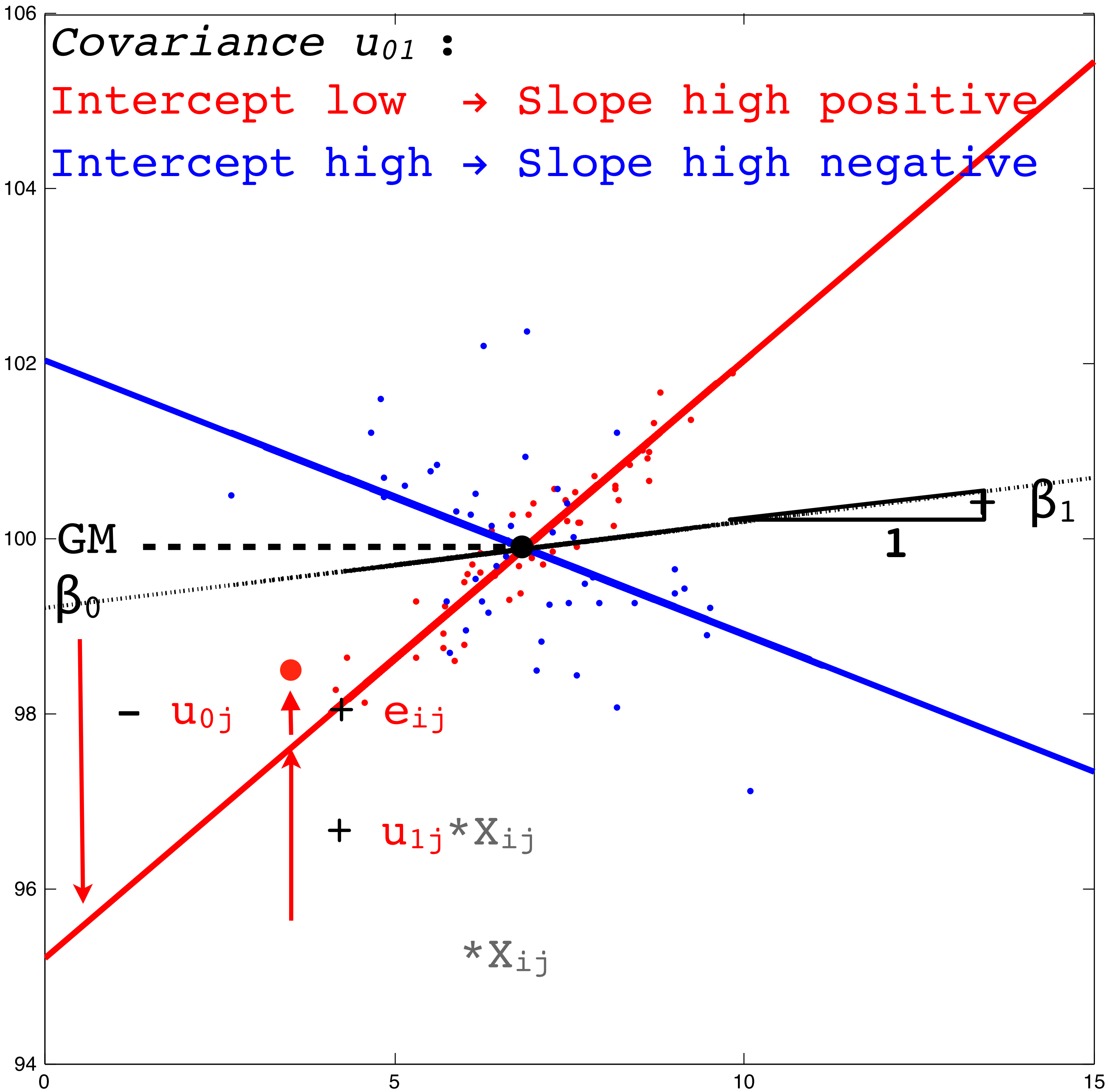
Convenient naming:
→ Fixed parameters are the same for every data point
→ Random parameters vary for every datapoint

Effects of CENTERING and STANDARDIZING:

#1 uncentered, unstandardized



Fixed part: $\beta_0 + \beta_1 * X_{ij}$

Intercept: $\beta_0 + \beta_1*(X_{ij}=0)$
Slope: β_1 (overall slope)

Random part: u_{0j} $u_{1j}*X_{ij}$ u_{01} e_{ij}

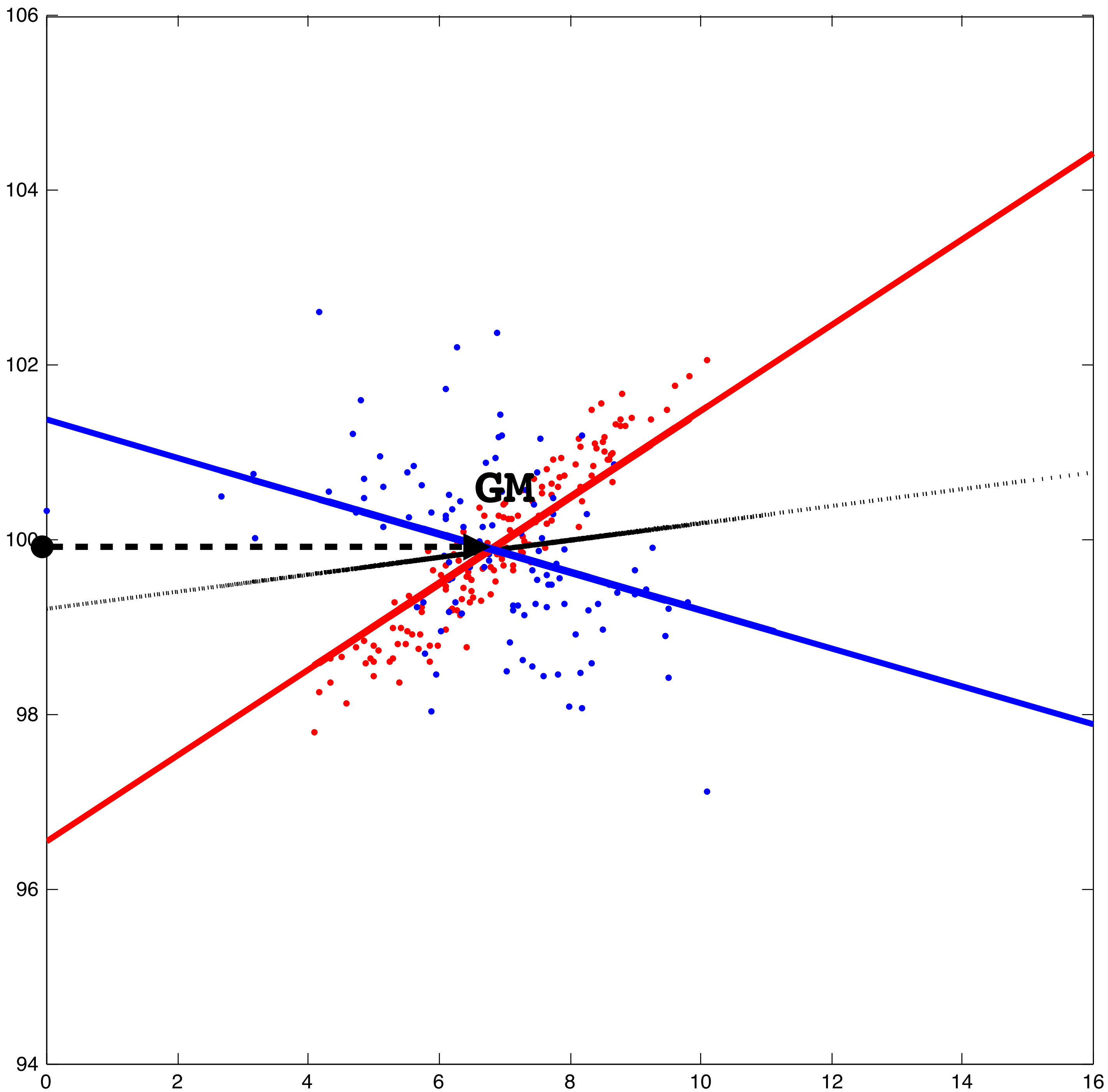
Intercept: $\beta_0 - u_{0j}$
Slope: $(\beta_1 + u_{1j})*X_{ij}$

Y_{ij} for data point X_{ij}:
$\beta_0 - u_{0j} + (\beta_1 + u_{1j})*X_{ij} + e_{ij}$

Intercept: $\beta_0 + u_{0j}$
Slope: $(\beta_1 - u_{1j})*X_{ij}$

Y_{ij} for data point X_{ij}:
$\beta_0 + u_{0j} + (\beta_1 - u_{1j})*X_{ij} + e_{ij}$

