be different!

To sum up:

- In this course, we use ONLY effect (-1,1) coding
- In your data analysis, be sure you know which coding is being used
- For more about this, see the Kugler, Dziak, and Trail chapter in Collins & Kugler (2018)

The regression model

$$\hat{Y} = \underset{\text{Intercept}}{\beta_0} + \underbrace{\beta_1 X_{MI} + \beta_2 X_{PEER} + \beta_3 X_{TEXT}}_{\text{Main effects}} + \underbrace{\beta_4 X_{MI \times PEER} + \beta_5 X_{MI \times TEXT} + \beta_6 X_{PEER \times TEXT} + \beta_7 X_{MI \times PEER \times TEXT}}_{\text{Interactions}}$$

The βs correspond to effect estimates.

Recall: Definition of optimization of an intervention used in MOST

Optimization of a multicomponent intervention is the process of identifying an intervention that provides the best expected outcome obtainable within key constraints imposed by the need for affordability, scalability, and/or efficiency.

The $\hat{Y}$s are estimates of the **expected outcome** associated with each experimental condition.

The regression model

$$\hat{Y} = \underbrace{\beta_0}_{\text{Intercept}} + \underbrace{\boxed{\beta_1 X_{MI}} + \beta_2 X_{PEER} + \beta_3 X_{TEXT}}_{\text{Main effects}} + \underbrace{\beta_4 X_{MI \times PEER} + \beta_5 X_{MI \times TEXT} + \beta_6 X_{PEER \times TEXT} + \beta_7 X_{MI \times PEER \times TEXT}}_{\text{Interactions}}$$

The βs correspond to effect estimates.

Coding effects

Through tedious but not difficult algebra you can show that

$$\beta_1 = Main\ Effect_{MI}/2$$

The regression model

$$\hat{Y} = \underbrace{\beta_0}_{\text{Intercept}} + \underbrace{\beta_1 X_{MI} + \beta_2 X_{PEER} + \beta_3 X_{TEXT}}_{\text{Main effects}} + \underbrace{\boxed{\beta_4 X_{MI \times PEER}} + \beta_5 X_{MI \times TEXT} + \beta_6 X_{PEER \times TEXT} + \beta_7 X_{MI \times PEER \times TEXT}}_{\text{Interactions}}$$

The βs correspond to effect estimates.

Coding effects

Through tedious but not difficult algebra you can show that for a 2^k experiment (**BUT THERE IS AN IMPORTANT CAVEAT****):

$$\beta_4 = Interaction_{MI \times PEER}/2$$

****IMPORTANT CAVEAT:** This equation holds for one definition of the interaction. See sections 3.14 and 4.7 in Collins (2018).

In this lesson you learned how to:

- Describe the regression model for a factorial experiment
- Explain the difference between dummy coding and effect coding

In the next lesson you will learn how to:

- Interpret main effects and interaction effects

References cited

- Collins, L.M. (2018). *Optimization of behavioral, biobehavioral, and biomedical interventions: The multiphase optimization strategy (MOST).* New York: Springer.
- Kugler, K.C., Dziak, J.J., & Trail, J. (2018). Coding and interpretation of effects in analysis of data from a factorial experiment. In L.M. Collins & K.C. Kugler (Eds.), *Optimization of behavioral, biobehavioral, and biomedical interventions: Advanced topics.* New York: Springer.