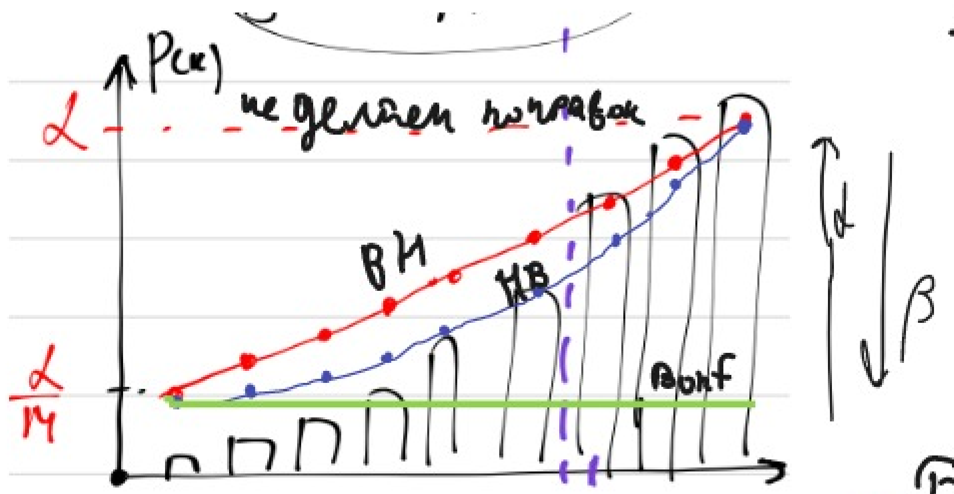
Формулы

	H_0 не отвергается	H_0 отвергается	Итого
H_0 верна	V	U	m_0
H_0 неверна	S	T	$m - m_0$
Итого	R	$m - R$	m

- FWER $= P\{U > 0\} = E[I\{U > 0\}] = \frac{m_0 \cdot \alpha}{m}$
- FDR $= E[\frac{U}{max\{U+T,1\}}] =$
- поправка Бонферрони $\alpha_{ind} = \frac{\alpha}{m}$
- поправка Холма-Бонферрони $\alpha_k = \frac{\alpha}{m+1-k}$
- процедура Бенджамини-Хохберга $\alpha_k = \frac{\alpha \cdot k}{m}$



k	α_{HB}	α_{BH}
1	α/M	α/M
2	$2\alpha/M$	$\alpha/(M-1)$

$$\frac{2\alpha(M-1) - M\alpha}{M(M-1)} = \frac{M\alpha - 2\alpha}{M(M-1)}$$

1) вопрос от бизнеса про очень большую выборку
2) 99% → $\alpha = 0,01$
3) $\mu_t = \mu_{t-1}$ → разобьем на группу
4) считаем средние
5) p-value
6) применяем поправки на много H_0.

Экспертным мнением
1) DBSCAN
2) по количеству сегментов

$m = \# H_0$

- Бонферрони
- Холм-Бонферрони
- Бенджамин-Хохберг
- -- α исходная

→
2/4 не отв
0/4 не отв.
0/4 не отв
0/4 не отв

2.1 Равенство средних в 4 группах студентов

Условие задачи

Допустим, на факультете экономических наук у нас есть 4 типа студентов:

- Мужского пола, учащиеся на бюджете
- Мужского пола, учащиеся на коммерческой основе
- Женского пола, учащиеся на бюджете
- Женского пола, учащиеся на коммерческой основе

Мы хотим проверить гипотезу о том, что средний балл по курсу "Математическая статистика"для всех 4 типов студентов будет равен среднему значению за прошлый год (контролируем инфляцию оценок!). На полученных данных мы проверили уже известные нам гипотезы для конкретных выборок и получили следующие значения p-value:

- $p_1 = 0.01$ $< \alpha_{ind}$
- $p_2 = 0.04$ $>$
- $p_3 = 0.02$ $>$
- $p_4 = 0.005$ $<$

H_1, H_4 отверг.

H_2, H_3 не отверг

мб есть инфляция

Реализуйте алгоритм Бонферрони, Холма-Бонферрони и Бенджамини-Хокберга для осуществления выводов касательно выдвинутых гипотез на заданном исходном уровне значимости $\alpha = 0.05$

Решение

Алгоритм Бонферрони

$$\alpha_{ind} = \frac{\alpha}{m} = \frac{0.05}{4} = 0.0125$$

Сравнивая с заданными полученными значениями p-value для каждой из групп, мы отвергаем нулевую гипотезу для групп 1 и 4, не отвергаем для групп 2 и 3, даже несмотря на то, что p-value для данной группы ниже заявленного уровня значимости для нашей исходной задачи.

Алгоритм Холма-Бонферрони

Запишем вектор из имеющихся p-значений, отсортированных по возрастанию:

$$p_v = [0.005, 0.01, 0.02, 0.04]$$

Посчитаем индивидуальные значения α для каждого типа студентов:

- $\alpha_1 = \frac{0.05}{4+1-1} = 0.0125$ $> p_{(1)} = p_4$
- $\alpha_2 = \frac{0.05}{4+1-2} = 0.0167$ $> p_{(2)} = p_1$
- $\alpha_3 = \frac{0.05}{4+1-3} = 0.025$ $> p_{(3)} = p_3$
- $\alpha_4 = \frac{0.05}{4+1-4} = 0.05$ $> p_{(4)} = p_2$

$i = \overline{1,4}$

$\forall i \; H_0^i$ отвергается

Попарно сравниваем p-value и α_k для всех выборок, начиная с наименьшего p_k, останавливаясь в том случае, когда $p_k \geq \alpha_k$. Таким образом, получается, что с помощью модифицированного алгоритма Холма-Бонферрони мы отвергаем гипотезу о равенстве среднего балла в группах 1, 3, 4 среднему за прошлый год. При этом не отвергаем гипотезу о равенстве среднего в группе 2 среднему за прошлый год в силу обозначенной поправки, хоть и p-value для данной группы ниже заявленного уровня значимости для нашей исходной задачи.

