Lesson 11: Interactions, Part 1

Nicky Wakim

2026-02-23

Learning Objectives

This time:

interactions → math

1. Define confounders and effect modifiers, and how they interact with the main relationship we model.
   ↳ type of interaction
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.

Next 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.

Regression analysis process

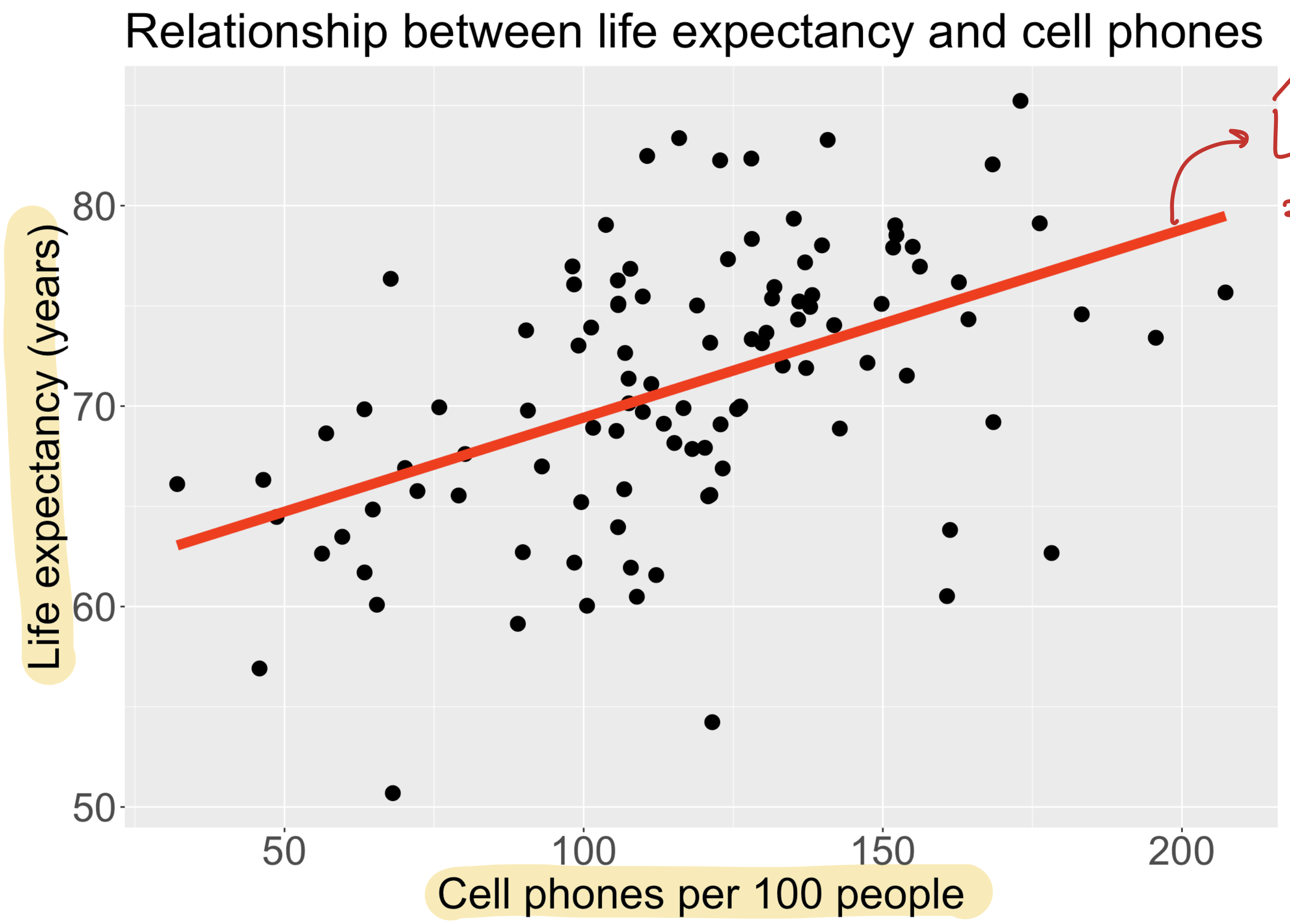
Model Selection
- Building a model
- Selecting variables & how to rep
- Prediction vs interpretation
- Comparing potential models

Model Fitting
- Find best fit line
- Using OLS in this class
- Parameter estimation
- Categorical covariates
- Interactions

Model Evaluation
- Evaluation of model fit
- Testing model assumptions
- Residuals
- Transformations
- Influential points
- Multicollinearity

Model Use (Inference)
- Inference for coefficients
- Hypothesis testing for coefficients
- Inference for expected Y given X
- Prediction of new Y given X

Recall our data and the main relationship



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

Learning Objectives

This 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.

Next time:

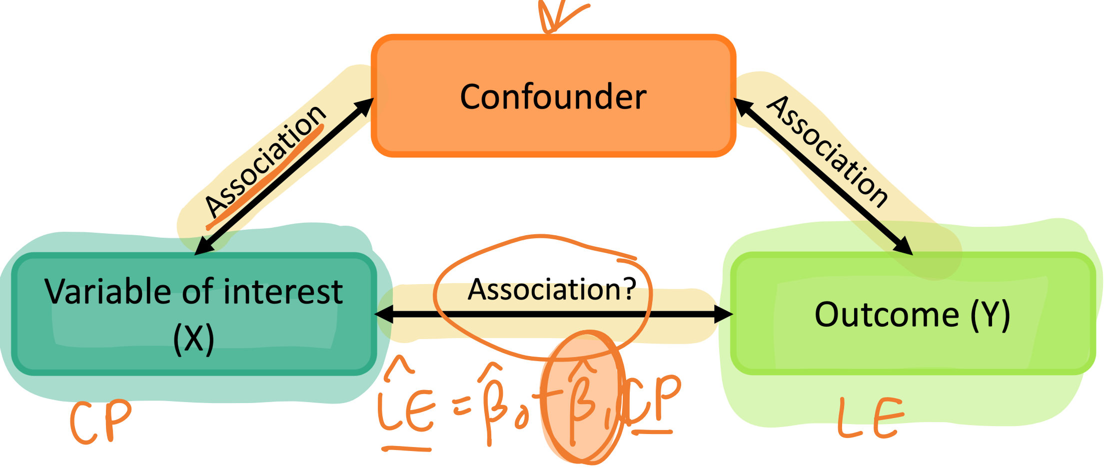
4. Interpret the interaction component of a model with **two categorical covariates**, and how the main variable's effect changes.
5. Interpret the interaction component of a model with **two continuous covariates**, and how the main variable's effect changes.

What is a confounder?

- A **confounding variable**, or **confounder**, is a factor/variable that wholly or partially *accounts for the observed effect of the risk factor on the outcome*

  main variable

- A confounder must be…
  - Related to the outcome Y, but not a consequence of Y
  - Related to the explanatory variable X, but not a consequence of X



- A classic example: We found an association between ice cream consumption and sunburn!
  - If we adjust for a potential confounder, temperature/hot weather, we may see that the association between ice and sunburn is not as large
- Another example: We found an association between socioeconomic status (SES) and lung cancer!
  - If we adjust for a potential confounder, exposure to air pollution, we may see that the association between SES and lung cancer decreases

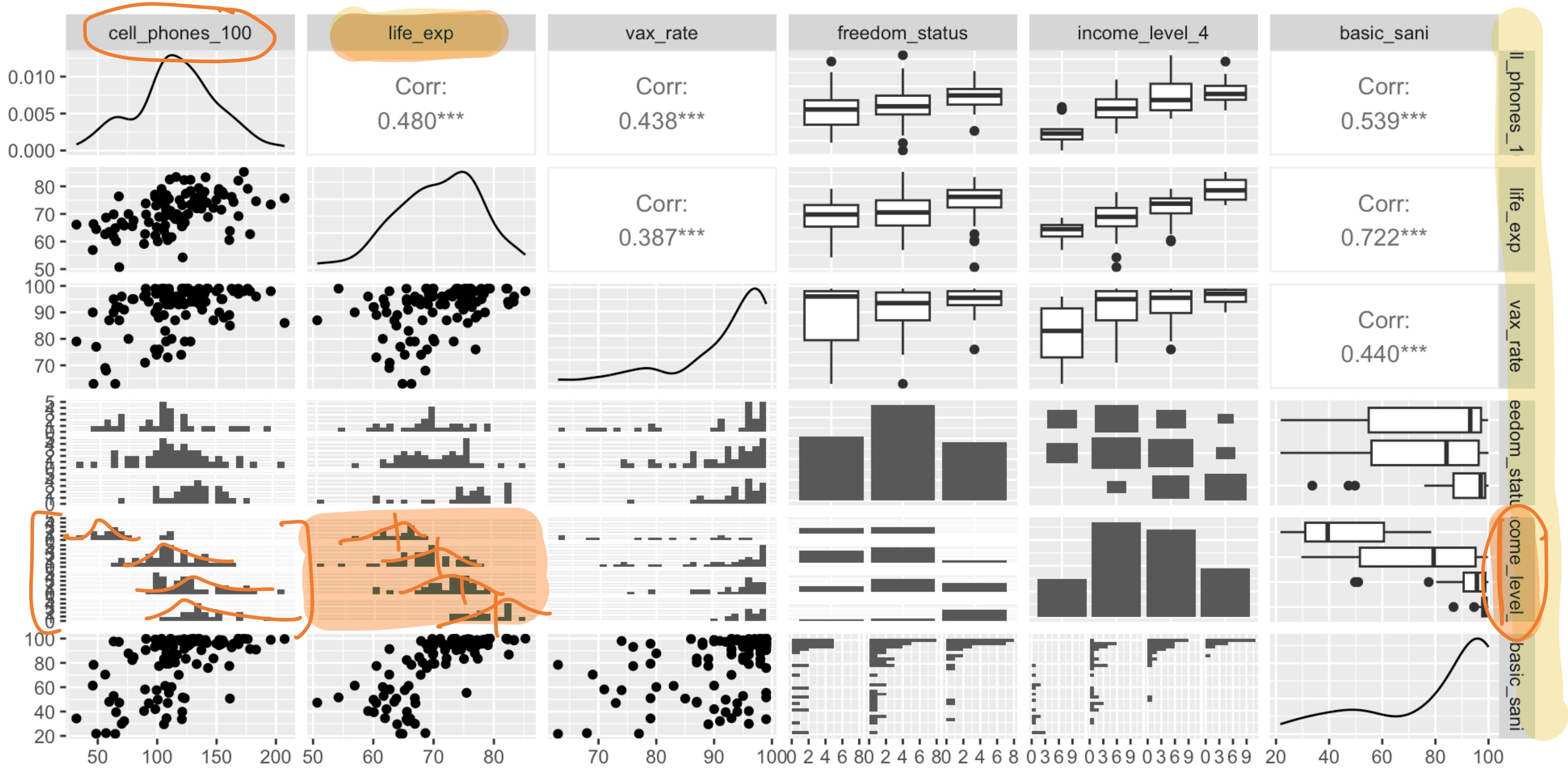
Proxies and confounders: the good and the harmful

- *This is totally my own tangent*
- A **proxy variable** is used to stand-in or represent another variable that is harder to measure
- Sometimes a confounder can be used as a proxy if it is hard to measure you explanatory variable/variable of interest
- Proxies can be helpful statistically while harmful socially OR helpful for both!

