Lesson 12: Interactions, Part 2

Nicky Wakim

2026-02-25

Learning Objectives

Last time:

1. Define confounders and effect modifiers, and how they interact with the main relationship we model.
2. Interpret the interaction component of a model with **a binary categorical covariate and continuous covariate**, and how the main variable's effect changes.
3. Interpret the interaction component of a model with **a multi-level categorical covariate and continuous covariate**, and how the main variable's effect changes.

This time:

4. Interpret the interaction component of a model with **two categorical covariates**, and how the main variable's effect changes.
5. Interpret the interaction component of a model with **two continuous covariates**, and how the main variable's effect changes.
6. Report results for a best-fit line (with confidence intervals) at different levels of an effect measure modifier

Learning Objectives

Last time:

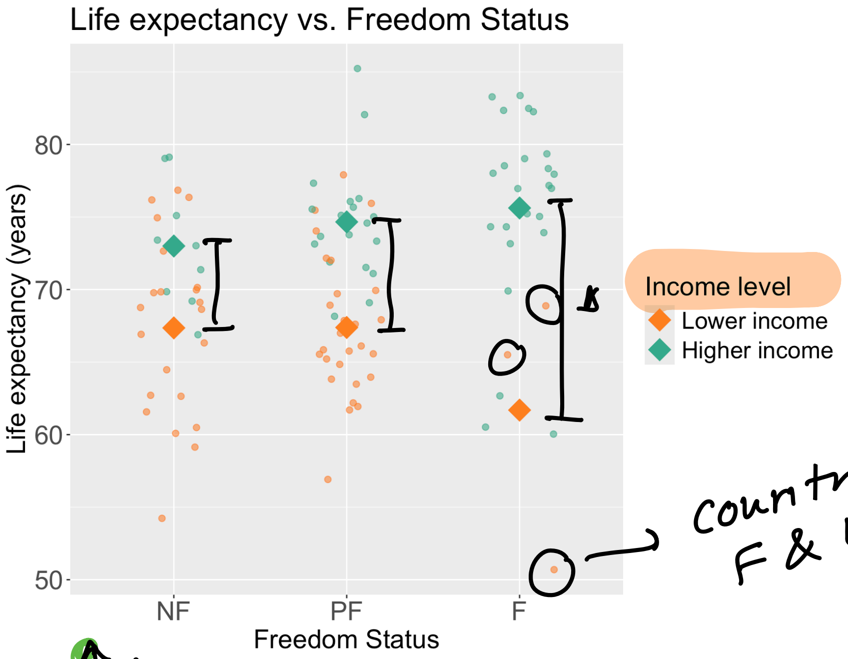
1. Define confounders and effect modifiers, and how they interact with the main relationship we model.
2. Interpret the interaction component of a model with **a binary categorical covariate and continuous covariate**, and how the main variable's effect changes.
3. Interpret the interaction component of a model with **a multi-level categorical covariate and continuous covariate**, and how the main variable's effect changes.

This time:

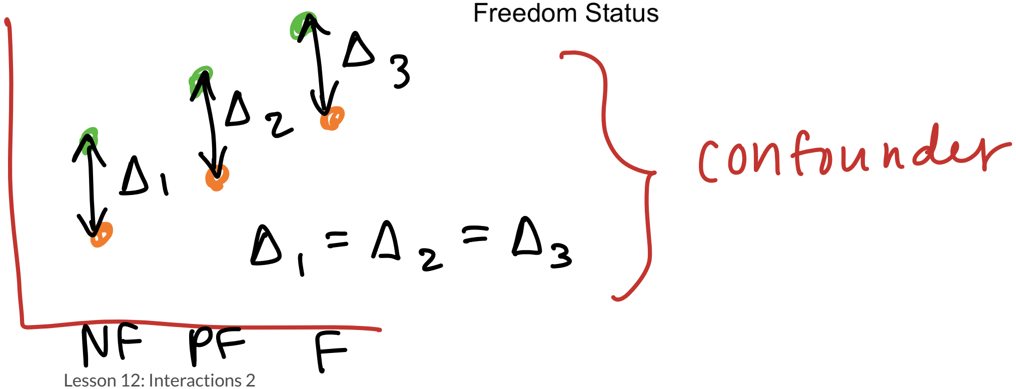
4. Interpret the interaction component of a model with **two categorical covariates**, and how the main variable's effect changes.
5. Interpret the interaction component of a model with **two continuous covariates**, and how the main variable's effect changes.
6. Report results for a best-fit line (with confidence intervals) at different levels of an effect measure modifier

Do we think income level can be an effect modifier for freedom status?

