Centering in Cross Sectional Data

Andy Grogan-Kaylor

24 May 2021 17:59:11

# Background

# Introduction

These notes represent a brief discussion of centering with cross sectional data. Since so much of my current work focuses on cross national work on parenting and child development, I use these ideas as my substantive example.

Consider a cross-national data set where we are attempting to understand predictors of *behavior problems* as a function of *per capita income* and *parental use of physical punishment*.

# Simulate Some Data

Show / Hide Data Simulation Code

. clear all

. set obs 100  
Number of observations (\_N) was 0, now 100.

. generate income = runiform(10000, 70000)

. generate physical\_punishment = rbinomial(1,.3)

. generate country = int(\_n/10) + 1

. generate e = rnormal(0,1)

. generate behavior\_problems = 110 + -.0001 \* income + 10 \* physical\_punishment + e // plausible regression relationship

. list in 1/10, abb(20) // list out some data  
  
 ┌──────────────────────────────────────────────────────────────────────────┐  
 │ income physical\_punishment country e behavior\_problems │  
 ├──────────────────────────────────────────────────────────────────────────┤  
 1. │ 40915.21 0 1 .2154569 106.1239 │  
 2. │ 58346.25 0 1 -.4418385 103.7235 │  
 3. │ 66061.07 0 1 1.762688 105.1566 │  
 4. │ 52075.53 0 1 -.4707811 104.3217 │  
 5. │ 27207.43 0 1 .2849969 107.5643 │  
 ├──────────────────────────────────────────────────────────────────────────┤  
 6. │ 38323.49 1 1 .0082076 116.1759 │  
 7. │ 25597.57 0 1 -.1955867 107.2447 │  
 8. │ 36809.94 0 1 -.0347036 106.2843 │  
 9. │ 22501.57 0 1 .9727817 108.7226 │  
 10. │ 43147.42 0 2 .2931708 105.9784 │  
 └──────────────────────────────────────────────────────────────────────────┘