Effects of CENTERING and STANDARDIZING:
#2 Grand Mean centering



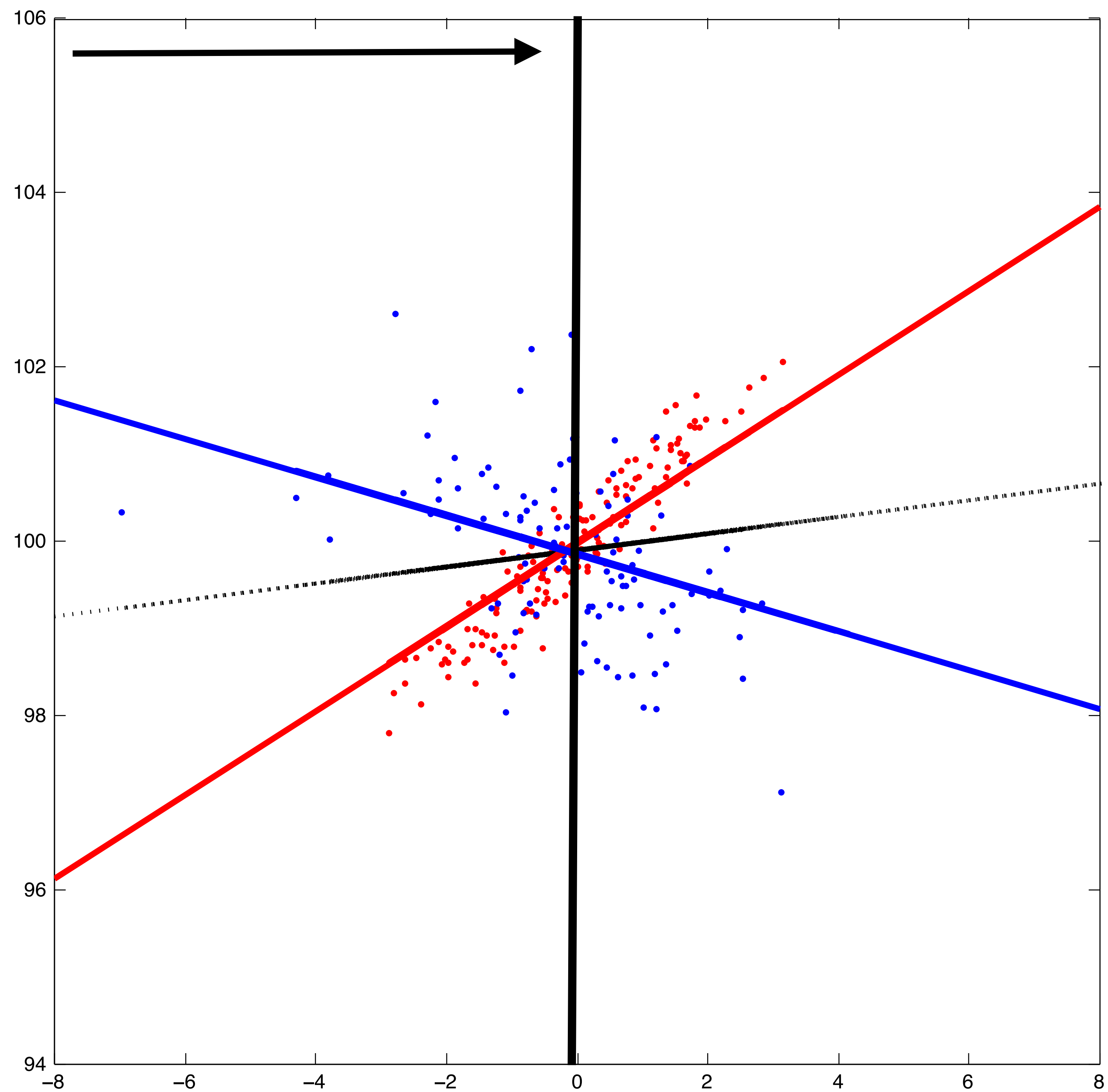
The mean of a distribution is a *location parameter* that measures the *central tendency* (here located at X=7).

Centering the variable X :

$$X_{GMcent} = \left(X_i - \overline{X}\right) \quad for \quad i = 1,\ldots,n$$

is a linear transformation of the data that just shifts the *location* of every data point with respect to the x-axis. The central tendency of X will now be located at X=0 instead of X=7.

Effects of CENTERING and STANDARDIZING:
#2 Grand Mean centering



The mean of a distribution is a *location parameter* that measures the *central tendency* (here located at X=7).

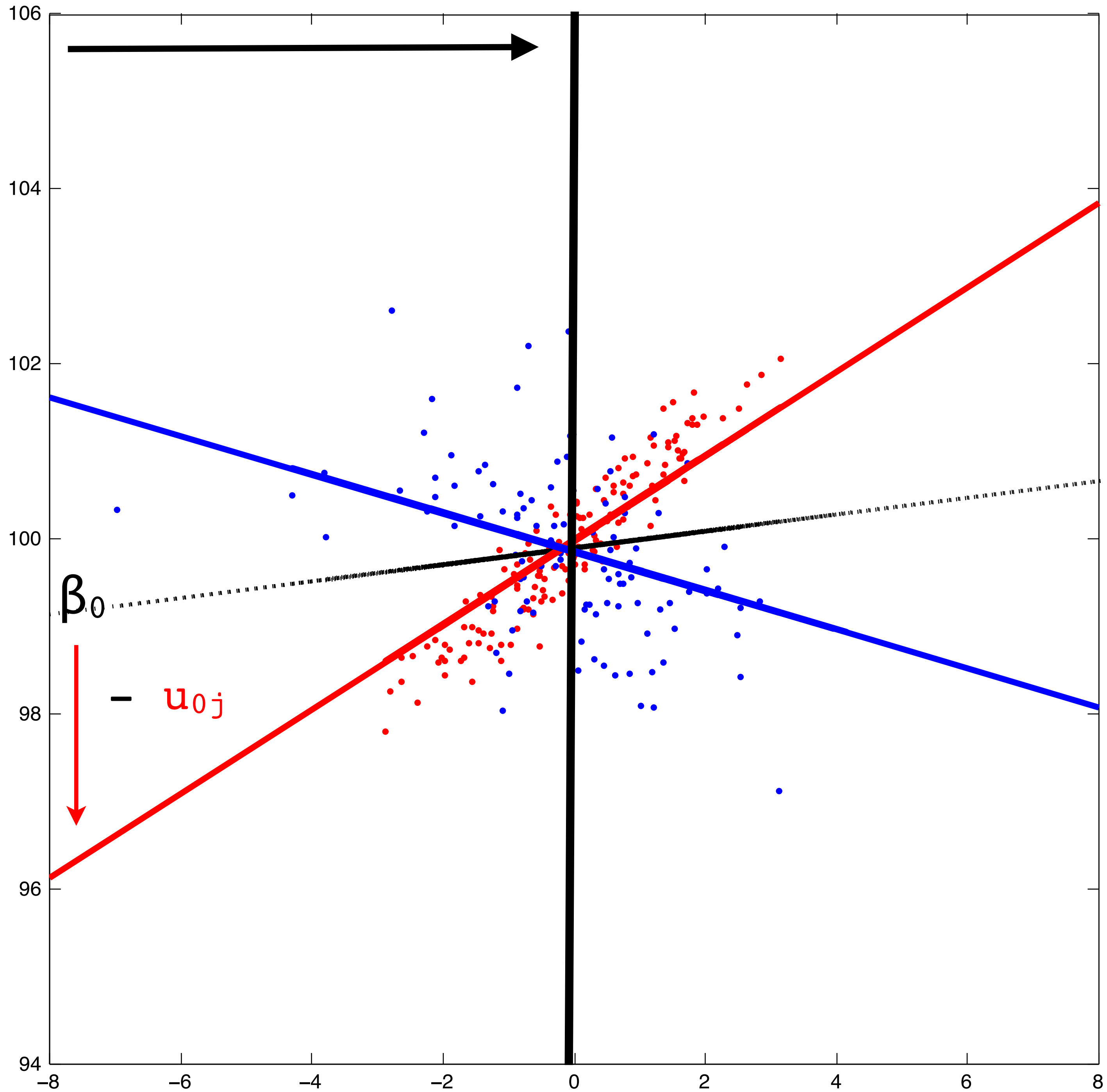
Centering the variable X :

$$X_{GMcent} = \left(X_i - \overline{X}\right) \quad for \quad i = 1,\ldots,n$$

is a linear transformation of the data that just shifts the *location* of every data point with respect to the x-axis. The central tendency of X will now be located at X=0 instead of X=7.

Does the location of X=0 influence model estimates?

Effects of CENTERING and STANDARDIZING:
#2 Grand Mean centering



Values of intercepts change:

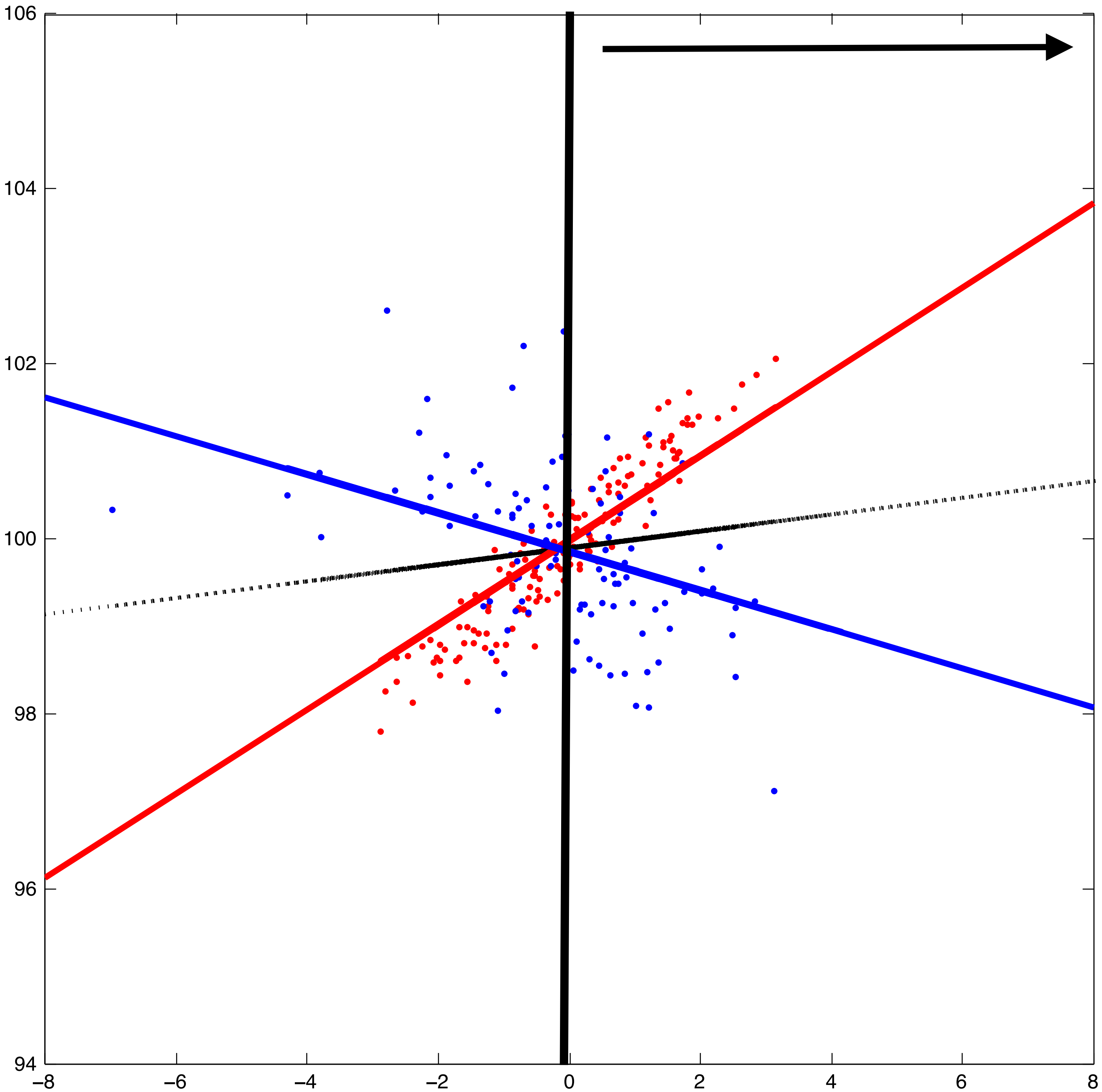
β_0 changes and so will u_{0j}

In this example the intercept variance disappeared!(the same intercepts for both groups)

This means intercepts will now not covary with slopes, a fanning in pattern (u_{01}=neg) is now $u_{01}=0$.

But this was also the case in the uncentered version at the mean of X (7)! We were interpreting model outcomes at values of X which are far beyond the observed values. Stay within the range of your data!

Effects of CENTERING and STANDARDIZING:
#2 Grand Mean centering



Intercepts will change:

If we move the location of X=0 even further values will change again.

Intercept-slope covariance will change:

In this example:

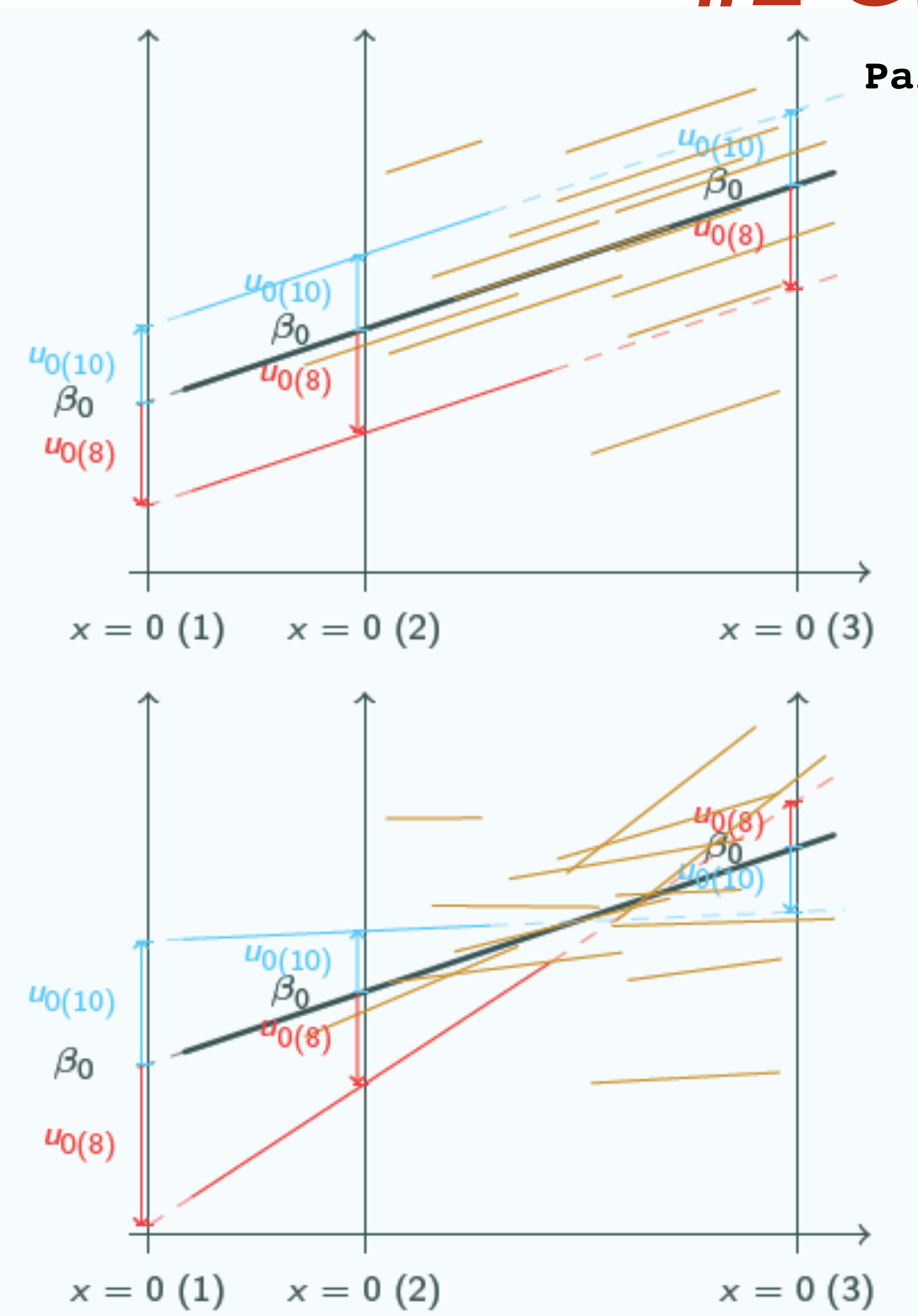
u_{01}=neg → u_{01}=0 → u_{01}=pos

Slopes will not change:

A slope represents the amount of change in Y if we move one unit in X. Slopes will only change if we change the units of X.