- Taking a break from cell phones to demonstrate interactions for two categorical variables
- We can start by visualizing the relationship between life expectancy and freedom status *by income level*
- Questions of interest: Does the effect of freedom status on life expectancy differ depending on income level?
  - This is the same as: Is income level an effect modifier for freedom status?
- Let's run an interaction model to see!



country F & LI



Model with interaction between a *multi-level categorical and binary variables*

ref grps
- low inc
- Not free

Model we are fitting:

$$
\begin{aligned}
LE = &\beta_0 + \beta_1 I(\text{high income}) + \beta_2 I(\text{FS} = \text{PF}) + \beta_3 I(\text{FS} = \text{F}) + \\
&\beta_4 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + \beta_5 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F}) + \epsilon
\end{aligned}
$$

main effects

- LE as life expectancy
- $I(\text{high income})$ as indicator of high income
- $I(\text{FS} = \text{PF})$ and $I(\text{FS} = \text{F})$ as the indicator for each freedom status

In R:

```
m_int_wr_inc = gapm %>%
  lm(formula = life_exp ~ income_level_2 + freedom_status + income_level_2*freedom_status)
m_int_wr_inc = gapm %>%
  lm(formula = life_exp ~ income_level_2*freedom_status)
```

Displaying the regression table and writing fitted regression equation

```
tidy_m_int_wr_inc = tidy(m_int_wr_inc, conf.int=T)
tidy_m_int_wr_inc %>% gt() %>% tab_options(table.font.size = 35) %>% fmt_number(decimals =
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	67.355	1.179	57.126	0.000	65.015	69.695
income_level_2Higher income	5.645	2.188	2.580	0.011	1.303	9.987
freedom_statusPF	0.029	1.588	0.018	0.985	−3.123	3.181
freedom_statusF	−5.665	3.404	−1.664	0.099	−12.419	1.089
income_level_2Higher income:freedom_statusPF	1.633	2.744	0.595	0.553	−3.813	7.078
income_level_2Higher income:freedom_statusF	8.287	4.026	2.058	0.042	0.299	16.275

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 I(\text{high income}) + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \widehat{\beta}_4 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F})$$

$$\widehat{LE} = 67.35 + 5.64 \cdot I(\text{high income}) + 0.03 \cdot I(\text{FS} = \text{PF}) - 5.67 \cdot I(\text{FS} = \text{F}) + 1.63 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + 8.29 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F})$$

Poll Everywhere Question

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 I(\text{high income}) + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \widehat{\beta}_4 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F})$$

$$\widehat{LE} = 67.35 + 5.64 \cdot I(\text{high income}) + 0.03 \cdot I(\text{FS} = \text{PF}) - 5.67 \cdot I(\text{FS} = \text{F}) + 1.63 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + 8.29 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F})$$

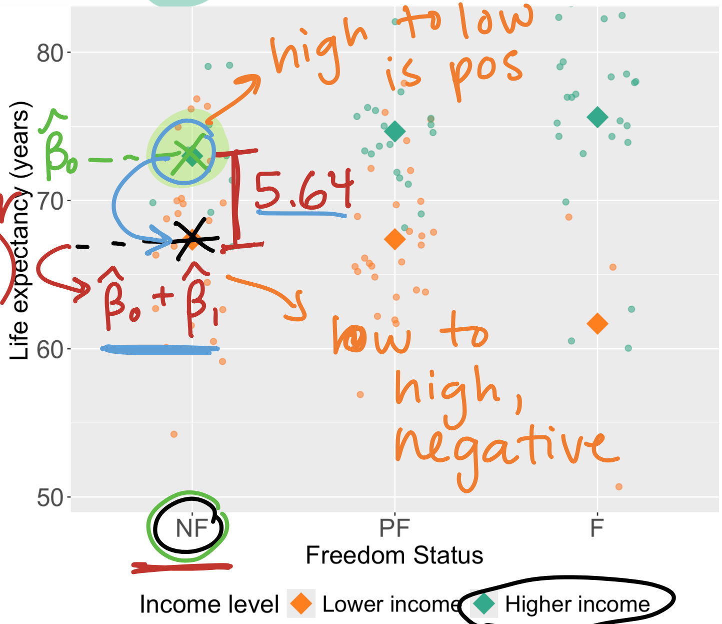
13:22 Wed Feb 25

Join by Web PollEv.com/nickywakim275

What would happen to our fitted interaction coefficients if we make high income the reference instead?

$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 I(\text{lower inc})$

Magnitude and sign of estimate would change	34%
Magnitude and sign of estimate would not change	0%
Magnitude would change and sign of estimate would not change	6%
✅ Magnitude would not change and sign of estimate would change	60%



$\widehat{\beta}_0$

5.64

$\widehat{\beta}_0 + \widehat{\beta}_1$

