

Regression Models w/ k-group & Quant Variables

- Sources of data for this model
- Variations of this model
- Main effects version of the model
 - Interpreting the regression weights
 - Plotting and interpreting the model
- Interaction version of the model
 - Composing the interaction terms
 - Testing the interaction term = testing homogeneity of regression slope assumption
 - Interpreting the regression weights
 - Plotting and interpreting the model

As always, “the model doesn’t care where the data come from”. Those data might be ...

- a measured k-group variable (e.g., single, married, divorced) and a measured quant variable (e.g., age)
- a manipulated k-group variable (Tx1 vs. Tx2 vs. Cx) and a measured quant variable (e.g., age)
- a measured k-group variable (e.g., single, married, divorced) and a manipulated quant variable (e.g., 0, 1, 2, 5, 10 practices)
- a manipulated binary k-group variable (Tx1 vs. Tx2 vs. Cx) and a manipulated quant variable (e.g., 0, 1, 2, 5, 10 practices)

Like nearly every model in the ANOVA/regression/GLM family – this model was developed for and originally applied to experimental designs with the intent of causal interpretability !!!

As always, causal interpretability is a function of design (i.e., assignment, manipulation & control procedures) – not statistical model or the constructs involved !!!

There are two important variations of this model

1. Main effects model

- Terms for the k-group variable & quant variable
- No interaction – assumes regression slope homogeneity
- b-weights for k-group & quant variables each represent main effect of that variable

2. Interaction model

- Terms for k-group variable & quant variable
- Term for interaction - does not assume reg slp homogen !!
- b-weights for k-group & quant variables each represent the simple effect of that variable when the other variable = 0
- b-weight for the interaction term represented how the simple effect of one variable changes with changes in the value of the other variable (e.g., the extent and direction of the interaction)



Models with a centered quantitative predictor & a dummy coded k-category predictor

This is called a main effects model → there are no interaction terms.

$$y' = b_1X + b_2Z_1 + b_3Z_2 + a$$

a → regression constant

- expected value of y if all predictors = 0
- mean of the control group (G3)
- height of control group Y-X regression line

b₁ → regression weight for centered quant predictor

- expected direction and extent of change in Y for a 1-unit increase in X after controlling for the other variable(s) in the model
- main effect of X
- slope of Y-X regression line for all groups

b₂ → regression weight for dummy coded comparison of G1 vs G3

- expected direction and extent of change in Y for a 1-unit increase in Z₁, after controlling for the other variable(s) in the model
- main effect of Z₁
- Y-X reg line height difference for G1 & G3

b₃ → regression weight for dummy coded comparison of G2 vs. G3

- expected direction and extent of change in Y for a 1-unit increase in Z₂, after controlling for the other variable(s) in the model
- main effect of Z₂
- Y-X reg line height difference for G2 & G3

Group	Z ₁	Z ₂
1	1	0
2	0	1
3*	0	0

To plot the model we need to get separate regression formulas for each Z group. We start with the multiple regression model...