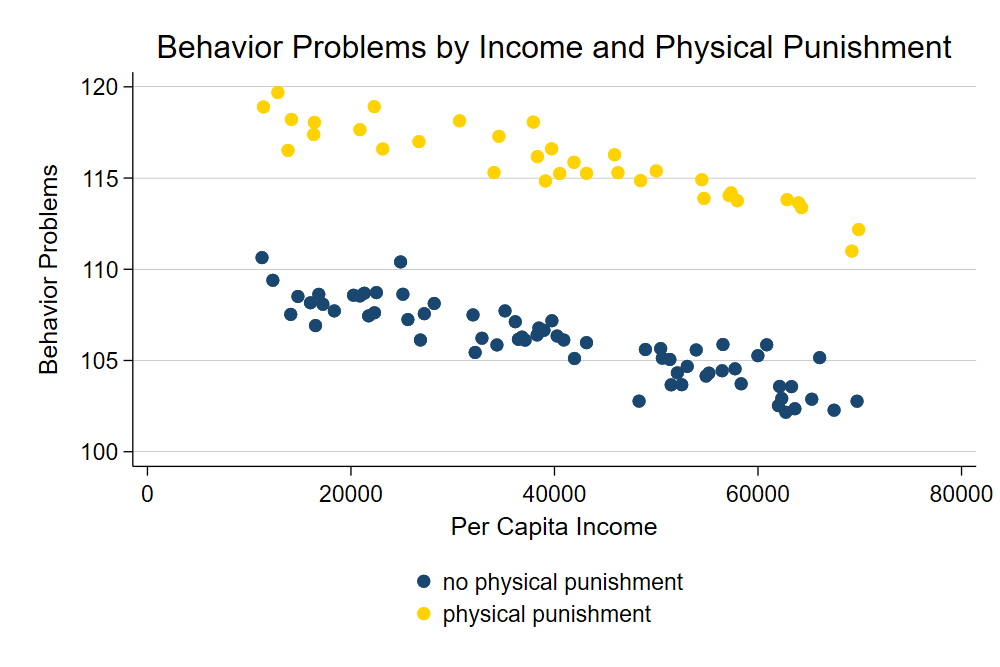
# Uncentered Data

## Equation

## Graph

. twoway (scatter behavior\_problems income if physical\_punishment ==0) ///  
> (scatter behavior\_problems income if physical\_punishment == 1), ///  
> legend(order(1 "no physical punishment" 2 "physical punishment") pos(6)) ///  
> title("Behavior Problems by Income and Physical Punishment") ///  
> xtitle("Per Capita Income") ///  
> ytitle("Behavior Problems") ///  
> scheme(michigan)

. graph export myscatter.png, width(1000) replace  
file myscatter.png saved as PNG format



Scatterplot

## Multilevel Model

. mixed behavior\_problems income physical\_punishment || country:  
  
Performing EM optimization:   
  
Performing gradient-based optimization:   
  
Iteration 0: log likelihood = -132.47696   
Iteration 1: log likelihood = -132.22777   
Iteration 2: log likelihood = -132.2277   
Iteration 3: log likelihood = -132.2277   
  
Computing standard errors:  
  
Mixed-effects ML regression Number of obs = 100  
Group variable: country Number of groups = 11  
 Obs per group:  
 min = 1  
 avg = 9.1  
 max = 10  
 Wald chi2(2) = 3030.41  
Log likelihood = -132.2277 Prob > chi2 = 0.0000  
  
────────────────────┬────────────────────────────────────────────────────────────────  
 behavior\_problems │ Coefficient Std. err. z P>|z| [95% conf. interval]  
────────────────────┼────────────────────────────────────────────────────────────────  
 income │ -.0001068 5.31e-06 -20.10 0.000 -.0001172 -.0000964  
physical\_punishment │ 9.763856 .1916716 50.94 0.000 9.388186 10.13953  
 \_cons │ 110.3409 .2426648 454.71 0.000 109.8653 110.8165  
────────────────────┴────────────────────────────────────────────────────────────────  
  
─────────────────────────────┬────────────────────────────────────────────────  
 Random-effects parameters │ Estimate Std. err. [95% conf. interval]  
─────────────────────────────┼────────────────────────────────────────────────  
country: Identity │  
 var(\_cons) │ 1.87e-17 1.79e-13 0 .  
─────────────────────────────┼────────────────────────────────────────────────  
 var(Residual) │ .8242157 .1165617 .6246875 1.087474  
─────────────────────────────┴────────────────────────────────────────────────  
LR test vs. linear model: chibar2(01) = 8.5e-13 Prob >= chibar2 = 1.0000

We note that -1.068 is the effect of every additional $10,000 of per capita income. 9.764. Notably, for this handout, 110.341 is the level of behavior problems for a child who did *not recieve physical punishment* living in a family with **$0** *income*.

# Grand Mean Centering

Grand mean centering helps us to have more meaningful intercepts of our continuous variables.

