# Regression Models w/ 2 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 …

- 2 measured quant variable (e.g., age & experience)
- 2 manipulated quant variables (e.g., 4, 16, 32, 64 practices & % receiving feedback 0, 25, 50, 75, 100)
- a measured quant variable (e.g., age) and a manipulated quant variable (e.g., 4, 16, 32, 64 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 both quant variables
   - No interaction – assumes regression slope homgeneity
   - b-weights for the 2 quant variables each represent main effect of that variable
2. Interaction model
   - Terms for both quant variables
   - Term for interaction - does not assume reg slp homogen !!
   - b-weights for 2 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 2 centered quantitative predictors

$$y' = b_1X_1 + b_2X_2 + a$$

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

a → regression constant
- expected value of Y if all predictors = 0

$b_1$ → regression weight for centered quant predictor $X_1$
- expected direction and extent of change in Y for a 1-unit increase in $X_1$ after controlling for the other variable(s) in the model
- main effect of $X_1$
- Slope of Y-$X_1$ regression line for all values of $X_2$

$b_2$ → regression weight for centered quant predictor $X_2$
- expected direction and extent of change in Y for a 1-unit increase in $X_2$ after controlling for the other variable(s) in the model
- main effect of $X_2$
- Height difference of Y-$X_1$ regression line for different values of $X_2$

By the way – we can plot this model with either $X_1$ or $X_2$ on the x-axis and plot that Y-$X_2$ regression line for different values of $X_1$ – our choice

## Models with 2 centered quantitative predictors

$$y' = b_1X_1 + b_2X_2 + a$$

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

Same idea …
- we want to plot these models so that we can see how both of the predictors are related to the criterion

Different approach …
- when the second predictor was binary or had k-categories, we plotted the Y-X regression line for each Z group
- now, however, we don't have any groups – both the $X_1$ & $X_2$ variables are centered quantitative variables
- what we'll do is to plot the Y-$X_1$ regression line for different $X_2$ values
- the most common approach is to plot the Y-$X_1$ regression line for…
  - the mean of $X_1$
  - +1 std above the mean of $X_1$
  - -1 std below the mean of $X_1$

} We'll plot 3 lines

To plot the model we need to get separate regression formulas for each chosen value of $X_2$. Start with the multiple regression model..

Model ➔ $y' = b_1X + b_2X_1 + a$

For $X_2$ = 0 (the mean of centered Z)
Substitute the 0 in for $X_2$
Simplify the formula