Model → $y' = b_1X + b_2Z_1 + b_3Z_2 + a$

Group	Z ₁	Z ₂
1	1	0
2	0	1
3*	0	0

For the Comparison Group coded Z₁ = 0 & Z₂ = 0

Substitute the Z code values

$$y' = b_1X + b_2*0 + b_3*0 + a$$

Simplify the formula

$$y' = b_1X + a$$

← slope ← height

For the Target Group coded Z₁ = 1 & Z₂ = 0

Substitute the Z code values

$$y' = b_1X + b_2*1 + b_3*0 + a$$

Simplify the formula

$$y' = b_1X + (b_2 + a)$$

← slope ← height

For the Target Group coded Z₁ = 0 & Z₂ = 1

Substitute the Z code values

$$y' = b_1X + b_2*0 + b_3*1 + a$$

Simplify the formula

$$y' = b_1X + (b_3 + a)$$

← slope ← height

Plotting & Interpreting Models

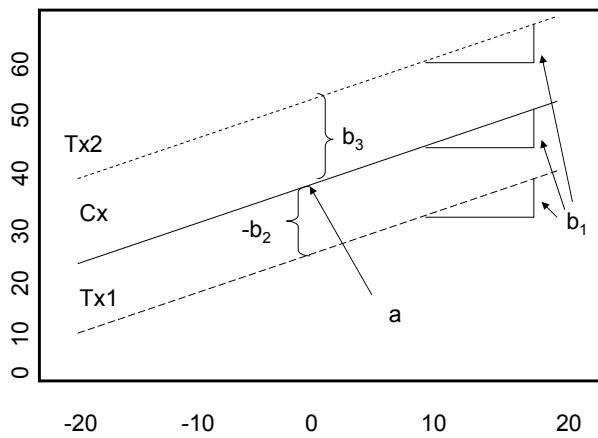
with a centered quantitative predictor & a dummy coded k-category predictor

$$y' = b_1X + -b_2Z_1 + b_3Z_2 + a$$

This is called a main effects model → no interaction → the regression lines are parallel.

$$X_{cen} = X - X_{mean}$$

$$Z_1 = Tx1 \text{ vs. } Cx(0) \quad Z_2 = Tx2 \text{ vs. } Cx(0)$$



a = ht of Cx line
→ mean of Cx

b₁ = slp of Cx line
Cx slp = Tx₁ slp = Tx₂ slp
No interaction

b₂ = htdif Cx & Tx₁
→ Cx & Tx₁ mean dif

b₃ = htdif Cx & Tx₂
→ Cx & Tx₂ mean dif

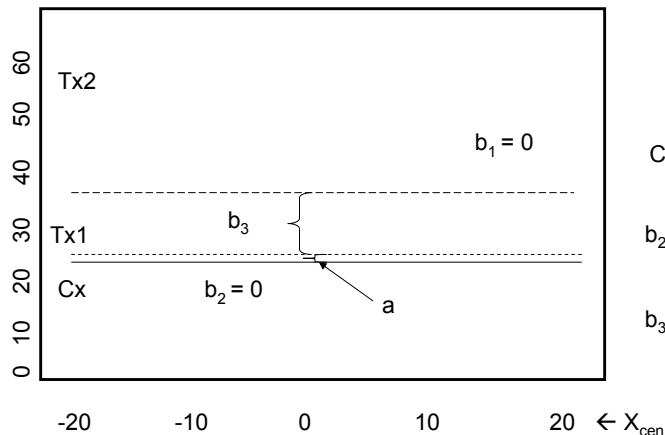
Plotting & Interpreting Models
with a centered quantitative predictor & a dummy coded k-category predictor

$$y' = b_1X + b_2Z_1 + b_3Z_2 + a$$

This is called a main effects model → no interaction → the regression lines are parallel.

$$X_{cen} = X - X_{mean}$$

$$Z_1 = Tx1 \text{ vs. } Cx(0) \quad Z_2 = Tx2 \text{ vs. } Cx(0)$$



- a = ht of Cx line → mean of Cx
- b_1 = slp of Cx line
Cx slp = Tx_1 slp = Tx_2 slp
No interaction
- b_2 = htdif Cx & Tx_1
→ Cx & Tx_1 mean dif
- b_3 = htdif Cx & Tx_2
→ Cx & Tx_2 mean dif

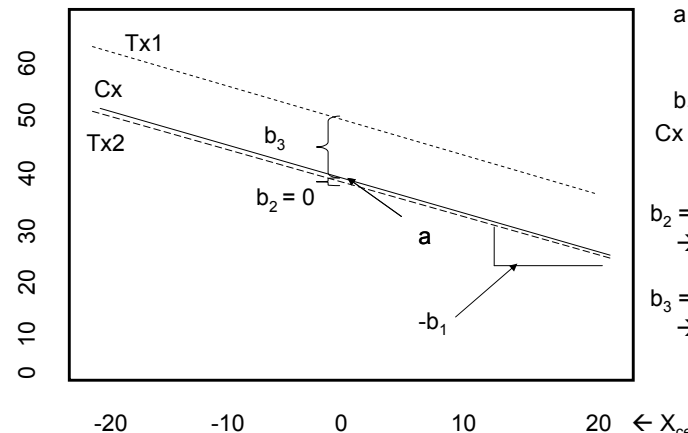
Plotting & Interpreting Models
with a centered quantitative predictor & a dummy coded k-category predictor

$$y' = -b_1X + b_2Z_1 + b_3Z_2 + a$$

This is called a main effects model → no interaction → the regression lines are parallel.