high to low is pos

low to high, negative

Ref

I(Lower income)

Comparing fitted regression *means* for each freedom status

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 I(\text{high income}) + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \\ & \widehat{\beta}_4 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F})\end{aligned}$$

$$\begin{aligned}\widehat{LE} =& 67.35 + 5.64 \cdot I(\text{high income}) + 0.03 \cdot I(\text{FS} = \text{PF}) - 5.67 \cdot I(\text{FS} = \text{F}) + \\ & 1.63 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + 8.29 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F})\end{aligned}$$

Not free

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 I(\text{high income}) + \\ & \widehat{\beta}_2 \cdot 0 + \widehat{\beta}_3 \cdot 0 + \\ & \widehat{\beta}_4 \cdot I(\text{high income}) \cdot 0 + \\ & \widehat{\beta}_5 \cdot I(\text{high income}) \cdot 0 \\ \widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 I(\text{high income})\end{aligned}$$

$I(FS = PF) = 0$

$I(FS = F) = 0$

Partly free

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 I(\text{high income}) + \\ & \widehat{\beta}_2 \cdot 1 + \widehat{\beta}_3 \cdot 0 + \\ & \widehat{\beta}_4 \cdot I(\text{high income}) \cdot 1 + \\ & \widehat{\beta}_5 \cdot I(\text{high income}) \cdot 0 \\ \widehat{LE} =& (\widehat{\beta}_0 + \widehat{\beta}_2) + \\ & (\widehat{\beta}_1 + \widehat{\beta}_4) I(\text{high income})\end{aligned}$$

Free

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 I(\text{high income}) + \\ & \widehat{\beta}_2 \cdot 1 + \widehat{\beta}_3 \cdot 0 + \\ & \widehat{\beta}_4 \cdot I(\text{high income}) \cdot 1 + \\ & \widehat{\beta}_5 \cdot I(\text{high income}) \cdot 0 \\ \widehat{LE} =& (\widehat{\beta}_0 + \widehat{\beta}_3) + \\ & (\widehat{\beta}_1 + \widehat{\beta}_5) I(\text{high income})\end{aligned}$$

Comparing fitted regression *means* for each income level

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 I(\text{high income}) + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \\ & \widehat{\beta}_4 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F}) \\ \widehat{LE} =& 67.35 + 5.64 \cdot I(\text{high income}) + 0.03 \cdot I(\text{FS} = \text{PF}) - 5.67 \cdot I(\text{FS} = \text{F}) + \\ & 1.63 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + 8.29 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F})\end{aligned}$$

For lower income countries: $I(\text{high income}) = 0$

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 \cdot 0 + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \\ & \widehat{\beta}_4 \cdot 0 \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 \cdot 0 \cdot I(\text{FS} = \text{F}) \\ \widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F})\end{aligned}$$

For higher income countries: $I(\text{high income}) = 1$

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 \cdot 1 + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \\ & \widehat{\beta}_4 \cdot 1 \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 \cdot 1 \cdot I(\text{FS} = \text{F}) \\ \widehat{LE} =& \left(\widehat{\beta}_0 + \widehat{\beta}_1\right) + \left(\widehat{\beta}_2 + \widehat{\beta}_4\right) I(\text{FS} = \text{PF}) + \\ & \left(\widehat{\beta}_3 + \widehat{\beta}_5\right) I(\text{FS} = \text{F})\end{aligned}$$

- Example interpretation: The America's effect on mean life expectancy increases $\widehat{\beta}_5$ comparing high income to low income countries.