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

slope ↗ ↖ height

For $X_2$ = +1 std
Substitute the std value in for $X_2$
Simplify the formula

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

slope ↗ ⟶

For $X_2$ = -1 std
Substitute the std value in for $X_2$
Simplify the formula

$$y' = b_1X + -b_2*std + a$$
$$y' = b_1X + (-b_2*std + a)$$

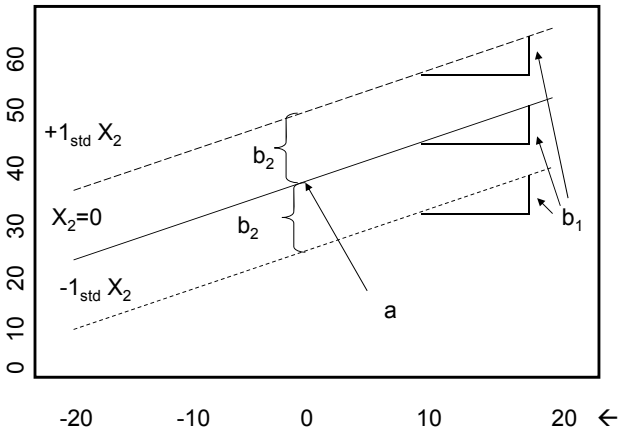
slope ↗ height ⟶

## Plotting & Interpreting Models with 2 centered quantitative predictors

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

$y' = b_1X_{1cen} + b_2X_{2cen} + a$

$X_{1cen} = X_1 - X_{1mean}$ $X_{2cen} = X_2 - X_{2mean}$




a = ht of $X_{2mean}$ line

$b_1$ = slp of $X_{2mean}$ line

0 slp = +1std slp = -1std slp
No interaction

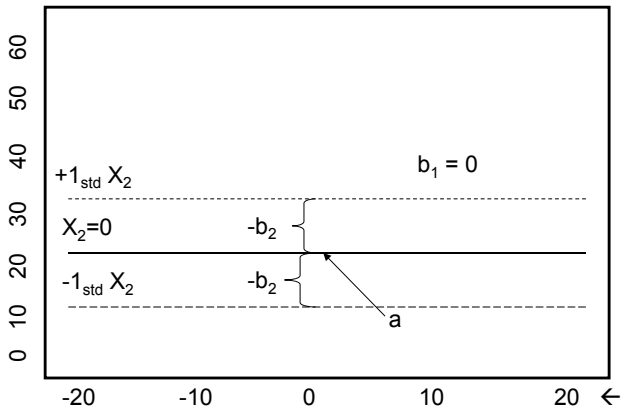
$b_2$ = htdifs among $X_2$-lines

## Plotting & Interpreting Models with 2 centered quantitative predictors

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

$y' = b_1X_{1cen} + b_2X_{2cen} + a$

$X_{1cen} = X_1 - X_{1mean}$ $X_{2cen} = X_2 - X_{2mean}$




a = ht of $X_{2mean}$ line

$b_1$ = slp of $X_{2mean}$ line

0 slp = +1std slp = -1std slp
No interaction

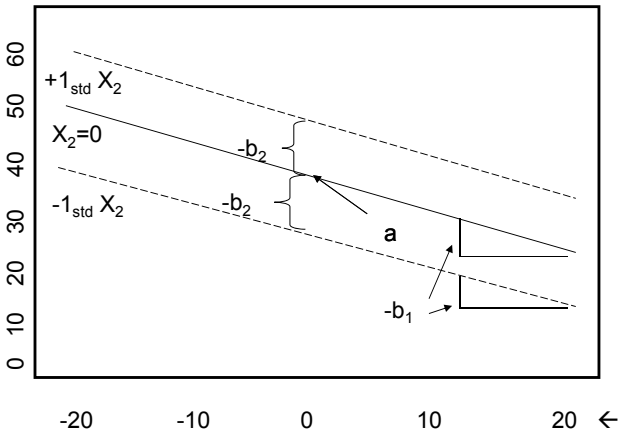
$b_2$ = htdifs among $X_2$-lines

## Plotting & Interpreting Models with 2 centered quantitative predictors

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

$y' = b_1X_{1cen} + b_2X_{2cen} + a$

$X_{1cen} = X_1 - X_{1mean}$ $X_{2cen} = X_2 - X_{2mean}$




a = ht of $X_{2mean}$ line

$b_1$ = slp of $X_{2mean}$ line

0 slp = +1std slp = -1std slp
No interaction

$b_2$ = htdifs among $X_2$-lines

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 2 centered quantitative variable

So, if you have exp_cen centered at its mean and age_cen centered at its mean, you would compute the interaction as…

```
compute  exp_age_int  =  exp_cen  *  age_cen.
```

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 (2 centered quant variables) 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, 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 one Y-X regression line for each 1-unit increase in the other X, 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, both variables are quantitative variables ($X_1$ & $X_2$)

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

- how does Y-$X_1$ regression line slope differ as $X_2$ increases by 1
- how does Y-$X_2$ regression line slope differ as $X_1$ increases by 1

## Interpreting the interaction regression weight, cont.

Example:

perf' = 8.2*#pract + 4.5*Age + 4.0Pr_Age + 42.3

We can describe the interaction regression weight 2 ways:

1. The expected direction and extent of change in the Y-$X_1$ regression slope for each 1-unit increase in $X_2$, holding…

   The slope of the performance-practice regression line increases by 4 with each 1-unit increase in age.

2. The expected direction and extent of change in the Y-$X_2$ regression slope for each 1-unit increase in $X_1$, holding…

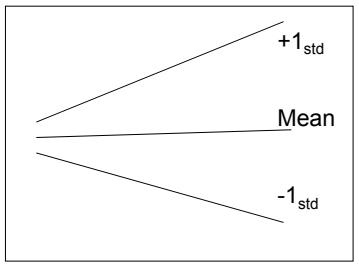
   The slope of the performance-age regression line increases by 4 with each 1-unit increase in practice.

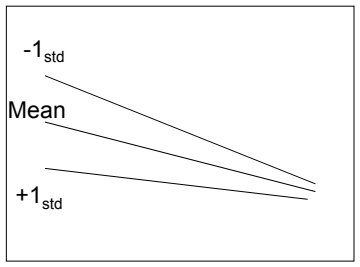
## Interpreting the interaction regression weight, cont.

perf' = 8.2*#pract + 4.5*Age + 4.0Pr_AGE + 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 all are positive, negative or some of each !!! Both of the plots below show a + interaction regression weight.