$$X_{cen} = X - X_{mean}$$

$$Z_1 = Tx1 \text{ vs. } Cx(0) \quad Z_2 = Tx2 \text{ vs. } Cx(0)$$



- a = ht of Cx line → mean of Cx
- b_1 = slp of Cx line
Cx slp = Tx_1 slp = Tx_2 slp
No interaction
- b_2 = htdif Cx & Tx_1
→ Cx & Tx_1 mean dif
- b_3 = htdif Cx & Tx_2
→ Cx & Tx_2 mean dif

Models with Interactions

As in Factorial ANOVA, an interaction term in multiple regression is a “non-additive combination”

- there are two kinds of combinations – additive & multiplicative
- main effects are “additive combinations”
- an interaction is a “multiplicative combination”

In SPSS you have to compute the interaction term – as the product of each dummy code for the k-group variable & the centered quantitative variable

Group	Mar1	Mar2
divorced	1	0
married	0	1
single*	0	0

The 3-group variable coded as on the right and a centered quant variable age_cen, then you would compute 2 interaction terms as ...

compute age_mar1_int = mar1 * age_cen.

compute age_mar2_int = mar2 * age_cen

Testing the interaction/regression homogeneity assumption...

There are two “nearly always equivalent” ways of testing the significance of the interaction term:

1. The t-test of the interaction terms will tell whether or not $b=0$ for each.
2. A nested model comparison, using the $R^2\Delta$ F-test to compare the main effect model (dummy-coded binary variable & centered quant variable) with the full model (also including the interaction product terms)

These may not be equivalent – it is possible for one of the interaction terms to have a significant b , but the $R^2\Delta$ to be nonsignificant.

Retaining H_0 : means that

- the interaction does not contribute to the model, after controlling for the main effects
- which can also be called regression homogeneity.

Interpreting the interaction regression weight

If the interaction contributes, we need to know how to interpret the regression weight for the interaction term.

We are used to regression weight interpretations that read like, “The direction and extent of the expected change in Y for a 1-unit change in X , holding all the other variables in the model constant at 0.”

Remember that an interaction in a regression model is about how the slope between the criterion and one predictor is different for different values of another predictor. So, the interaction regression weight interpretation changes just a bit...

An interaction regression weight tells the direction and extent of change in the slope of the Y - X regression line for each 1-unit increase in that Z , holding all the other variables in the model constant at 0.

Notice that in interaction is about regression slope differences, not correlation differences – you already know how to compare corrs

Interpreting the interaction regression weight, cont.

Like interactions in ANOVA, interactions in multiple regression tell how the relationship between the criterion and one variable changes for different values of the other variable – i.e., how the simple effects differ.

Just as with ANOVA, we can pick either variable as the simple effect, and see how the simple effect of that variable is different for different values of the other variable.

The difference is that in this model, one variable is a quantitative variable (X) and the other is a k -groups variable (Z)

So, we can describe the interaction in 2 different ways – both from the same interaction regression weight!

- how does the Y - X regression line slope differ for the k groups?
- how does the Y - X regression line height (i.e., mean) differences among the groups differ for different values of X ?

Interpreting the interaction regression weight, cont.

Eg: FB_1 constant = 1 FB_2 intermittent = 1 No FB = comp
 $perf = 6 * \#pract + 4 * FB_1 + 3 * FB_2 + 4 * Pr_FB_1 + -2 * Pr_FB_2 + 42.3$

We can describe the interaction reg weight for Pr_FB_1 2 ways:

1. The expected direction and extent of change in the Y-X regression slope for each 1-unit increase in that Z, holding...

The slope of the performance-practice regression line for those with constant feedback has a slope 4 more than the slope of the regression line for those without feedback .

2. The expected direction and extent of change in mean difference between constant and No FG conditions for each 1-unit increase in X, holding

The mean performance difference between the feedback and no feedback groups will increase by 4 with each additional practice.