Effects of CENTERING and STANDARDIZING:

#2 Grand Mean centering



Parameter estimates that depend on the location of X=0

Random intercepts, fixed slopes:

- β_0

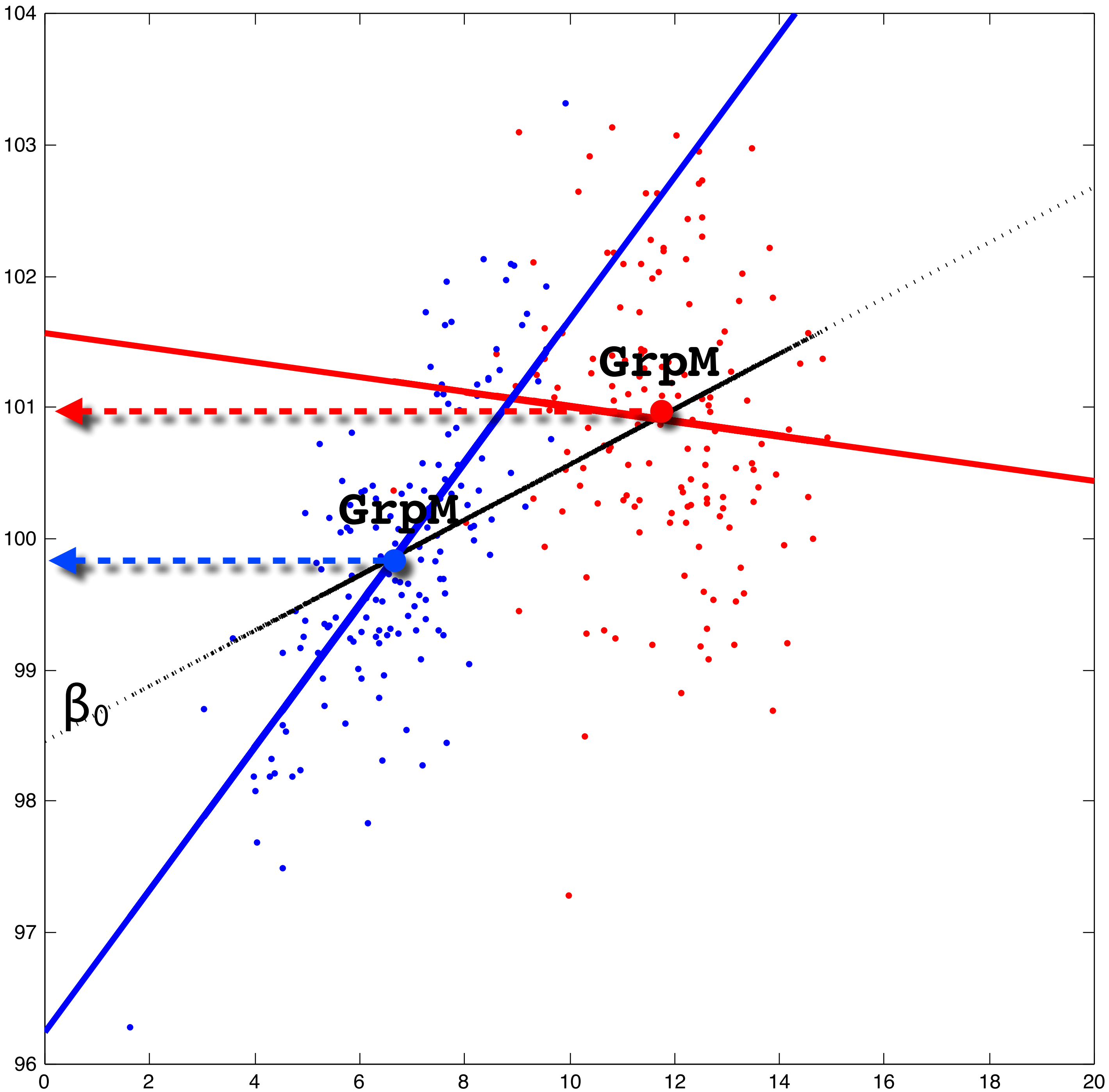
Random intercepts, random slopes:

- β_0
- σ^2_{u0}
- σ_{u01}

A model with Grand mean centered predictors is linearly equivalent to the same model with uncentered predictors

Effects of CENTERING and STANDARDIZING:

#3 Goup Mean centering



If there are *k* groups in the dataset, each individual *i* can be centered to their own group mean:

$$X_{GrMcent} = \left(X_{ij} - \overline{X}_j \right) \quad for \quad \begin{matrix} i = 1,\ldots,n \\ j = 1,\ldots,k \end{matrix}$$

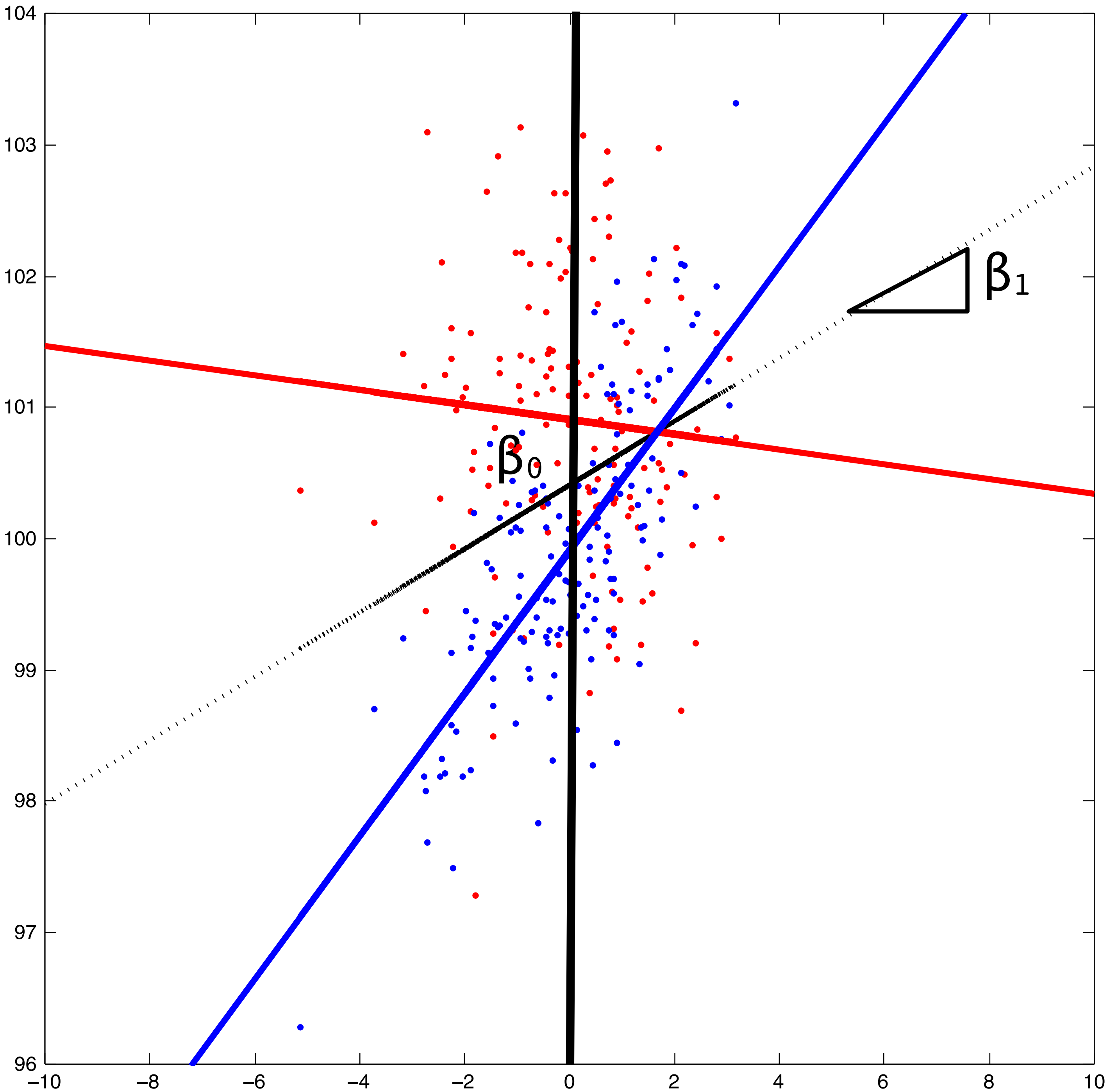
All the groups in the dataset will now have the same mean = 0 for X.

In other words: The correlation of the level 1 predictor (X) with all level 2 predictors has been removed.

(this may be necessary if correlation are very high)

Effects of CENTERING and STANDARDIZING:

#3 Goup Mean centering



If we now fit a model it will no longer be linearly numerically equivalent to either the uncentered or the grand-mean centered version of the model.

The model is now an evaluation of differences in Y at the group means of X.

For instance, if level1=students, level2=schools, Y=reading, X=arithmetic grpmn cntrd:

β_1 represents the average relation between Reading and Arithmetic in the population adjusted for mean arithmetic scores of individual schools.

Effects of CENTERING and STANDARDIZING:

#4 Standardizing

Standardize (with respect to unit, mean, standard deviation and range) to make the the relative magnitude of coefficients of predictors comparable, or to compare regressions on different dependent variables.

- Two methods:

1 Multiply unstandardized coefficients by ratio of SD's: s_x/s_y

2 Standardize variables before analysis

Standardizing before analysis and centering are similar except that there is a multiplicative factor in standardization. A SD is a *scale* parameter of a distribution. Obviously it will change values of slopes (Y change for 1 *unit* change in X). Though it is a linear transformation of the variable it will not be a linear transformation of the variance components of the random part.

If the random part is important for presentation and interpretation this can be a problem.

- Take care if you have dummy variables and/or interaction effects!