Таким образом, мы можем заключить, что для обоих полученных результатов $FWER \leq \alpha = 0.05$, однако с помощью поправки Холма-Бонферрони вероятность ошибки второго рода (вероятность не отвергнуть одну или более неверную гипотезу) для полученного в этом пункте вывода ниже, нежели в первом пункте (что, в целом, и так интуитивно понятно).

Алгоритм Бенджамини-Хокберга

По вектору отсортированных p-value сравним с индивидуальными значениями α:

$$\alpha_k = \frac{\alpha \cdot k}{m} \qquad p_v = [0.005, 0.01, 0.02, 0.04]$$

- $\alpha_1 = \frac{0.05 \cdot 1}{4} = 0.0125$ $> p_1$
- $\alpha_1 = \frac{0.05 \cdot 2}{4} = 0.025$ $> p_2$
- $\alpha_1 = \frac{0.05 \cdot 3}{4} = 0.0375$ $> p_3$
- $\alpha_1 = \frac{0.05 \cdot 4}{4} = 0.05$ $> p_4$

H_0^i отвергается

В рамках аналогичного попарного сравнения можем убедиться, что во для каждой исследуемой группы $p_k < \alpha_k$. Таким образом, в соответствии с алгоритмом Бенджамини-Хокберга для всех групп гипотеза о равенстве среднего в группе среднему за прошлый год отвергается.

В соответствии со знанием с лекции, делаем вывод, что $FDR \leq \alpha = 0.05$. Но действительно ли это так для нашей задачи? Какое необходимое условие упущено из внимания?

miro

2.2 Связь типа личности и интеллекта

Условие задачи

Чтобы продемонстрировать значимость поправок Бонферрони, рассмотрим следующий пример. Исследователь хочет выяснить, связана ли личность с интеллектом. Участники заполняют опросник личности, который измеряет пять черт: экстраверсию, невротизм, добросовестность, покладистость и открытость. Кроме того, участники проходят три теста на интеллект: тест на вербальные способности, тест на числовые способности и тест на абстрактные способности. Полученные p-значения для гипотезы об отсутствии корреляции между тестами и опросами представлены в таблице:

	Экстраверсия	Невротизм	Добросовестность	Покладистость	Открытость
Вербальные	0.10	0.29	0.004	0.56	0.15
Числовые	0.08	0.002	0.09	0.09	0.35
Абстрактные	0.04	0.57	0.04	0.15	0.30

Поправка Бонферрони

Уровень α для каждого теста становится равным $\frac{0.05}{0.15} = 0.003$. В данном случае только одно из значений p меньше 0.003. Следовательно, исследователь может сделать общий вывод из своего исследования: личность связана со способностями, ведь один из тестов не отвергается! Корректировка Бонферрони гарантирует, что вероятность того, что хотя бы один из 15 тестов может привести к ошибке первого рода, составляет заданный уровень значимости. Следовательно, поправка Бонферрони гарантирует, что вероятность того, что исследователи сделают ложный вывод о связи личности со способностями, не превысит 0.05.

Тем не менее, поправка Бонферрони заметно снижает мощность проводимого теста, то есть поправка уменьшает вероятность того, что будет допущена ошибка второго рода (не отвергнута ложная гипотеза).

$\alpha = 0,05$

$m = 15$

$\downarrow$

$r_{ij} \quad \forall\, i = \overline{1,3},\ j = \overline{1,5}$

$H_0^i: \ r_{ij} = 0$

исследователь хотят:

$\exists\, H_0^i$ отвергнута $\Rightarrow H\{$связь способностей и черт характера$\}$ не отв.

$\alpha_{Bon} = \frac{0,05}{15} = 0,003$

только одна ~~не~~ отверг.

$\Rightarrow$ и это всё равно хорошо

Процедура Холма-Бонферрони

Отсортируем полученные p-value по возрастанию:

k	p_k	α_k
1	0.002	0.0033
2	0.004	0.0036
3	0.04	0.0038
4	0.04	0.0042
5	0.08	0.0045
6	0.09	0.005
7	0.09	0.0056
8	0.10	0.0063
9	0.15	0.0071
10	0.15	0.0083
11	0.29	0.01
12	0.30	0.0125
13	0.35	0.0167
14	0.56	0.025
15	0.57	0.05

same

плевать, что дальше, все остальные не отвергаем

Таким образом, в результате полученных значений α_{ind} мы можем сделать вывод о том, что уже после второго попарного сравнения $p_k > \alpha_k$, что соотносится с исходным тестом Бонферрони и даёт нам возможность сделать аналогичный вывод: не отвергается лишь гипотеза о наличии связи между опросом по невротизму и результатом теста на работу с цифрами.