Let's take a look back at the plot

confounders → parallel w/ consistent diff

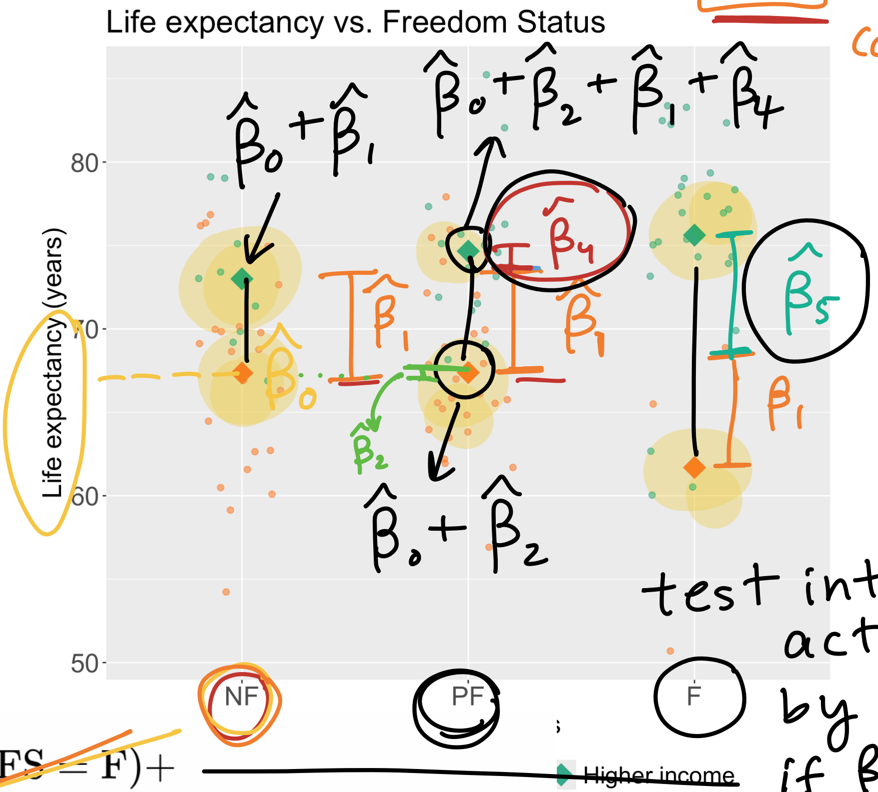
interaction

For lower income countries: $I(\text{high income}) = 0$

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F})$$

For higher income countries: $I(\text{high income}) = 1$

$$\widehat{LE} = (\widehat{\beta}_0 + \widehat{\beta}_1) + (\widehat{\beta}_2 + \widehat{\beta}_4) I(\text{FS} = \text{PF}) + (\widehat{\beta}_3 + \widehat{\beta}_5) I(\text{FS} = \text{F})$$



test interaction by seeing if $\beta_4 = 0$ & $\beta_5 = 0$

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 I(\text{high income}) + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \widehat{\beta}_4 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F})$$

$$\widehat{LE} = 67.35 + 5.64 \cdot I(\text{high income}) + 0.03 \cdot I(\text{FS} = \text{PF}) - 5.67 \cdot I(\text{FS} = \text{F}) + 1.63 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + 8.29 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F})$$

Interpretation for interaction between two categorical variables

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 \cdot 1 + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \\ & \widehat{\beta}_4 \cdot 1 \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 \cdot 1 \cdot I(\text{FS} = \text{F})\end{aligned}$$

$$\begin{aligned}\widehat{LE} =& \left[\widehat{\beta}_0 + \widehat{\beta}_1 \cdot I(\text{high income})\right] + \underbrace{\left[\widehat{\beta}_2 + \widehat{\beta}_4 \cdot I(\text{high income})\right]}_{\text{the effect of PF countries}} I(\text{FS} = \text{PF}) + \\ & \underbrace{\left[\widehat{\beta}_3 + \widehat{\beta}_5 \cdot I(\text{high income})\right]}_{\text{effect of free countries}} I(\text{FS} = \text{F})\end{aligned}$$

- Interpretation:
  - $\widehat{\beta}_1$ = mean change in life expectancy comparing high income to low income countries, for countries/territories that are not free
  - $\widehat{\beta}_4$ = mean change in life expectancy comparing high income to low income countries, for countries/territories that are partly free
  - $\widehat{\beta}_5$ = mean change in life expectancy comparing high income to low income countries, for countries/territories that are free

Test interaction between two categorical variables (1/2)

- We run an F-test for a group of coefficients (β_4, β_5) in the below model (see lesson 9)

$$
\begin{aligned}
LE =&\beta_0 + \beta_1 I(\text{high income}) + \beta_2 I(\text{FS} = \text{PF}) + \beta_3 I(\text{FS} = \text{F}) + \\
&\beta_4 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + \beta_5 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F}) + \epsilon
\end{aligned}
$$

Null H_0	Alternative H_1
$\beta_4 = \beta_5 = 0$	$\beta_4 \neq 0$ and/or $\beta_5 \neq 0$

Null / Smaller / Reduced model

$$
\begin{aligned}
LE =&\beta_0 + \beta_1 I(\text{high income}) + \\
&\beta_2 I(\text{FS} = \text{PF}) + \beta_3 I(\text{FS} = \text{F}) + \\
&\epsilon
\end{aligned}
$$

FS & inc as confounders

Alternative / Larger / Full model

$$
\begin{aligned}
LE =&\beta_0 + \beta_1 I(\text{high income}) + \beta_2 I(\text{FS} = \text{PF}) + \\
&\beta_3 I(\text{FS} = \text{F}) + \beta_4 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{PF}) + \\
&\beta_5 \cdot I(\text{high income}) \cdot I(\text{FS} = \text{F}) + \epsilon
\end{aligned}
$$

FS & inc interaction

Test interaction between two categorical variables (2/2)

- Fit the reduced and full model