Within group, between group and contextual effects of for level 1 predictors

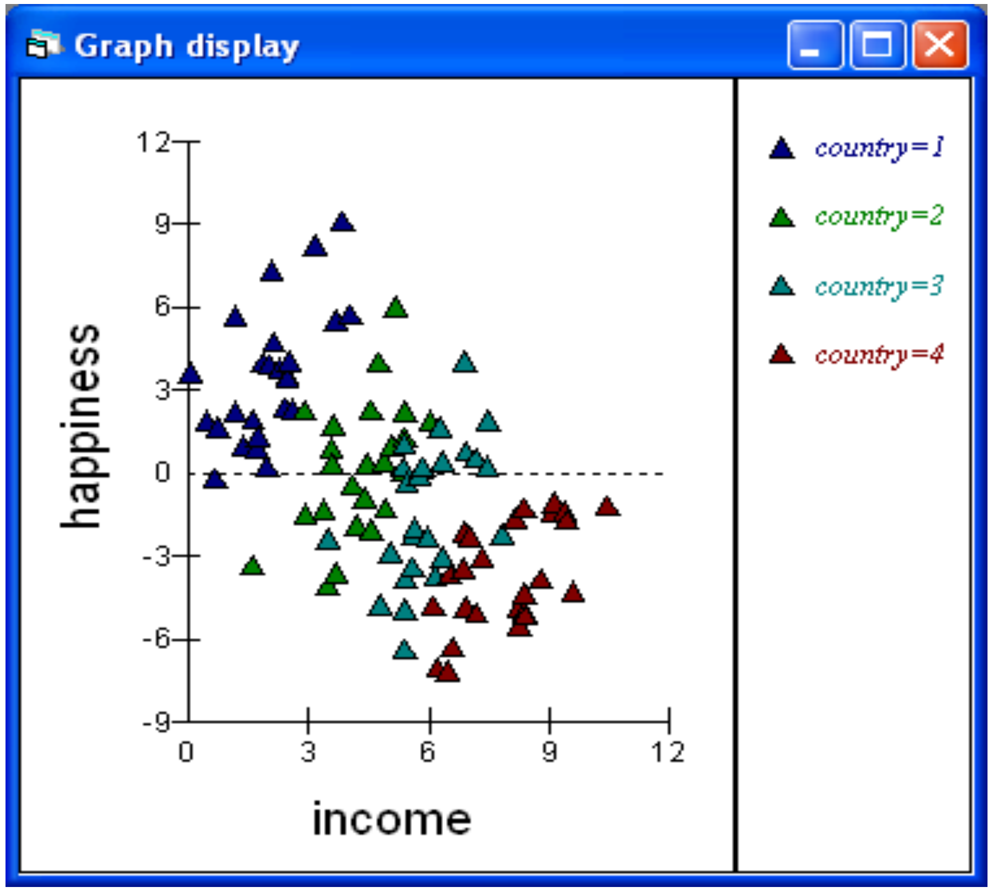
Often contextual variables are constructed by aggregating level 1 predictors, country level income, family level aggression etc

We will simulate a small data set with 100 individuals from four groups, for the sake of argument lets say that our response is happiness, we have a predictor variable that is income and the groups are 4 different countries. Lets simulate the mean income in each country as 2,4,6,8. We then simulate our response as

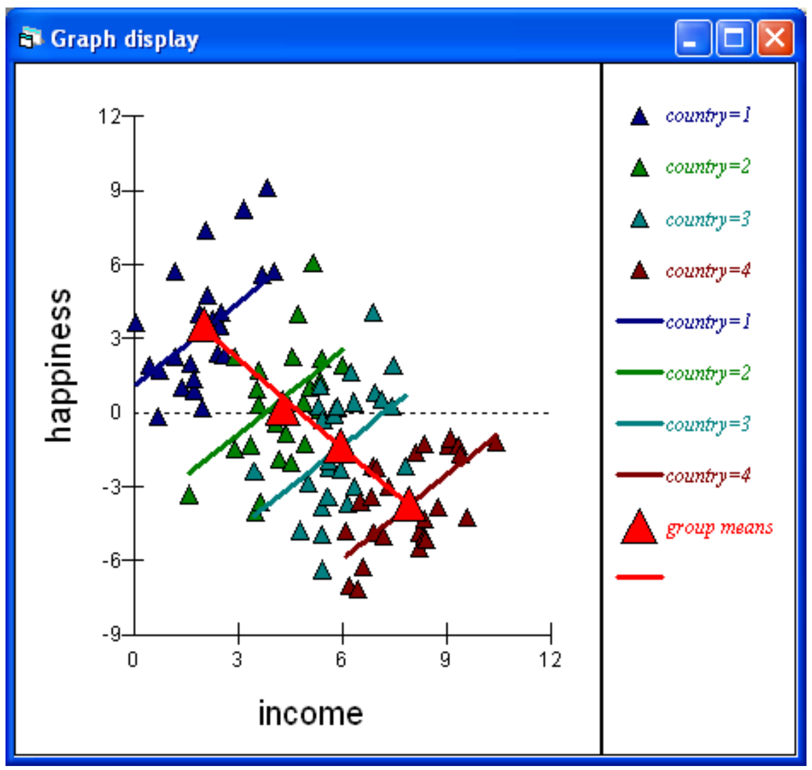
$$y_{ij} = 5 + 1 \times \text{income}_{ij} + (-2) \times \text{av\_income}_j + u_j + e_{ij}$$

$$e_{ij} \sim N(0,2)$$

Note that this model does not include a country level random effect. Therefore the only differences between country happiness levels are produced by differing incomes in the countries.

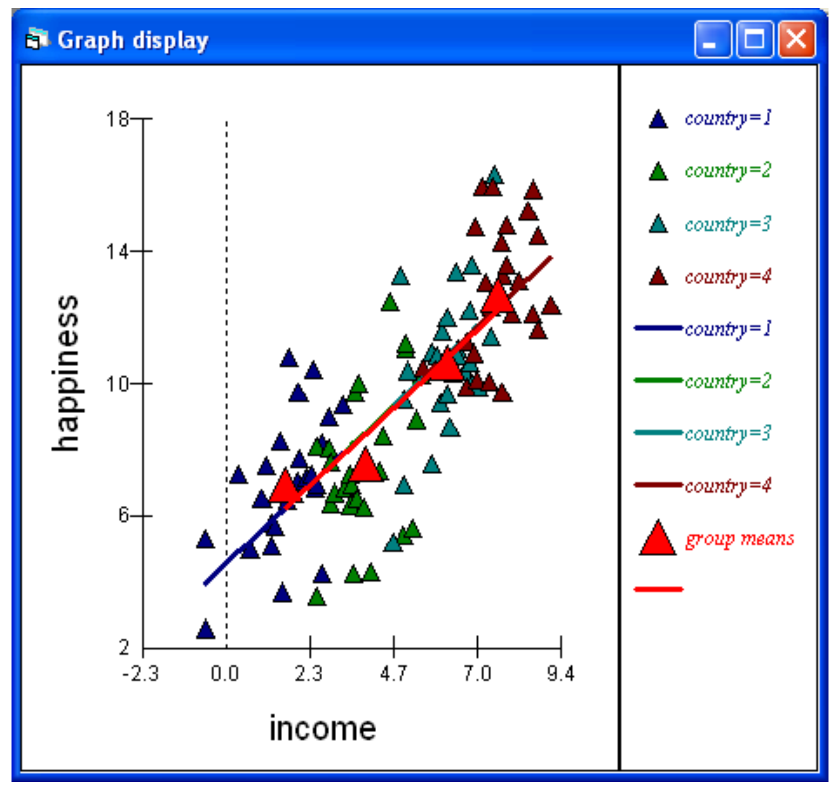


Examples of different contextual effects



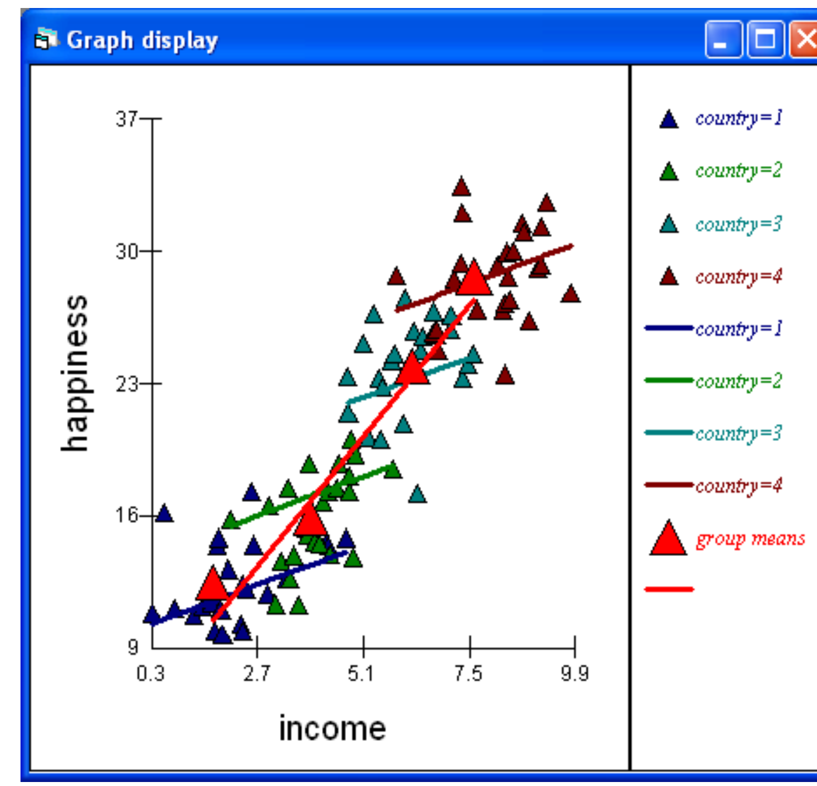
W=1,B=
-1,C=-2

Contextual effect pulls down intercept as av_income increases



W = 1,
B=1,C=0

Contextual effect 0 so intercept unchanged as av_income increases



W=1,B=3,C
=2

Contextual effect pushes up intercept as av_income increases

Context in multilevel

- Difference in “BETWEEN” regression and “WITHIN” regression is captured by one statistical model!
- In the paper all kinds of rules of thumb are used from the era predating ML modeling.