Interpreting the interaction regression weight, cont.

Eg: FB_1 constant = 1 FB_2 intermittent = 1 No FB = comp
 $perf = 6 * \#pract + 4 * FB_1 + 3 * FB_2 + 4 * Pr_FB_1 + -2 * Pr_FB_2 + 42.3$

We can describe the interaction reg weight for Pr_FB_2 2 ways:

1. The expected direction and extent of change in the Y-X regression slope for each 1-unit increase in that Z, holding...

The slope of the performance-practice regression line for those with intermittent feedback has a slope 2 less than the slope of the regression line for those without feedback .

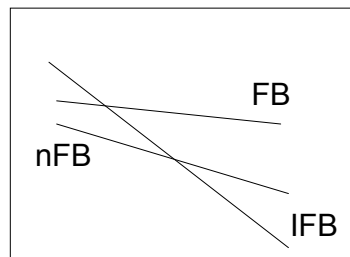
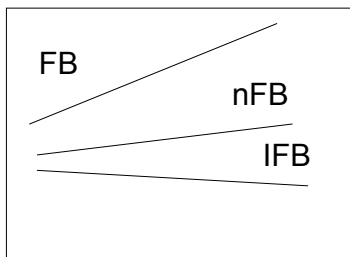
2. The expected direction and extent of change in mean difference between intermittent and No FG conditions for each 1-unit increase in X, holding ...

The mean performance difference between the feedback and no feedback groups will decrease by 2 with each additional practice.

Interpreting the interaction regression weight, cont.

$perf = 6 * \#pract + 4 * FB_1 + 3 * FB_2 + 2 * Pr_FB_1 + -2 * Pr_FB_2 + 42.3$

Be sure to notice that Pr_FB_1 is “more” and Pr_FB_2 is “less” -- neither says whether each is positive, negative or one of each !!! Both of the plots below show FB with a “more positive” slope that nFB and IFB with a “less positive” than nFB



Models with a centered quantitative predictor & a dummy coded k-category predictor & their interaction

Group	Z ₁	Z ₂
1	1	0
2	0	1
3*	0	0

$$y' = b_1X + b_2Z_1 + b_3Z_2 + b_4XZ_1 + b_5XZ_2 + a$$

a → regression constant

- expected value of y if all predictors = 0
- mean of the control group (G3)
- height of control group Y-X regression line

b₁ → regression weight for centered quant predictor

- expected direction and extent of change in Y for a 1-unit increase in X after controlling for the other variable(s) in the model
- simple effect of X when X=0 (G3)
- slope of quant-criterion regression for

b₂ → regression weight for dummy coded comparison of G1 vs G3

- expected direction and extent of change in Y for a 1-unit increase in Z₁ after controlling for the other variable(s) in the model
- simple effect of Z₁ when X = 0 (the centered mean)
- Y-X reg line height difference of G1 & G3 when X = 0

b₃ → regression weight for dummy coded comparison of G2 vs. G3

- expected direction and extent of change in Y for a 1-unit increase in Z₂ after controlling for the other variable(s) in the model
- simple effect of Z₂ when X=0 (the centered mean)
- Y-X reg line height difference of G2 & G3 when X=0

Next page...

Models with a centered quantitative predictor & a dummy coded k-category predictor & their interaction

Group	Z ₁	Z ₂
1	1	0
2	0	1
3*	0	0

$$y' = b_1X + b_2Z_1 + b_3Z_2 + b_4XZ_1 + b_5XZ_2 + a$$

b₄ → regression weight for interaction term involving Z₁

- expected direction and extent of change in the Y-X regression slope for each 1-unit increase in Z₁
- expected direction and extent of change in mean difference between G1 & G3 for each 1-unit increase in X
- Y-X reg line slope difference of groups G1 & G3

b₅ → regression weight for interaction term involving Z₂

- expected direction and extent of change in the Y-X regression slope for each 1-unit increase in Z₂
- expected direction and extent of change in mean difference between G2 & G3 for each 1-unit increase in X
- Y-X reg line slope difference of groups G1 & G3

To plot the model we need to get separate regression formulas for each Z group. We start with the multiple regression model...