- Examples
  - Bad: BMI serving as a measurement for physical health or diet
    - Many studies show how harmful, mentally and physically, it is to equate BMI to health
  - Interesting: Using occurrence of online search queries as a proxy for public health risk perception
  - Helpful contextualization: Using race as a proxy for systemic racism, and thus a way to identify how to and who needs resources
- In our lab, I discuss using ~~sex assigned at birth~~ in our model

Exploratory approach to identifying confounders

```r
gapm2 %>% ggpairs()
```



Including a confounder in the model

- In the following model we have two variables, X_1 and X_2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

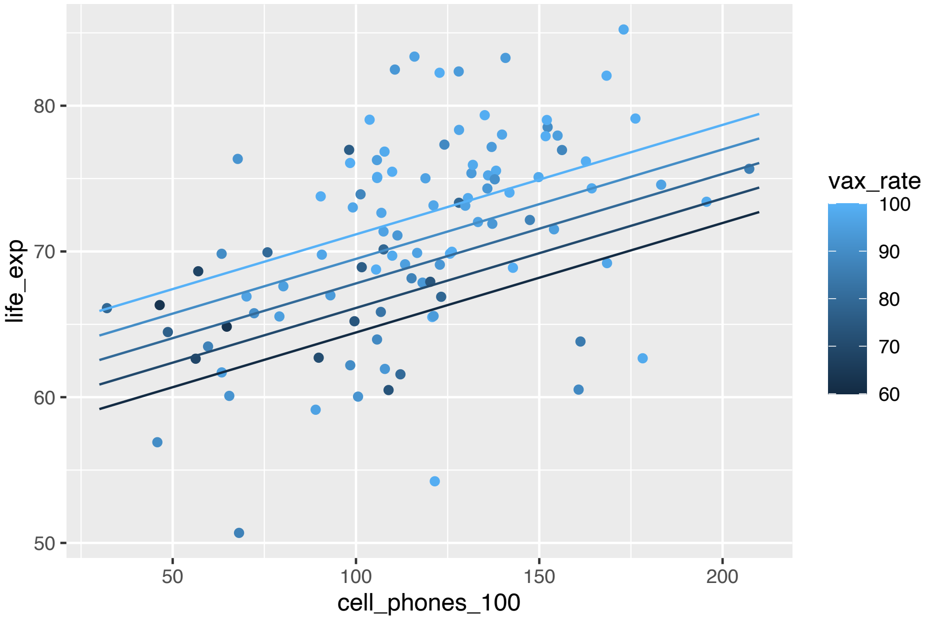
- And we assume that every level of the confounder, there is parallel slopes
- Note: to interpret β_1, we did not specify any value of X_2; only specified that it be held constant
  - Implicit assumption: effect of X_1 is equal across all values of X_2
- The above model assumes that X_1 and X_2 do not *interact* (with respect to their effect on Y)
  - Epidemiology: no "effect modification" → have NOT included how X_1's effect on Y changes w/ X_2
  - Meaning the effect of X_1 is the same regardless of the values of X_2
  - This model is often called a **"main effects model"**

↳ No interaction
confounders included

Where have we modeled a confounder before?

- We have seen a plot of life expectancy vs. cell phones with different levels of food supply colored (Lesson 8)
- In our plot and the model, we treat food supply as a **confounder**
- If ~~food supply~~ vax rate is a confounder in the relationship between life expectancy and cell phones, then we only use main effects in the model:

$$\text{LE} = \beta_0 + \beta_1 \text{CP} + \beta_2 \text{VR} + \epsilon$$



Poll everywhere question 1

$I(X_2 = \text{red}) = 1$
green is ref.

13:38 Mon Feb 23 92%

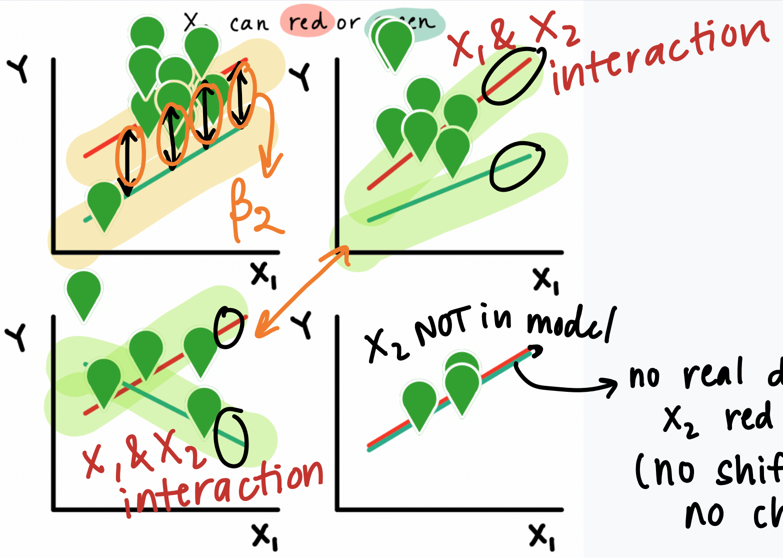
Join by Web **PollEv.com/nickywakim275**

$$Y = \beta_0 + \beta_1 X_1 + \underbrace{\beta_2 X_2}_{\beta_2} + \varepsilon$$

if X_2 is red

$$Y = (\beta_0 + \beta_2) + \beta_1 X_1$$

If X_1 is a continuous variable, and we are interested in the relationship between Y and X_1, which of the following pictures shows X_2 as a confounder? X_2 is a categorical variable with two groups: red and green.

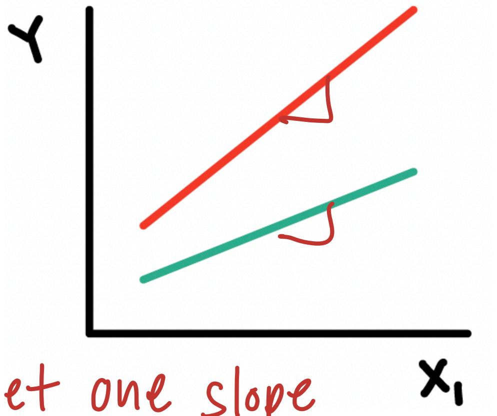


no real diff b/w
X_2 red vs. green
(no shift in lines,
no change in slope)

What is an effect modifier?/interaction

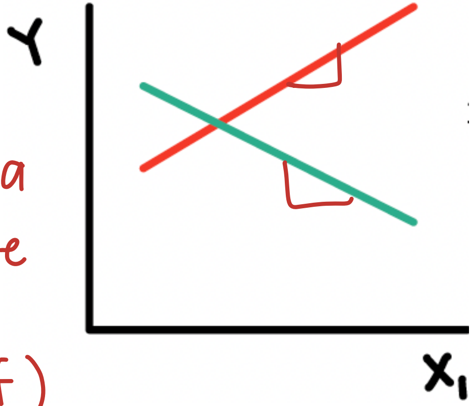
- An additional variable in the model
  - Outside of the main relationship between Y and X_1 that we are studying
- An effect modifier will change the effect of X_1 on Y depending on its value of eff mod
  - Aka: as the effect modifier's values change, so does the association between Y and X_1
  - So the coefficient estimating the relationship between Y and X_1 changes with another variable
- Example: A breast cancer education program (the exposure) that is much more effective in reducing breast cancer (outcome) in rural areas than urban areas.
  - Location (rural vs. urban) is the EMM

effect meas modifier



if X_2 red, we get one slope (X_1's effect)

if X_2 is green, we get a diff slope (X_1's eff is diff)



