Statistics 2

Multiple regression: Interaction effects

Casper Albers & Jorge Tendeiro

Lecture 5, 2019 – 2020

university of groningen

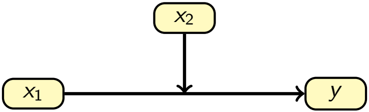
Overview

Literature for this lecture

Read:

Agresti, Section 11.4 - 11.5

Statistical interaction



- There is interaction between x_1 and x_2 if the effect of x_1 on y changes according to the level of x_2.
- In other words: x_2 moderates the effect of x_1 on y.

Example:

- y = annual income
- x_1 = number of years of education
- x_2 = years of working experience

It has been verified that the effect of education (x_1) on income (y) varies with the years of working experience (x_2): The annual income as a function of education increases with years of working experience.

Thus, education (x_1) and years of working experience (x_2) interact.

x_1, x_2: Predictors.

y: Response variable.

With no interaction

The combined impact of x_1 and x_2 on y equals the sum of their separated effects (usual multiple regression):

$$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2.$$

With interaction

The combined impact of x_1 and x_2 on y is larger/smaller than the sum of their separated effects:

$$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \underbrace{x_1 x_2}_{x_3}, \qquad \text{for } \beta_3 \neq 0.$$

The new predictor $x_3 = x_1 x_2$ is called the cross-product of x_1 and x_2.

So, what changes?

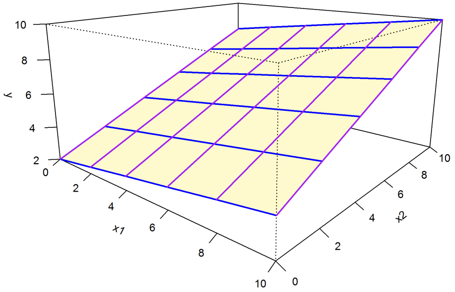
With no interactions...

$$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2$$

- β_1, the regression of y on x_1, is constant over all values of x_2.
- β_2, the regression of y on x_2, is constant over all values of x_1.

Example: $\widehat{y} = 2 + .2x_1 + .6x_2$

- $\beta_1 = .2$: Amount by which $\widehat{y}$ changes when x_1 increases 1 unit and x_2 is kept fixed.
- $\beta_2 = .6$: Amount by which $\widehat{y}$ changes when x_2 increases 1 unit and x_1 is kept fixed.



- "*β_1 is constant over all values of x_2*"

  All **blue** lines are parallel, i.e, have the same slope β_1:

$$\widehat{y} = \mathbf{.2}x_1 + (.6x_2 + 2).$$

- "*β_2 is constant over all values of x_1*"

  All **purple** lines are parallel, i.e, have the same slope β_2:

$$\widehat{y} = \mathbf{.6}x_2 + (.2x_1 + 2).$$

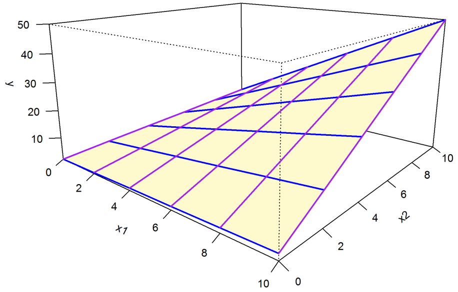
With interactions...

$$E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \qquad \text{for } \beta_3 \neq 0$$

- The regression of y on x_1 is conditional on, or moderated by, the value of x_2.
- The regression of y on x_2 is conditional on, or moderated by, the value of x_1.

Example: $\widehat{y} = 2 + .2x_1 + .6x_2 + .4x_1x_2$

- $\beta_1 = .2$: Amount by which $\widehat{y}$ changes when x_1 increases 1 unit **and** $\mathbf{x_2 = 0}$.
- $\beta_2 = .6$: Amount by which $\widehat{y}$ changes when x_2 increases 1 unit **and** $\mathbf{x_1 = 0}$.
- $\beta_3 = .4$: Tougher!...