Group	Z ₁	Z ₂
1	1	0
2	0	1
3*	0	0

$$y' = b_1X + b_2Z_1 + b_3Z_2 + b_4XZ_1 + b_5XZ_2 + a$$

Gather all "Xs" together

$$y' = b_1X + b_4XZ_1 + b_5XZ_2 + b_2Z_1 + b_3Z_2 + a$$

Factor out "X"

$$y' = (b_1 + b_4Z_1 + b_5Z_2)X + (b_2Z_1 + b_3Z_2 + a)$$

slope

height

Now we apply this formula for each group – changing the values of Z₁ & Z₂ to represent each group in turn

We need to get separate regression formulas for each Z group.

Start with $\rightarrow y' = (b_1 + b_4Z_1 + b_5Z_2)X + (b_2Z_1 + b_3Z_2 + a)$

For the Comparison Group coded $Z_1 = 0$ & $Z_2 = 0$

$$y' = (b_1 + b_4 \cdot 0 + b_5 \cdot 0)X + (b_2 \cdot 0 + b_3 \cdot 0 + a)$$

$$y' = (b_1)X + a$$

slope \rightarrow $(b_1)X$ a \leftarrow height

For the Group 1 coded $Z_1 = 1$ & $Z_2 = 0$

$$y' = (b_1 + b_4 \cdot 1 + b_5 \cdot 0)X + (b_2 \cdot 1 + b_3 \cdot 0 + a)$$

$$y' = (b_1 + b_4)X + (b_2 + a)$$

slope \rightarrow $(b_1 + b_4)X$ $(b_2 + a)$ \leftarrow height

For the Group 2 coded $Z_1 = 0$ & $Z_2 = 1$

$$y' = (b_1 + b_4 \cdot 0 + b_5 \cdot 1)X + (b_2 \cdot 0 + b_3 \cdot 1 + a)$$

$$y' = (b_1 + b_5)X + (b_3 + a)$$

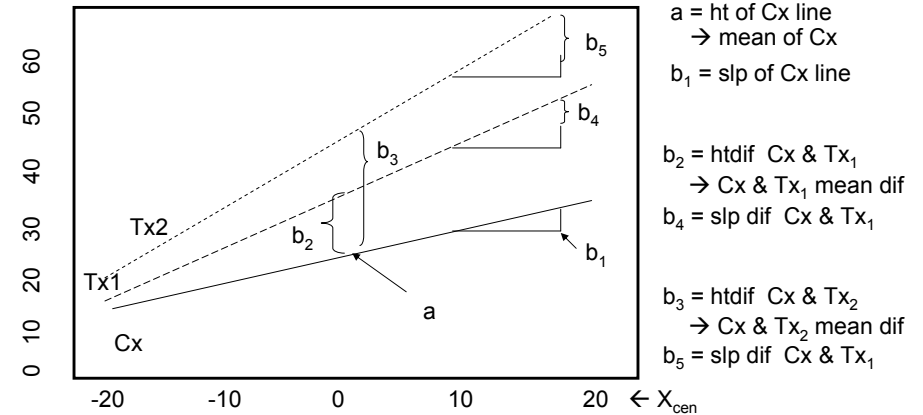
slope \rightarrow $(b_1 + b_5)X$ $(b_3 + a)$ \leftarrow height

Plotting & Interpreting Models with a centered quantitative predictor & a dummy coded k-category predictor & their Interaction

$$y' = b_1X_{cen} + b_2Z_1 + b_3Z_2 + b_4XZ_1 + b_5XZ_2 + a$$

$$X_{cen} = X - X_{mean} \quad Z_1 = Tx1 \text{ vs. } Cx(0) \quad Z_2 = Tx2 \text{ vs. } Cx(0)$$

$$XZ_1 = X_{cen} * Z_1 \quad XZ_2 = X_{cen} * Z_2$$

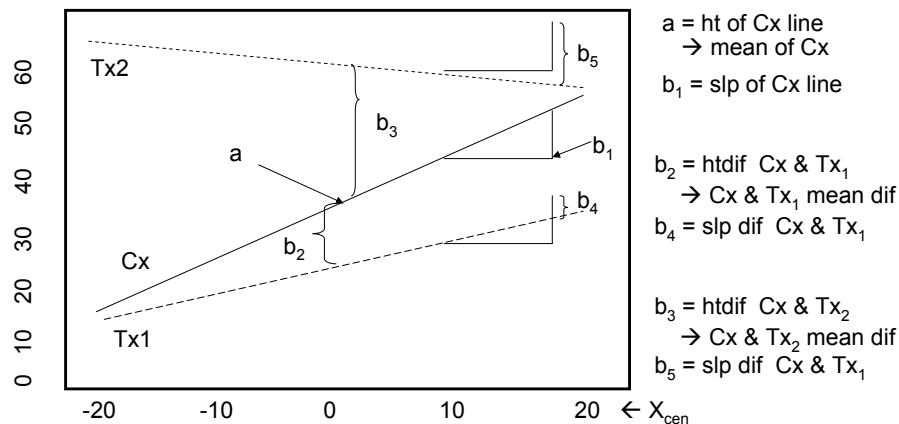


Plotting & Interpreting Models with a centered quantitative predictor & a dummy coded k-category predictor & their Interaction

$$y' = b_1X_{cen} + b_2Z_1 + b_3Z_2 + b_4XZ_1 + b_5XZ_2 + a$$

$$X_{cen} = X - X_{mean} \quad Z_1 = Tx1 \text{ vs. } Cx(0) \quad Z_2 = Tx2 \text{ vs. } Cx(0)$$

$$XZ_1 = X_{cen} * Z_1 \quad XZ_2 = X_{cen} * Z_2$$



Plotting & Interpreting Models with a centered quantitative predictor & a dummy coded k-category predictor & their Interaction

$$y' = b_1X_{cen} + b_2Z_1 + b_3Z_2 + b_4XZ_1 + b_5XZ_2 + a$$

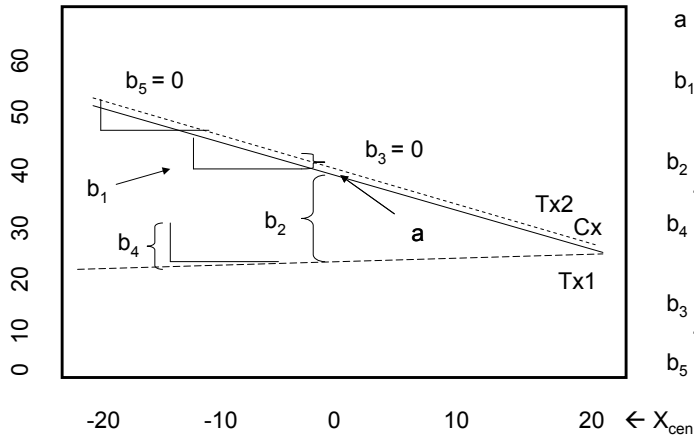
$$X_{cen} = X - X_{mean}$$

$$Z_1 = Tx1 \text{ vs. } Cx(0)$$

$$Z_2 = Tx2 \text{ vs. } Cx(0)$$

$$XZ_1 = X_{cen} * Z_1$$

$$XZ_2 = X_{cen} * Z_2$$



a = ht of Cx line
 → mean of Cx
 b_1 = slp of Cx line

b_2 = htdif Cx & Tx₁
 → Cx & Tx₁ mean dif
 b_4 = slp dif Cx & Tx₁

b_3 = htdif Cx & Tx₂
 → Cx & Tx₂ mean dif
 b_5 = slp dif Cx & Tx₁

So, what do the significance tests from this model tell us and what do they not tell us about the model we have plotted?

We know whether or not the slope of the comparison group is = 0 (t-test of the quant variable weight).

We know whether or not the slope of **each** target group is different from the slope of the comparison group (t-test of the interaction term weight).

But, there is no t-test to tell us whether or not the slope of the Y-X regression line for **either** group = 0.

We know whether or not the mean of **each** target is different from the mean of the comparison group when X = 0 (its mean; t-test of the binary variable weight).

But, there is no test of the group mean differences at any other value of X.

- This is important when there is an interaction, because the interaction tells us the group means differ for different values of X.