Regression Models w/ 2-group & Quant Variables

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

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

- a measured binary variable (e.g., ever- vs. never-married) and a measured quant variable (e.g., age)
- a manipulated binary variable (Tx vs. Cx) and a measured quant variable (e.g., age)
- a measured binary variable (e.g., ever- vs. never-married) and a manipulated quant variable (e.g., 0, 1, 2, 5, or 10 practices)
- a manipulated binary variable (Tx vs. Cx) and a manipulated quant variable (e.g., 0, 1, 2, 5, or 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 binary variable & quant variable
   - No interaction – assumes regression slope homogeneity
   - b-weights for binary & quant variables each represent main effect of that variable
2. Interaction model
   - Terms for binary variable & quant variable
   - Term for interaction - does not assume reg slp homogen !!
   - b-weights for binary & 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 binary predictor

$$y' = b_1X + b_2Z + a$$

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

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

b_1 → 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 both groups

b_2 → regression weight for dummy coded binary predictor

- expected direction and extent of change in Y for a 1-unit increase in Z, after controlling for the other variable(s) in the model
- main effect of Z
- group Y-X regression line height difference

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

Model → $y' = b_1X + b_2Z + a$

For the Comparison Group coded Z = 0

Substitute the 0 in for Z $\quad y' = b_1X + b_2{*}0 + a$

Simplify the formula $\quad y' = \underbrace{b_1}_{\text{slope}}X + \underbrace{a}_{\text{height}}$

For the Target Group coded Z = 1

Substitute the 1 in for Z $\quad y' = b_1X + b_2{*}1 + a$

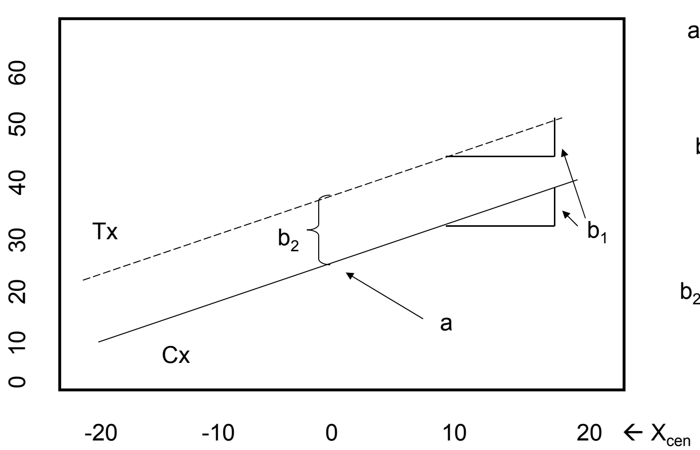
Simplify the formula $\quad y' = \underbrace{b_1}_{\text{slope}}X + \underbrace{(b_2 + a)}_{\text{height}}$

Plotting & Interpreting Models with a centered quantitative predictor & a dummy coded binary predictor

$$y' = b_1X + b_2Z + a$$

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

$X_{cen} = X - X_{mean}$ $\qquad$ Z = Tx1 vs. Cx(0)



a = ht of Cx line
→ mean of Cx

b_1 = slp of Cx line

Cx slp = Tx slp
No interaction

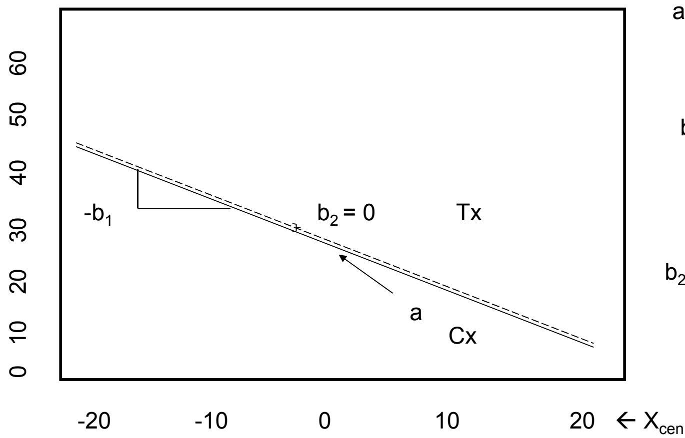
b_2 = htdif Cx & Tx
→ Cx & Tx mean dif

Plotting & Interpreting Models
with a centered quantitative predictor & a dummy coded binary predictor

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

$$y' = -b_1X + -b_2Z + a$$

$X_{cen} = X - X_{mean}$ Z = Tx1 vs. Cx(0)



a = ht of Cx line
→ mean of Cx

b_1 = slp of Cx line
Cx slp = Tx slp
No interaction

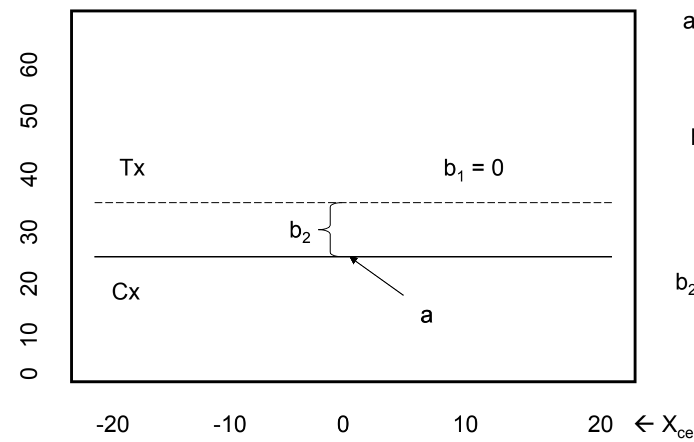
b_2 = htdif Cx & Tx
→ Cx & Tx mean dif

Plotting & Interpreting Models
with a centered quantitative predictor & a dummy coded binary predictor

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

$$y' = b_1X + b_2Z + a$$

$X_{cen} = X - X_{mean}$ Z = Tx1 vs. Cx(0)



a = ht of Cx line
→ mean of Cx

b_1 = slp of Cx line
Cx slp = Tx slp
No interaction

b_2 = htdif Cx & Tx
→ Cx & Tx 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 the binary variable dummy code & the centered quantitative variable

So, if you have sex_dc (0=male & 1=female) and age_cen centered at its mean, you would compute the interaction as…

```
compute  age_sex_int  =  sex_dc  *  age_cen.
```

- males will have age_sex_int values of 0
- females will have age_sex_int values = their age_cen values

Testing the interaction/regression homogeneity assumption…

There are two equivalent ways of testing the significance of the interaction term:

1. The t-test of the interaction term will tell whether or not b=0
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 term)

These are equivalent because $t^2 = F$, both with the same df & p.

Retaining H0: means that

- the interaction term 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 to the model, 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 increase 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 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 binary grouping 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 2 groups?

• how does the Y-X regression line height difference differ for different values of X (how does the mean difference differ for different values of X)?

Interpreting the interaction regression weight, cont.

Example: FB = feedback 0 = no feedback 1 = feedback

perf' = 8.2*#pract + 4.5*FB + 4.0Pr_FB + 42.3

We can describe the interaction regression weight 2 ways:

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

   The slope of the performance-practice regression line for those with feedback (coded 1) has a slope 4 more than the slope of the regression line for those without feedback (coded 0).

2. The expected direction and extent of change in group mean difference for each 1-unit increase in X, holding …

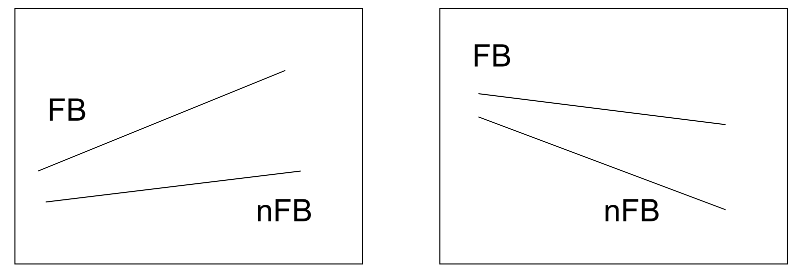
   The mean performance difference between the feedback and no feedback groups will increase by 4 with each additional practice.

Interpreting the interaction regression weight, cont.

perf' = 8.2*#pract + 4.5*FB + 4.0Pr_FB + 42.3

The slope of the performance-practice regression line for those with feedback (coded 1) has a slope 4 more than the slope of the regression line for those without feedback (coded 0).

Be sure to notice that it says "more" -- it doesn't say whether both are positive, negative or one of each !!! Both of the plots below show FB with a "more positive" slope that nFB



Models with a centered quantitative predictor, a dummy coded binary predictor & their interaction

$y' = b_1X + b_2Z + b_3XZ + a$

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

b_1 → 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 Z = 0 (comparison group)
- slope of Y-X regression line for the comparison group (Z coded 0)

b_2 → regression weight for dummy coded binary 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 Z when X = 0
- Y-X reg line height difference of groups when X = 0 (the centered mean)

b_3 → regression weight for interaction term
- expected direction and extent of change in the Y-X regression slope for each 1-unit increase in Z, after controlling for the other variable(s) in the model
- expected direction and extent of change in group mean difference for each 1-unit increase in X, after controlling for the other variable(s) in the model
- Y-X reg line slope difference of groups

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

Model → $y' = b_1X + b_2Z + b_3XZ + a$

Gather all "Xs" together $y' = b_1X + b_3XZ + b_2Z + a$

Factor out "X" $y' = \underbrace{(b_1 + b_3Z)}_{\text{slope}}X + \underbrace{(b_2Z + a)}_{\text{height}}$

For the Comparison Group coded Z = 0

Substitute the 0 in for Z $y' = (b_1 + b_3{*}0)X + (b_2{*}0 + a)$

Simplify the formula $y' = b_1X + a$

slope height

For the Target Group coded Z = 1

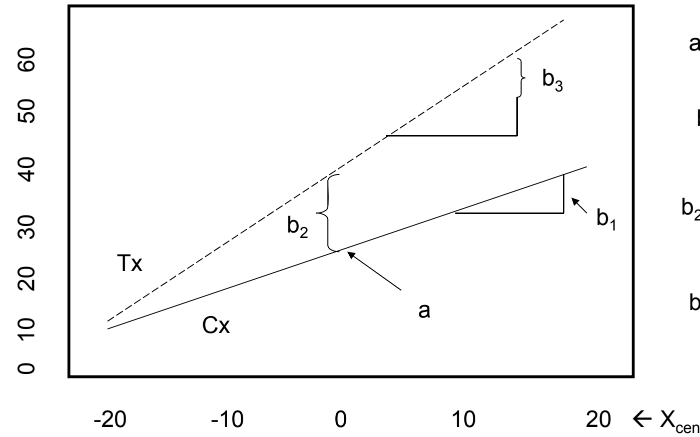
Substitute the 1 in for Z $y' = (b_1 + b_3{*}1)X + (b_2{*}1 + a)$

Simplify the formula $y' = \underbrace{(b_1 + b_3)}_{\text{slope}}X + \underbrace{(b_2 + a)}_{\text{height}}$

Plotting Models with a centered quantitative predictor, a dummy coded binary predictor & their interaction

$$y' = b_1X + b_2Z + b_3XZ + a$$

$X_{cen} = X – X_{mean}$ Z = Tx1 vs. Cx(0) $XZ = X_{cen} * Z$



a = ht of Cx line
→ mean of Cx

b_1 = slp of Cx line

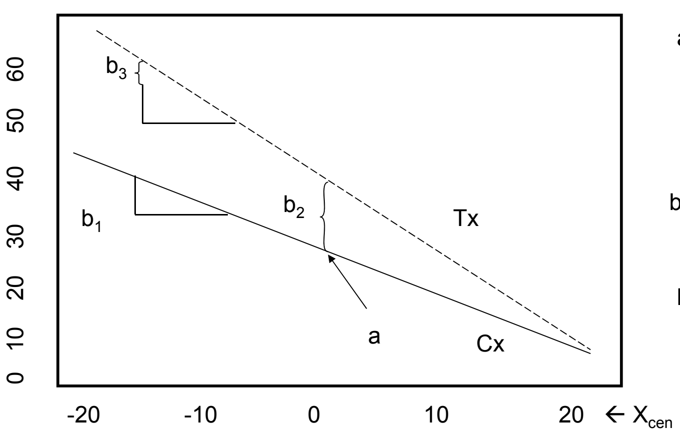
b_2 = htdif Cx & Tx
→ Cx & Tx mean dif at X=0

b_3 = slpdif Cx & Tx

Plotting Models with a centered quantitative predictor, a dummy coded binary predictor & their interaction

$$y' = b_1X + b_2Z + b_3XZ + a$$

$X_{cen} = X – X_{mean}$ Z = Tx1 vs. Cx(0) $XZ = X_{cen} * Z$



a = ht of Cx line
→ mean of Cx

b_1 = slp of Cx line

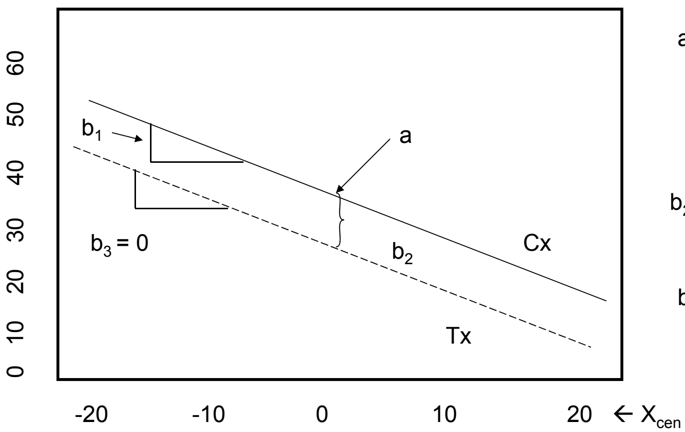
b_2 = htdif Cx & Tx
→ Cx & Tx mean dif at X=0

b_3 = slpdif Cx & Tx

Plotting Models with a centered quantitative predictor, a dummy coded binary predictor & their interaction

$$y' = b_1X + b_2Z + b_3XZ + a$$

$X_{cen} = X - X_{mean}$ Z = Tx1 vs. Cx(0) $XZ = X_{cen} * Z$



a = ht of Cx line
→ mean of Cx

b_1 = slp of Cx line

b_2 = htdif Cx & Tx
→ Cx & Tx mean dif at X=0

b_3 = slpdif Cx & Tx

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 group coded 0 is = 0 (t-test of the quant variable weight).

We know whether or not the slope of the group coded 1 is different from the slope of the group coded 0 (t-test of the interaction term weight)

But, there is no t-test to tell us if the slope of the Y-X regression line for the group coded 1 = 0.

We know whether or not the mean of the group coded 1 is different from the mean of the group coded 0, when X =0 (its mean; t-test of the binary variable weight.

But, there is no test of the group mean difference 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.