How do we include an effect modifier in the model?

interaction

- Interactions!!
- We can incorporate interactions into our model through product terms: (simplest representation X_1 & X_2 continuous)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon$$

$I(X_2 = red)$ $I(X_2 = red)$

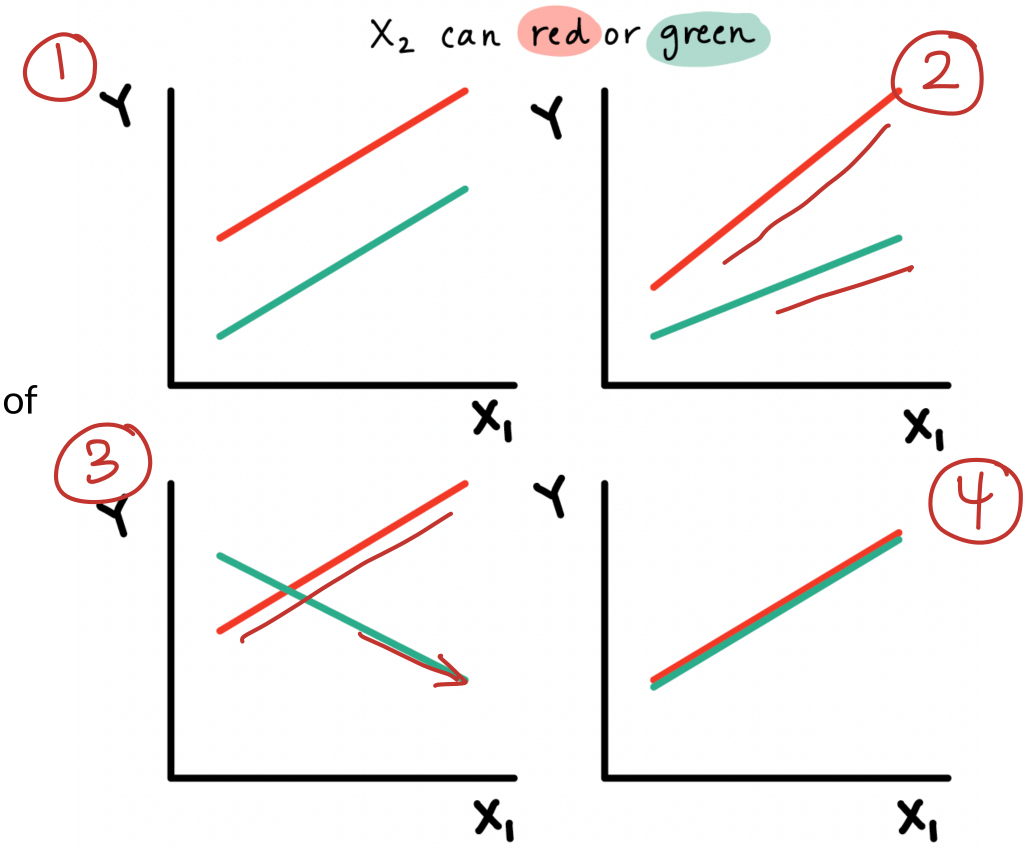
- Terminology:
  - main effect parameters: β_1, β_2
    - The main effect models estimate the *average* X_1 and X_2 effects
  - interaction parameter: β_3

Types of interactions / non-interactions

- Common types of interactions:
  - Synergism: X_2 strengthens the X_1 effect 2
  - Antagonism: X_2 weakens the X_1 effect 3
- If the interaction coefficient is not significant
  - No evidence of effect modification, i.e., the effect of X_1 does not vary with X_2 1

    but X_2 is confounder
- If the main effect of X_2 is also not significant 4
  - No evidence that X_2 is a confounder

    X_1 alone



Learning Objectives

This time:

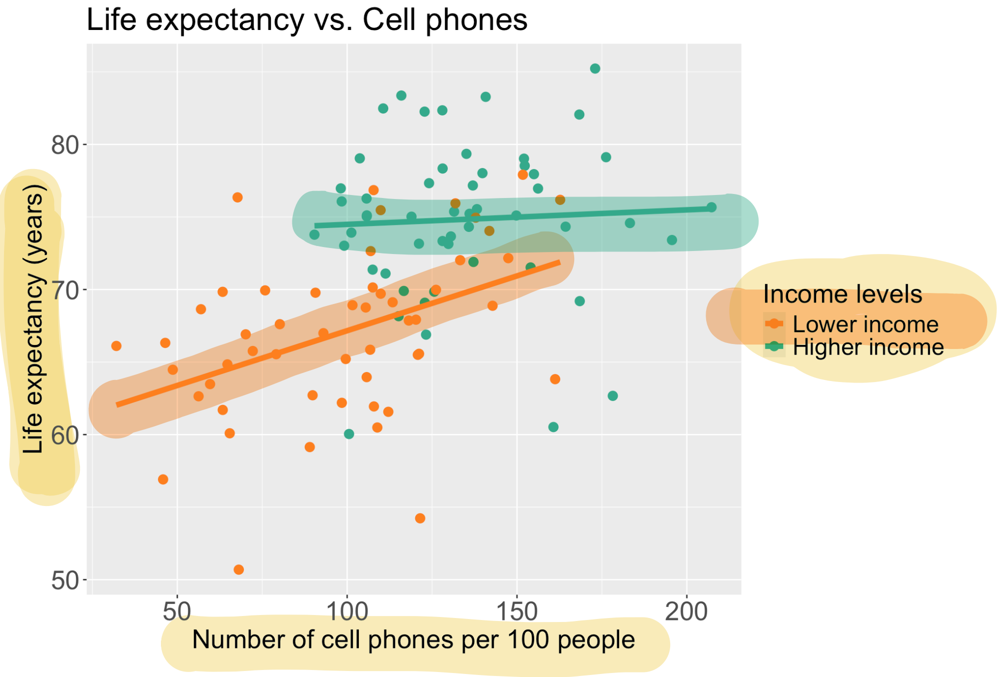
1. Define confounders and effect modifiers, and how they interact with the main relationship we model.
2. Interpret the interaction component of a model with **a binary categorical covariate and continuous covariate**, and how the main variable's effect changes.
3. Interpret the interaction component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

Next 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.

