

# Регрессионный анализ: панельные данные и каузальность

Модели со смешанными эффектами: основы

## Вопрос

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

## Вопрос

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

## Ответ

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

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

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

$e_{ij}$  – ошибка на индивидуальном уровне

## Вопрос

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

## Вопрос

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

## Ответ

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, с чем это связано? Наша задача на последующих этапах – снизить долю необъясненной как межгрупповой, так и внутригрупповой вариации.

## Вопрос

Как это сделать: снизить межгрупповую и внутригрупповую вариацию?

## Вопрос

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)

## Вопрос

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)



## Вопрос

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:

## Вопрос

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:

## Ответ

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

$u_{1j}$  – межгрупповая изменчивость в характере взаимосвязи

## Вопрос

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?

## Вопрос

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?

## Ответ

$u_{1j}$  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-модераторов

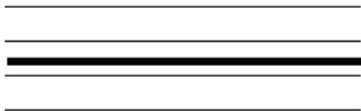
$y_{ij} = \gamma_{00} + \gamma_{01} * Z_j + \gamma_{10} * X_{ij} + \gamma_{11} * X_{ij} * Z_j + u_{0j} + u_{1j} * X_{ij} + e_{ij}$ , где

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

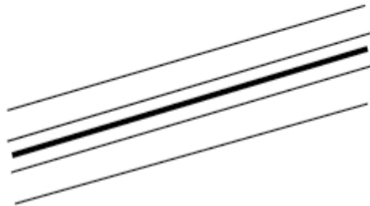
## НЕМНОГО КАРТИНОК

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

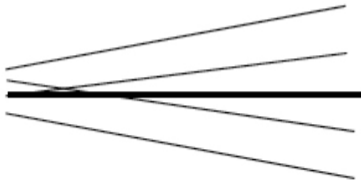
## A. No Fixed, No Random



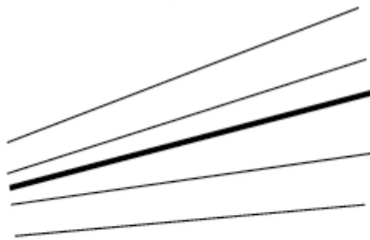
## B. Yes Fixed, No Random



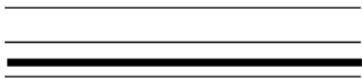
## C. No Fixed, Yes Random



## D. Yes Fixed, Yes Random



## A. No Fixed, No Random



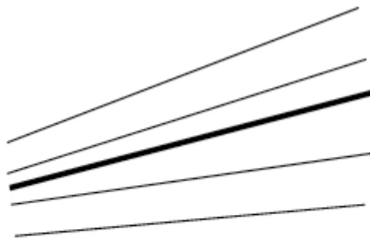
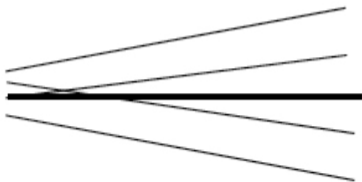
## B. Yes Fixed, No Random



### 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  $y_{ij}$ .



## Вопрос

Обозначьте допущения модели со смешанными эффектами.

## Вопрос

Обозначьте допущения модели со смешанными эффектами.

## Ответ

- ①  $E(u_j) = E(e_{ij}) = 0$
- ② нормальное распределение  $u_j, e_{ij}$
- ③ нет строгого допущения о гомоскедастичности  $e_{ij}, u_j$ .
- ④  $Cov(u_j, e_{ij}) = 0$
- ⑤ случайные эффекты для константы и наклона могут быть скоррелированы

## Вопрос

Что содержательно означает корреляция между случайными эффектами?

## Вопрос

Что содержательно означает корреляция между случайными эффектами?

## Ответ

Пример положительной корреляции: при более высоких значениях «стартовых» условий наблюдаем более выраженную положительную взаимосвязь зависимой переменной и ключевого предиктора.

Когда такое может быть? Пример: те страны, у которых изначально менее благоприятные экономические условия, долго восстанавливаются после экономического кризиса.