- "*The regression of y on x_1 is conditional on the value of x_2*"
  The **blue** lines are not parallel, i.e, their slope varies with x_2:
  $$\widehat{y} = \mathbf{(.2 + .4x_2)}x_1 + .6x_2 + 2.$$
- "*The regression of y on x_2 is conditional on the value of x_1*"
  The **purple** lines are not parallel, i.e, their slope varies with x_1:
  $$\widehat{y} = \mathbf{(.6 + .4x_1)}x_2 + .2x_1 + 2.$$

Interpreting regression coefficients

Therefore:

> *Interpretation of the regression coefficients changes when interactions are added to the model.*

Once more, for $\widehat{y} = a + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2$:

- b_1 is amount by which y changes when x_1 increases 1 unit **and $\mathbf{x_2 = 0}$**.
- b_2 is amount by which y changes when x_2 increases 1 unit **and $\mathbf{x_1 = 0}$**.

Problems

1. How to interpret b_1, b_2 when x_1 and/or x_2 have no meaningful zero?
2. How to interpret b_3?

Interpreting b_1, b_2

Q: How to interpret b_1, b_2 when x_1 and/or x_2 have no meaningful zero?
A: Use centered predictors: Replace x_1, x_2, x_1x_2 by $x_1^c, x_2^c, x_1^c x_2^c$, where

$$x_1^c = (x_1 - \overline{x}_1) \text{ and } x_2^c = (x_2 - \overline{x}_2).$$

So, for $\widehat{Y} = a + b_1 x_1^c + b_2 x_2^c + b_3 x_1^c x_2^c$:

- b_1 is amount by which y changes when x_1 (or x_1^c) increases 1 unit **and** $x_2 = \overline{x}_2$.
- b_2 is amount by which y changes when x_2 (or x_2^c) increases 1 unit **and** $x_1 = \overline{x}_1$.

Note: Interpretation of b_1, b_2 remains conditional on a specific value of the other predictor, but now such a value exists always (namely $\overline{x}_2, \overline{x}_1$ resp.).

Centering – example

x_1	x_2	$x_1 x_2$
12	6	72
16	10	160
17	9	153
20	6	120
24	11	264
19	6	114
$\overline{x}_1 = 18$	$\overline{x}_2 = 8$	

x_1^c $= (x_1 - \overline{x}_1)$	x_2^c $= (x_2 - \overline{x}_2)$	$x_1^c x_2^c$
-6	-2	12
-2	2	-4
-1	1	-1
2	-2	-4
6	3	18
1	-2	-2

Interpreting b_1, b_2

How do the values of b_1, b_2 change before/after centering x_1, x_2?

- When no interaction is present: b_1, b_2 remain the same.
  (Revisit the 3D plot – parallel lines $\longleftarrow$)

$$\widehat{y} = a + b_1 x_1 + b_2 x_2$$

$$\widehat{y} = \widetilde{a} + b_1 x_1^c + b_2 x_2^c$$

  (the intercept does change, though)

- When interaction is present: b_1, b_2 change.
  (Revisit the 3D plot – nonparallel lines $\longleftarrow$)

Interpreting b_3

Q: How to interpret b_3, the regression coefficient of $x_1x_2/x_1^cx_2^c$?

A: b_3 reflects the part of $x_1x_2/x_1^cx_2^c$ from which x_1 and x_2 have been controlled (i.e., partialled out).

To see this, notice how you can *conceptually* estimate b_3:

- Partial x_1, x_2 from x_1x_2: $x_1x_2 = u + u_1x_1 + u_2x_2 + \text{res}$.
- Regress y on res: $\widehat{y} = v + b_3\text{res}$.

Example: $\widehat{y} = 2 + .2x_1 + .6x_2 + .4x_1x_2 \rightarrow$

y increases $b_3 = .4$ units when the part of x_1x_2 that is independent of x_1 and x_2 increases 1 unit.