Cross-level interaction:
Level2 * Level1

Interpreting interactions

- Re-write your equations (see also session 1)

$Y' = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$
$Y' = (\beta_0 + \beta_2 X_2) + (\beta_1 + \beta_3 X_2) X_1$ *(X_2 as a moderator)*
Let $X_2 = 0$, Y' (at $X_2 = 0$) $= \beta_0 + \beta_1 X_1$
Thus β_1 is the regression coefficient of X_1 for cases with $X_2 = 0$.

- Centered variables? What does X=0 mean?

- Standardized variables? (See notes on Hox ch. 3 and 4)

Interactions in regular regression

- Interactions between predictors in multiple regression models (or moderator effects):

- Categorical and continuous variable
- Two continuous variables

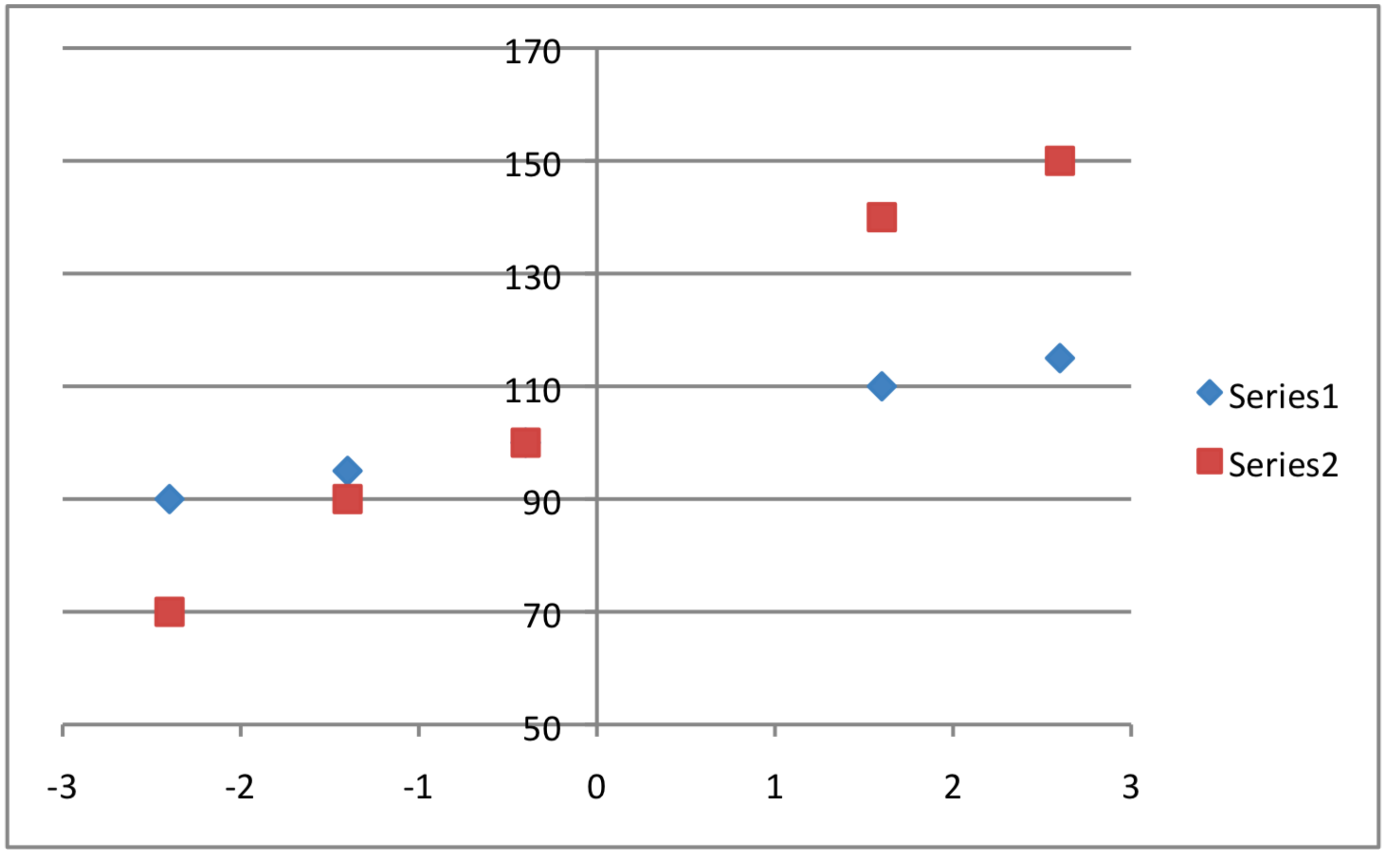
- General form: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 XZ_i + \varepsilon_i$

- General approach: Start without the interactions, add them in a second model

- Center variables

Categorical / Nominal x Continuous

- Create dummy variable(s) for nominal variable
- Center other predictor(s)
- Create interaction term(s): Series1 x centered predictor



Categorical / Nominal x Continuous

- Model 1: $Y_i = \beta_0 + \beta_1$ $xcent_i + \beta_2$ $dummy_i + \varepsilon_i$
- Model 2: $Y_i = \beta_0 + \beta_1$ $xcent_i + \beta_2$ $dummy_i + \beta_3$ $xcent{*}dummy_i + \varepsilon_i$

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	102.000	5.697		17.904	.000
	xcent	10.640	2.172	.866	4.899	.002
	dummy	8.000	8.057	.176	.993	.354
2	(Constant)	102.000	1.181		86.349	.000
	xcent	5.000	.637	.407	7.851	.000
	dummy	8.000	1.671	.176	4.789	.003
	interaction	11.279	.901	.649	12.523	.000

a. Dependent Variable: y

Categorical / Nominal x Continuous

- Regression equations, rearrange terms for easier interpretation:
$\beta_0 + \beta_1$ xcent$_i$ + β_2 dummy$_i$ + β_3 xcent*dummy$_i$ =
$\beta_0 + \beta_2$ dummy$_i$ + ($\beta_1 + \beta_3$ dummy)*xcent

- Dummy = 0: **Y' = 102 + 5 * xcent**

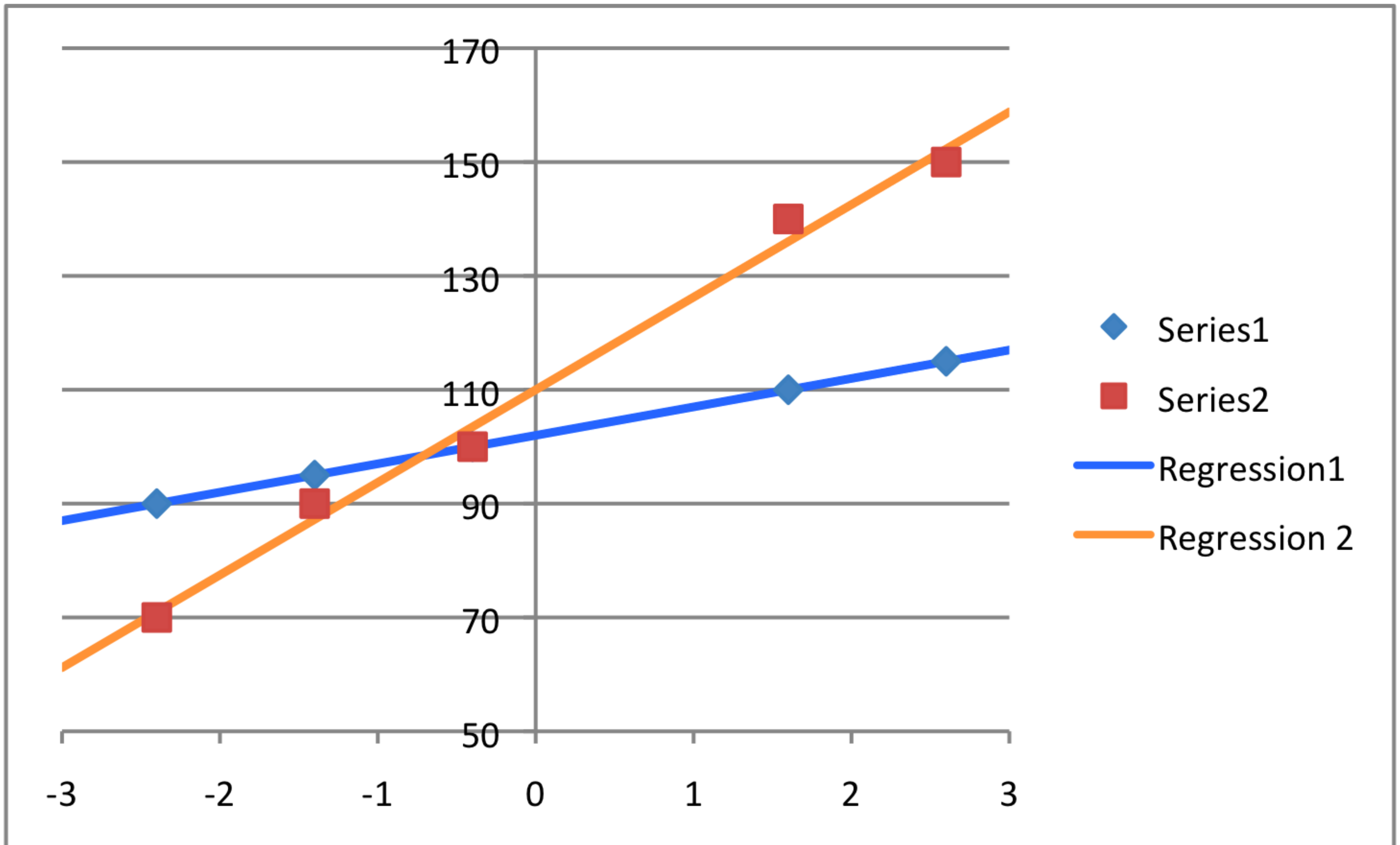
- Dummy = 1: Y' = 102 + 8 + (5 + 11.279) * xcent = **110 + 16.270 * xcent**