## Вопрос

Каков механизм получения оценки «уникального» значения характера взаимосвязи (или «стартовых» условий) для  $j$ -ой страны?

## Вопрос

Каков механизм получения оценки «уникального» значения характера взаимосвязи (или «стартовых» условий) для  $j$ -ой страны?

## Ответ (продолжение на след. слайде)

Представьте, что мы разделили массив на  $N$  подвыборок (всего  $N$  стран) и на каждой подвыборке оценили взаимосвязь ключевого предиктора и зависимой переменной. Полученные оценки коэффициентов при предикторе обозначим  $\hat{\alpha}_{1j}$ . Ту же взаимосвязь мы оценили на всей выборке, соответствующая оценка (средний характер взаимосвязи по всей выборке =  $\hat{\gamma}_{10}$ ).

## Ответ (продолжение)

$$\hat{b}_{1j} = w_j * \hat{\alpha}_{1j} + (1 - w_j) * \hat{\gamma}_{10},$$

где  $\hat{b}_{1j}$  – характер взаимосвязи предиктора и зависимой переменной в  $j$ -ой стране;

$w_j$  – «вес» для оценки по отдельной подвыборке;

$(1 - w_j)$  – «вес» для среднего по всей выборке, насколько средний характер взаимосвязи по всей выборке значим в формировании оценки  $\hat{b}_{1j}$

## Вопрос

В каком случае оценка взаимосвязи по отдельной подвыборке имеет больший вес в формировании оценки  $\hat{b}_{1j}$ ?



## Вопрос

В каком случае оценка взаимосвязи по отдельной подвыборке имеет больший вес в формировании оценки  $\hat{b}_{1j}$ ?

## Ответ

$$w_j = \frac{Var(u_{0j})}{Var(u_{0j}) + \frac{Var(e_{ij})}{n_j}}$$

- ❶ при более высокой межгрупповой вариации
- ❷ при большем размере подвыборки (количества наблюдений, доступных для j-ой страны)

при прочих равных условиях оценка взаимосвязи по отдельной подвыборке становится более значимой при формировании оценки  $\hat{b}_{1j}$ .

## BLUP

Также уникальную оценку коэффициента для отдельной единицы анализа можно переписать как:

$$\hat{b}_{1j} = \hat{\alpha}_{1j} + \hat{\gamma}_{10} - w_j * \hat{\gamma}_{10} = \hat{\gamma}_{10} + w_j(\hat{\alpha}_{1j} - \hat{\gamma}_{10}),$$

где  $w_j(\hat{\alpha}_{1j} - \hat{\gamma}_{10})$  – BLUP (предсказанное значение случайного эффекта)

## Вопрос

В контексте моделирования со смешанными эффектами нас в ряде случаев интересует не то, что мы в среднем наблюдаем по выборке, а страновая специфика. В связи с этим возникает необходимость предсказания характера взаимосвязи ключевого предиктора и зависимой переменной в отдельно взятой стране (или «стартовых» условий в отдельно взятой стране). Что показывает BLUP-значение?

## Вопрос

В контексте моделирования со смешанными эффектами нас в ряде случаев интересует не то, что мы в среднем наблюдаем по выборке, а страновая специфика. В связи с этим возникает необходимость предсказания характера взаимосвязи ключевого предиктора и зависимой переменной в отдельно взятой стране (или «стартовых» условий в отдельно взятой стране). Что показывает BLUP-значение?

## Ответ

BLUP – предсказанное значение отклонения в характере взаимосвязи (или отклонения в «стартовых условиях») в  $j$ -ой стране в отличие от того, что мы наблюдаем в среднем по всей выборке. ( $\hat{u}_j$ )

## Вопрос

Как осуществляется тестирование альтернативных спецификаций моделей со смешанными эффектами?

## Вопрос

Как осуществляется тестирование альтернативных спецификаций моделей со смешанными эффектами?

## Ответ

- 1 тест отношения правдоподобий (для вложенных моделей)
- 2 информационные критерии