```r
m_int_wr_inc_red = gapm %>%
  lm(formula = life_exp ~ income_level_2 + freedom_status)
m_int_wr_inc_full = gapm %>%
  lm(formula = life_exp ~ income_level_2*freedom_status)
```

▶ Display the ANOVA table with F-statistic and p-value

term	df.residual	rss	df	sumsq	statistic	p.value
life_exp ~ income_level_2 + freedom_status	101.000	3,160.771	NA	NA	NA	NA
life_exp ~ income_level_2 * freedom_status	99.000	3,027.855	2.000	132.915	2.173	0.119

- **Conclusion:** There is not a significant interaction between freedom status and income level (p = 0.119).

Learning Objectives

Last time:

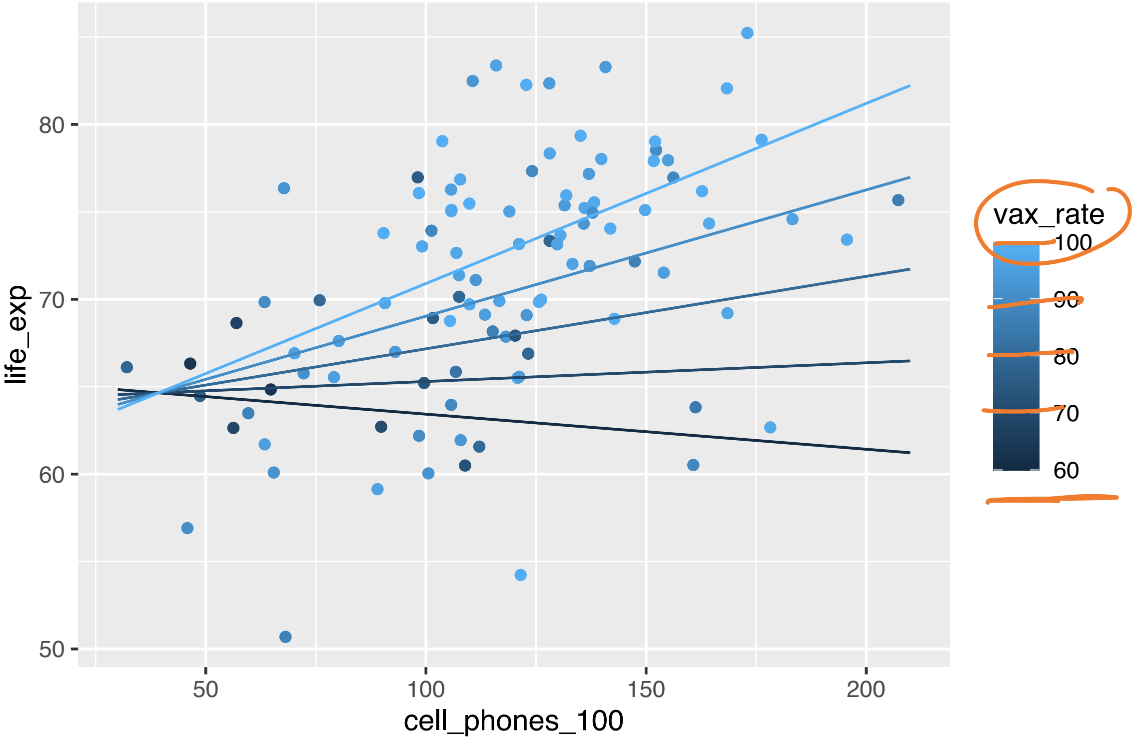
1. Define confounders and effect modifiers, and how they interact with the main relationship we model.
2. Interpret the interaction component of a model with **a binary categorical covariate and continuous covariate**, and how the main variable's effect changes.
3. Interpret the interaction component of a model with **a multi-level categorical covariate and continuous covariate**, and how the main variable's effect changes.

This time:

4. Interpret the interaction component of a model with **two categorical covariates**, and how the main variable's effect changes.
5. Interpret the interaction component of a model with **two continuous covariates**, and how the main variable's effect changes.
6. Report results for a best-fit line (with confidence intervals) at different levels of an effect measure modifier

Do we think vaccination rate is an effect modifier for cell phones?

- We can start by visualizing the relationship between life expectancy and cell phones *by vaccination rate*
- Questions of interest: Does the effect of cell phones on life expectancy differ depending on vaccination rate?
  - This is the same as: Is vaccination rate is an effect modifier for cell phones? Is vaccination rate an effect modifier of the association between life expectancy and cell phones?
- Let's run an interaction model to see!



Model with interaction between *two continuous variables*

Model we are fitting:

$$LE = \beta_0 + \beta_1 CP^c + \beta_2 VR^c + \beta_3 CP^c \cdot VR^c + \epsilon$$

- LE as life expectancy
- CP^c as the **centered** around the mean number of cell phones per 100 people (continuous variable)
- VR^c as the **centered** around the mean vaccination rate (continuous variable)

▶ Code to center CP and VR

In R:

