

Практическая эконометрика.

Лекция 3. Тест Манна-Уитни.

Множественное тестирование гипотез

преподаватели 2023: Ольга Сучкова, Алексей Замниус, Анна Ставнийчук. При поддержке Георгия Калашнова

21 сентября 2023 г.

Тест Манна-Уитни (по Демешеву - ссылка)

- ▶ Подходит для любого распределения признака и любого размера выборки
- ▶ Игнорирует распределения признака, считает, сколько раз признак в тритмент-группе превосходит признак в контрольной группе
- ▶ На доске – игрушечный пример: $N(T=1)=3$, $N(T=0)=2$. Значения при $T=1$: 2, 4, 7. Значения при $T=0$: 5, 6. Рассчитаем U-статистику, p-значение
- ▶ На доске – примеры тримент-эффектов, когда критерий работает «хорошо» или «плохо».
- ▶ В русскоязычной Википедии Манн-Уитни описан неверно!

Выводы о тестах на равенство средних

- ▶ Если ситуация идеальная, т.е. большая (10000) выборка и распределение признака нормальное, то делаем $N(0,1)$
- ▶ Если выборка маленькая (100) и что-то не так с распределением, то бутстрап
- ▶ Если выборка совсем маленькая (30), то тест Уэлча
- ▶ Если распределение не нормальное, то тест Манна-Уитни (универсальный?)
- ▶ Помним про MDE (след раз – про снижение дисперсии)
- ▶ Все тесты оценивают АТЕ, но эффект может быть неоднородным (нужны другие методы: crf, мета-лёрнеры)

Table of Contents

Коррекция Бонферони

Важное из прошлой лекции

- ▶ Работаем с концепцией потенциальных исходов (модель Рубина)
- ▶ Предпосылки - безусловная независимость и SUTVA
- ▶ ATE, ATT, ATnT
- ▶ В идеальном эксперименте оценка эффекта - по разнице в средних Y (парная регрессия Y от T)

Несколько «нестандартных» ситуаций

- ▶ Несколько разных триментов, а не 0 и 1
- ▶ Проверяем результаты для многих метрик
- ▶ Проверяем гетерогенность эффекта и сравниваем несколько подгрупп

Везде много попарных сравнений...

- ▶ Менеджер верит, что эксперимент влияет хотя бы на что-то. Аналитик провёл 40 тестов, все они провалились. А давайте проверим еще 41й эффект с метрикой «заработная плата через 2 года после выпуска в подгруппе людей до 23 лет». Там-то эффект есть?
- ▶ Пример не из экономики - «зомби-лосось» (см. Филипп Ульянкин – ссылка)

Почему так нельзя

H_0 : treatment ни на что не влияет

$$P(\text{test1} | H_{0\text{test1}}) = 0.05$$

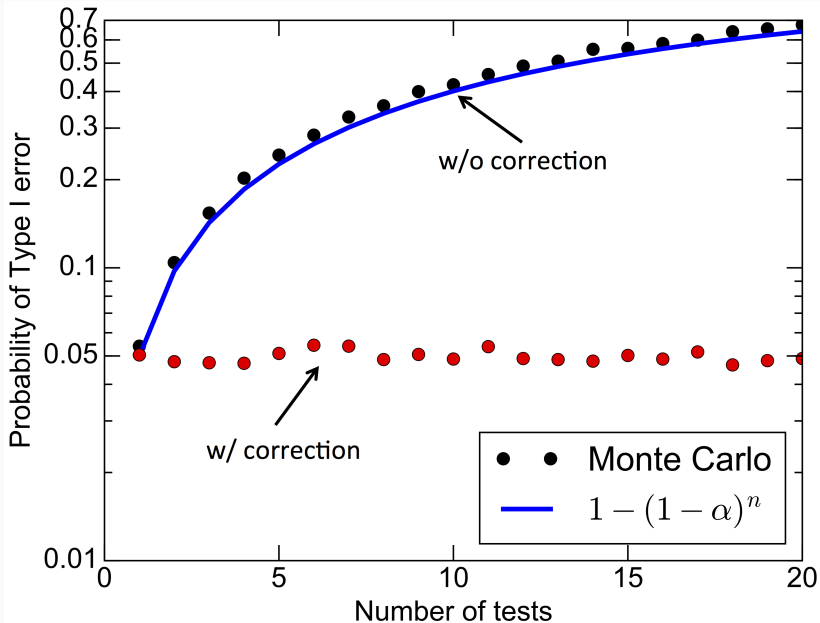
$$P(\text{test2} | H_{0\text{test2}}) = 0.05$$

$$P(\text{test3} | H_{0\text{test3}}) = 0.05$$

Family-wise error (FWER): вероятность отвергнуть верную нулевую гипотезу хотя бы в одном тесте.

Предположим test1, test2 и test3 распределены независимо, тогда

$$\text{FWER} = 1 - (1 - P(t1 | H_0))(1 - P(t2 | H_0))(1 - P(t3 | H_0)) > 0.05$$



Как правильно тестировать несколько гипотез за раз

- ▶ H_0 : верны все гипотезы H_{0ti}
- ▶ H_1 : верна хоть одна из гипотез H_{1ti}

J-test: на доске доверительный эллипс и квадрат

Коррекция Бонферрони

Просто поделить уровень значимости на количество тестов

$$\alpha' = \frac{\alpha}{m}$$

$$\text{FWER} \leq \sum_{i=1}^m P(\text{test}_i | H_0) = m\alpha' = \alpha$$

Восходящая процедура Хольма

1. Отсортировать р-значения тестов по возрастанию:

$$p_{(1)}, p_{(2)}, \dots, p_{(m)}$$

2. Найти минимальный k такой, что $p_{(k)} > \frac{\alpha}{m+1-k}$
3. отвергнуть все гипотезы с индексом $i < k$

Примеры процедур

- ▶ Коррекция Бонферрони
- ▶ Коррекция Сидака (альтернатива Бонферрони)
- ▶ Восходящая процедура Хольма
- ▶ Нисходящая процедура Сидака (альтернатива Хольму)
- ▶ Восходящая процедура Хольма с коррекцией на зависимость Romano, Joseph P and Michael Wolf (2005)

- ▶ Процедуры как Бонферрони применяют, если не хотят получить ложно-положительный результат. Например, когда это дорого обойдётся.
- ▶ Пример, когда можно обойтись без Бонферрони, - в тестах на выбросы (ссылка с объяснением), (ссылка с примером статьи). Это зависит от взгляда на выборку.