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	102.000	5.697		17.904	.000
	xcent	10.640	2.172	.866	4.899	.002
	dummy	8.000	8.057	.176	.993	.354
2	(Constant)	102.000	1.181		86.349	.000
	xcent	5.000	.637	.407	7.851	.000
	dummy	8.000	1.671	.176	4.789	.003
	interaction	11.279	.901	.649	12.523	.000

a. Dependent Variable: y

Categorical / Nominal x Continuous



Recall the definition of a moderator variable:

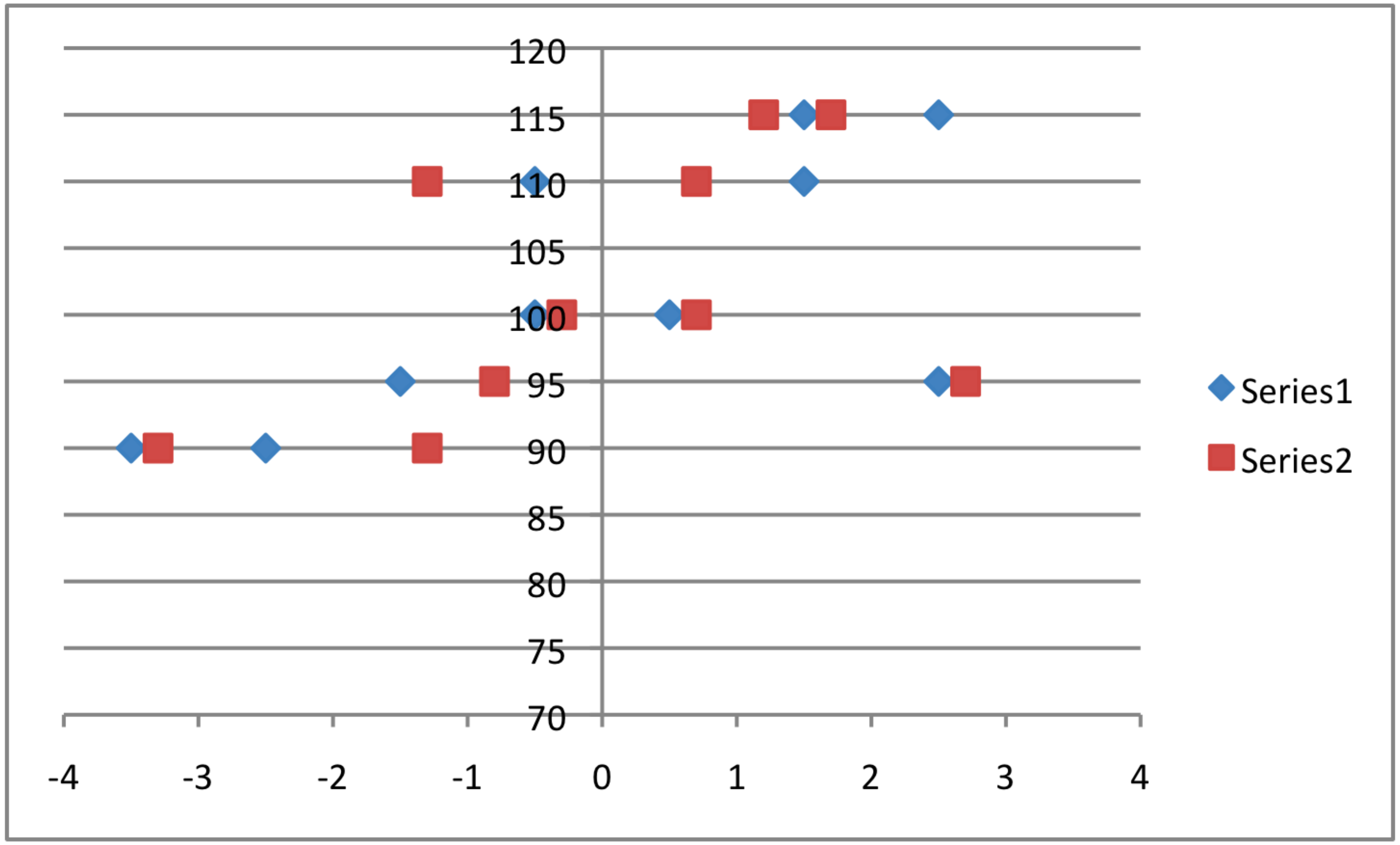
The relationship between X and Y changes for different levels of the moderator:

The slope

Test whether slopes are significant for each group

Continuous x Continuous

- Center predictors
- Create interaction term(s): x1cent * x2cent



Continuous x Continuous

- Model 1: $Y_i = \beta_0 + \beta_1\ x1cent_i + \beta_2\ x2cent_i + \varepsilon_i$

- Model 2: $Y_i = \beta_0 + \beta_1\ x1cent_i + \beta_2\ x2cent_i + \beta_3\ x1cent{*}x2cent + \varepsilon_i$

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	102.000	1.971		51.748	.000
	x1cent	9.014	2.846	1.907	3.167	.016
	x2cent	-7.280	3.362	-1.304	-2.165	.067
2	(Constant)	104.502	2.528		41.345	.000
	x1cent	8.679	2.661	1.836	3.262	.017
	x2cent	-7.416	3.132	-1.329	-2.368	.056
	interaction	-.820	.570	-.300	-1.440	.200

a. Dependent Variable: y

Continuous x Continuous

- Regression equations, rearrange terms for easier interpretation, take x2 as moderator:

$$\beta_0 + \beta_1\, x1cent_i + \beta_2\, x2cent_i + \beta_3\, x1cent*x2cent =$$
$$\beta_0 + \beta_2\, x2cent_i + (\beta_1 + \beta_3\, x2cent_i)*x1cent_i$$

