Посібник до вивчення дисципліни “Статистичний аналіз даних”

Book · November 2018

CITATIONS
0

READS
1,276

1 author:

Yana Bondarenko
Oles Honchar Dnipro National University
43 PUBLICATIONS **37** CITATIONS

SEE PROFILE

Міністерство освіти і науки України

Дніпровський національний університет імені Олеся Гончара

Я. С. Бондаренко, С. В. Кравченко

ПОСІБНИК ДО ВИВЧЕННЯ ДИСЦИПЛІНИ “СТАТИСТИЧНИЙ АНАЛІЗ ДАНИХ”

Дніпро

2018

УДК 519.23 (075.8)
Б81

Рецензенти: канд. фіз.-мат. наук, доц. В.М. Трактинська,
канд. фіз.-мат. наук, доц. Н.І. Послайко

Б81 Бондаренко, Я. С. Посібник до вивчення дисципліни “Статистичний аналіз даних” [Текст] / Я.С. Бондаренко, С.В. Кравченко. – Д: Ліра, 2018. – 40 с.

Викладено основні поняття і методи множинної перевірки статистичних гіпотез. Теоретичні положення проілюстровано модельними експериментами. Досліджено математичну модель мультиваріантного тестування цільової сторінки сайту.

Для студентів механіко-математичного факультету ДНУ спеціальності “Статистика”.

Рекомендовано до друку вченою радою
механіко-математичного факультету
Дніпровського національного університету імені Олеся Гончара
протокол №3 від 20.11.2018 року

Навчальне видання

Яна Сергіївна Бондаренко
Сергій Вікторович Кравченко

Посібник до вивчення дисципліни
“Статистичний аналіз даних”

Друкується за авторською редакцією

Підписано до друку 20.11.18. Формат 60×84/16. Папір друкарський.
Друк плоский. Ум. друк. арк. 2,33. Тираж 20 пр. Зам. № 332.

Друкарня «Ліра», вул. Наукова, 5, м. Дніпро, 49107. Свідоцтво про внесення до Державного реєстру серія ДК №6042 від 26.02.2018 р.

ВСТУП

Для розуміння ефекту множинної перевірки статистичних гіпотез розглянемо задачу про пошук екстрасенсів. У 1950 році парапсихолог Джозеф Райн (Joseph Rhine, 1895–1980) проводив експеримент, який полягав у виявленні людей з екстрасенсорними здібностями. Тисяча осіб брала участь в експерименті. Кожна людина повинна була відгадати колір карти у послідовності з десяти карт. Дев'ять із десяти карт відгадали дванадцять осіб, всі десять карт – дві особи. У наступних експериментах жодна з них не підтвердила свої здібності. Чому?

Сформулюємо та розв’яжемо поставлену задачу як задачу перевірки статистичних гіпотез.

Вибір нульової гіпотези. $H_0 -$ людина не має екстрасенсорних здібностей, $H_1 -$ людина має екстрасенсорні здібності.

Визначення випадкової величини та її розподілу. Стохастичний експеримент полягає у спостереженні випадкової величини ξ – числа правильно вгаданих карт, $\xi = 0, 1, 2, \dots, 10$. Випадкова величина ξ має біноміальний розподіл с параметрами $(10, p)$, де $p -$ імовірність «успіху» («успіх» – колір карти визначений правильно).

Якщо гіпотеза H_0 справедлива, то ξ має біномний розподіл

$$P_0\{\xi = k\} = C_{10}^k (p_0)^k (1 - p_0)^{10-k}, \qquad k = 0, 1, 2, \dots, 10,$$

де $p_0 = 0{,}5$ – імовірність «успіху» за умови, що людина не має екстрасенсорних здібностей. Обидва результати експерименту – колір карти визначений правильно або неправильно – рівноймовірні для звичайної людини.

Якщо гіпотеза H_1 справедлива, то ξ має біномний розподіл

$$P_1\{\xi = k\} = C_{10}^k (p_1)^k (1 - p_1)^{10-k}, \qquad k = 0, 1, 2, \dots, 10,$$

де p_1 – імовірність «успіху» за умови, що людина має екстрасенсорні здібності.

Вибір критерію (вибір критичної множини). Природно відхиляти гіпотезу H_0, якщо ξ набуває великих значень і не відхиляти, якщо ξ набуває малих значень. Тоді критична множина визначається так:

$$S = \{x: x \geq l\}.$$

Якщо $\xi \in S$, тобто $\xi \geq l$, то нульова гіпотеза H_0 відхиляється.

Якщо $\xi \notin S$, тобто $\xi < l$, то нульова гіпотеза H_0 не відхиляється.

Вибір рівня значущості. Оберемо $\alpha = 0{,}05$ і визначимо l як найменше ціле число, для якого справедлива нерівність:

$$P(S|H_0) = P\{\xi \geq l\,|H_0\} = \sum_{k=l}^{10} C_{10}^k (0{,}5)^k (0{,}5)^{10-k} \leq \alpha,$$

зокрема,

$$P(S\,|H_0) = C_{10}^9 \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right)^1 + C_{10}^{10} \left(\frac{1}{2}\right)^{10} \left(\frac{1}{2}\right)^0 = 11\left(\frac{1}{2}\right)^{10} = 0{,}0107 \leq 0{,}05.$$

Отже, $l = 9$ і критерій набуває вигляду $S = \{x: x \geq 9\}$.

Якщо $\xi \geq 9$, то H_0 відхиляється – людина має екстрасенсорні здібності. Якщо $\xi < 9$, то H_0 не відхиляємо – людина не має екстрасенсорних здібностей. Правильне визначення дев'яти або десяти карт із десяти карт говорить про те, що людина – екстрасенс.

Імовірності помилок першого та другого роду. Імовірності помилок першого та другого роду наведені в таблиці 1.1.1.

Таблиця 1.1.1. Імовірності помилок I та II роду

	Справедлива гіпотеза	
Гіпотеза, яка приймається	H_0 – людина не має екстрасенсорних здібностей	H_1 – людина має екстрасенсорні здібності
H_0 – людина не має екстрасенсорних здібностей	$1 - 11\left(\frac{1}{2}\right)^{10}$	$1 - \sum_{k=9}^{10} C_{10}^k (p_1)^k (1-p_1)^{10-k}$
H_1 – людина має екстрасенсорні здібності	$11\left(\frac{1}{2}\right)^{10}$	$\sum_{k=9}^{10} C_{10}^k (p_1)^k (1-p_1)^{10-k}$

Ефект множинної перевірки гіпотез. У експерименті брала участь тисяча осіб. Імовірність того, що з 1000 осіб *жодна* людина не відгадає дев'ять або десять карт із десяти карт становить

$$\left(1 - 11\left(\frac{1}{2}\right)^{10}\right)^{1000} = 0{,}00002.$$

Тоді ймовірність того, що з 1000 осіб *хоча б одна* людина випадково відгадає дев'ять або десять карт із десяти карт становить

$$1 - \left(1 - 11\left(\frac{1}{2}\right)^{10}\right)^{1000} = 0{,}99998.$$

Якщо ми подивимось, як веде себе ця ймовірність в залежності від кількості осіб, то побачимо, що вона зростає дуже швидко.

Дійсно, якщо в експерименті беруть участь 100 осіб, то ми знайдемо *хоча б одного* екстрасенса з імовірністю більше 0,5:

$$1 - \left(1 - 11\left(\frac{1}{2}\right)^{10}\right)^{100} = 0{,}66041.$$

Для 200 осіб імовірність знайти *хоча б одного* екстрасенса дорівнює

$$1 - \left(1 - 11\left(\frac{1}{2}\right)^{10}\right)^{200} = 0{,}88468.$$

Для 500 осіб імовірність знайти *хоча б одного* екстрасенса дорівнює

$$1 - \left(1 - 11\left(\frac{1}{2}\right)^{10}\right)^{500} = 0{,}99548.$$

Той факт, що ми за допомогою цієї статистичної процедури знаходимо екстрасенсів є прикладом *ефекту множинної перевірки гіпотез*. При одночасній перевірці великої кількості статистичних гіпотез, імовірність зробити *хоча б одну* помилку першого роду велика.

1. Множинна перевірка статистичних гіпотез

1.1. Математична постановка задачі множинної перевірки гіпотез

1.1.1. Задача перевірки однієї статистичної гіпотези

Нехай $\xi(\omega) = \left(\xi_1(\omega), \dots, \xi_n(\omega)\right)$ – реалізація вибірки $\xi = \left(\xi_1, \dots, \xi_n\right)$ з невідомого розподілу F. Відносно розподілу F висувається нульова гіпотеза $H_0: F = G$ або, що теж саме, $\xi = \left(\xi_1, \dots, \xi_n\right)$ – вибірка з розподілу G, де G повністю визначений. Альтернативна гіпотеза $H_1: F \neq G$.

Для перевірки нульової гіпотези будемо використовувати статистику $T = T\left(\xi_1, \dots, \xi_n\right)$, яка є деякою функцією від вибірки. Нам відомий розподіл статистики T при справедливій нульовій гіпотезі H_0. За цим розподілом ми розраховуємо *рівень значущості, що досягається*, тобто ймовірність здобути таке або ще більше значення статистики, ніж ми отримали в експерименті:

$$p = P\{T \geq t | H_0\}.$$

Рівень значущості, що досягається p, порівнюється з рівнем значущості α (як правило, $\alpha = 0{,}05$). Якщо $p \leq \alpha$, то нульова гіпотеза $H_0: F = G$ відхиляється на користь альтернативної гіпотези, у супротивному разі – не відхиляється.

Процедура перевірки нульової гіпотези побудована так, що рівень значущості α обмежує зверху ймовірність помилки першого роду.

1.1.2. Задача множинної перевірки статистичних гіпотез

Нехай $\xi^1(\omega) = \left(\xi_1^1(\omega), \dots, \xi_n^1(\omega)\right), \dots, \xi^m(\omega) = \left(\xi_1^m(\omega), \dots, \xi_n^m(\omega)\right)$ – реалізації вибірок $\xi^1 = \left(\xi_1^1, \dots, \xi_n^1\right), \dots, \xi^m = \left(\xi_1^m, \dots, \xi_n^m\right)$ з невідомих розподілів $F_1, \dots, F_m$. Відносно розподілів $F_1, \dots, F_m$ висувається m нульових гіпотез $H_0^1: F_1 = G_1, \dots, H_0^m: F_m = G_m$. Альтернативні гіпотези: $H_1^1: F_1 \neq G_1, \dots, H_1^m: F_m \neq G_m$.

Для перевірки нульових гіпотез будемо використовувати статистики $T^1, \dots, T^m$, які є деякими функціями від вибірок $\xi^1 = (\xi_1^1, \dots, \xi_n^1), \dots, \xi^m = (\xi_1^m, \dots, \xi_n^m)$.

Нам відомі розподіли статистик T^i при справедливих нульових гіпотезах H_0^i. За цими розподілами ми розраховуємо *рівні значущості, що досягаються*:

$$p_i = P\{T^i \geq t | H_0^i\}, \qquad i = 1,2, \dots, m.$$

Рівні значущості, що досягаються p_i, порівнюються з рівнем значущості α (як правило, $\alpha = 0{,}05$). Якщо $p_i \leq \alpha$, то нульова гіпотеза $H_0^i: F_i = G_i$ відхиляється на користь альтернативної гіпотези, у супротивному разі – не відхиляється.

У таблиці 1.1.2 наведено кількість справедливих і несправедливих, відхилених і невідхилених нульових гіпотез $H_0^i,\ i = 1,2, \dots m$.

Табл. 1.1.2. Кількість справедливих і несправедливих, відхилених і невідхилених гіпотез

	Кількість справедливих гіпотез H_0^i	Кількість несправедливих гіпотез H_0^i	Разом
Кількість невідхилених гіпотез H_0^i	U	T	$m - R$
Кількість відхилених гіпотез H_0^i	V	S	R
Разом	m_0	$m - m_0$	m

У таблиці 1.1.2 нам відома лише загальна кількість гіпотез m. Єдиний параметр, яким ми можемо керувати, – це R, кількість гіпотез, які ми відхиляємо, при цьому величина, яка нас хвилює найбільше, – це V, кількість справедливих гіпотез, які ми відхиляємо. Ми хочемо, щоб ця величина була якомога меншою. Але єдине, що ми можемо робити, так це

перерозподіляти гіпотези з другого рядка в перший рядок таблиці. Якщо ми хочемо припускати незначну кількість помилок першого роду, нам необхідно відхиляти менше гіпотез.

Означення. **Груповою ймовірністю помилки першого роду Family-Wise Error Rate** називають імовірність зробити хоча б одну помилку першого роду:

$$FWER = P(V > 0).$$

Цю величину ми хочемо контролювати, тобто хочемо побудувати таку статистичну процедуру, щоб імовірність припустити хоча б одну помилку першого роду не перевищувала α:

$$FWER = P(V > 0) \leq \alpha.$$

Як цього можна досягти? Єдине, що у нас є – це рівні значущості $\alpha_1, \alpha_2, \dots, \alpha_m$, на яких перевіряються гіпотези $H_0^1, H_0^2, \dots, H_0^m$. Наша задача обрати $\alpha_1, \alpha_2, \dots, \alpha_m$ так, щоб забезпечити обмеження $FWER$ числом α.