Do we think income level is an effect modifier for cell phones?

- Let's say we only have two income groups: low income and high income
- We can start by visualizing the relationship between life expectancy and cell phones *by income level*
- Questions of interest: Is the effect of number of cell phones on life expectancy differ depending on income level?
  - This is the same as: Is income level is an effect modifier for cell phones?
  - "effect of cell phones" differing = different slopes between CP and LE depending on the income group
- Let's run an interaction model to see!



Model with interaction between a *binary categorical and continuous variables*

Model we are fitting:

$$LE = \beta_0 + \underbrace{\beta_1 \overset{X_1}{CP} + \beta_2 \overset{X_2}{I(\text{high income})}}_{\text{main effects}} + \underbrace{\beta_3 \overset{X_1}{CP} \cdot \overset{X_2}{I(\text{high income})}}_{\text{interaction}} + \epsilon$$

- LE as outcome
- CP as continuous variable that is our main variable of interest
- $I(\text{high income})$ as the indicator that income level is "high income" (binary categorical variable)

```
m_int_inc2 = gapm %>%
  lm(formula = life_exp ~ cell_phones_100 + income_level_2 +
       cell_phones_100*income_level_2)
```

OR

```
m_int_inc2 = gapm %>%
  lm(formula = life_exp ~ cell_phones_100*income_level_2)
```

↳ automatically includes main effects

Displaying the regression table and writing fitted regression equation

▶ Code to display regression table with interaction

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	59.609	2.406	24.776	0.000	54.836	64.382
cell_phones_100	0.076	0.023	3.247	0.002	0.029	0.122
income_level_2Higher income	13.879	4.438	3.127	0.002	5.075	22.683
cell_phones_100:income_level_2Higher income	−0.066	0.036	−1.829	0.070	−0.137	0.006

X_1 X_2 $I(\text{high income})$

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

$$\widehat{LE} = 59.61 + 0.076 \cdot CP + 13.879 \cdot I(\text{high income}) - 0.066 \cdot CP \cdot I(\text{high income})$$

how income level changes CP's effect

Poll Everywhere Ques

14:15 Mon Feb 23

Join by Web PollEv.com/nickywakim275

x_1 x_2 $I(\text{high incone})$

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

$$\widehat{LE} = 59.61 + 0.076 \cdot CP + 13.879 \cdot I(\text{high income}) - 0.066 \cdot CP \cdot I(\text{high income})$$

how income level changes CP's effect

when high inc

$-0.066\,CP\,(1)$

Based only on the coefficient estimate for the interaction term: $\beta_3 = -0.066$... What can we say about the effect of number of cell phones?

Cell phones' effect is strengthened for countries with high income (compared to low income).	4%
Cell phones' effect is weakened for countries with high income (compared to low income). ✓	75%
Cell phones' effect is strengthened for countries with low income (compared to high income). ✓	0%
Cell phones' effect is weakened for countries with low income (compared to high income).	21%

CP's effect:

0.01

when high inc

CP's eff when low inc:

0.076

Comparing fitted regression lines for each income level

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP + \widehat{\beta}_2 I(\text{high income}) + \widehat{\beta}_3 CP \cdot I(\text{high income})$$
$$\widehat{LE} = 59.61 + 0.076 \cdot CP + 13.879 \cdot I(\text{high income}) - 0.066 \cdot CP \cdot I(\text{high income})$$

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

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP + \widehat{\beta}_2 \cdot 0 + \widehat{\beta}_3 CP \cdot 0$$
$$\widehat{LE} = 59.61 + 0.076 \cdot CP + 13.879 \cdot 0 - 0.066 \cdot CP \cdot 0$$
$$\widehat{LE} = 59.61 + 0.076 \cdot CP$$

$$\hat{LE} = \underbrace{\hat{\beta}_0}_{\text{intercept}} + \underbrace{\hat{\beta}_1}_{\text{slope}} CP$$

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

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP + \widehat{\beta}_2 \cdot 1 + \widehat{\beta}_3 CP \cdot 1$$
$$\widehat{LE} = 59.61 + 0.076 \cdot CP + 13.879 \cdot 1 - 0.066 CP \cdot 1$$
$$\widehat{LE} = (59.61 + 13.879) + (0.076 - 0.066) \cdot CP$$
$$\widehat{LE} = 73.49 + 0.01 \cdot CP$$