Note: This interpretation of b_3:

- Requires that x_1 and x_2 are both entered as predictors.
- Is insensitive to using centered predictors (unlike b_1, b_2, see previous slide), so $b_{3,x_1x_2} = b_{3,x_1^cx_2^c}$.

Simple regression equations. Simple slopes

Consider the following regression model with centered predictors (we dropped the 'C' superscript for convenience):

$$\boxed{\widehat{y} = 2.2x_1 + 2.6x_2 + .4x_1x_2 + 16.}$$

The regression of y on x_1 when $x_2 = x_2^*$ is

$$\widehat{y} = (2.2 + .4x_2^*)x_1 + 2.6x_2^* + 16. \tag{1}$$

The regression of y on x_2 when $x_1 = x_1^*$ is

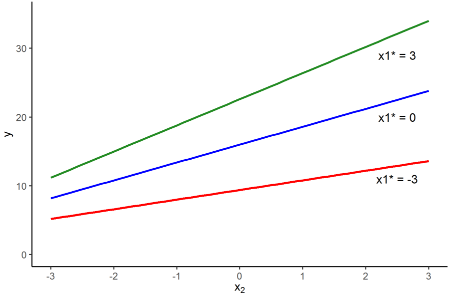
$$\widehat{y} = (2.6 + .4x_1^*)x_2 + 2.2x_1^* + 16. \tag{2}$$

- (1) and (2): Simple regression equations.
- $(2.2 + .4x_2^*)$ and $(2.6 + .4x_1^*)$: Simple slopes.

- You can plot simple regression lines to assess interactions graphically.
- Do so for several values of the moderator (i.e, the conditioning predictor x_1^*/x_2^*) (often $\{-1sd, 0, 1sd\}$).
- Plot only in the meaningful range of the predictors.

Example: $\widehat{y} = (2.6 + .4x_1^*)x_2 + 2.2x_1^* + 16$, for $x_1^* = -3, 0, 3$.

x_1^*	simple regression equation
-3	$\widehat{y} = 1.4x_2 + 9.4$
0	$\widehat{y} = 2.6x_2 + 16$
3	$\widehat{y} = 3.8x_2 + 22.6$



Simple regression equations. Simple slopes

Two types of plots of simple regression equations:

- Lines in plot are not parallel.
  $\rightarrow$ Simple slopes change for different values of the moderator.
  $\rightarrow$ Interaction.

- Lines in plot are parallel.
  $\rightarrow$ Simple slopes are constant for different values of the moderator.
  $\rightarrow$ No interaction.

Centering does not change the values of simple slopes for corresponding values of the conditioning predictor.

Post hoc probing of interactions

- Goal: Create CIs/significance tests for simple slopes.
- Use t-tests with $df = n - p - 1$ ($p =$ number of predictors).
- SE's are given (by software): No need to compute them manually.

Regression of y on x_1 at $x_2 = x_2^*$
$\widehat{y} = \underbrace{(b_1 + b_3 x_2^*)}_{b \text{ at } x_2^*} x_1 + b_2 x_2^* + a$

$$\boxed{t = \frac{b_1 + b_3 x_2^*}{SE_{b \text{ at } x_2^*}} \sim t(n-p-1)df}$$

Regression of y on x_2 at $x_1 = x_1^*$
$\widehat{y} = \underbrace{(b_2 + b_3 x_1^*)}_{b \text{ at } x_1^*} x_2 + b_1 x_1^* + a$

$$\boxed{t = \frac{b_2 + b_3 x_1^*}{SE_{b \text{ at } x_1^*}} \sim t(n-p-1)df}$$

Example: Physical endurance

Example from the textboob by Cohen, Cohen, West, and Aiken (2003)[1].

- y = ENDUR = physical endurance (in nr. of minutes jogging)
- x_1 = AGE = age (in years)
- x_2 = EXER = exercise (in nr. of years of vigorous physical exercise)

Sample size: $n = 245$.