Для розв'язання поставленої задачі використовують метод Бонферроні та спадні процедури множинної перевірки гіпотез – метод Холма та метод Шидака.

1.2. Метод Бонферроні

1.2.1. Статистична процедура методу

Одним з найбільш простих та відомих способів контролю групової ймовірності помилки першого роду є метод, названий на честь італійського математика Карло Еміліо Бонферроні (Carlo Emilio Bonferroni, 1892–1960). В методі Бонферроні рівні значущості, що досягаються p_i, здобуті при перевірці гіпотез H_0^i, $i = 1,2, \dots, m$, порівнюються з рівнями значущості α_i:

$$\alpha_1 = \alpha_2 = \dots = \alpha_m = \frac{\alpha}{m}.$$

Альтернативний спосіб полягає в тому, щоб перетворюються всі рівні значущості, що досягаються p_i, домножаючи їх на m:

$$\tilde{p}_i = \min(1, mp_i), \qquad i = 1,2,\dots,m.$$

Ці модифіковані рівні значущості, що досягаються $\tilde{p}_i$, порівнюються з рівнем значущості α. За такої процедури контролюється величина *FWER*.

Теорема 1.2.1. Якщо гіпотези $H_0^1,\dots,H_0^m$ відхиляються при $p_i \le \alpha/m$, то $FWER \le \alpha$.

Доведення. Для $m_0 < m$ здобуваємо:

$$FWER = P\{V > 0\} = P\left\{\bigcup_{i=1}^{m_0}\left\{p_i \le \frac{\alpha}{m}\right\}\right\} \le \sum_{i=1}^{m_0} P\left\{p_i \le \frac{\alpha}{m}\right\} \le \sum_{i=1}^{m_0}\frac{\alpha}{m} \le \alpha.$$

Зауваження. В ідеалі бажано, щоб імовірність зробити хоча б одну помилку першого роду в точності дорівнювала α. Але при використанні методу Бонферроні *FWER* не просто менша за α, а набагато менша ніж α. Зменшення кількості помилок першого роду призводить до збільшення кількості помилок другого роду, тобто потужність статистичної процедури знижується.

1.2.2. Модельний експеримент

Для ілюстрації процедури методу Бонферроні згенеруємо 150 вибірок з нормального розподілу $N_{0;1}$ та 50 вибірок з нормального розподілу $N_{1;1}$. Обсяги всіх вибірок дорівнюють 20. Для кожної вибірки перевіряється гіпотеза про рівність середнього значення нулеві $H_0^i: a_i = 0, i = 1,\dots 200$ проти двосторонньої альтернативи $H_1^i: a_i \neq 0,\ i = 1,\dots 200$ за допомогою критерію Стьюдента. Результати перевірки гіпотез без поправки на множинну перевірку наведені в таблиці 1.2.2.1.

Таблиця 1.2.2.1. Результати експерименту 1

	# справедливих H_0^i	# несправедливих H_0^i	Разом
# прийнятих H_0^i	142	2	144
# відхилених H_0^i	8	48	56
Разом	150	50	200

На рисунку 1.2.2.1 рівні значущості, що досягаються p_i, позначені блакитними точками для справедливих гіпотез і – червоними точками для несправедливих гіпотез. Ми відхилили 48 несправедливих гіпотез (48 червоних точок знаходяться нижче прямої $\alpha = 0{,}05$) і не відхилили лише 2 несправедливі гіпотези (дві червоні точки знаходяться вище прямої $\alpha = 0{,}05$). Разом з тим ми відхилили 8 справедливих гіпотез, тобто зробили 8 помилок першого роду (вісім блакитних точок знаходяться нижче прямої $\alpha = 0{,}05$) і не відхили 142 справедливі гіпотези (142 блакитні точки знаходяться вище прямої $\alpha = 0{,}05$).

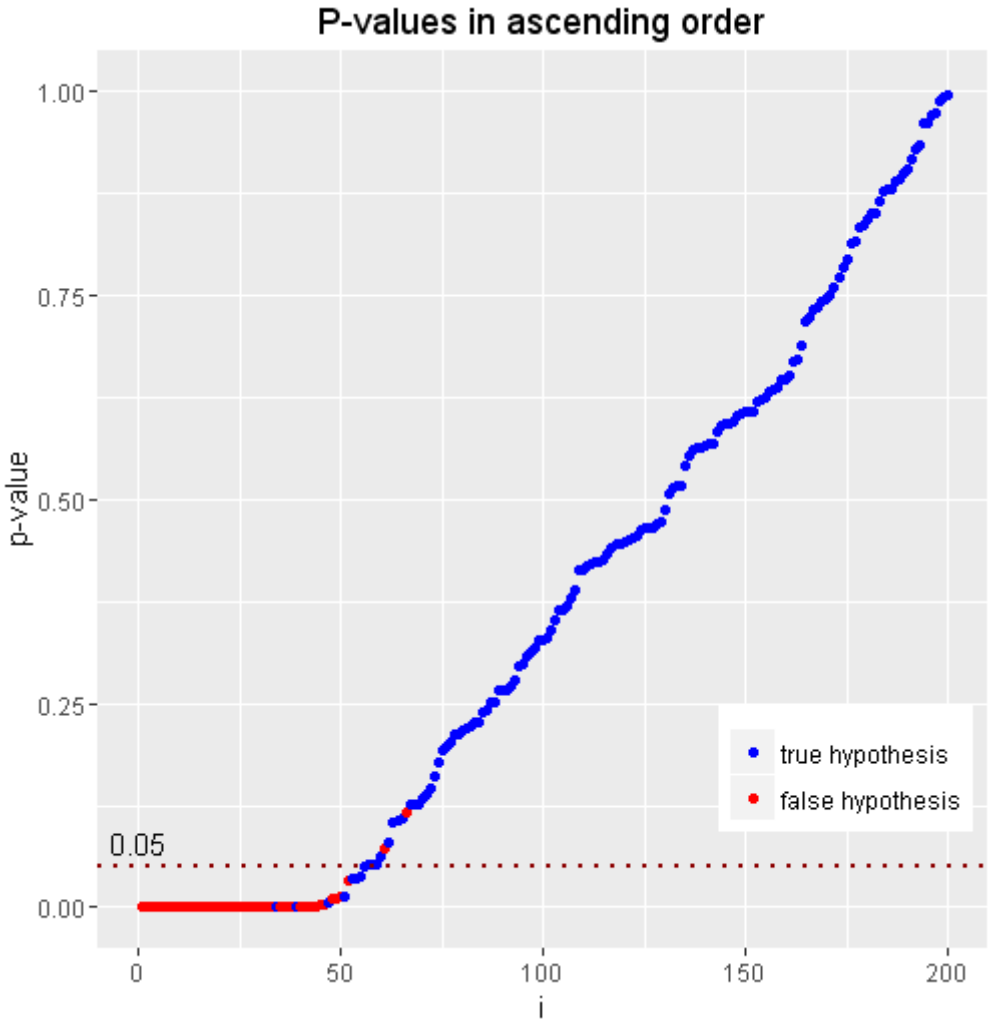


Рис. 1.2.2.1. Результати експерименту 1

Результати перевірки гіпотез згідно з методом Бонферроні наведені в таблиці 1.2.2.2. На рисунках 1.2.2.2, 1.3.4.1, 1.3.4.2, 1.4.4.1, 1.4.4.2 рівні значущості, що досягаються $\tilde{p}_i$, позначатимемо блакитними точками для справедливих гіпотез і – червоними точками для несправедливих гіпотез.

Таблиця 1.2.2.2. Результати експерименту 2

	# справедливих H_0^i	# несправедливих H_0^i	Разом
# прийнятих H_0^i	150	28	178
# відхилених H_0^i	0	22	22
Разом	150	50	200

Ми не відхилили жодної справедливої нульової гіпотези, тобто не припустили жодної помилки першого роду (немає блакитних точок нижче прямої $\alpha = 0{,}05$), але, разом з тим, ми втратили можливість відхилити 28 несправедливих гіпотез із 50 (28 червоних точок знаходяться вище прямої $\alpha = 0{,}05$). Статистична процедура методу Бонферроні має низьку потужність, тобто ймовірність відхилити несправедливі гіпотези низька.

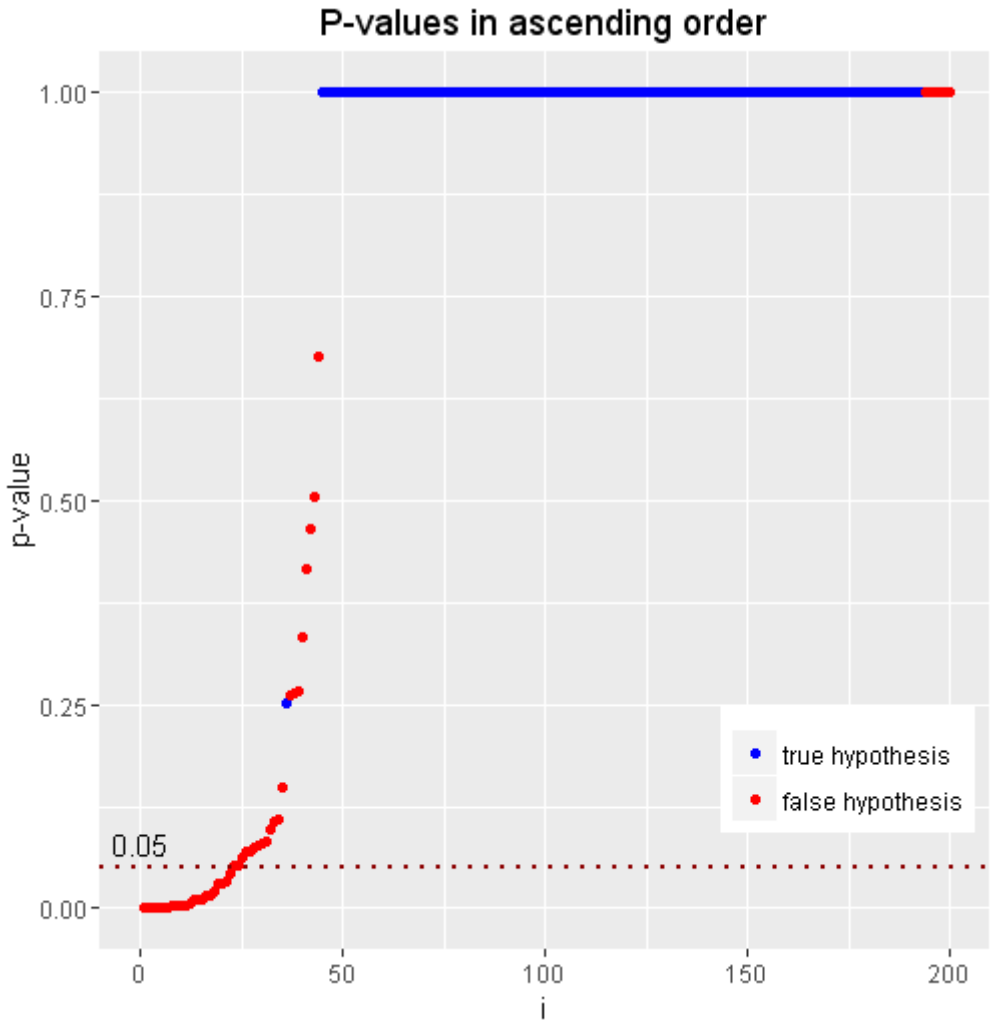


Рис. 1.2.2.2. Результати експерименту 2

1.3. Спадна процедура множинної перевірки гіпотез

1.3.1. Статистична процедура методів

Ідея спадної процедури множинної перевірки гіпотез полягає в тому, що нульові гіпотези відхиляються послідовно, починаючи з найбільш значущих, тобто рух йде за спаданням значущості. В загальному вигляді вона представлена так: здобуті рівні значущості, що досягаються впорядковуються за зростанням: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$. Всі гіпотези позначаються так, щоб їх номери відповідали номерам рівнів значущості, що досягаються, в цьому варіаційному ряді: $H_{(1)}, H_{(2)}, \ldots, H_{(m)}$.

На першому кроці найменший рівень значущості, що досягається $p_{(1)}$, порівнюємо з рівнем значущості α_1. Якщо $p_{(1)} \geq \alpha_1$, то приймаємо всі нульові гіпотези $H_{(1)}, H_{(2)}, \ldots, H_{(m)}$ і процес зупиняється. Якщо $p_{(1)} < \alpha_1$, то відхиляємо гіпотезу $H_{(1)}$ і процедура продовжується. На другому кроці порівнюємо $p_{(2)}$ і α_2. Якщо $p_{(2)} \geq \alpha_2$, то приймаємо всі нульові гіпотези, що залишилися: $H_{(2)}, \ldots, H_{(m)}$ – і процедура завершується. Якщо $p_{(2)} < \alpha_2$, то відхиляємо гіпотезу $H_{(2)}$, і процедура продовжується. І так далі.

Методи Шидака та Холма побудовані за допомогою спадної процедури множинної перевірки гіпотез.

1.3.2. Метод Шидака

Статистична процедура множинної перевірки гіпотез, запропонована у 1967 році чеським математиком Збинеком Шидаком (Zbyněk Šidák, 1933-1999), полягає в тому, що рівні значущості, що досягаються p_i, здобуті при перевірці статистичних гіпотез H_0^i, $i = 1,2, \ldots, m$, порівнюються з рівнями значущості α_i:

$$\alpha_1 = \alpha_2 = \ldots = \alpha_m = 1 - (1 - \alpha)^{1/m}.$$

Альтернативний спосіб полягає в тому, що перетворюються всі рівні значущості, що досягаються: $\tilde{p}_{(i)} = 1 - (1 - p_i)^m$. Ці модифіковані рівні значущості, що досягаються $\tilde{p}_i$, порівнюються з рівнем значущості α. За такої процедури контролюється величина $FWER$ за умов, що статистики T^i, $i = 1,2, \dots, m$, незалежні.

1.3.3. Метод Холма

Для того, щоб розв'язання проблеми низької потужності методу Бонферроні, шведський математик Стур Холм (Sture Holm) у 1979 році запропонував рівні значущості $\alpha_1, \alpha_2, \dots, \alpha_m$ взяти не рівними, а різними:

$$\alpha_1 = \frac{\alpha}{m}, \dots, \alpha_i = \frac{\alpha}{m - i + 1}, \dots, \alpha_m = \alpha.$$

Альтернативний спосіб полягає в тому, щоб перетворюються всірівні значущості, що досягаються: $\tilde{p}_{(i)} = min\left(1, max\left((m - i + 1)p_{(i)}, \tilde{p}_{(i-1)}\right)\right)$. Ці модифіковані рівні значущості, що досягаються $\tilde{p}_i$, порівнюються з рівнем значущості α.

Метод Холма потужніший за метод Бонферроні, тобто метод завжди відхиляє не менше гіпотез, ніж метод Бонферроні, оскільки його рівні значущості α_i не менші, ніж в методі Бонферроні. Метод Холма контролює $FWER$ без додаткових припущень.

1.3.4. Модельний експеримент

Для ілюстрації статистичної процедури методів Холма та Шидака згенеруємо 150 вибірок з нормального розподілу $N_{0;1}$ та 50 вибірок з нормального розподілу $N_{1;1}$. Обсяги всіх вибірок дорівнюють 20. Для кожної вибірки перевіряється гіпотеза про рівність середнього значення нулеві $H_0^i: a_i = 0, i = 1, \dots 200$ проти двосторонньої альтернативи $H_1^i: a_i \neq 0, i = 1, \dots 200$ за допомогою критерію Стьюдента.

Результати перевірки гіпотез із застосуванням методу Шидака наведені в таблиці 1.3.4.1 та рисунку 1.3.4.1.

Таблиця 1.3.4.1. Результати експерименту 3

	# справедливих H_0^i	# несправедливих H_0^i	Разом
# прийнятих H_0^i	150	23	173
# відхилених H_0^i	0	27	27
Разом	150	50	200

Ми не відхилили жодної справедливої нульової гіпотези, тобто не припустили жодної помилки першого роду (немає блакитних точок нижче прямої $\alpha = 0{,}05$), але, разом з тим, ми втратили можливість відхилити 23 несправедливі гіпотези із 50 (23 червоні точки знаходяться вище прямої $\alpha = 0{,}05$).

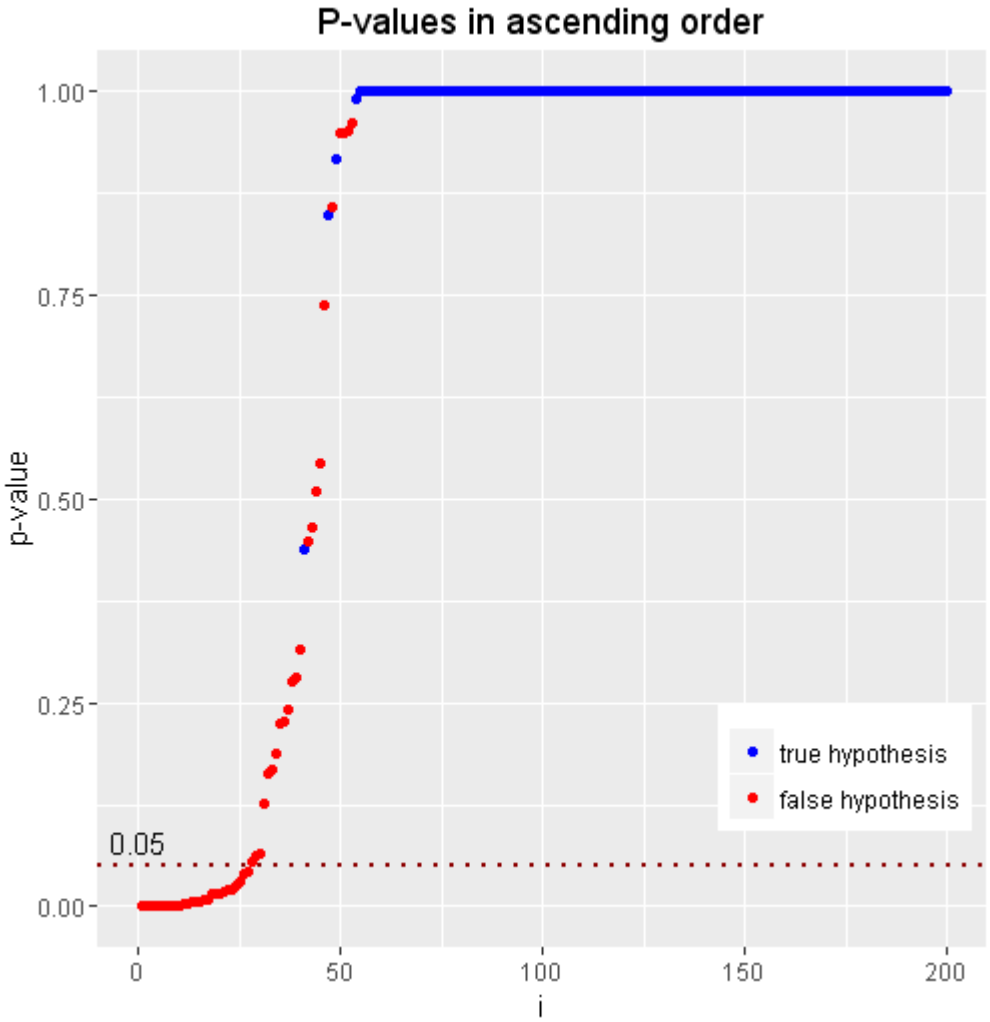


Рис. 1.3.4.1. Результати експерименту 3

Результати перевірки гіпотез із застосуванням методу Холма наведені в таблиці 1.3.4.2.

Таблиця 1.3.4.2. Результати експерименту 4

	# справедливих H_0^i	# несправедливих H_0^i	Разом
# прийнятих H_0^i	150	20	170
# відхилених H_0^i	0	30	30
Разом	150	50	200

Ми не відхилили жодної справедливої нульової гіпотези, тобто не зробили жодної помилки першого роду (немає блакитних точок нижче прямої $\alpha = 0{,}05$). Але, вже на відміну від методу Бонферроні, 20 гіпотез із 50 несправедливих гіпотез було прийнято (20 червоних точок знаходяться вище прямої $\alpha = 0{,}05$).

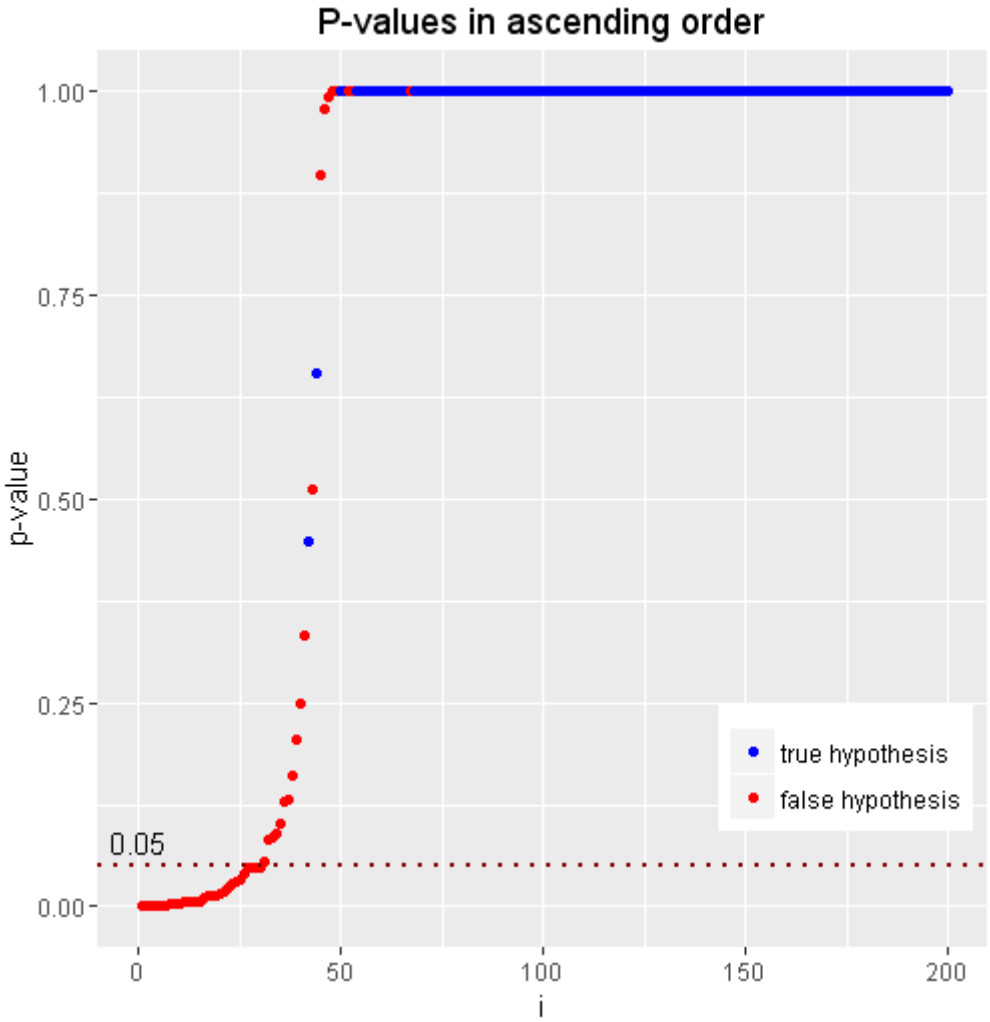


Рис. 1.3.4.2. Результати експерименту 4

Відмінність між методами Холма і Бонферроні не така велика. У відхиленні всіх несправедливих нульових гіпотез метод Холма нам не допоміг, але цей метод дозволив без додаткових припущень відхилити ще вісім несправедливих гіпотез, а цього достатньо, щоб почати використовувати цей метод.

1.4. Висхідна процедура множинної перевірки гіпотез

1.4.1. Статистична процедура методів

При перевірці сотень або тисяч гіпотез, можна трохи збільшити кількість помилок першого роду заради того, щоб збільшити потужність процедури і відхилити більше несправедливих гіпотез, тобто здійснити менше помилок другого роду. В таких випадках доцільно використовувати не FamilyWise Error Rate, а очікувану частку помилкових відхилень гіпотез **False Discovery Rate**:

$$FDR = M\left(\frac{V}{max(R,1)}\right),$$

де V – кількість відхилених справедливих гіпотез; R – кількість відхилених гіпотез.

Висхідні методи, які контролюють FDR, працюють з тим самим варіаційним рядом рівнів значущості, що досягаються, що й низхідні методи: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$, але процедура починається з іншого кінця цього ряду. На першому кроці найбільший рівень значущості, що досягається $p_{(m)}$, порівнюємо з рівнем значущості α_m. Якщо $p_{(m)} \leq \alpha_m$, то всі нульові гіпотези $H_{(1)}, H_{(2)}, \dots, H_{(m)}$ відхиляються, і процедура зупиняється. Якщо $p_{(m)} > \alpha_m$, то гіпотеза $H_{(m)}$ приймається, і процедура продовжується. На другому кроці порівнюємо $p_{(m-1)}$ і α_{m-1}. Якщо $p_{(m-1)} \leq \alpha_{m-1}$, то всі нульові гіпотези $H_{(1)}, H_{(2)}, \dots, H_{(m-1)}$ відхиляються, і процедура зупиняється. Якщо $p_{(m-1)} > \alpha_{m-1}$, то приймаємо гіпотезу $H_{(m-1)}$ і процедура продовжується. І так далі. Висхідна процедура завжди відхиляє не менше гіпотез, ніж спадна для одних і тих самих α_i.