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

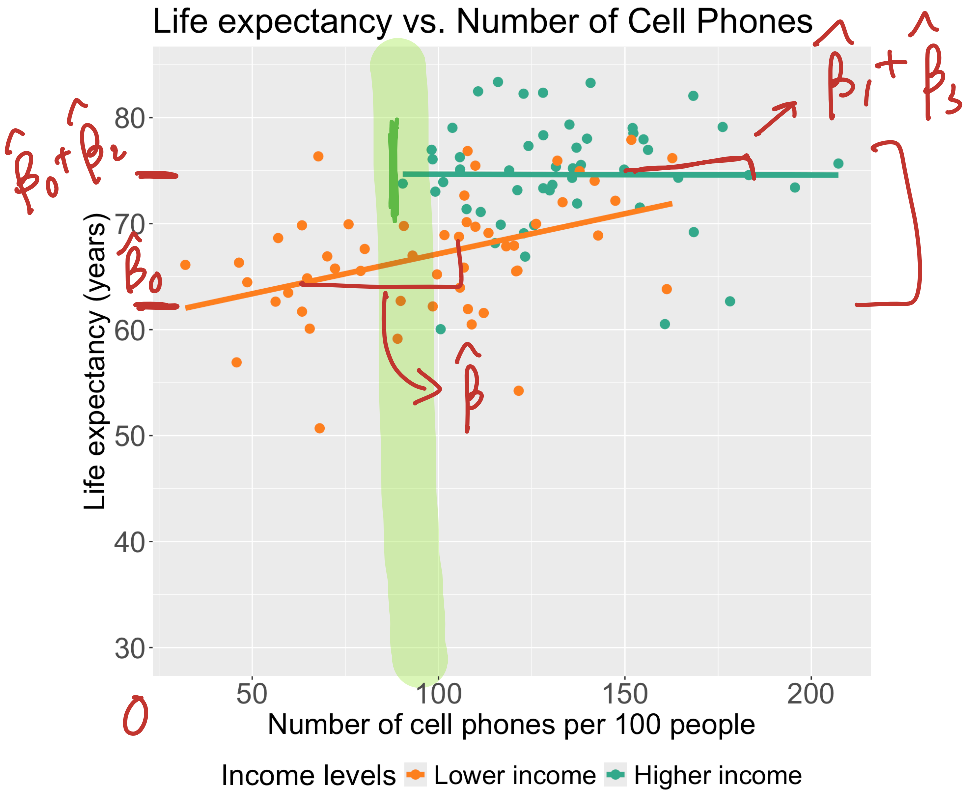
Let's take a look back at the plot

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

$$\widehat{LE} = \widehat{\beta}_0 + \widehat{\beta}_1 CP$$
$$\widehat{LE} = 59.61 + 0.076 \cdot CP$$

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

$$\widehat{LE} = (\widehat{\beta}_0 + \widehat{\beta}_2) + (\widehat{\beta}_1 + \widehat{\beta}_3)CP$$
$$\widehat{LE} = (59.61 + 13.879) + (0.076 - 0.066) \cdot CP$$
$$\widehat{LE} = 73.49 + 0.01 \cdot CP$$



Poll Everywhere Question 3

14:26 Mon Feb 23 73%

Join by Web PollEv.com/nickywakim275

73.49

Use the regression line for high income for this question: $\widehat{LE} = (59.61 + 13.879) + (0.076 - 0.066)CP$. What is the intercept for the line (we may omit the 95% CI for now)? What does it mean?

slope

0.01: For every 1 additional cell phone per 100 people, the mean difference in life expectancy in high income countries is 0.01 years.

12%

✓ 73.49: The mean life expectancy for high income countries with 0 cell phones per 100 people is 73.49 years.

77%

59.61: The mean life expectancy for high income countries with 0 cell phones per 100 people is 59.61 years.

4%

PAUSE: Centering continuous variables when including interactions

- For the high income group, the mean life expectancy had a regression line with a small intercept

$$
\begin{aligned}
\widehat{LE} &= (\widehat{\beta}_0 + \widehat{\beta}_2) + (\widehat{\beta}_1 + \widehat{\beta}_3)CP \\
\widehat{LE} &= (59.61 + 13.879) + (0.076 - 0.066) \cdot CP \\
\widehat{LE} &= 73.49 + 0.01 \cdot CP
\end{aligned}
$$

- Intercept of 73.49 is misleading because
  - Makes you think some of the life expectancies for high income countries are lower than that of low income countries (depending on the CP)
  - There are no high income countries with CP less than ~80
- Other online sources about when and when not to center:
  - The why and when of centering continuous predictors in regression modeling
  - When not to center a predictor variable in regression

Centering a variable

- Centering a variable means that we will subtract the mean or median (or other measurement of center) from the measured value
- Mean centered:

$$X_i^c = X_i - \overline{X}$$

- Median centered:

$$X_i^c = X_i - \text{median } X$$

- Centering the continuous variables in a model (when they are involved in interactions) helps with:
  - Interpretations of the coefficient estimates
  - Correlation between the main effect for the variable and the interaction that it is involved with
    - To be discussed in future lecture: leads to multicollinearity issues

It’ll be helpful to center number of cell phones

- Centering number of cell phones:

$$CP^c = CP - \overline{CP}$$

- Centering in R:

mean

```
gapm = gapm %>%
  mutate(CP_c = cell_phones_100 - median(cell_phones_100))
```

- I’m going to print the mean so I can use it for my interpretations

```
(mean_CP = mean(gapm$cell_phones_100))
```

```
[1] 116.5234
```

- Now all intercept values (in each respective freedom status) will be the mean life expectancy when number of cell phones per 100 people is 116.52
- We will used center CP for the rest of the lecture

Displaying the regression table and writing fitted regression equation AGAIN

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

▶ Code to display regression table with interaction

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	68.420	0.851	80.361	0.000	66.731	70.109
CP_c	0.076	0.023	3.247	0.002	0.029	0.122
income_level_2Higher income	6.236	1.220	5.111	0.000	3.816	8.657
CP_c:income_level_2Higher income	−0.066	0.036	−1.829	0.070	−0.137	0.006

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

$$\widehat{LE} = 68.42 + 0.076 \cdot CP^c + 6.236 \cdot I(\text{high income}) - 0.066 \cdot CP^c \cdot I(\text{high income})$$

Interpretation for interaction between binary categorical and continuous variables

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

$$\widehat{LE} = \underbrace{\left[\widehat{\beta}_0 + \widehat{\beta}_2 \cdot I(\overset{0\;1}{\text{high income}})\right]}_{\text{intercept}} + \underbrace{\left[\widehat{\beta}_1 + \widehat{\beta}_3 \cdot I(\overset{0\;1}{\text{high income}})\right]}_{\text{CP's effect (slope)}} CP^c$$

- Interpretation:
  - β_3 = mean change in number of cell phones's effect, comparing higher income to lower income levels
    - AKA: the change in slopes (for line between CP and LE) comparing high income to low income
  - where the "number of cell phones effect" = change in mean life expectancy per one additional cell phone (slope) with income level held constant, i.e. "adjusted number of cell phones effect"
- In summary, the interaction term can be interpreted as "difference in adjusted cell phones' effect comparing higher income to lower income levels"
- It will be helpful to test the interaction to round out this interpretation!!