Fit the regression model with interaction for centered x_1^c, x_2^c:

1. Compute x_1^c, x_2^c (i.e., center the predictors).
   (*Note*: $\overline{AGE} = 49.18$, $\overline{EXER} = 10.67$.)
2. Compute $x_1^c x_2^c$ (i.e., multiply x_1^c and x_2^c).
3. Estimate the model:

$$\boxed{\widehat{\text{ENDUR}} = a + b_1\,\text{AGE}^c + b_2\,\text{EXER}^c + b_3\,\text{AGE}^c \times \text{EXER}^c}$$

[1] Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/ Correlation Analysis for the Behavioral Sciences* (3rd ed.). Mahwah: NJ.

Example: Physical endurance

Model Summary

Model	R	R^2	Adjusted R^2	RMSE
1	0.454	0.206	0.196	9.700

20.6% of the variance of 'ENDUR' is accounted for by all predictors jointly.

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	5887	3	1962.41	20.86	$<$ .001
	Residual	22674	241	94.08		
	Total	28561	244			

$\mathcal{H}_0 : R^2 = 0$,
or equivalently,
$\mathcal{H}_0 : \beta_{\text{AGE}} = \beta_{\text{EXER}} = 0$,
is rejected at $\alpha = 5\%$ ($F(3, 241) = 20.86$, $p < .001$).

Example: Physical endurance

Coefficients

	Unstandardized	Standard Error	Standardized	t	p	95% CI Lower	95% CI Upper
(Intercept)	25.889	0.647		40.037	$<$ .001	24.615	27.162
AGE	-0.262	0.064	-0.244	-4.085	$<$ .001	-0.388	-0.135
EXER	0.973	0.137	0.429	7.124	$<$ .001	0.704	1.242
AGE_EXER	0.047	0.014	0.201	3.476	$<$ .001	0.020	0.074

Estimated regression model

$$\boxed{\widehat{\text{ENDUR}} = 25.889 - 0.262\,\text{AGE}^c + 0.973\,\text{EXER}^c + 0.047\,\text{AGE}^c \times \text{EXER}^c}$$

Interpretation

- $b_{\text{AGE}} = -0.262$: Endurance decreases .262 minutes per increase of 1 year of age, at the mean of 'EXER' (10.67 years of exercise).
- $b_{\text{EXER}} = 0.973$: Endurance increases .973 minutes per increase of 1 year of age, at the mean of 'AGE' (49.18 years).

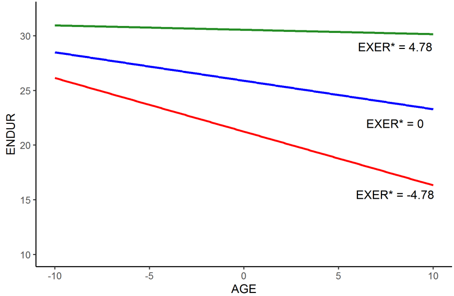
Example: Physical endurance

- Simple regression equation (SRE) of ENDUR on AGE, for fixed EXER:

$$\boxed{\widehat{\text{ENDUR}} = (-.262 + .047\text{EXER}^{C*})\text{AGE} + .973\text{EXER}^{C*} + 25.889}$$

- SRE of ENDUR on AGE, for $\text{EXER}^{C*} = -1sd, 0, 1sd$ ($sd_{\text{EXER}} = 4.78$):

EXER^{C*}	SRE
-4.78	$\widehat{ENDUR} = -.49\text{AGE}^c + 21.24$
0	$\widehat{ENDUR} = -.26\text{AGE}^c + 25.89$
4.78	$\widehat{ENDUR} = -.04\text{AGE}^c + 30.54$



Endurance decreases less and less with age as exercise increases.

Comparing regression models

Multiple regression is particularly complex when p is large.

- Partial regression effects can vary a lot according to what other predictors are in the model.
- Previously useful predictors can be redundant once more predictors are added to the model.
- Interaction effects complicate things further.

We can compare pairs of nested regression models, such that one model is a special case of the other.

- Complete model: More general model.
- Reduced model: Special case of the complete model.