1.4.2. Метод Бенджаміні-Хохберга

У 1995 році ізраїльські математики Йоав Бенджаміні (Yoav Benjamini) та Йозеф Хохберг (Yosef Hochberg) запропонували рівні значущості $\alpha_1, \alpha_2, \dots, \alpha_m$ змінювати лінійно:

$$\alpha_1 = \frac{\alpha}{m}, \dots, \alpha_i = \frac{\alpha i}{m}, \dots, \alpha_m = \alpha.$$

Альтернативний спосіб полягає в тому, що перетворюються всі рівні значущості, що досягаються p_i:

$$\tilde{p}_{(i)} = \min\left(1, \frac{m p_{(i)}}{i}, \tilde{p}_{(i+1)}\right).$$

Ці модифіковані рівні значущості, що досягаються $\tilde{p}_i$, порівнюються з рівнем значущості α.

Метод Бенджаміні-Хохберга забезпечує контроль над FDR тільки за умов незалежності статистик T^i, $i = 1,2, \dots, m$, які використовуються при перевірці гіпотез. Ця вимога достатньо сильна.

1.4.3. Метод Бенджаміні-Хохберга-Іекутіелі

У 2001 році ізраїльські математики Йоав Бенджаміні (Yoav Benjamini) та Даніель Іекутіелі (Daniel Yekutieli) удосконалили метод Бенджаміні-Хохберга й запропонували рівні значущості $\alpha_1, \alpha_2, \dots, \alpha_m$ змінювати лінійно за формулами:

$$\alpha_1 = \frac{\alpha}{m\sum_{j=1}^{m} 1/j}, \dots, \alpha_i = \frac{\alpha i}{m\sum_{j=1}^{m} 1/j}, \dots, \alpha_m = \frac{\alpha}{\sum_{j=1}^{m} 1/j}.$$

Альтернативний спосіб полягає в тому, що перетворюються всі рівні значущості, що досягаються p_i:

$$\tilde{p}_{(i)} = \min\left(1, \frac{m p_{(i)}}{i}\sum_{j=1}^{m}\frac{1}{j}, \tilde{p}_{(i+1)}\right).$$

Ці модифіковані рівні значущості, що досягаються $\tilde{p}_i$, порівнюються з рівнем значущості α. Метод Бенджаміні-Хохберга-Іекутіелі контролює False Discovery Rate без додаткових припущень.

1.4.4. Модельний експеримент

Для ілюстрації статистичної процедури методів Бенджаміні-Хохберга та Бенджаміні-Хохберга-Іекутіелі згенеруємо 150 вибірок з нормального розподілу $N_{0;1}$ та 50 вибірок з нормального розподілу $N_{1;1}$. Обсяги всіх вибірок дорівнюють 20. Для кожної вибірки перевіряється гіпотеза про рівність середнього значення нулеві $H_0^i: a_i = 0, i = 1, \dots 200$ проти двосторонньої альтернативи $H_1^i: a_i \neq 0, i = 1, \dots 200$ за допомогою критерію Стьюдента.

Результати перевірки гіпотез із застосуванням методу Бенджаміні-Хохберга наведені в таблиці 1.4.4.1.

Таблиця 1.4.4.1. Результати експерименту 5

	# справедливих H_0^i	# несправедливих H_0^i	Разом
# прийнятих H_0^i	148	5	153
# відхилених H_0^i	2	45	47
Разом	150	50	200

Ми відхилили 2 справедливі нульові гіпотези, тобто зробили 2 помилки першого роду (дві блакитні точки знаходяться нижче прямої $\alpha = 0{,}05$) і лише 5 несправедливих гіпотез із 50 було прийнято (5 червоних точок знаходяться вище прямої $\alpha = 0{,}05$).

Результати перевірки гіпотез із застосуванням методу Бенджаміні-Хохберга-Іекутіелі наведені в таблиці 1.4.4.2.

Таблиця 1.4.4.2. Результати експерименту 6

	# справедливих H_0^i	# несправедливих H_0^i	Разом
# прийнятих H_0^i	149	8	157
# відхилених H_0^i	1	42	43
Разом	150	50	200

Ми відхилили одну справедливу нульову гіпотезу, тобто зробили 1 помилку першого роду, і 8 несправедливих гіпотез із 50 було прийнято.

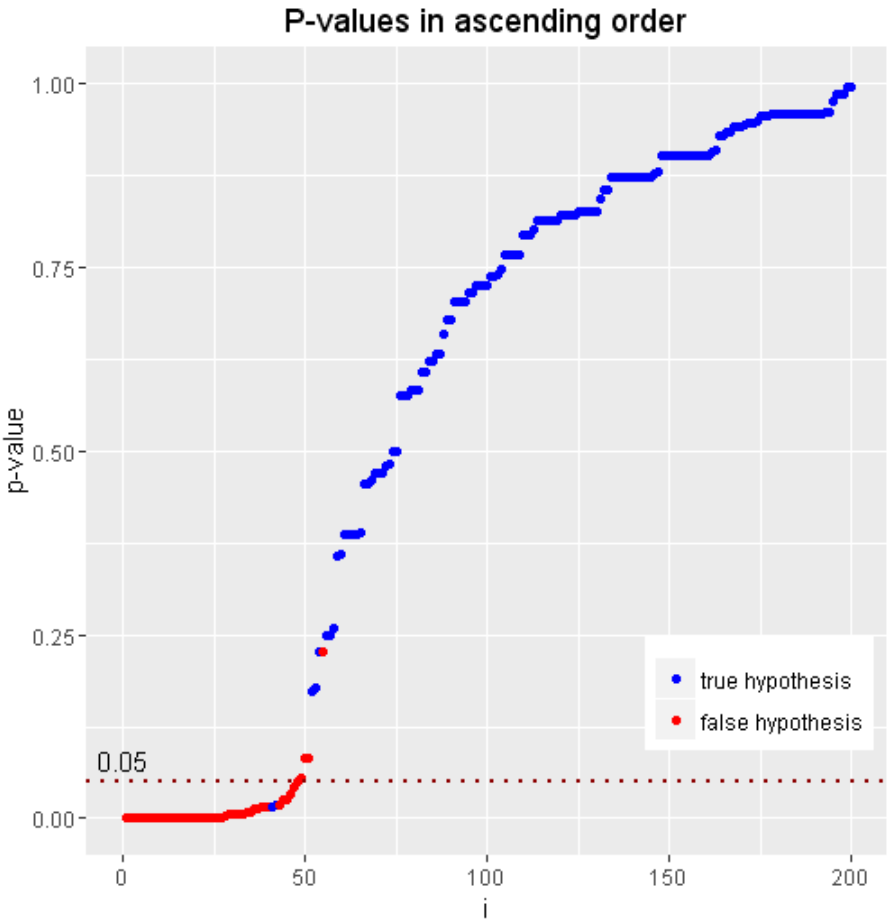


Рис. 1.5.4.1. Результати експерименту 5

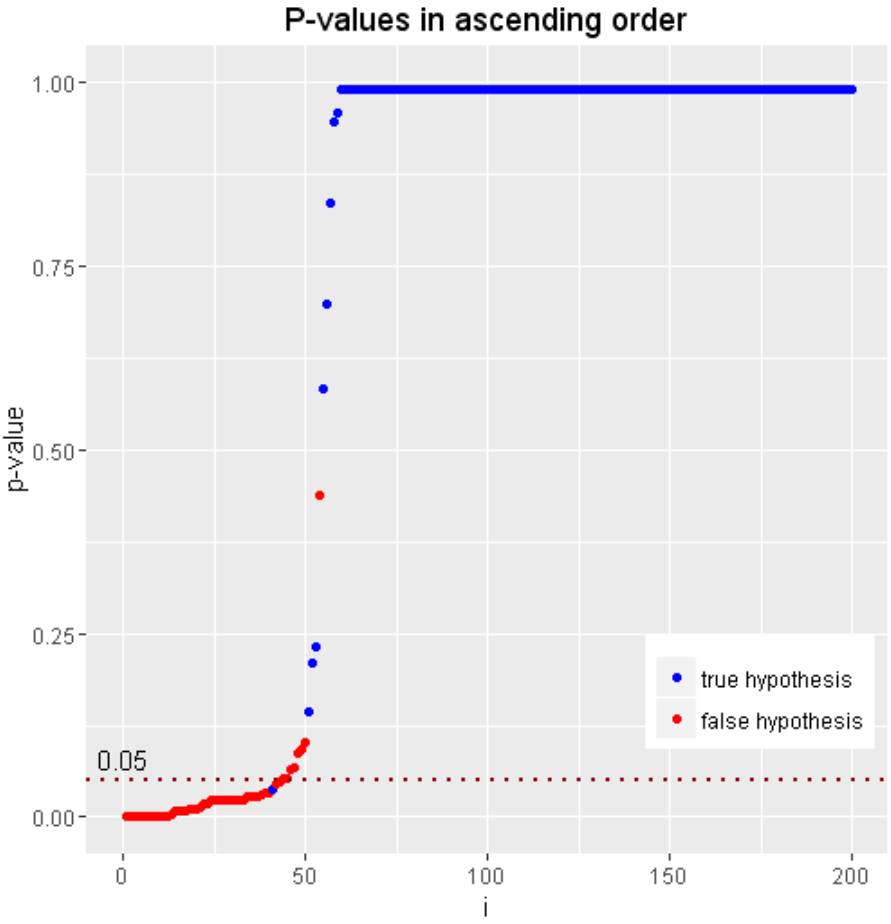


Рис. 1.5.4.2. Результати експерименту 6

2. Мультиваріантне тестування

Діяльність кожної компанії направлена на отримання прибутку. В умовах, коли значна частина цільової аудиторії здобуває інформацію про продукти та послуги через інтернет і робить покупки онлайн, кожній компанії важливо мати ефективний сайт. З метою збільшення прибутку сайт компанії підлягає постійній оптимізації. Об'єктом оптимізації є коефіцієнт конверсії – процент відвідувачів сайту, які виконали певну цільову дію, що є бажаною для власника сайту.

Одним із способів оптимізації коефіцієнту конверсії є проведення мультиваріантного тестування. На цільовій сторінці сайту обирається декілька елементів та створюються їхні модифікації. Далі утворюються усі можливі комбінації модифікацій цих елементів, тим самим створюються варіації сторінки. Потік відвідувачів рівномірно розподіляється між варіаціями цільової сторінки. Кожному відвідувачу пропонується до перегляду одна з таких варіацій та відслідковується поведінка відвідувача: відвідувач виконав цільову дію чи ні.

При мультиваріантному тестуванні одночасно тестуються усі можливі комбінації елементів цільової сторінки, що дозволяє оцінити вплив кожного елементу та їх взаємодію на коефіцієнт конверсії. За результатами тестування обирають варіацію цільової сторінки (оптимальну комбінацію елементів), яка здобула найбільший коефіцієнт конверсії. Зазначимо, що мультиваріантне тестування коштовне та потребує часу.

Математичною основою мультиваріантного тестування є теорія ймовірностей та математична статистика.

Сервіси для мультиваріантного тестування Optimizely, Visual Website Optimizer (VWO), Convert, Adobe Target, Google Optimize & Optimize 360 використовують класичні (частотні), байєсівські, бандитські статистичні методи. Ми розглянемо частотний підхід (Frequentist Approach) до розв'язання задачі мультиваріантного тестування цільової сторінки.

2.1. Термінологія інтернет-маркетингу

Цільова сторінка (Landing Page) – сторінка, на яку переходить відвідувач, залучений за допомогою реклами, систем пошуку, поштової розсилки.

Заклик до дії (Call-to-Action) – елемент цільової сторінки (кнопка, текстове посилання, зображення), який спонукає відвідувача на цільову дію. Зовнішній вигляд елементу має виділятися серед контенту сторінки, оскільки саме він перетворює відвідувача на користувача.

Цільова дія (Conversion Action) – дія відвідувача, яка значуща для власника сайту (скачування безкоштовного контенту, реєстрація, підписка на розсилку, покупка товару, заказ послуги).

Унікальний відвідувач (Unique Visitor) – відвідувач з унікальними характеристиками (IP-адреса, браузер, реєстраційні дані), який зайшов на цільову сторінку протягом деякого проміжку часу (доба/тиждень/місяць).

Користувач (User) – відвідувач, який взаємодіє з цільовою сторінкою.

Оригінальна сторінка (Baseline) – базовий варіант цільової сторінки.

Варіація (Variation) – альтернативний варіант цільової сторінки.

Конверсія (Conversion) – цільова дія, яку відвідувач виконав на сторінці сайту.

Коефіцієнт конверсії (Conversion Rate) – відношення числа конверсій до загальної кількості відвідувачів.

Оптимізація цільової сторінки (Landing Page Optimization) – процес вдосконалення елементів цільової сторінки з метою максимізації коефіцієнту конверсії.

A/B тестування (A/B testing) – метод оптимізації оригінальної цільової сторінки. Оригінальна сторінка та її варіація мають бути однакові за виключенням одного елементу, вплив якого на сприйняття відвідувачів перевіряється під час тестування.

2.2. Постановка задачі

Нехай в m групах відвідувачів проводяться незалежні випробування Бернуллі з імовірністю успіху $p_1, \dots, p_m$ в кожній групі в одному випробуванні. Імовірності успіху $p_1, \dots, p_m$ невідомі. Нехай перша група є базовою (baseline). Відносно невідомих параметрів висуваються нульові гіпотези $H_0^{ij}: p_i = p_j,\ i = 1,\ j = 2, \dots, m$ – базова та j–та групи мають однакову ймовірність успіху (базова група та j–та група мають однакові коефіцієнти конверсії). Альтернативні гіпотези двосторонні: $H_0^{ij}: p_i \neq p_j$. Необхідно за реалізаціями вибірок з розподілів Бернуллі з параметрами $p_1, \dots, p_m$ дійти висновку – відхиляти $H_0^{ij}: p_i = p_j$ чи ні.