Poll Everywhere Question 4

14:38 Mon Feb 23 69%

Join by Web PollEv.com/nickywakim275

Use the regression line for high income for this question: $\widehat{LE} = (59.61 + 13.879) + (0.076 - 0.066)CP$. What is the ~~intercept~~ SLOPE for the line (we may omit the 95% CI for now)? What does it mean?

0.01: For every 1 additional cell phone per 100 people, the mean difference in life expectancy in high income countries is 0.01 years, adjusting for vax rate.

64%

0.076: For every 1 additional cell phone per 100 people, the mean difference in life expectancy is 0.076 years, adjusting for income level.

11%

0.01: For every 1 additional cell phone per 100 people, the mean difference in life expectancy is 0.01 years, adjusting for income level.

14%

$$\widehat{LE} = \hat{\beta}_0 + \hat{\beta}_1 CP^c + \hat{\beta}_2 I(\text{high inc}) + \hat{\beta}_3 CP^c I(HI) + \hat{\beta}_4 VR$$

interaction

confounder

Test interaction between binary categorical and continuous variables

- We run an F-test for a single coefficient (β_3) in the below model (see Lesson 10, MLR: Using the F-test)

$$LE = \beta_0 + \beta_1 CP^c + \beta_2 I(\text{high income}) + \beta_3 CP^c \cdot I(\text{high income}) + \epsilon$$

Null H_0	Alternative H_1
$\beta_3 = 0$	$\beta_3 \neq 0$

Null / Smaller / Reduced model	Alternative / Larger / Full model
$LE = \beta_0 + \beta_1 CP^c + \beta_2 I(\text{high income}) + \epsilon$	$LE = \beta_0 + \beta_1 CP^c + \beta_2 I(\text{high income}) + \beta_3 CP^c \cdot I(\text{high income}) + \epsilon$

- I'm going to be skipping steps so please look back at Lesson 10 for full steps (required in HW 4)

Test interaction between binary categorical and continuous variables

- Fit the reduced and full model

```r
m_int_inc_red = gapm %>% lm(formula = life_exp ~ CP_c + income_level_2)
m_int_inc_full = gapm %>%
  lm(formula = life_exp ~ CP_c + income_level_2 + CP_c*income_level_2)
```

▶ Code to display ANOVA table for testing interaction

anova(m_int_inc_red, m_int_inc_full)

term	df.residual	rss	df	sumsq	statistic	p.value
life_exp ~ CP_c + income_level_2	102.000	2,957.645	NA	NA	NA	NA
life_exp ~ CP_c + income_level_2 + CP_c * income_level_2	101.000	2,862.847	1.000	94.798	3.344	0.070

- **Conclusion:** There is not a significant interaction between cell phones and income level (p = 0.07)
  - **If significant, we say more:** For higher income levels, for every one additional cell phone per 100 people, the mean life expectancy increases 0.01 years. For lower income levels, for every one additional cell phone per 100 people, the mean life expectancy increases 0.076 years. Thus, the effect of cell phones is seven times more for high income than low income levels.

Learning Objectives

This time:

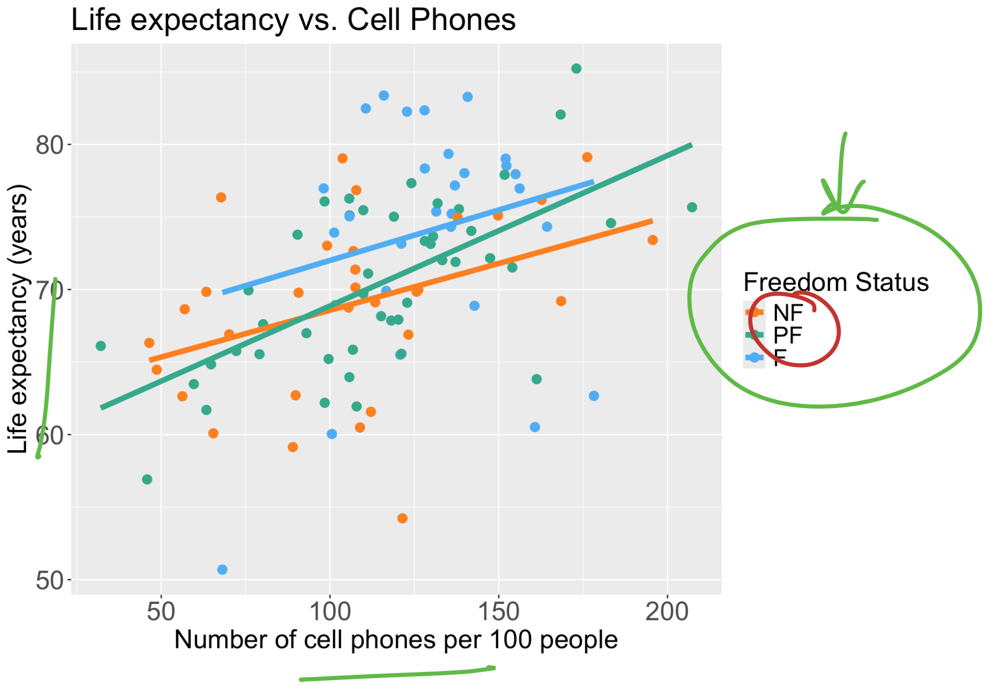
1. Define confounders and effect modifiers, and how they interact with the main relationship we model.
2. Interpret the interaction component of a model with **a binary categorical covariate and continuous covariate**, and how the main variable's effect changes.
3. Interpret the interaction component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

Next 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.

Do we think freedom status is an effect modifier for cell phones?

