

Центрирование в модели с фиксированными эффектами twoways: пояснение

В статье J. Kropko, R. Kubines, 2020 авторы дали формулу для получения оценки коэффициента при предикторе в модели с фиксированными эффектами на пространственные единицы и на временные периоды (модель twoways):

$$\hat{\beta}_{TW} = \frac{\sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}) (x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{x})}{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{x})^2}$$

Давайте разберем, откуда здесь появляется общее среднее в формуле (\bar{x})

1. Вспомним, что внутригрупповое преобразование в модели twoways предполагает центрирование по пространственному и временному измерениям. Можно, к примеру, сначала вычесть из значений исходных переменных групповое среднее по пространственным единицам, далее – из полученных отклонений от группового среднего вычесть среднее по временным периодам. *Примечание: порядок может быть обратным, то есть, сначала можно вычесть групповое среднее по временным периодам, потом – вычесть среднее по пространственным единицам. Результат будет идентичным*
2. В результате первого шага центрирования получаем:

$$y_{it}^* = y_{it} - \frac{\sum_{t=1}^T y_{it}}{T} = y_{it} - \bar{y}_{i.}$$

3. Второй шаг внутригруппового преобразования предполагает центрирование по второму измерению: в данном случае, по временным периодам. Нужно вычислить среднее от уже полученных на предыдущем шаге отклонений от среднего, то есть, посчитаем среднее по временным периодам от $(y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it})$.

Обозначим это среднее как $\tilde{y}_{.t}$:

$$\tilde{y}_{.t} = \frac{\sum_{i=1}^N y_{it}^*}{N} = \frac{\sum_{i=1}^N (y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it})}{N} = \frac{\sum_{i=1}^N y_{it}}{N} - \frac{\sum_{i=1}^N \sum_{t=1}^T y_{it}}{NT} = \bar{y}_{.t} - \bar{y}$$

Соответственно, в результате второго шага центрирования – по временным периодам – получается:

$$y_{it}^{**} = y_{it}^* - (\bar{y}_{.t} - \bar{y}) = y_{it}^* - \bar{y}_{.t} + \bar{y} = y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}$$

Аналогичным образом применяем двухэтапное центрирование для предикторов, получая $(x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{x})$