2.3. Z–критерій для двох часток

Частота успіху $\hat{p}_i = \mu_i / n_i$ є незміщеною, спроможною оцінкою для параметра p_i базової групи, частота успіху $\hat{p}_j = \mu_j / n_j$ є незміщеною, спроможною оцінкою для параметра p_j $j-$ тої групи. Оцінка

$$\hat{p} = \frac{\hat{p}_i n_i + \hat{p}_j n_j}{n_i + n_j}$$

виступає об'єднаною частотою успіху в двох групах.

Для перевірки нульових гіпотез скористаємося z–критерієм. Якщо гіпотезу $H_0^{ij}:\ p_i = p_j$ відхиляти при

$$Z = \frac{|\,\hat{p}_i - \hat{p}_j|}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \geq z_{1-\alpha/2},$$

і не відхиляти в супротивному разі, то з імовірністю α гіпотеза буде відхилятися, коли вона справедлива (альтернатива двостороння $H_0^{ij}: p_i \neq p_j$), $z_{1-\alpha/2}$ – $(1-\alpha/2)-$ квантиль стандартного нормального розподілу.

Або, що те ж саме, гіпотеза H_0^{ij}: $p_i = p_j$ відхиляється, якщо

$$\hat{p}_i - \hat{p}_j \notin \left(- z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}, z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \right)$$

і не відхиляється в супротивному разі.

Або, що те ж саме, за розподілом статистики Z обчислюється рівень значущості, що досягається:

$$p_{ij} = P\{Z \geq z | H_0^{ij}\} = 2\left(1 - N_{0;1}(z)\right).$$

Якщо $p_{ij} \leq \alpha$, то нульова гіпотеза $H_0^{ij}: p_i = p_j$ відхиляється і не відхиляється в супротивному разі.

2.4. Число унікальних відвідувачів, необхідне для тестування

Про розподіл статистики Z. Якщо нульова гіпотеза $H_0^{ij}: p_i = p_j$ справедлива, то статистика Z має стандартний нормальний розподіл при великих обсягах вибірок n_i та n_j:

$$Z = \frac{\hat{p}_i - \hat{p}_j}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} = \frac{\hat{p}_i - \hat{p}_j}{SE}.$$

Якщо альтернативна гіпотеза $H_1^{ij}: p_i - p_j = \theta$ справедлива, то статистика Z_1 має стандартний нормальний розподіл при великих обсягах вибірок n_i та n_j:

$$Z_1 = \frac{\hat{p}_i - \hat{p}_j - \theta}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} = \frac{\hat{p}_i - \hat{p}_j - \theta}{SE}.$$

Потужність критерію. Потужністю критерію називають імовірність відхилити нульову гіпотезу, коли справедлива альтернативна гіпотеза.

$$P\{S|H_1\} = P\{|\hat{p}_i - \hat{p}_j| \geq z_{1-\alpha/2} SE | H_1\} =$$

$$= P\{\hat{p}_i - \hat{p}_j \geq z_{1-\alpha/2}SE | H_1\} + P\{\hat{p}_i - \hat{p}_j \leq -z_{1-\alpha/2}SE | H_1\} =$$

$$= P\{\hat{p}_i - \hat{p}_j - \theta \geq z_{1-\alpha/2}SE - \theta\} + P\{\hat{p}_i - \hat{p}_j - \theta \leq -z_{1-\alpha/2}SE - \theta\} =$$

$$= P\left\{\frac{\hat{p}_i - \hat{p}_j - \theta}{SE} \geq z_{1-\alpha/2} - \frac{\theta}{SE}\right\} + P\left\{\frac{\hat{p}_i - \hat{p}_j - \theta}{SE} \leq -z_{1-\alpha/2} - \frac{\theta}{SE}\right\} =$$

$$= P\left\{Z_1 \geq z_{1-\alpha/2} - \frac{\theta}{SE}\right\} + P\left\{Z_1 \leq -z_{1-\alpha/2} - \frac{\theta}{SE}\right\}.$$

Нехай потужність критерію відома та дорівнює $1-\beta, \beta > 0$. Тоді відкидаючи останній доданок отримуємо наближено:

$$P\left\{Z_1 \geq z_{1-\alpha/2} - \frac{\theta}{SE}\right\} \approx 1-\beta,$$

$$1 - P\left\{Z_1 < z_{1-\alpha/2} - \frac{\theta}{SE}\right\} \approx 1-\beta,$$

$$P\left\{Z_1 < z_{1-\alpha/2} - \frac{\theta}{SE}\right\} \approx \beta,$$

За означенням функції розподілу маємо

$$N_{0;1}\left(z_{1-\alpha/2} - \frac{\theta}{SE}\right) \approx \beta.$$

Позначимо через

$$z_\beta \approx z_{1-\frac{\alpha}{2}} - \frac{\theta}{SE}.$$

За означенням z_β є β – квантилем $N_{0;1}$ розподілу. Звідси

$$z_{1-\alpha/2} - z_\beta \approx \frac{\theta}{SE}.$$

Припустимо, що обсяги вибірок в базовій групі та *j*–тій групі рівні, тобто $n_i = n_j = n$, тоді

$$z_{1-\alpha/2} - z_\beta \approx \frac{\theta}{\sqrt{\frac{2\hat{p}(1-\hat{p})}{n}}}.$$

Враховуючи, що

$$-z_\beta = z_{1-\beta},$$

знаходимо мінімальний обсяг вибірки базової групи (або j −тої групи), який забезпечує задані ймовірність помилки першого роду α, потужність

критерію $1-\beta$, різницю коефіцієнтів конверсії θ та базовий коефіцієнт конверсії $\hat{p}$ при перевірці простої гіпотези H_0^{ij}: $p_i = p_j$ проти простої альтернативи H_1^{ij}: $p_i \neq p_j$:

$$n \approx \frac{2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \hat{p}(1-\hat{p})}{\theta^2}.$$

Число унікальних відвідувачів, необхідне для проведення тестування, дорівнює mn.

2.5. Реалізація A/B тестування

Розглянемо частинний випадок мультиваріантного тестування – A/B-тестування. При A/B тестуванні першій (базовій) групі відвідувачів пропонують до перегляду сторінку A, другій (експериментальній) групі відвідувачів пропонують до перегляду сторінку B (варіацію сторінки A). Для збереження чистоти експерименту необхідно ідентифікувати відвідувачів під час A/B тестування та при повторних відвідуваннях сайту пропонувати їм до перегляду ту саму сторінку A або B, яку вони бачили попереднього разу.

Висувається нульова гіпотеза про однаковий рівень конверсії в обох групах H_0^{12}: $p_1 = p_2$. Моделюється потік відвідувачів. Кожен відвідувач з імовірністю ½ може потрапити до першої групи й з імовірністю ½ може потрапити до другої групи. Після того, як відвідувач опинився в одній із двох груп, моделюється його поведінка. Поведінка відвідувача однозначно визначається двома подіями: успіх – відвідувач виконав цільову дію, неуспіх – відвідувач не виконав цільову дію. Якщо відвідувач опинився в першій групі, то подія успіх відбувається з імовірністю p_1. Якщо відвідувач опинився в другій групі, то подія успіх відбувається з імовірністю p_2. Для перевірки гіпотези H_0^{12}: $p_1 = p_2$ скористаємося z–критерієм для двох часток.

Спочатку проведемо A/A тестування, тобто порівняємо одну й ту ж цільову сторінку з метою перевірки точності інструменту тестування.

На рисунку 2.5.1 зображено результат реалізації A/A тестування при параметрах $\alpha = 0{,}05$, $\beta = 0{,}2$, $\hat{p} = 0{,}1$. Знаходження траєкторії в межах довірчого інтервалу означає невідхилення нульової гіпотези H_0^{12}: $p_1 = p_2$. Коефіцієнти конверсії в першій та другій групах однакові.

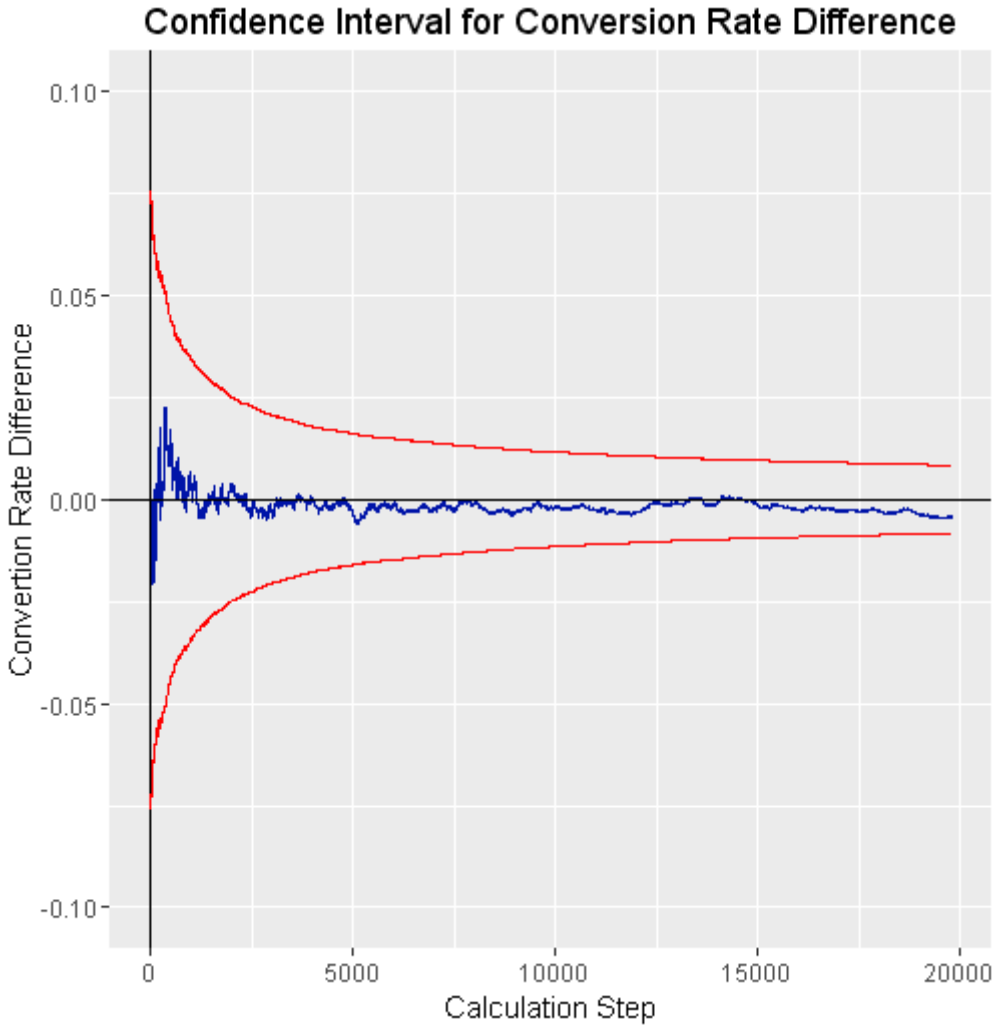


Рис. 2.5.1. Довірчий інтервал для різниці коефіцієнтів конверсії

На рисунку 2.5.2 зображено результат реалізації A/A тестування при параметрах $\alpha = 0{,}05$, $\beta = 0{,}2$, $\hat{p} = 0{,}1$. Перетин траєкторією межі довірчого інтервалу – помилка першого роду (нульова гіпотеза H_0^{12}: $p_1 = p_2$ відхиляється, коли вона справедлива).

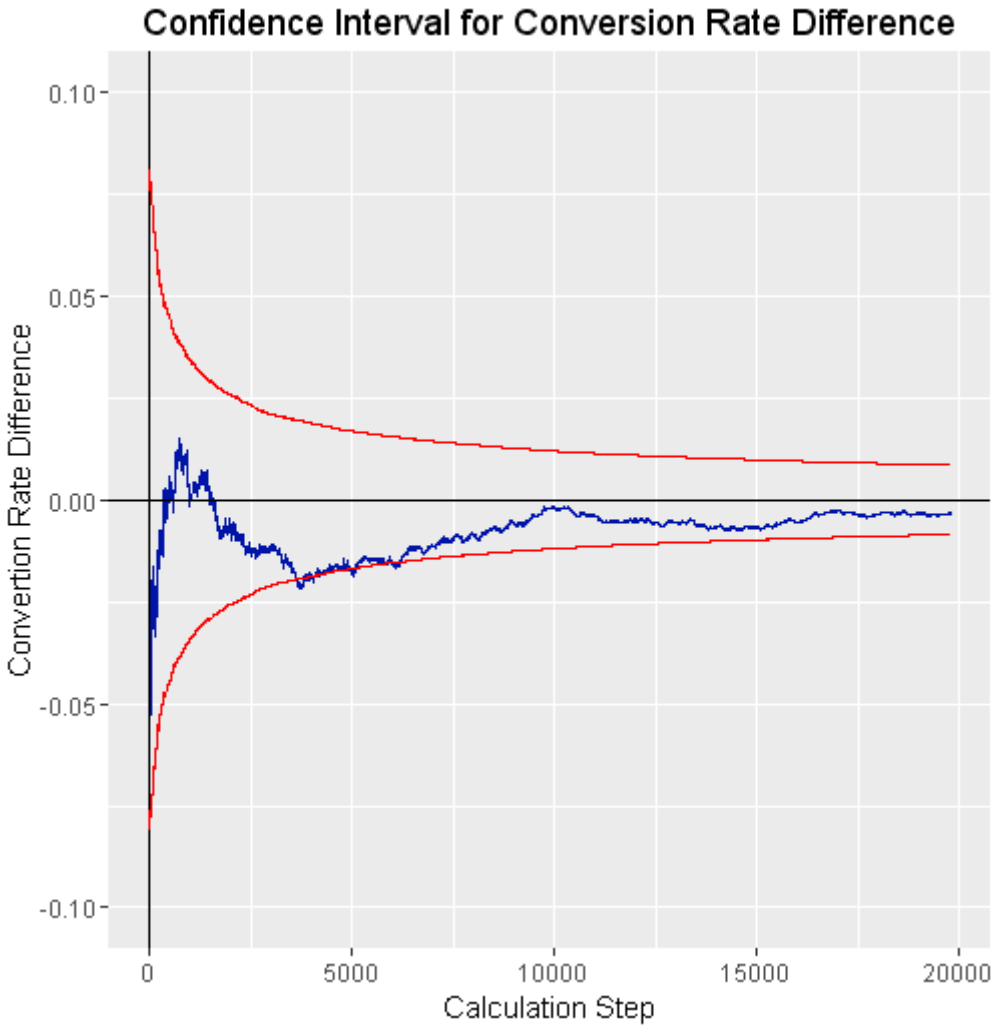


Рис. 2.5.2. Довірчий інтервал для різниці коефіцієнтів конверсії

На рисунку 2.5.3 зображено результат реалізації A/B тестування при параметрах $\alpha = 0{,}05$, $\beta = 0{,}2$, $\hat{p} = 0{,}1$ $\theta = 0{,}02$. Мінімальне число унікальних відвідувачів базової групи дорівнює

$$n \approx \frac{2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \hat{p}(1-\hat{p})}{\theta^2} = \frac{2(1{,}96 + 0{,}84)^2 \cdot 0{,}1 \cdot 0{,}9}{0{,}02^2} = 3532.$$

Знаходження траєкторії поза межами довірчого інтервалу при мінімальному числі унікальних відвідувачів $n = 3532$ в одній із груп означає відхилення гіпотези H_0^{12}: $p_1 = p_2$. Коефіцієнти конверсії в першій та другій групах різні.

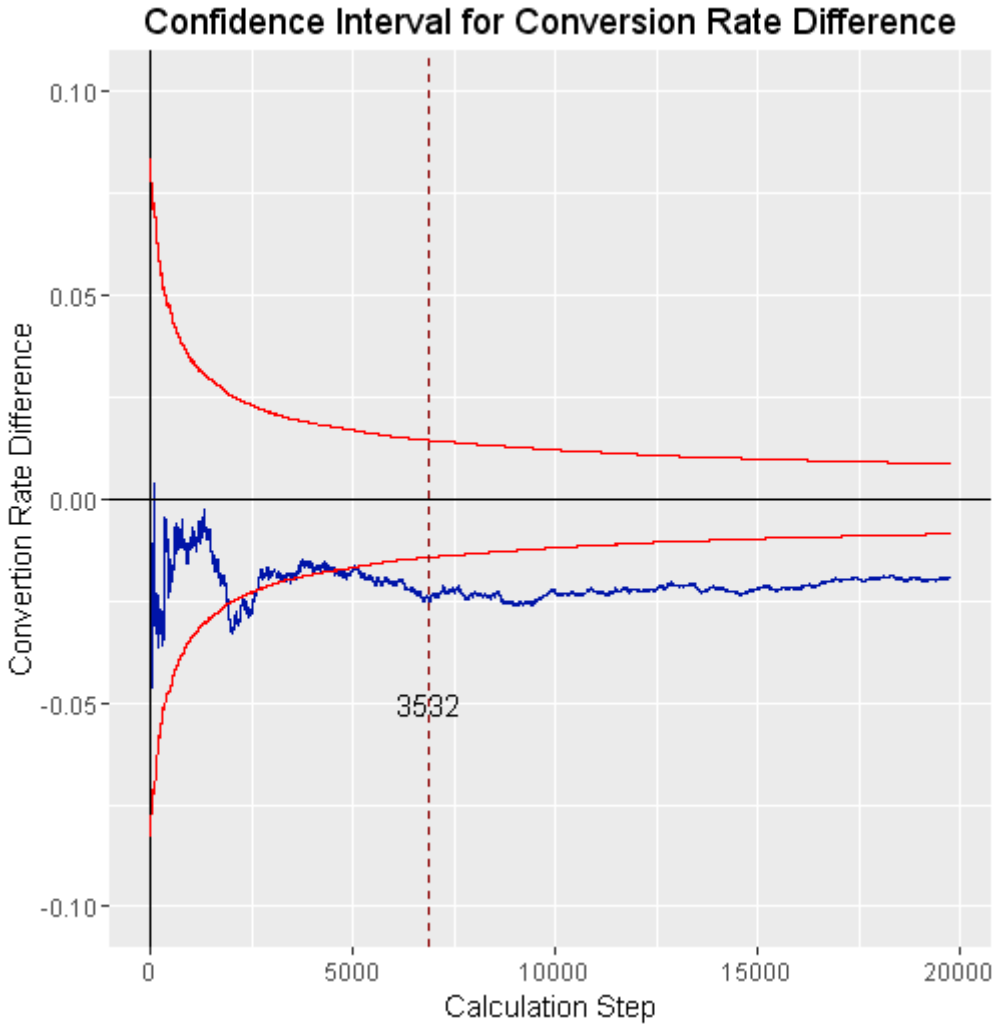


Рис. 2.5.3. Довірчий інтервал для різниці коефіцієнтів конверсії

2.6. Реалізація мультиваріантного тестування

При мультиваріантному тестуванні одночасно тестуються усі можливі комбінації елементів сторінки та оцінюється вплив кожного елементу та їх взаємодія на рівень конверсії. Групам відвідувачів для перегляду пропонують сторінки $A_1, A_2, \ldots, A_m$. Для збереження чистоти експерименту необхідно ідентифікувати відвідувачів під час тестування та при повторних відвідуваннях сайту пропонувати їм до перегляду ту саму сторінку, яку вони бачили попереднього разу.

Висувається $m-1$ нульова гіпотеза про однаковий рівень конверсії в базовій та j–тій групах, $j = 2, \ldots, m$. Моделюється потік відвідувачів

сайту. Кожен відвідувач з імовірністю $1/m$ може потрапити на одну зі сторінок. Після того, як відвідувач опинився на одній із сторінок, моделюється його поведінка. Поведінка відвідувача однозначно визначається двома подіями: успіх – відвідувач виконав цільову дію, неуспіх – відвідувач не виконав цільову дію. Якщо відвідувачу запропоновано для перегляду сторінка A_i, то подія успіх відбувається з імовірністю p_i.

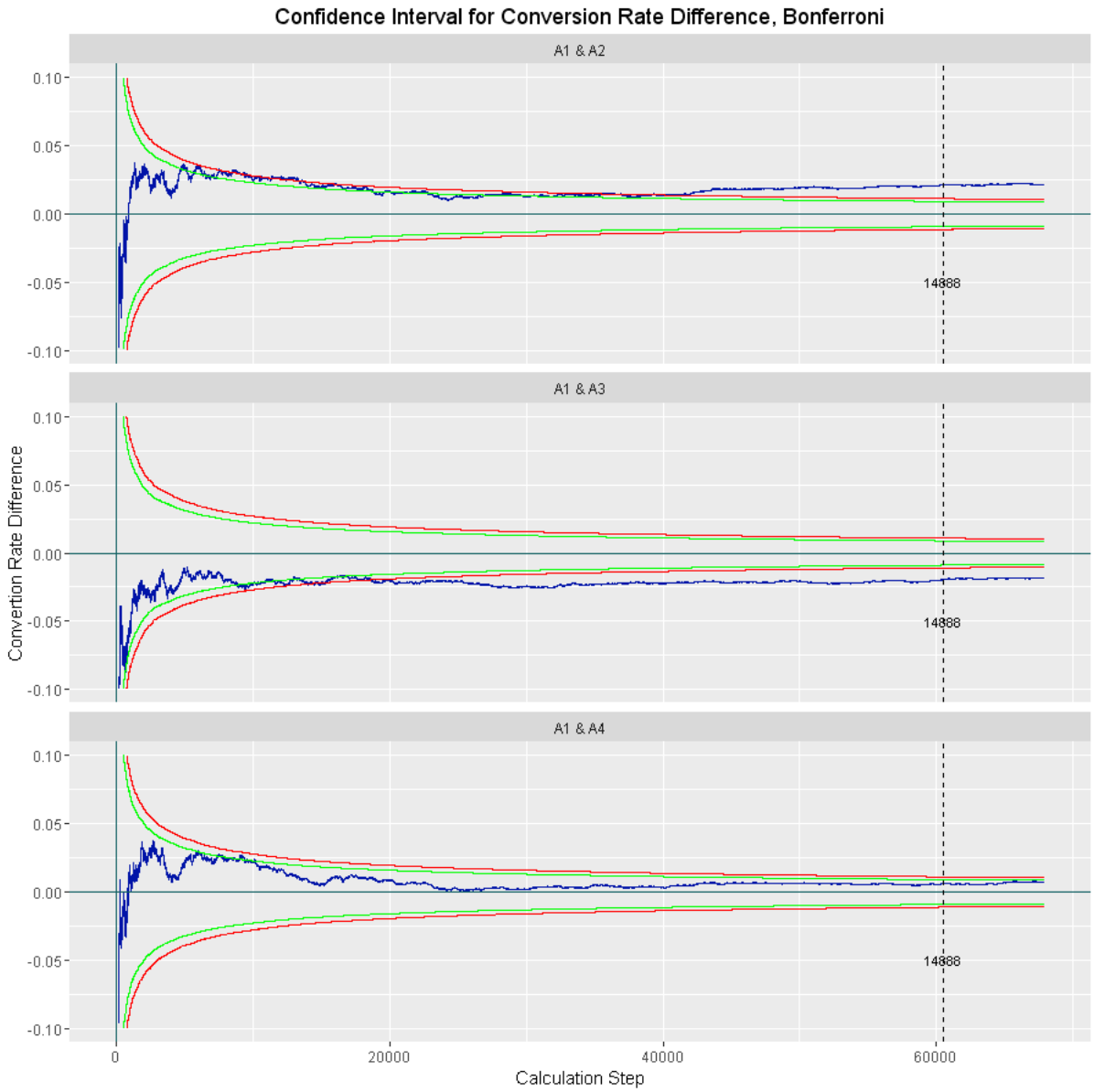
При перевірці однієї нульової гіпотези H_0^{ij}: $p_i = p_j$ ймовірність помилки першого роду обмежена зверху рівнем значущості α. Тоді при одночасній перевірці $m-1$ нульових гіпотез H_0^{ij}: $p_i = p_j$ імовірність припустити принаймні одну помилку першого роду обмежена зверху величиною $1-(1-\alpha)^{m-1}$, яка стає досить великою при достатньо малих m. Для усунення цього ефекту – ефекту множинних порівнянь – в розділі 1 розглянута низка статистичних процедур, які різняться своєю потужністю та умовами застосування в різних ситуаціях.

На рис. 2.6.1 зображено реалізацію мультиваріантного тестування при таких параметрах: $\beta = 0{,}2$, $\hat{p} = 0{,}2$, $\theta = 0{,}015$. Для контролю $FWER$ використано метод Бонферроні ($\alpha_i = 0{,}05/3,\ i = 1{,}2{,}3$). Мінімальне число унікальних відвідувачів базової групи є розв'язком рівняння:

$$\frac{1}{3}\sum_{i=1}^{3}\left(1 - N_{0;1}\left(z_{1-\alpha_i/2} - \frac{\theta}{SE}\right)\right) = 1-\beta.$$

Знаходження різниці коефіцієнтів конверсії поза межами довірчого інтервалу при числі відвідувачів $n = 14888$ означає відхилення гіпотези H_0^{ij}. Довірчі інтервали з поправкою Бонферроні зображено червоним кольором. Довірчі інтервали без поправки зображено зеленим кольором.

Метод Бонферроні – універсальний метод, він не залежить від характеру гіпотез, що перевіряються, та їх взаємозв'язків. Але цей метод має суттєвий недолік: потужність методу знижується, якщо число статистичних гіпотез, що перевіряються, збільшується.