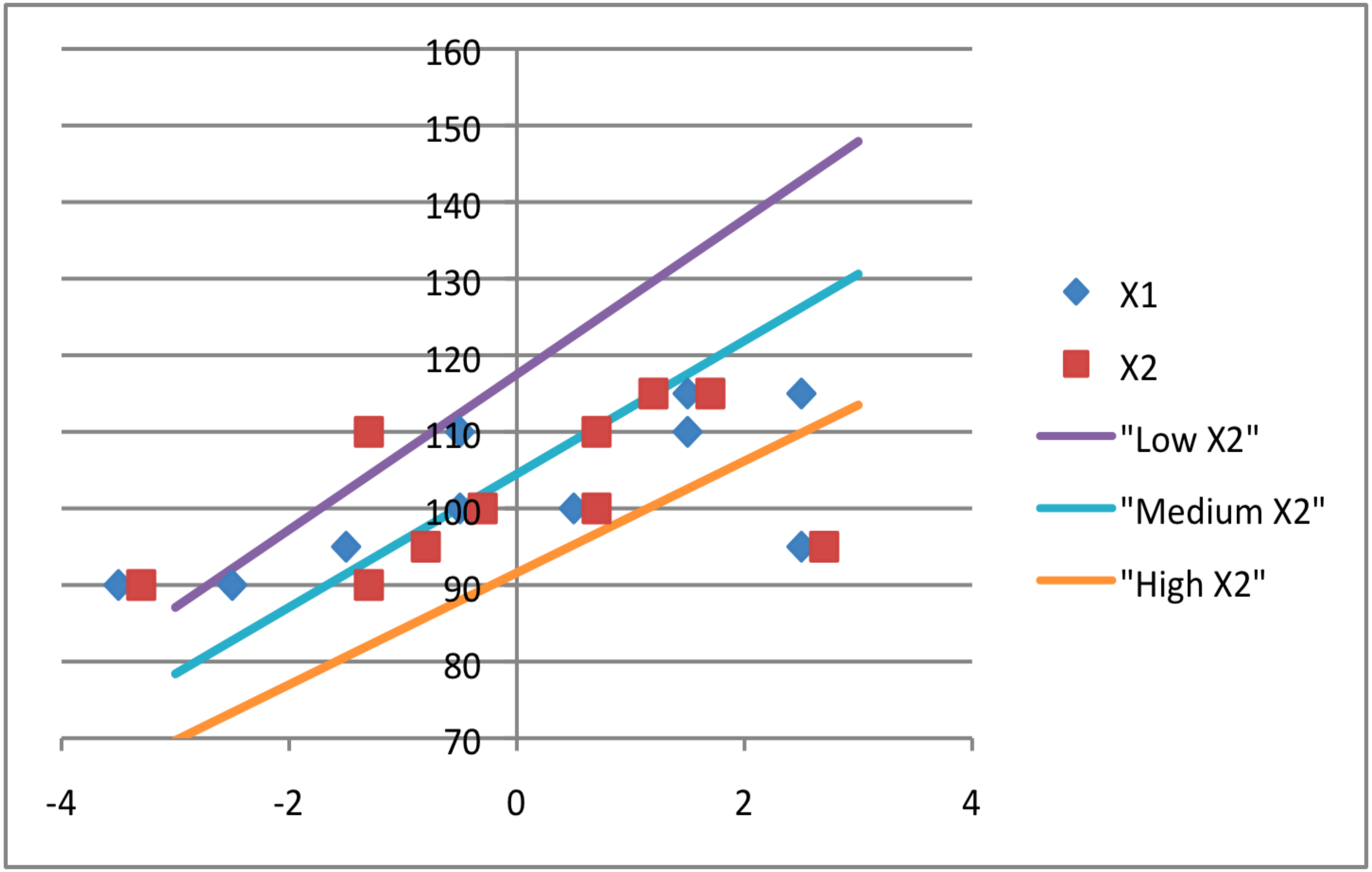
- -1 SD of x2cent (-1.75): Y' = 104.5 + -7.4* -1.75 + (8.7 -0.82*-1.75) * x1cent = **117.5 + 10.14 * x1cent**
- Mean of x2cent (0): Y' = **104.5 + 8.7 * x1cent**
- +1 SD of x2cent (1.75): Y' = 104.5 + -7.4*1.75 + (8.7 -0.82*1.75) * x1cent = **91.6 + 7.3 * x1cent**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	102.000	1.971		51.748	.000
	x1cent	9.014	2.846	1.907	3.167	.016
	x2cent	-7.280	3.362	-1.304	-2.165	.067
2	(Constant)	104.502	2.528		41.345	.000
	x1cent	8.679	2.661	1.836	3.262	.017
	x2cent	-7.416	3.132	-1.329	-2.368	.056
	interaction	-.820	.570	-.300	-1.440	.200

a. Dependent Variable: y

Continuous x Continuous



Interaction was not significant

Cross-level interaction in multilevel models

Recall the definition of a moderator variable:

The relationship between X and Y is different for different levels of the moderator: The slope

Adding cross-level interactions = explain slope variance

Contextual effects explaining level 2 variance

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + e_{ij}$$
$$\beta_{0j} = \beta_0 + u_{0j}$$
$$\beta_{1j} = \beta_1 + u_{1j}$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) \quad \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix}$$

Contextual variables explain level 2 variation in the intercept and cross-level interactions of contextual variables with x_1 explain level variability in the slope coefficient

This can be seen by re-arranging

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_2 x_{2j} + \beta_3(x_2 x_1)_{ij} + e_{ij}$$
$$\beta_{0j} = \beta_0 + u_{0j}$$
$$\beta_{1j} = \beta_1 + u_{1j}$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) \quad \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix}$$

$\longrightarrow$

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + e_{ij}$$
$$\beta_{0j} = \beta_0 + \beta_2 x_{2j} + u_{0j}$$
$$\beta_{1j} = \beta_1 + \beta_3 x_{2j} + u_{1j}$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) \quad \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix}$$

Cross-level interaction

Level 1: pupils ($n = 2000$)

dependent variable: popularity (1 – 10)

explanatory variable: gender (0 = boy, 1 = girl)

Level 2: classes ($n = 200$)

explanatory variable: teacher experience (2 – 25)

Cross-level interaction:

M0: intercept only (empty model)

$$Y_{ij} = \beta_{0ij}CONS_{ij}, \qquad \beta_{0ij} = \beta_0 + u_{0j} + e_{ij}$$

M1: effects of gender and teacher experience

$$Y_{ij} = \beta_{0ij}CONS_{ij} + \beta_{1j}GIRL_{ij} + \beta_2 TEXP_j$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{ij}, \qquad \beta_{1j} = \beta_1 + u_{1j}$$

M2: M1 plus cross-level interaction

$$Y_{ij} = \beta_{0ij}CONS_{ij} + \beta_{1j}GIRL_{ij} + \beta_2 TEXP_j + \beta_3 GIRL_{ij}{*}TEXP_j$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{ij}, \qquad \beta_{1j} = \beta_1 + u_{1j}$$

	Model M0	Model M1
Fixed Part		
Intercept	5.31 (0.10)	3.34 (0.16)
Girl		0.84 (0.06)
Teacher Exp.		0.11 (0.01)
Girl * Teacher Exp.		
Random part		
Level 1 (pupil)	0.64 (0.02)	0.39 (0.01)
Level 2 (class)		
- intercept	0.87 (0.13)	0.40 (0.06)
- slope		0.27 (0.05)
- intercept-slope covariance		0.02 (0.04)
Deviance	5112.7	4261.2

Fixed part of model M1 (averaged over classes):

$Y'_{ij} = 3.34 + 0.84GIRL_{ij} + 0.11TEXP_j$

for boys (GIRL = 0): $Y'_{ij} = 3.34 + 0.11TEXP_j$
for girls (GIRL = 1): $Y'_{ij} = (3.34 + 0.84)+ 0.11TEXP_j$

gender and teacher experience have independent effects on popularity

Random part of model M1

pupil-level variance decreased from 0.64 to 0.39
% of variance = 100 * (0.64 - 0.39) / 0.64 = 39%

intercept variance decreased from 0.87 to 0.40
% of variance = 100 * (0.87 - 0.40) / 0.87 = 54%

popularity difference between boys and girls differs by class
variance of these differences = 0.27
s.d. = $\sqrt{0.27} = 0.52$

67% of gender differences will be between 0.32 and 1.36

	Model M0	Model M1	Model M2
Fixed Part			
Intercept	5.31 (0.10)	3.34 (0.16)	3.31 (0.16)
Girl		0.84 (0.06)	1.33 (0.13)
Teacher Exp.		0.11 (0.01)	0.11 (0.01)
Girl * Teacher Exp.			-0.03 (0.01)
Random part			
Level 1 (pupil)	0.64 (0.02)	0.39 (0.01)	0.39 (0.01)
Level 2 (class)			
- intercept	0.87 (0.13)	0.40 (0.06)	0.40 (0.06)
- slope		0.27 (0.05)	0.22 (0.04)
- intercept-slope covariance		0.02 (0.04)	0.02 (0.04)
Deviance	5112.7	4261.2	4245.9

Fixed part of model M2

$Y'_{ij} = 3.31 + 1.33GIRL_{ij} + 0.11TEXP_j - 0.03GIRL_{ij}*TEXP_j$

effect of pupil gender depends on teacher experience:

$Y'_{ij} = 3.31 + 0.11TEXP_j + (1.33 – 0.03TEXP_j)*GIRL_{ij}$

teacher experience	gender difference
2	1.33 – 0.03*2 = 1.27
10	1.33 – 0.03*10 = 1.03
15	1.33 – 0.03*15 = 0.88
25	1.33 – 0.03*25 = 0.58

Or: effect of teacher experience depends on gender of pupil:

$Y'_{ij} = 3.31 + 1.33GIRL_{ij} + (0.11– 0.03GIRL_{ij})*TEXP_j$

Random part of model M2

variance of gender difference decreased from .27 to .22

% of variance = 100*(0.27 – 0.22) / 0.27 = 18.5%

Interpreting interactions

- Re-write your equations!

$Y' = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$
$Y' = (\beta_0 + \beta_2 X_2\ ) + (\beta_1 + \beta_3 X_2) X_1$ *(X_2 as a moderator)*
Let $X_2 = 0$, Y' (at $X_2 = 0$) $= \beta_0 + \beta_1 X_1$

Thus β_1 is the regression coefficient of X_1 for cases with $X_2 = 0$.

Next:
- Effects of centering (group vs. grand mean) on interpretation
- Effect of standardizing variables on interpretation
- Modelling the variance function