Homework for Chapter 16: Fixed Effects

*How Does It Work?*

1. You observe the number of vacations taken by Zac and Skylar in 2012, 2013, and 2014. In those years, Zac took 3, 7, and 5 vacations, respectively. Skylar took 2, 6, and 10.
   1. Isolate the numbers that represent the variation *between* Zac and Skylar in their vacation-taking.

Answer:

mean\_Zac = (3+7+5)/3=5

mean\_Skylar=(2+6+10)/3=6

Difference of mean\_Zac and mean\_Skylar=|5-6|=1

* 1. Isolate the variation *within* Zac and within Skylar in their vacation-taking.

Answer:

|  |  |  |  |
| --- | --- | --- | --- |
|  | 2012 | 2013 | 2014 |
| Within\_Zac | -2 | 2 | 0 |
| Within\_Skylar | -4 | 0 | 4 |

* 1. (Difficult!) We perform a fixed effects analysis of the effect of vacations on happiness. A vacation increases Zac’s happiness by 1 “happiness point,” but it increases Skylar’s happiness by 2 “happiness points.” Will our fixed effects estimate likely give us an answer closer to 1, closer to 2, or exactly 1.5?

Answer: It will be closer to 2 (Skylar’s effect) because Skylar has more within variation. (Fixed effects will focus a lot heaveily on people with a lot of within variation)

1. You are interested in the effect of cultural events on the levels of trust in a city. Perhaps big events like concerts bring people together and they can trust each other more. You plan to look at the relationship between trust and number of events in a given year, with fixed effects for city. Draw a causal diagram for this research question with at least four back door paths. Which paths will be closed by fixed effects, and which will remain open?

Answer:

图示

描述已自动生成

The path from the number of cultural events to the levels of trust via geography, the path from the number of cultural events to the levels of trust via the size of population, and the path from the number of cultural events to the levels of trust via general political preference, will be closed by fixed effects; while the path from the number of cultural events to the levels of trust via pandemic will remain open.

1. Classify each of the following forms of variation as “between variation”, “within variation”, or a combination of both.
   1. (Individual = person) How a child’s height changes as they age. (within variation)
   2. (Individual = person) In a data set tracking many people over many years, the variation in the number of children a person has in a given year. (within variation)
   3. (Individual = city) Overall, Paris, France has more restaurants than Paris, Texas. (between variation)
   4. (Individual = genre) The average pop music album sells more copies than the average jazz album (between variation)
   5. (Individual = genre) Miles Davis’ Kind of Blue sold very well for a jazz album. (It could be a combination of between variation and within variation: the phrase “jazz album” has two meanings: first, Kind of Blue sells very well compared with other jazz albums (within variation; second, Kind of Blue, as a jazz album, sells better than some albums of other genres (between variation), also implying that jazz albums sells worse than other albums of different genres)
   6. (Individual = genre) Michael Jackson’s Thriller, a pop album, sold many more copies than Kind of Blue, a jazz album. (between variation)
2. Why does the process of taking each observation relative to its individual-level mean have the effect of “controlling for individual”?

Answer: Because taking each observation relative to its individual level mean means that we are calculating the changes compared to their typical levels (means), which means that we are getting rid of the between variation of individual and focusing on the within-variation. Therefore, we are controlling for individual.

Besides, graphically, if we take each observation relative to tis individual-level mean is drawing subset graphics in which the mean become (0,0), when we put these subset graphics together based on the point (0,0), we could find that we are controlling for individual.

*How is it Performed?*

1. You are interested in the effect of cultural events on the levels of trust in a city. You run a regression of trust levels (on a 0-100 scale) on the number of cultural events with city fixed effects and get a coefficient on cultural events of 3.6. Assume that there are still some back doors open, so do not interpret the result causally. Interpret the 3.6, explaining it in an English sentence.

Answer: For a given city, in a year where the number of cultural events is one unit higher than it typically is for that city, then we could expect the trust levels to be a 3.6-unit higher than it typically is for that city.

1. You are interested in the effect of cultural events on the levels of trust in a city. You run a regression of trust levels (on a 0-100 scale) on the number of cultural events with city and year fixed effects and get a coefficient on cultural events of 2.4. Assume that there are still some back doors open, so do not interpret the result causally. Interpret the 2.4, explaining it in an English sentence.

Answer: For a given city and a given year, when the number of cultural events is one unit higher than it typically is, we could expect the trust levels to be a 3.6-unit higher than it typically is.

1. Two-way fixed effects with terms for both individual and time are often referred to as “controlling for individual and time effects”. Why might a researcher want to do this rather than just taking individual fixed effects and adding a linear/polynomial/etc. term for time?

Answer: Taking two-way fixed effects with terms for both individual and time is better than taking individual fixed effects with adding a term for time because the one set of fixed effects (individual fixed effects) and the one set of fixed effects (year fixed effects) will influence each other. So it is better to take two-way fixed effects which refers to as “controlling for individual and time effects.

1. Which of the following explains why random effects is likely to do a better job of estimating the individual-level effects than fixed effects, if its assumptions hold?

Answer: c.

* 1. Because it makes the assumption that the individual effects are unrelated to the other predictors, which breaks that back door and thus reduces bias.
  2. Because random effects allows some amount of between variation into the model, and some of the real individual effect is that between variation. (The individual effect would be the within-variaiton?).
  3. Because it uses the information from the entire data set to estimate each individual effect, rather than relying on only a few observations per individual.
  4. It won’t. Enforcing Durbin-Wu-Hausman makes both methods produce the same estimates anyway.

Coding (which includes any how-the-pros-do-it questions)