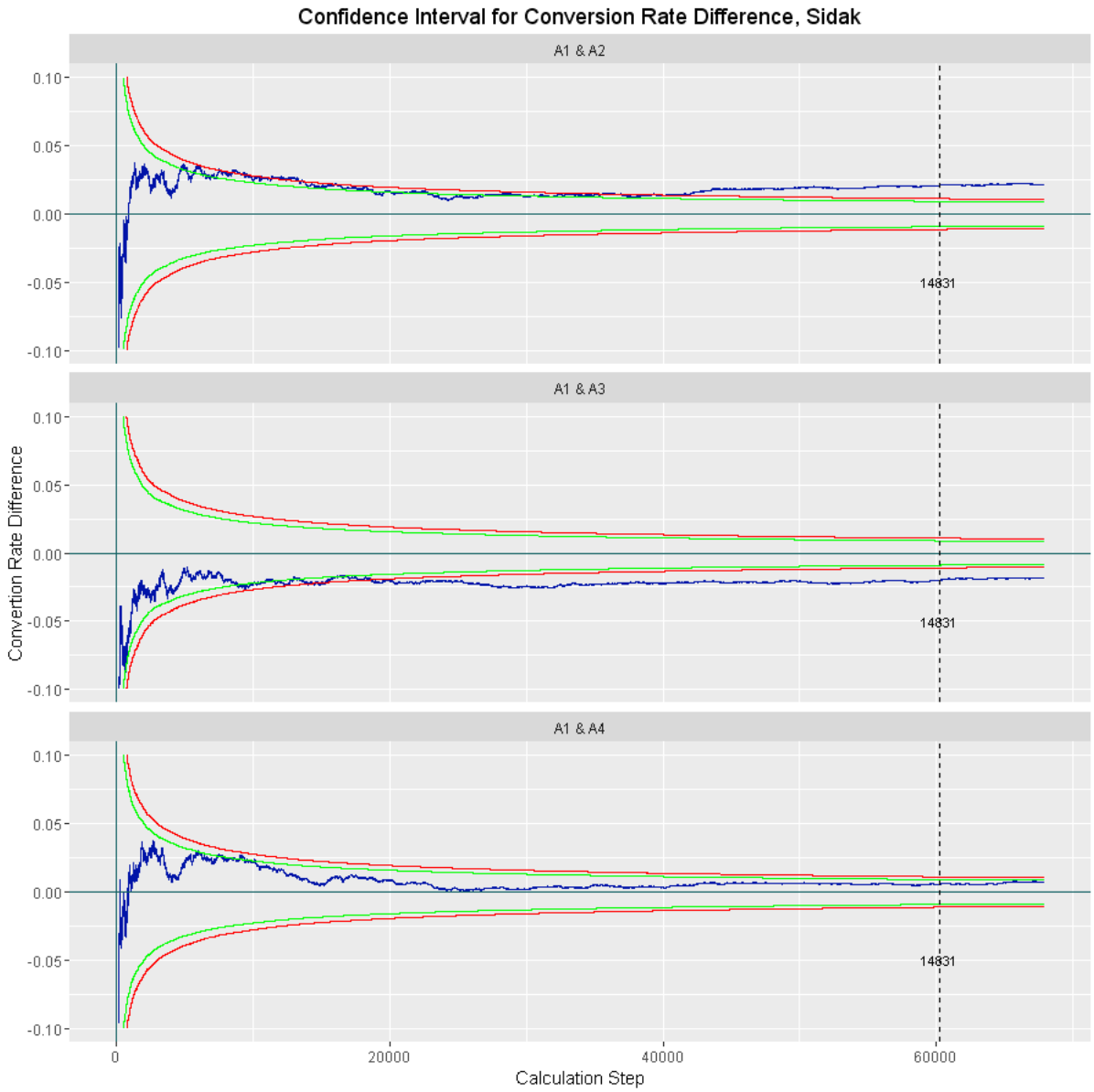


2.6.1. Довірчі інтервали для різниці коефіцієнтів конверсії

На рис. 2.6.2 зображено реалізацію мультиваріантного тестування при таких параметрах: $\beta = 0{,}2$, $\hat{p} = 0{,}2$, $\theta = 0{,}015$. Для контролю $FWER$ використано метод Шидака ($\alpha_i = 1 - (1 - 0{,}05)^{1/3}$, $i = 1{,}2{,}3$). Мінімальне число унікальних відвідувачів базової групи є розв'язком рівняння:

$$\frac{1}{3}\sum_{i=1}^{3}\left(1 - N_{0;1}\left(z_{1-\alpha_i/2} - \frac{\theta}{SE}\right)\right) = 1 - \beta.$$

Знаходження різниці коефіцієнтів конверсії поза межами довірчого інтервалу при числі відвідувачів $n = 14831$ означає відхилення гіпотези H_0^{ij}. Довірчі інтервали з поправкою Шидака зображено червоним кольором. Довірчі інтервали без поправки зображено зеленим кольором.



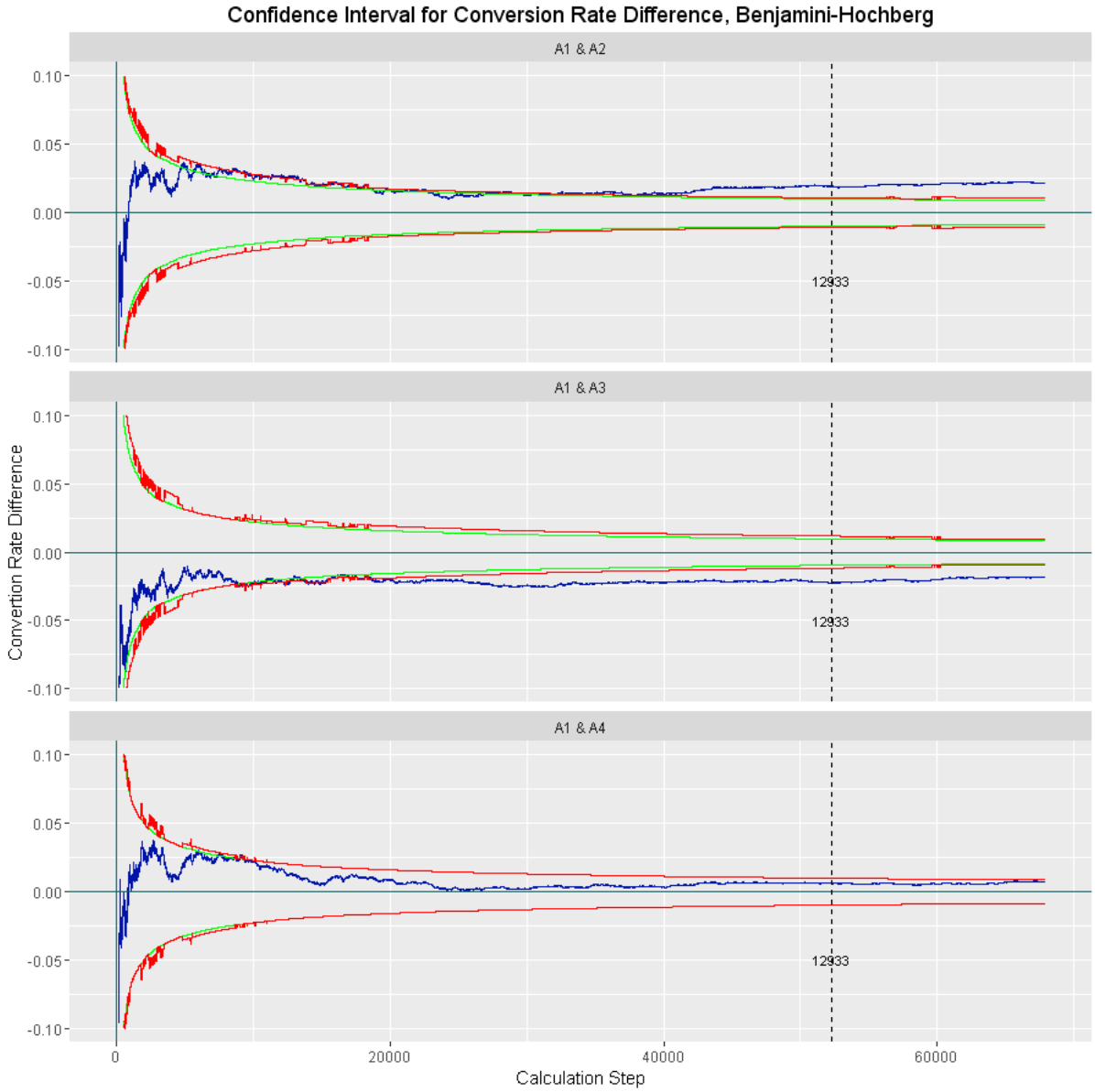
2.6.2. Довірчі інтервали для різниці коефіцієнтів конверсії

На рис. 2.6.3 зображено реалізацію мультиваріантного тестування при таких параметрах: $\beta = 0{,}2$, $\hat{p} = 0{,}2$, $\theta = 0{,}015$. Для контролю False Discovery Rate було використано поправку Бенджаміні-Хохберга ($\alpha_1 = 0{,}05/3$, $\alpha_2 = 2 \cdot 0{,}05/3$, $\alpha_3 = 0{,}05$). Мінімальне число унікальних відвідувачів базової групи є розв'язком рівняння:

$$\frac{1}{3}\sum_{i=1}^{3}\left(1 - N_{0;1}\left(z_{1-\alpha_i/2} - \frac{\theta}{SE}\right)\right) = 1 - \beta.$$

Знаходження різниці коефіцієнтів конверсії поза межами довірчого інтервалу при числі відвідувачів $n = 12933$ означає відхилення гіпотези

H_0^{ij}. Довірчі інтервали з поправкою Бенджаміні-Хохберга зображено червоним кольором. Довірчі інтервали без поправки зображено зеленим кольором. Довірчі інтервали не є монотонними, це пов'язано зі статистичною процедурою методу.



2.6.3. Довірчі інтервали для різниці коефіцієнтів конверсії

Результати тестування при параметрах $\beta = 0{,}2$, $\hat{p} = 0{,}2$, $\theta = 0{,}015$ наведено в таблиці 2.6.1. Статистична значущість Statistical Significance – це імовірність того, що різниця в коефіцієнтах конверсії між базовою та j–тою варіацією цільової сторінки не випадкова.

Переможцем є варіація цільової сторінки обрана згідно з метрикою Improvement (покращення – це відносна різниця у вдосконаленні варіації цільової сторінки відносно базового варіанту цільової сторінки):

$$Improvement = \frac{Variation\ Conv.Rate\ \% - Baseline\ Conv.Rate\ \%}{Baseline\ Conv.Rate\ \%}$$

Таблиця 2.6.1. Результати мультиваріантного тестування

Hypothesis	Conv.Rate	Improvement	Multiple Comparison Adjustment Method		
			Bonferroni	Šidák	Benjamini Hochberg
			Statistical Significance		
H_0^{12}	$\hat{p}_1 = 0{,}2068$ $\hat{p}_2 = 0{,}2249$	8,75 %	0,9972	-	-
H_0^{13}	$\hat{p}_1 = 0{,}2068$ $\hat{p}_3 = 0{,}1852$	-10,44 %	0,9998	-	-
H_0^{14}	$\hat{p}_1 = 0{,}2068$ $\hat{p}_4 = 0{,}2132$	3,09 %	0,3059	-	-
H_0^{12}	$\hat{p}_1 = 0{,}2068$ $\hat{p}_2 = 0{,}2249$	8,75 %	-	0,9998	-
H_0^{13}	$\hat{p}_1 = 0{,}2068$ $\hat{p}_3 = 0{,}1852$	-10,44 %	-	0,9973	-
H_0^{14}	$\hat{p}_1 = 0{,}2068$ $\hat{p}_4 = 0{,}2132$	3,09 %	-	0,4541	-
H_0^{12}	$\hat{p}_1 = 0{,}2068$ $\hat{p}_2 = 0{,}2249$	8,75 %	-	-	0,9991
H_0^{13}	$\hat{p}_1 = 0{,}2068$ $\hat{p}_3 = 0{,}1852$	-10,44 %	-	-	0,9999
H_0^{14}	$\hat{p}_1 = 0{,}2068$ $\hat{p}_4 = 0{,}2132$	3,09 %	-	-	0,7686

Переможцем стала варіація цільової сторінки з коефіцієнтом конверсії 22,49% та покращенням 8,75%. Множинні порівняння із застосуванням методу Бенджаміні-Хохберга призводять до найменших допустимих помилок при відхиленні нульових гіпотез H_0^{ij}, а саме 0,0009; 0,0001 для гіпотез H_0^{12}, H_0^{13} відповідно.