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



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

Model we are fitting: w/ Not free as ref grp

$$LE = \beta_0 + \beta_1 CP^c + \beta_2 I(\text{FS} = \text{PF}) + \beta_3 I(\text{FS} = \text{F}) + \quad \text{main effects}$$
$$\beta_4 CP^c \cdot I(\text{FS} = \text{PF}) + \beta_5 CP^c \cdot I(\text{FS} = \text{F}) + \epsilon$$

each indicator × CP's

- LE as life expectancy
- CP^c as centered number of cell phones (continuous variable)
- $I(\text{FS} = \text{PF})$ and $I(\text{FS} = \text{F})$ as the indicator for freedom status (with NF as reference group)

In R:

```
m_int_wr = gapm %>%
  lm(formula = life_exp ~ CP_c + freedom_status + CP_c*freedom_status)
```

OR

```
m_int_wr = gapm %>% lm(formula = life_exp ~ CP_c * freedom_status)
```

Displaying the regression table and writing fitted regression equation

- Code to display regression table with interaction

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	69.645	1.102	63.177	0.000	67.457	71.832
CP_c	0.065	0.029	2.263	0.026	0.008	0.121
freedom_statusPF	0.953	1.407	0.677	0.500	−1.840	3.745
freedom_statusF	3.519	1.706	2.063	0.042	0.135	6.904
CP_c:freedom_statusPF	0.039	0.038	1.036	0.303	−0.036	0.114
CP_c:freedom_statusF	0.005	0.056	0.091	0.928	−0.105	0.115

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \\ & \widehat{\beta}_4 CP^c \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 CP^c \cdot I(\text{FS} = \text{F}) \\ \widehat{LE} =& 69.645 + 0.065 \cdot CP^c + 0.953 \cdot I(\text{FS} = \text{PF}) + 3.519 \cdot I(\text{FS} = \text{F}) + \\ & 0.039 \cdot CP^c \cdot I(\text{FS} = \text{PF}) + 0.005 \cdot CP^c \cdot I(\text{FS} = \text{PF})\end{aligned}$$

Comparing fitted regression lines for each freedom status

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \\ & \widehat{\beta}_4 CP^c \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 CP^c \cdot I(\text{FS} = \text{F}) \\ \widehat{LE} =& 69.645 + 0.065 \cdot CP^c + 0.953 \cdot I(\text{FS} = \text{PF}) + 3.519 \cdot I(\text{FS} = \text{F}) + \\ & 0.039 \cdot CP^c \cdot I(\text{FS} = \text{PF}) + 0.005 \cdot CP^c \cdot I(\text{FS} = \text{PF})\end{aligned}$$

Not free

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 \cdot 0 + \\ & \widehat{\beta}_3 \cdot 0 + \widehat{\beta}_4 CP^c \cdot 0 + \\ & \widehat{\beta}_5 CP^c \cdot 0 \\ \widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 CP^c\end{aligned}$$

$I(FS=PF)=0$
$I(FS=F)=0$

Partly free

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 \cdot 1 + \\ & \widehat{\beta}_3 \cdot 0 + \widehat{\beta}_4 CP^c \cdot 1 + \\ & \widehat{\beta}_5 CP^c \cdot 0 \\ \widehat{LE} =& (\widehat{\beta}_0 + \widehat{\beta}_2) + (\widehat{\beta}_1 + \widehat{\beta}_4) CP^c\end{aligned}$$

$I(PF)=1$, $I(F)=0$

Free

$$\begin{aligned}\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 \cdot 0 + \\ & \widehat{\beta}_3 \cdot 1 + \widehat{\beta}_4 CP^c \cdot 0 + \\ & \widehat{\beta}_5 CP^c \cdot 1 \\ \widehat{LE} =& (\widehat{\beta}_0 + \widehat{\beta}_3) + (\widehat{\beta}_1 + \widehat{\beta}_5) CP^c\end{aligned}$$

$I(PF)=0$, $I(F)=1$

Interpretation for interaction between multi-level categorical and continuous variables

$$\begin{aligned}
\widehat{LE} =& \widehat{\beta}_0 + \widehat{\beta}_1 CP^c + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F}) + \\
& \widehat{\beta}_4 CP^c \cdot I(\text{FS} = \text{PF}) + \widehat{\beta}_5 CP^c \cdot I(\text{FS} = \text{F}) \\
\widehat{LE} =& \left[\widehat{\beta}_0 + \widehat{\beta}_2 I(\text{FS} = \text{PF}) + \widehat{\beta}_3 I(\text{FS} = \text{F})\right] + \\
& \underbrace{\left[\widehat{\beta}_1 + \widehat{\beta}_4 \cdot I(\text{FS} = \text{F}) + \widehat{\beta}_5 \cdot I(\text{FS} = \text{F})\right]}_{\text{CP's effect}} CP^c
\end{aligned}$$

- Interpretation:
  - β_4 = mean change in cell phones's effect, comparing countries that are partly free to countries that are not free
  - β_5 = mean change in cell phones's effect, comparing countries that are free to countries that are not free
- It will be helpful to test the interaction to round out this interpretation!!

Test interaction between multi-level categorical & continuous variables

- We run an F-test for a group of coefficients (β_4 and β_5) in the below model

$$
\begin{aligned}
LE = &\beta_0 + \beta_1 CP^c + \beta_2 I(\text{FS} = \text{PF}) + \beta_3 I(\text{FS} = \text{F}) + \\
&\underbrace{\cancel{\beta_4 CP^c \cdot I(\text{FS} = \text{PF})}}_{0} + \underbrace{\cancel{\beta_5 CP^c \cdot I(\text{FS} = \text{F})}}_{0} + \epsilon
\end{aligned}
$$

Null H_0

$\beta_4 = \beta_5 = 0$

Alternative H_1

$\beta_4 \neq 0$ and/or $\beta_5 \neq 0$

Null / Smaller / Reduced model

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

no interactions

Alternative / Larger / Full model

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

Test interaction between multi-level categorical & continuous variables

$F = \dfrac{\frac{SSE_R - SSE_F}{\Delta df}}{\frac{SSE_F}{df_F}}$

- Fit the reduced and full model

```
m_int_fs_red = gapm %>% lm(formula = life_exp ~ CP_c + freedom_status)
m_int_fs_full = gapm %>%
  lm(formula = life_exp ~ CP_c + freedom_status + CP_c*freedom_status)
```

▶ Code to display ANOVA table for testing interaction

F-stat

term	df.residual	rss	df	sumsq	statistic	p.value
life_exp ~ CP_c + freedom_status	101.000	3,517.244	NA	NA	NA	NA
life_exp ~ CP_c + freedom_status + CP_c * freedom_status	99.000	3,475.580	2.000	41.664	0.593	0.554

- Conclusion: There is not a significant interaction between number of cell phones and freedom status (p = 0.554)
- Freedom status is NOT an effect measure modifier of CP on LE

my full model does NOT sig explain MORE variation than red