```r
m_int_vr = gapm %>% lm(formula = life_exp ~ CP_c + VR_c + CP_c*VR_c)
```

OR

```r
m_int_vr = gapm %>% lm(formula = life_exp ~ CP_c*VR_c)
```

Displaying the regression table and writing fitted regression equation

```
1 tidy_m_vr = tidy(m_int_vr, conf.int=T)
2 tidy_m_vr %>% gt() %>% tab_options(table.font.size = 35) %>% fmt_number(decimals = 5)
```

row names

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	70.56972	0.61989	113.84216	0.00000	69.34002	71.79942
CP_c	0.07666	0.01832	4.18404	0.00006	0.04032	0.11301
VR_c	0.23765	0.08409	2.82598	0.00568	0.07083	0.40447
CP_c:VR_c	0.00308	0.00192	1.59912	0.11292	−0.00074	0.00689

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 VR^c + \widehat{\beta}_3 CP^c \cdot VR^c$$

$$\widehat{LE} = 70.57 + 0.08 \cdot CP^c + 0.24 \cdot VR^c + 0.003 \cdot CP^c \cdot VR^c$$

Comparing fitted regression lines for various vaccination rates

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 VR^c + \widehat{\beta}_3 CP^c \cdot VR^c$$
$$\widehat{LE} = 70.57 + 0.08 \cdot CP^c + 0.24 \cdot VR^c + 0.003 \cdot CP^c \cdot VR^c$$

@ 91.45% : $VR^c = 0$
$\hookrightarrow = VR - \overline{VR}$
$= 91.45 - 91.45 = 0$

To identify different lines, we need to pick example vaccination rates:

mean rate

Vaccination rate of 86.45 % −5

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 \cdot (-5) + \widehat{\beta}_3 CP^c \cdot (-5)$$
$$\widehat{LE} = (\widehat{\beta}_0 - 5\widehat{\beta}_2) + (\widehat{\beta}_1 - 5\widehat{\beta}_3) CP^c$$

Vaccination rate of 91.45 %

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 \cdot 0 + \widehat{\beta}_3 CP^c \cdot 0$$
$$\widehat{LE} = (\widehat{\beta}_0) + (\widehat{\beta}_1) CP^c$$

@ mean VR

Vaccination rate of 96.45 % +5

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 \cdot 5 + \widehat{\beta}_3 CP^c \cdot 5$$
$$\widehat{LE} = (\widehat{\beta}_0 + 5\widehat{\beta}_2) + (\widehat{\beta}_1 + 5\widehat{\beta}_3) CP^c$$

$VR^c = 96.45 - \overline{VR} = 96.45 - 91.45 = 5$

Poll Everywhere Question??

13:49 Wed Feb 25 84%

by Web **PollEv.com/nickywakim275**

Which of the following is the correct interpretation of $\widehat{\beta}_1 = 0.08$ in the following model: $$\widehat{LE} = \widehat\beta_0 + \widehat\beta_1 CP^c + \widehat\beta_2 VR^c + \widehat\beta_3 CP^c \cdot VR^c

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 VR^c + \widehat{\beta}_3 CP^c VR^c$$

~~The mean change in cell phones's effect is 0.08 years for every 1% increase in vaccination rate.~~ 9% 0

~~The mean change in cell phones's effect is 0.003 years for every 1% increase in vaccination rate.~~ 3%

At a vaccination rate 0%, for every one additional cell phone per 100 people, the mean increase in life expectancy is 0.08 years (95% CI: 0.04, 0.11) 35%

$\rightarrow VR^c = 0 \rightarrow$ left w/ $\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP^c$

At a vaccination rate 91.45%, for every one additional cell phone per 100 people, the mean increase in life expectancy is 0.08 years (95% CI: 0.04, 0.11) 53%

Interpretation for interaction between two continuous variables

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 VR^c + \widehat{\beta}_3 CP^c \cdot VR^c$$

$$\widehat{LE} = \left[\widehat{\beta}_0 + \widehat{\beta}_2 \cdot VR^c\right] + \underbrace{\left[\widehat{\beta}_1 + \widehat{\beta}_3 \cdot VR^c\right]}_{\text{CP's effect}} CP$$

$VR^c = 0$

$VR^c = 5$

- Interpretation:
  - β_3 = mean change in cell phones's effect, for every 1 % increase in vaccination rate
- In summary, the interaction term can be interpreted as "difference in adjusted effect of number of cell phones for every 1 % increase in vaccination rate"
- It will be helpful to test the interaction to round out this interpretation!!

Test interaction between two continuous variables (1/2)

- We run an F-test for a single coefficients (β_3) in the below model (see lesson 9)

$$LE = \beta_0 + \beta_1 CP^c + \beta_2 VR^c + \beta_3 CP^c \cdot VR^c + \epsilon$$

Null H_0

$$\beta_3 = 0$$

Alternative H_1

$$\beta_3 \neq 0$$

Null / Smaller / Reduced model

$$LE = \beta_0 + \beta_1 CP^c + \beta_2 VR^c + \epsilon$$

Alternative / Larger / Full model

$$LE = \beta_0 + \beta_1 CP^c + \beta_2 VR^c + \beta_3 CP^c \cdot VR^c + \epsilon$$

Test interaction between two continuous variables (2/2)

- Fit the reduced and full model

```
m_int_vr_red = gapm %>%
  lm(formula = life_exp ~ CP_c + VR_c)
m_int_vr_full = gapm %>%
  lm(formula = life_exp ~ CP_c + VR_c + CP_c*VR_c)
```

▶ Display the ANOVA table with F-statistic and p-value

term	df.residual	rss	df	sumsq	statistic	p.value
life_exp ~ CP_c + VR_c	102.000	3,480.371	NA	NA	NA	NA
life_exp ~ CP_c + VR_c + CP_c * VR_c	101.000	3,394.429	1.000	85.942	2.557	0.113

- Conclusion: There is not a significant interaction between cell phones and vaccination rate (p = 0.113). Vaccination rate is not an effect modifier of the association between cell phones and life expectancy.

Learning Objective

Last time:

1. Define confounders and effect modifiers, and how they interact with the main relationship we model.
2. Interpret the interaction component of a model with **a binary categorical covariate and continuous covariate**, and how the main variable's effect changes.
3. Interpret the interaction component of a model with **a multi-level categorical covariate and continuous covariate**, and how the main variable's effect changes.

This time:

4. Interpret the interaction component of a model with **two categorical covariates**, and how the main variable's effect changes.
5. Interpret the interaction component of a model with **two continuous covariates**, and how the main variable's effect changes.
6. Report results for a best-fit line (with confidence intervals) at different levels of an effect measure modifier

How to find the confidence interval for each slope?

- In the example with VR and CP, we showed:

Best-fit line for vaccination rate of 96.45 %

$VR^c = 5$

$$\widehat{LE} = \underbrace{(\widehat{\beta}_0 + 5\widehat{\beta}_2)}_{\text{intercept}} + \underbrace{(\widehat{\beta}_1 + 5\widehat{\beta}_3)}_{\text{slope}} CP^c$$

- Often, we want to report the estimate of the combined coefficients: $\widehat{\beta}_1 + 5\widehat{\beta}_3$
  - This allows us to make a statement like: "At a vaccination rate of 96.45%, mean life expectancy increases $(\widehat{\beta}_1 + 5\widehat{\beta}_3)$ years for every one additional cell phone per 100 people (95% CI: __, __)."

- We can calculate $\widehat{\beta}_1 + 5\widehat{\beta}_3$ by using the values of the estimated coefficients
- BUT we always want to have a **95% confidence interval** when we report this combined estimate!!

Getting a 95% confidence interval requires linear combinations!

- If we want a confidence interval for $\widehat{\beta}_1 + 5\widehat{\beta}_3$, then we would use the formula:

estimated value

$$\underbrace{\left(\widehat{\beta}_1 + 5\widehat{\beta}_3\right)}_{\text{estimated value}} \pm t^* \times \underbrace{\left[SE_{(\beta_1+5\beta_3)}\right]}_{\text{SE for estimated value}}$$

SE for estimated value

- The hard part is figuring out what $SE_{(\beta_1+5\beta_3)}$ (or $\text{Var}(\beta_1 + 5\beta_3)$) equals
- We need to go back to variance of linear combinations (BSTA 511/611, EPI 525):

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

or

$$\text{Var}(aX - bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) - 2ab\text{Cov}(X, Y)$$

$$\text{Var}(\hat{\beta}_1 + 5\hat{\beta}_3) = \text{Var}(\hat{\beta}_1) + 5^2\,\text{Var}(\hat{\beta}_3) + 2 \cdot 5 \cdot \text{COV}(\hat{\beta}_1, \hat{\beta}_3)$$

Calculating $SE_{(\beta_1+5\beta_3)}$ "by hand" (REFERENCE)

- A helpful function that returns the variance-covariance matric of all the coefficients in model `m_int_vr`:

```r
vcov(m_int_vr)
```

```
              (Intercept)          CP_c          VR_c     CP_c:VR_c
(Intercept)  0.3842647156 -2.466140e-04 -1.098844e-02 -4.873521e-04
CP_c        -0.0002466140  3.357335e-04 -5.354918e-04  1.872476e-06
VR_c        -0.0109884449 -5.354918e-04  7.071783e-03  8.343243e-05
CP_c:VR_c   -0.0004873521  1.872476e-06  8.343243e-05  3.700339e-06
```

$$\text{Var}(\beta_1) = 3.35733 \times 10^{-4}$$
$$\text{Var}(\beta_3) = 0.0070718$$
$$\text{Cov}(\beta_1, \beta_3) = -5.35492 \times 10^{-4}$$

$$\text{Var}(\beta_1 + 5\beta_3) = \text{Var}(\beta_1) + 5^2\text{Var}(\beta_3) + 2 \cdot 5 \cdot \text{Cov}(\beta_1, \beta_3)$$
$$\text{Var}(\beta_1 + 5\beta_3) = 3.35733 \times 10^{-4} + 5^2 \times 3.7 \times 10^{-6} + 2 \cdot 5 \cdot 1.872 \times 10^{-6}$$
$$\text{Var}(\beta_1 + 5\beta_3) = 4.46967 \times 10^{-4}$$
$$SE_{(\beta_1+5\beta_3)} = \sqrt{4.46967 \times 10^{-4}}$$
$$SE_{(\beta_1+5\beta_3)} = 0.0211416$$

We can use R and `estimable()` to find the estimate and CI

For $\widehat{\beta}_1 + 5\widehat{\beta}_3$: → $0\hat{\beta}_0 + 1\hat{\beta}_1 + 0\hat{\beta}_2 + 5\hat{\beta}_3$

interpret intercept: $\hat{\beta}_0 + 5\hat{\beta}_2$

$(1\hat{\beta}_0 + 0\hat{\beta}_1 + 5\hat{\beta}_2 + 0\hat{\beta}_3)$

```r
library(gmodels)
m_int_vr %>% estimable(
                          c("(Intercept)" = 0,  # beta0
                            "CP_c"        = 1,  # beta1
                            "VR_c"        = 0,  # beta2
                            "CP_c:VR_c"   = 5), # beta3
                          conf.int = 0.95)
```

1, 5, 0, 0

```
          Estimate Std. Error t value  DF     Pr(>|t|)   Lower.CI Upper.CI
(0 1 0 5) 0.09204476 0.02114159 4.35373 101 3.213735e-05 0.05010553 0.133984
```

$\hat{\beta}_1 + 5\hat{\beta}_3$ $\quad$ $SE(\hat{\beta}_1 + 5\hat{\beta}_3)$

Our conclusion: At a vaccination rate of 96.45 %, mean life expectancy increases 0.092 years for every one additional cell phone per 100 people (95% CI: 0.05, 0.134).

Another example: income (binary) and CP (1/2)

```
m_int_inc2 = gapm %>%
  lm(formula = life_exp ~ CP_c*income_level_2)
```

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 I(\text{high income}) + \widehat{\beta}_3 CP^c \cdot I(\text{high income})$$

$$\widehat{LE} = 68.408 + 0.076 \cdot CP^c + 6.247 \cdot I(\text{high income}) - 0.066 \cdot CP^c \cdot I(\text{high income})$$

For lower income countries: $I(\text{high income}) = 0$

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP$$

$$\widehat{LE} = 68.41 + 0.076 \cdot CP$$

For higher income countries: $I(\text{high income}) = 1$

$$\widehat{LE} = (\widehat{\beta}_0 + \widehat{\beta}_2) + (\widehat{\beta}_1 + \widehat{\beta}_3) CP$$

$$\widehat{LE} = (68.41 + 6.247) + (0.076 - 0.066) \cdot CP$$

$$\widehat{LE} = 74.65 + 0.01 \cdot CP$$

Another example: income (binary) and CP (2/2)

```
names(m_int_inc2$coefficients) # I just need to see the exact names
```

```
[1] "(Intercept)"                   "CP_c"
[3] "income_level_2Higher income"   "CP_c:income_level_2Higher income"
```

```
m_int_inc2 %>% estimable(
                   c("(Intercept)" = 0,        # beta0
                     "CP_c"       = 1,        # beta1
                     "income_level_2Higher income" = 0,      # beta2
                     "CP_c:income_level_2Higher income"  = 1),   # beta3
                   conf.int = 0.95)
```

$\hat{\beta}_1 + \hat{\beta}_3$

```
             Estimate Std. Error   t value  DF  Pr(>|t|)    Lower.CI   Upper.CI
(0 1 0 1) 0.01001064 0.02723878 0.3675143 101 0.7140044 -0.04402377 0.06404504
```

Our conclusion: For countries with high income, mean life expectancy increases 0.01 years for every one additional cell phone per 100 people (95% CI: -0.044, 0.064).

If our example had an effect measure modifier

- None of our examples had a significant interaction, so it's hard to demonstrate exactly how we would report this
- Let's say, **just for example**, that income had a significant interaction with CP
  - How would we report this to an audience??
- Here's how to report on an interaction/EMM:
  - We found that a country's income status (high or low) is a significant effect measure modifier on number of cell phones (*include p-value for interaction test here*). For countries with high income, mean life expectancy increases 0.01 years for every additional cell phone per 100 people (95% CI: -0.044, 0.064). For countries with low income, mean life expectancy increases 6.247 years for every additional cell phone per 100 people (95% CI: 3.825, 8.668).