Nested regression models

Example 1

- Complete model: $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$
- Reduced model: $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2$
- Test: $\mathcal{H}_0 : \beta_3 = \beta_4 = 0$ versus $\mathcal{H}_1$: Not all β_i are 0

Example 2

- Complete model: $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$
- Reduced model: $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2$
- Test: $\mathcal{H}_0 : \beta_3 = 0$ versus $\mathcal{H}_1 : \beta_3 \neq 0$

Example 3

- Complete model: $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_3$
- Reduced model: $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$
- Test: $\mathcal{H}_0 : \beta_4 = 0$ versus $\mathcal{H}_1 : \beta_4 \neq 0$

Nested regression models

Compare the reduced and complete nested models.

- Compare the complete model's R^2 (R_c^2) with the reduced model's R^2 (R_r^2).
- Necessarily, $R_r^2 \leq R_c^2$.
- The question is whether R_c^2 is large enough in comparison to R_r^2.

Test statistic

$$\boxed{F = \frac{(SSE_r - SSE_c)/df_1}{SSE_c/df_2} = \frac{(R_c^2 - R_r^2)/df_1}{(1-R_c^2)/df_2} \underset{\mathcal{H}_0}{\sim} F(df_1, df_2)}$$

- $df_1 =$ difference of the number of regression effects between the complete and reduced models.
- $df_2 =$ residual df for the complete model.

Example: Physical endurance

Complete model:

$$\boxed{\widehat{\text{ENDUR}} = a + b_1\,\text{AGE}^c + b_2\,\text{EXER}^c + b_3\,\text{AGE}^c \times \text{EXER}^c}$$

Reduced model:

$$\boxed{\widehat{\text{ENDUR}} = a + b_1\,\text{AGE}^c + b_2\,\text{EXER}^c}$$

Test

- $\mathcal{H}_0 : \beta_3 = 0$
- $\mathcal{H}_1 : \beta_3 \neq 0$

Example: Physical endurance

Model Summary

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	*F* Change	df1	df2	*p*
0	0.408	0.166	0.159	9.919	0.166	24.14	2	242	$<$ .001
1	0.454	0.206	0.196	9.700	0.040	12.08	1	241	$<$ .001

Note. Null model includes AGE, EXER

ANOVA

Model		Sum of Squares	df	Mean Square	*F*	*p*
0	Regression	4751	2	2375.34	24.14	$<$ .001
	Residual	23810	242	98.39		
	Total	28561	244			
1	Regression	5887	3	1962.41	20.86	$<$ .001
	Residual	22674	241	94.08		
	Total	28561	244			

- $df_1 = 3 - 2 = 1.$
- $df_2 = 241.$
- $F = \frac{(SSE_r - SSE_c)/df_1}{SSE_c/df_2} = \frac{(23810-22674)/1}{22674/241} = 12.08,$
  or, equivalently,
  $F = \frac{(R_c^2 - R_r^2)/df_1}{(1-R_c^2)/df_2} = \frac{(.206-.166)/1}{(1-.206)/241} = 12.08.$

Example: Physical endurance

Also note that, when (and only when) $df_1 = 1$, we have that

$$F_{1,df_2} = t^2_{df_2},$$

where

- F = test statistic used for model comparison.
- t = test statistic of the extra effect in the complete model.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
0	(Intercept)	26.531	0.634		41.865	$<$ .001
	AGE	-0.257	0.066	-0.240	-3.925	$<$ .001
	EXER	0.916	0.139	0.404	6.610	$<$ .001
1	(Intercept)	25.889	0.647		40.037	$<$ .001
	AGE	-0.262	0.064	-0.244	-4.085	$<$ .001
	EXER	0.973	0.137	0.429	7.124	$<$ .001
	AGE_EXER	0.047	0.014	0.201	3.476	$<$ .001

$$\boxed{t^2 = 3.476^2 = 12.08}$$

Conclusion

Reject $\mathcal{H}_0 : \beta_3 = 0$ and retain the complete model.

Next week: Partial correlation, standardized regression

Agresti, Sections 11.6 - 11.7