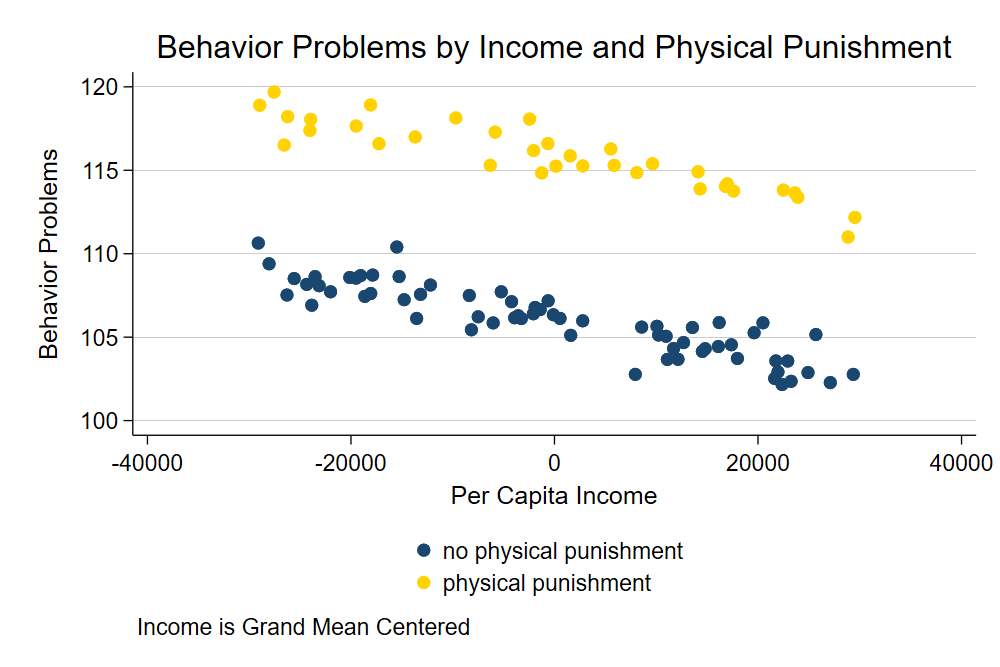
Essentially, we are going to create

. egen m\_income = mean(income) // grand mean of income

. generate c\_income = income - m\_income // grand mean centered income

. twoway (scatter behavior\_problems c\_income if physical\_punishment ==0) ///  
> (scatter behavior\_problems c\_income if physical\_punishment == 1), ///  
> legend(order(1 "no physical punishment" 2 "physical punishment") pos(6)) ///  
> title("Behavior Problems by Income and Physical Punishment") ///  
> caption("Income is Grand Mean Centered") ///  
> xtitle("Per Capita Income") ///  
> ytitle("Behavior Problems") ///  
> scheme(michigan)

. graph export myscatter2.png, width(1000) replace  
file myscatter2.png saved as PNG format



Scatterplot With Grand Mean Centering

In a graph, we see that *grand mean centering* has transformed the intercept so now the term is the level of behavior problems for the *average* child who did *not recieve physical punishment*.

# Group Mean Centering

In group mean centering, we are doing something slightly different. We are creating a mean for each *group*, which in this data is *country*: , where is the index for *group* or *country*.

. bysort country: egen m\_g\_income = mean(income) // GROUP mean of income

. generate c\_g\_income = income - m\_g\_income // GROUP mean centered income

. bysort country: egen m\_g\_physical\_punishment = mean(physical\_punishment) // GROUP mean of physical punishment

. generate c\_g\_physical\_punishment = physical\_punishment - m\_g\_physical\_punishment // GROUP mean centered physical punishment

## Multilevel Model

Interestingly, *group mean centering* has many implications. Here I focus on how employing different variables might provide *conceptually* or *theoretically* different results. For the sake of parismony, in the brief discussion below I focus on these *conceptual* or *theoretical* differences, and do not provide output. I use the quietly prefix to suppress output.

### Covariate, and Group Mean

One parameterization of the multilevel model is to enter the *covariate* and its *group level mean* i.e. and .

This first parameterization focuses on *individual scores on covariates* and their *country level means*.

What is the effect of *income*, *country level mean income*, *physical punishment* and *country level mean of physical punishment* on *behavior problems*?

. quietly: mixed behavior\_problems income m\_g\_income physical\_punishment m\_g\_physical\_punishment || country:

### Group Mean Centered Covariate, and Group Mean

A second, equally valid, but conceptually different parameterization of the multilevel model is to enter the *covariate deviated from its group level mean* and the *group level mean* i.e. and .

This second parameterization focuses on how *individuals differ from their country level means*, and *country level means*.

What is the effect of *income deviated from its country level mean*, *country level mean income*, *physical punishment deviated from its country level punishment*, and *country level mean of physical punishment* on *behavior problems*?

. quietly: mixed behavior\_problems c\_g\_physical\_punishment m\_g\_income c\_g\_physical\_punishment m\_g\_physical\_punishment || country: