

# Методы анализа неоднородных данных и паттерн анализ

Модели со смешанными эффектами

Условимся, что работаем с пространственным массивом данных, в котором есть более и менее крупные единицы анализа (к примеру, данные по индивидам «вложенные» в страны). Рассмотрим в терминах МЕ-моделей общий вид ANOVA-модели и поясним спецификацию.

Условимся, что работаем с пространственным массивом данных, в котором есть более и менее крупные единицы анализа (к примеру, данные по индивидам «вложенные» в страны). Рассмотрим в терминах МЕ-моделей общий вид ANOVA-модели и поясним спецификацию.

$$y_{ij} = \gamma_{00} + u_{0j} + e_{ij}, \text{ где}$$

$\gamma_{00}$  – среднее значение зависимой переменной

$u_{0j}$  – межгрупповая изменчивость (в значении зависимой переменной)

$e_{ij}$  – ошибка на индивидуальном уровне

## Зачем нам нужна такая модель?

Определить, достаточно ли вариации объясняется на втором уровне. Для этого понадобится рассчитать внутриклассовый коэффициент корреляции (ICC)

## Зачем нам нужна такая модель?

Определить, достаточно ли вариации объясняется на втором уровне. Для этого понадобится рассчитать внутриклассовый коэффициент корреляции (ИСС)

$$\text{ИСС} = \frac{\text{Var}(u_{0j})}{\text{Var}(u_{0j}) + \text{Var}(e_{ij})}$$

ИСС показывает долю вариации зависимой переменной, объясняемой межгрупповыми различиями (страновой спецификой). В случае если наблюдаем очень маленькие или, наоборот, очень высокие значения ИСС, это свидетельствует о том, что нет необходимости в моделях, учитывающих неоднородность данных (зависимость единиц внутри групп).

Что содержательно включает в себя случайный эффект ( $u_{0j}$ )?

Что содержательно включает в себя случайный эффект ( $u_{0j}$ )?

Это страновые характеристики, неучтенные в модели. Включением  $u_{0j}$  в модель мы ничего не объясняем, но в явном виде моделируем межгрупповую вариацию. Если «стартовые» условия различаются, то мы должны понять, ПОЧЕМУ? Наша задача на последующих этапах – снизить долю необъясненной как межгрупповой, так и внутригрупповой вариации.

Как это сделать: снизить межгрупповую и внутригрупповую вариацию?



Как это сделать: снизить межгрупповую и внутригрупповую вариацию?

## Ответ

$y_{ij} = \gamma_{00} + \gamma_{01} * Z_j + \gamma_{10} * X_{ij} + u_{0j} + e_{ij}$ , где

$\gamma$  – фиксированные эффекты (то, что мы в среднем наблюдаем по всей выборке: либо среднее значение зависимой переменной в стартовых условиях ( $\gamma_{00}$ ), либо среднее значение характера взаимосвязи)

Как это сделать: снизить межгрупповую и внутригрупповую вариацию?

## Ответ

$y_{ij} = \gamma_{00} + \gamma_{01} * Z_j + \gamma_{10} * X_{ij} + u_{0j} + e_{ij}$ , где

$\gamma$  – фиксированные эффекты (то, что мы в среднем наблюдаем по всей выборке: либо среднее значение зависимой переменной в стартовых условиях ( $\gamma_{00}$ ), либо среднее значение характера взаимосвязи)

$u_{0j}$  – межгрупповая изменчивость в «стартовых» условиях (то, что осталось необъясненного)

В предыдущей модели мы исходим из предположения о том, что у нас одинаковый характер взаимосвязи  $X_{ij}$  и  $y_{ij}$ . В ряде случаев это содержательно неоправданно. Поэтому мы обращаемся к такой модели:

В предыдущей модели мы исходим из предположения о том, что у нас одинаковый характер взаимосвязи  $X_{ij}$  и  $y_{ij}$ . В ряде случаев это содержательно неоправданно. Поэтому мы обращаемся к такой модели:

$y_{ij} = \gamma_{00} + \gamma_{01} * Z_j + \gamma_{10} * X_{ij} + u_{0j} + u_{1j} * X_{ij} + e_{ij}$ , где  $u_{1j}$  – межгрупповая изменчивость в характере взаимосвязи

На предыдущем шаге мы опять же смоделировали изменчивость по странам, однако не объяснили, а ПОЧЕМУ взаимосвязь  $X_{ij}$  и  $y_{ij}$  разная? Как можно уточнить модель?

На предыдущем шаге мы опять же смоделировали изменчивость по странам, однако не объяснили, а ПОЧЕМУ взаимосвязь  $X_{ij}$  и  $y_{ij}$  разная? Как можно уточнить модель?

$u_{1j}$  содержательно включает в себя страновые характеристики – потенциальное объяснение, почему взаимосвязь  $X_{ij}$  и  $y_{ij}$  разная. Давайте уменьшим долю необъясненной вариации, протестируем предикторы на страновом уровне в роли факторов-«посредников» (модерация).

$$y_{ij} = \gamma_{00} + \gamma_{01} * Z_j + \gamma_{10} * X_{ij} + \gamma_{11} * X_{ij} * Z_j + u_{0j} + u_{1j} * X_{ij} + e_{ij},$$

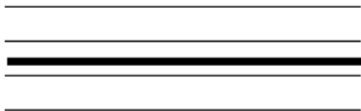
где

$u_{1j}$  – оставшаяся межгрупповая изменчивость во взаимосвязи

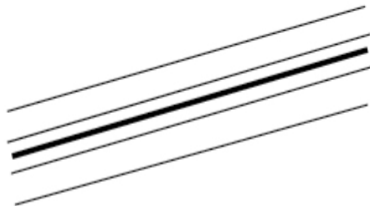
## НЕМНОГО КАРТИНОК

Изобразите схематично фиксированные и случайные эффекты для взаимосвязи (см. далее).

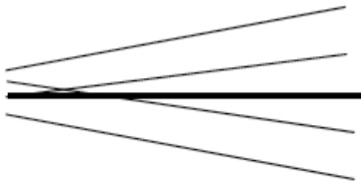
## A. No Fixed, No Random



## B. Yes Fixed, No Random



## C. No Fixed, Yes Random

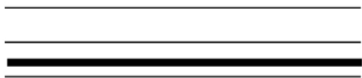


## D. Yes Fixed, Yes Random





## A. No Fixed, No Random



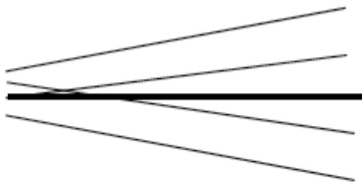
## B. Yes Fixed, No Random



### Примечание

Fixed, random – в данном случае относится к характеру взаимосвязи. На всех картинках – случайный эффект для стартовых условий.

Источник изображений: L.Hoffman, ICPSR Summer School



## Пояснения к картинкам

- Картинка А: ANOVA-модель (нет объясняющих переменных в модели, моделируются межгрупповые различия в стартовых условиях)
- Картинка В. Предположение об одинаковой взаимосвязи  $X_{ij}$  и  $y_{ij}$  во всех странах.
- Картинка С. Предположение о разной взаимосвязи  $X_{ij}$  и  $y_{ij}$ . В среднем по всей выборке взаимосвязь незначима, НО это НЕ значит, что можно выбрасывать фиксированный эффект из модели.
- Картинка D. Предположение о разной взаимосвязи  $X_{ij}$  и  $y_{ij}$ .