Models with 2 centered quantitative predictors & their Interaction

$$y' = b_1X_1 + b_2X_2 + b_3X_1X_2 + a$$

Same idea …

- we want to plot these models so that we can see how both of the predictors are related to the criterion

Different approach …

- when the second predictor was binary or had k-categories, we plotted the Y-X regression line for each Z group
- now, however, we don't have any groups – both the $X_1$ & $X_2$ variables are centered quantitative variables
- what we'll do is to plot the Y-$X_1$ regression line for different $X_2$ values
- the most common approach is to plot the Y-$X_1$ regression line for…
  - the mean of $X_1$
  - +1 std above the mean of $X_1$
  - -1 std below the mean of $X_1$

} We'll plot 3 lines

To plot the model we need to get separate regression formulas for each chosen value of $X_2$. Start with the multiple regression model..

Model →
$$y' = b_1X_1 + b_2X_2 + b_3X_1X_2 + a$$

Gather all "Xs" together
$$y' = (b_1 + b_3X_2)X_1 + (b_2X_2 + a)$$

Factor out "X"

slope: $(b_1 + b_3X_2)$ height: $(b_2X_2 + a)$

We will apply this reconfigured model with three different values of $X_2$, to get three Y-X regression lines to plot

To plot the model we need to get separate regression formulas for each chosen value of $X_2$. Start with the multiple regression model..

$$y' = (b_1 + b_3X_2)X_1 + (b_2X_2 + a)$$

For $X_2$ = 0 (the mean of centered $X_2$)
Substitute the 0 for $X_2$
Simplify the formula

$$y' = (b_1 + b_30)X_1 + (b_20 + a)$$

$$y' = b_1X_1 + a$$

slope: $b_1$ height: $a$

For $X_2$ = +1 std
Substitute the std for $X_2$
Simplify the formula

$$y' = \underbrace{(b_1 + b_3\text{std})X_1}_{\text{slope}} + \underbrace{(b_2\text{std} + a)}_{\text{height}}$$

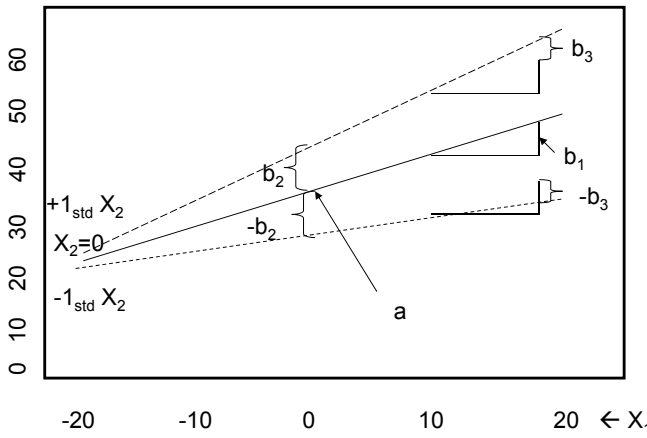
For $X_2$ = -1 std
Substitute the -std for $X_2$
Simplify the formula

$$y' = \underbrace{(b_1 + b_3{*}\text{-std})X_1}_{\text{slope}} + \underbrace{(b_2{*}\text{-std} + a)}_{\text{height}}$$

## Plotting & Interpreting Models with 2 centered quantitative predictors & their Interaction

$y' = b_1X_1 + b_2X_2 + b_3X_1X_2 + a$

$X_{1cen} = X_1 - X_{1mean}$ $X_{2cen} = X_2 - X_{2mean}$ $X_1X_2 = X_{1cen} * X_{2cen}$




a = ht of $X_{2mean}$ line

$b_1$ = slp of $X_{2mean}$ line

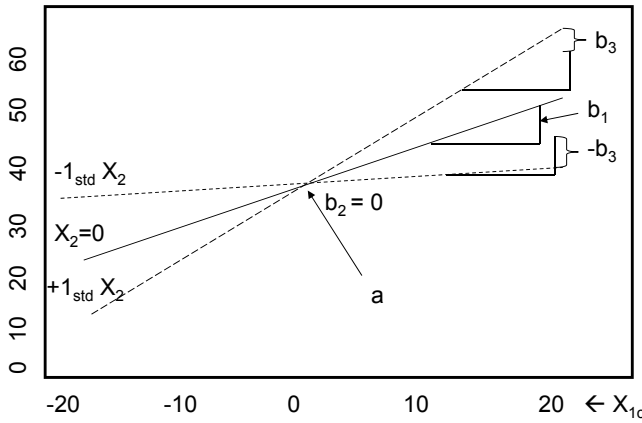
$b_2$ = htdifs among $X_2$ lines at $X_1 = 0$

$b_3$ = htdifs among $X_2$ lines


## Plotting & Interpreting Models with 2 centered quantitative predictors & their Interaction

$y' = b_1X_1 + b_2X_2 + b_3X_1X_2 + a$

$X_{1cen} = X_1 - X_{1mean}$ $X_{2cen} = X_2 - X_{2mean}$ $X_1X_2 = X_{1cen} * X_{2cen}$



a = ht of $X_{2mean}$ line

$b_1$ = slp of $X_{2mean}$ line

$b_2$ = htdifs among $X_2$ lines at $X_1 = 0$

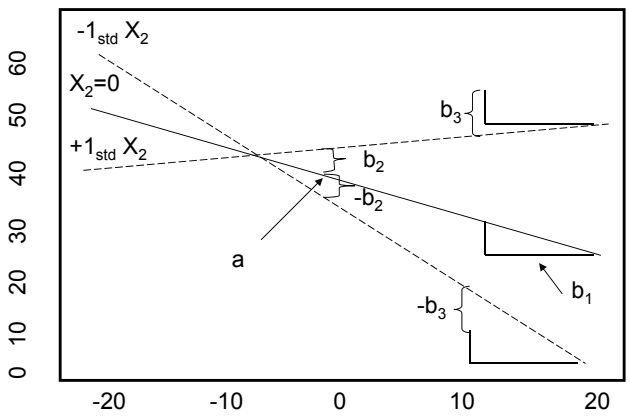
$b_3$ = htdifs among $X_2$ lines



## Plotting & Interpreting Models with 2 centered quantitative predictors & their Interaction

$y' = b_1X_1 + b_2X_2 + b_3X_1X_2 + a$

$X_{1cen} = X_1 - X_{1mean}$ $X_{2cen} = X_2 - X_{2mean}$ $X_1X_2 = X_{1cen} * X_{2cen}$



a = ht of $X_{2mean}$ line

$b_1$ = slp of $X_{2mean}$ line

$b_2$ = htdifs among $X_2$ lines at $X_1 = 0$

$b_3$ = htdifs among $X_2$ lines

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 Y-$X_1$ regression line = 0 for the mean of $X_2$ (t-test of the $X_1$ weight).

We know whether or not the slope of the Y-$X_1$ regression weight changes with the value of $X_2$ (t-test of the interaction term weight).

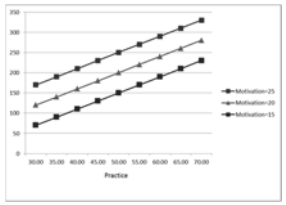
But, there is no t-test to tell us if the slope of the Y-$X_1$ regression line = 0 for any other value of $X_2$.

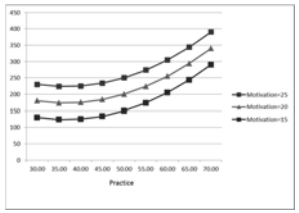
We know whether or not the height of the Y-$X_1$ regression line is different for different values of $X_2$ when $X_1$ = 0 (t-test of the $X_2$ weight).

But, there is no test of the Y-$X_1$ regression line height difference at any other value of $X_2$.

## QxQ Main & Interaction effects in this model.

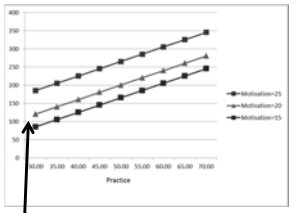
**Linear & Quadratic Main Effects**



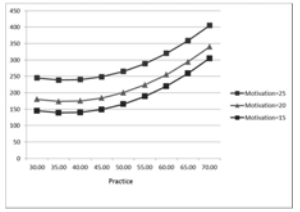


**Quadratic practice Linear motivation**

**Linear practice Linear motivation**



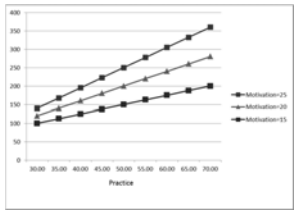
**Linear practice Quadratic motivation**



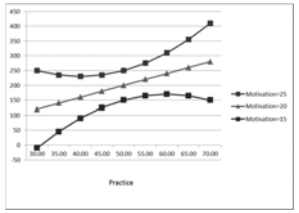
**Quadratic practice Quadratic motivation**

**Unequal height change from 15→20 line vs 20→25 line**

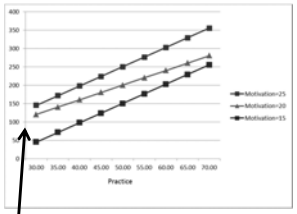
**Linear & quadratic Interactions**



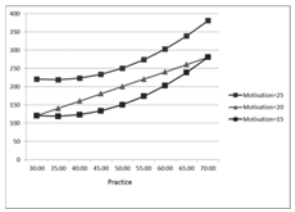
**Linear practice X Linear motivation**



**Quadratic practice X Linear motivation**



**Linear practice X Quadratic motivation**



**Quadratic practice X Quadratic motivation**

**Unequal height change from 15→20 lines vs 20→25 lines is dif for dif practice values**