Додаток 1. Модельний експеримент

```
m = 200; alpha = 0.05
df<-1:20
for (i in 1:150){df<-cbind(df, rnorm(20,mean=0,sd=1))}
for (i in 1:50){df<-cbind(df, rnorm(20,mean=1,sd=1))}
df<-as.data.frame(df)

V<-0; U<-0; p.value.1<-0

for (i in 2:151){
      t_test<-t.test(df[i], alternative="two.sided", mu=0)
      if (t_test$p.value<alpha) {V<-V+1} else {U<-U+1}
      p.value.1[i-1]<-t_test$p.value
}

T<-0; S<-0; p.value.2<-0

for (i in 152:201){
  t_test<-t.test(df[i], alternative="two.sided", mu=0)
      if (t_test$p.value<alpha) {S<-S+1} else {T<-T+1}
      p.value.2[i-151]<-t_test$p.value
}

m0<-U+V; R<-V+S
without_correction<-matrix(c(U, V, m0, T, S, m-m0, m-R, R, m), nrow=3,
                           ncol=3, byrow=FALSE)

dimnames(without_correction)<-list(c("# невідхилених Hi", "# відхилених
Hi", "Разом"), c("# справедливих Hi", "# несправедливих Hi", "Разом"))

print(without_correction)

df1<-as.data.frame(rbind(cbind(p.value.1, 0),cbind(p.value.2, 1)))
colnames(df1)<-c("pvalue", "Hi")

df1$Hi<-factor(df1$Hi)
df1$color[df1$Hi==0]<-"blue"
df1$color[df1$Hi==1]<-"red"
df1<-df1[order(df1$pvalue),]

library(ggplot2)
ggplot(df1, aes(x=1:200, y=pvalue, color=Hi))+geom_point(size=1.5)+
  scale_x_continuous(name = "i", limits = c(0, 200))+
  scale_y_continuous(name = "p-value", limits = c(0, 1))+
  scale_color_manual(values = c("blue", "red"),
                     labels = c("true hypothesis", "false hypothesis"))+
  geom_hline(linetype="longdash", yintercept=0.05, color = "red", size=1)+
  geom_text(aes(x=0, y=0.08), label=c(as.character(0.05)), color="black")+
  theme(legend.title = element_blank(),legend.position = c(0.84, 0.2))+
  ggtitle("P-values in ascending order")
```

Додаток 2. Статистична процедура методу Бонферроні

```
# bonferroni correction

m = 200; alpha = 0.05
df<-1:20
for (i in 1:150){df<-cbind(df, rnorm(20,mean=0,sd=1))}
for (i in 1:50){df<-cbind(df, rnorm(20,mean=1,sd=1))}
df<-as.data.frame(df)

V<-0; U<-0; p.value.1<-0
for (i in 2:151){
      t_test<-t.test(df[i], alternative="two.sided", conf.level = 0.95)
      if (t_test$p.value<alpha/m) {V<-V+1} else {U<-U+1}
      p.value.1[i-1]<-min(1,t_test$p.value*m)
}

T<-0; S<-0; p.value.2<-0

for (i in 152:201){
      t_test<-t.test(df[i], alternative="two.sided", conf.level = 0.95)
      if (t_test$p.value<alpha/m) {S<-S+1} else {T<-T+1}
      p.value.2[i-151]<-min(1,t_test$p.value*m)
}

m0<-U+V; R<-V+S

with_correction<-matrix(c(U, V, m0, T, S, m-m0, m-R, R, m), nrow=3,
                            ncol=3, byrow=FALSE)

dimnames(with_correction)<-list(c("# невідхилених Hi", "# відхилених Hi",
"Разом"), c("# справедливих Hi", "# несправедливих Hi", "Разом"))

print(with_correction)

df1<-as.data.frame(rbind(cbind(p.value.1, 0),cbind(p.value.2, 1)))
colnames(df1)<-c("pvalue", "Hi")

df1$Hi<-factor(df1$Hi)
df1$color[df1$Hi==0]<-"blue"
df1$color[df1$Hi==1]<-"red"
df1<-df1[order(df1$pvalue),]

library(ggplot2)
ggplot(df1, aes(x=1:200, y=pvalue, color = Hi))+
  geom_point(size = 1.5)+
  scale_x_continuous(name = "i", limits = c(0, 200))+
  scale_y_continuous(name = "p-value", limits = c(0, 1))+
  scale_color_manual(values = c("blue", "red"),
                     labels = c("true hypothesis", "false hypothesis"))+
  geom_hline(linetype="longdash", yintercept=0.05, color="red", size=1)+
  geom_text(aes(x=0, y=0.08), label=c(as.character(0.05)), color="black")+
  theme(legend.title = element_blank(),legend.position = c(0.84, 0.2))+
  ggtitle("P-values in ascending order")
```

Додаток 3. Статистична процедура методу Шидака

```
# sidak correction

m = 200; alpha = 0.05
df<-1:20
for (i in 1:150){df<-cbind(df, rnorm(20,mean=0,sd=1))}
for (i in 1:50){df<-cbind(df, rnorm(20,mean=1,sd=1))}
df<-as.data.frame(df)

V<-0; U<-0; p.value.1<-0
for (i in 2:151){
      t_test<-t.test(df[i], alternative="two.sided", conf.level = 0.95)
      if (t_test$p.value<1-(1-alpha)^(1/m)) {V<-V+1} else {U<-U+1}
      p.value.1[i-1]<-1-(1-t_test$p.value)^m
}

T<-0; S<-0; p.value.2<-0

for (i in 152:201){
      t_test<-t.test(df[i], alternative="two.sided", conf.level = 0.95)
      if (t_test$p.value<1-(1-alpha)^(1/m)) {S<-S+1} else {T<-T+1}
      p.value.2[i-151]<-1-(1-t_test$p.value)^m
}

m0<-U+V; R<-V+S

with_correction<-matrix(c(U, V, m0, T, S, m-m0, m-R, R, m), nrow=3,
                            ncol=3, byrow=FALSE)

dimnames(with_correction)<-list(c("# невідхилених Hi", "# відхилених Hi",
"Разом"), c("# справедливих Hi", "# несправедливих Hi", "Разом"))
print(with_correction)

df1<-as.data.frame(rbind(cbind(p.value.1, 0),cbind(p.value.2, 1)))
colnames(df1)<-c("pvalue", "Hi")

df1$Hi<-factor(df1$Hi)
df1$color[df1$Hi==0]<-"blue"
df1$color[df1$Hi==1]<-"red"
df1<-df1[order(df1$pvalue),]

library(ggplot2)
ggplot(df1, aes(x=1:200, y=pvalue, color=Hi))+geom_point(size=1.5)+
  scale_x_continuous(name = "i", limits = c(0, 200))+
  scale_y_continuous(name = "p-value", limits = c(0, 1))+
  scale_color_manual(values = c("blue", "red"),
                     labels = c("true hypothesis", "false hypothesis"))+
  geom_hline(linetype="dotted", yintercept=0.05, color="darkred", size=1)+
  geom_text(aes(x=0, y=0.08), label=c(as.character(0.05)), color="black")+
  theme(legend.title = element_blank(),legend.position = c(0.84, 0.2))+
  ggtitle("P-values in ascending order")
```

Додаток 4. Статистична процедура методу Холма

```
# holm correction

m = 200; alpha = 0.05
df<-1:20
for (i in 1:150){df<-cbind(df, rnorm(20,mean=0,sd=1))}
for (i in 1:50){df<-cbind(df, rnorm(20,mean=1,sd=1))}
df<-as.data.frame(df)

p.value.1<-0
for (i in 2:151){
      t_test<-t.test(df[i], alternative="two.sided", conf.level = 0.95)
      p.value.1[i-1]<-t_test$p.value
}
p.value.2<-0
for (i in 152:201){
      t_test<-t.test(df[i], alternative="two.sided", conf.level = 0.95)
      p.value.2[i-151]<-t_test$p.value
}
df1<-as.data.frame(rbind(cbind(p.value.1, 0),cbind(p.value.2, 1)))
colnames(df1)<-c("pvalue", "Hi")
df1$Hi<-factor(df1$Hi)
df1$color[df1$Hi==0]<-"blue"
df1$color[df1$Hi==1]<-"red"
df1<-df1[order(df1$pvalue),]

U<-150; V<-0; T<-50; S<-0

for (i in 1:200){if (df1[i,1]>=alpha/(m-i+1)){break}
            else {if (df1[i,2]==0){U<-U-1; V<-V+1} else {T<-T-1; S<-S+1}
            }
}
m0<-U+V; R<-V+S
with_correction<-matrix(c(U, V, m0, T, S, m-m0, m-R, R, m), nrow=3,
                            ncol=3, byrow=FALSE)

dimnames(with_correction)<-list(c("# невідхилених Hi", "# відхилених Hi",
"Разом"), c("# справедливих Hi", "# несправедливих Hi", "Разом"))
print(with_correction)

for (i in 1:200){df1[i,4]<-min(1,(m-i+1)*df1[i,1])}
colnames(df1)<-c("pvalue", "Hi", "color", "pvalue_modify")

library(ggplot2)
ggplot(df1, aes(x=1:200, y=pvalue_modify, color=Hi))+
geom_point(size=1.5)+
  scale_x_continuous(name = "i", limits = c(0, 200))+
  scale_y_continuous(name = "p-value", limits = c(0, 1))+
  scale_color_manual(values = c("blue", "red"),
                     labels = c("true hypothesis", "false hypothesis"))+
  geom_hline(linetype="longdash",yintercept=0.05, color="red", size=1)+
  geom_text(aes(x=0, y=0.08), label=c(as.character(0.05)), color="black")+
  theme(legend.title = element_blank(),legend.position = c(0.84, 0.2))+
  ggtitle("P-values in ascending order")
```

Додаток 5. Статистична процедура методу Бенджаміні-Хохберга

```
# benjamini-hochberg correction

m = 200; alpha = 0.05
df<-1:20
for (i in 1:150){df<-cbind(df, rnorm(20,mean=0,sd=1))}
for (i in 1:50){df<-cbind(df, rnorm(20,mean=1,sd=1))}
df<-as.data.frame(df)

p.value.1<-0
for (i in 2:151){
  t_test<-t.test(df[i], alternative="two.sided", conf.level = 0.95)
  p.value.1[i-1]<-t_test$p.value
}
p.value.2<-0
for (i in 152:201){
  t_test<-t.test(df[i], alternative="two.sided", conf.level = 0.95)
  p.value.2[i-151]<-t_test$p.value
}
df1<-as.data.frame(rbind(cbind(p.value.1, 0),cbind(p.value.2, 1)))
colnames(df1)<-c("pvalue", "Hi")
df1$Hi<-factor(df1$Hi)
df1$color[df1$Hi==0]<-"blue"
df1$color[df1$Hi==1]<-"red"
df1<-df1[order(df1$pvalue),]

U<-0; V<-150; T<-0; S<-50

for (i in 200:1){if (df1[i,1]<=alpha*i/m) {break}
            else {if (df1[i,2]==0) {U<-U+1; V<-V-1}
                            else {T<-T+1; S<-S-1}
                  }
}
m0<-U+V; R<-V+S
with_correction<-matrix(c(U, V, m0, T, S, m-m0, m-R, R, m), nrow=3,
                                ncol=3, byrow=FALSE)
dimnames(with_correction)<-list(c("# невідхилених Hi", "# відхилених Hi",
"Разом"), c("# справедливих Hi", "# несправедливих Hi", "Разом"))
print(with_correction)

for (i in 1:200){df1[i,4]<-min(1,(m/i)*df1[i,1])}
colnames(df1)<-c("pvalue", "Hi", "color", "pvalue_modify")

library(ggplot2)
ggplot(df1, aes(x=1:200, y=pvalue_modify, color = Hi))+
  geom_point(size = 1.5)+
  scale_x_continuous(name = "i", limits = c(0, 200))+
  scale_y_continuous(name = "p-value", limits = c(0, 1))+
  scale_color_manual(values = c("blue", "red"),
                     labels = c("true hypothesis", "false hypothesis"))+
  geom_hline(linetype="longdash",yintercept=0.05, color="red", size=1)+
  geom_text(aes(x=0, y=0.08), label=c(as.character(0.05)), color="black")+
  theme(legend.title = element_blank(),legend.position = c(0.84, 0.2))+
  ggtitle("P-values in ascending order")
```

Додаток 6. Статистична процедура методу Бенджаміні-Хохберга-Іекутіелі

```
# benjamini-hochberg-yekutieli correction

m = 200; alpha = 0.05
df<-1:20
for (i in 1:150){df<-cbind(df, rnorm(20,mean=0,sd=1))}
for (i in 1:50){df<-cbind(df, rnorm(20,mean=1,sd=1))}
df<-as.data.frame(df)

p.value.1<-0
for (i in 2:151){
  t_test<-t.test(df[i], alternative="two.sided", conf.level = 0.95)
  p.value.1[i-1]<-t_test$p.value
}
p.value.2<-0
for (i in 152:201){
  t_test<-t.test(df[i], alternative="two.sided", conf.level = 0.95)
  p.value.2[i-151]<-t_test$p.value
}
df1<-as.data.frame(rbind(cbind(p.value.1, 0),cbind(p.value.2, 1)))
colnames(df1)<-c("pvalue", "Hi")
df1$Hi<-factor(df1$Hi)
df1$color[df1$Hi==0]<-"blue"
df1$color[df1$Hi==1]<-"red"
df1<-df1[order(df1$pvalue),]

U<-0; V<-150; T<-0; S<-50; total<-0

for (j in 1:200){total<-total+1/j}
for (i in 200:1){if (df1[i,1]<=(alpha*i)/(m*total)) {break}
            else {if (df1[i,2]==0) {U<-U+1; V<-V-1} else {T<-T+1; S<-S-1}
                  }
}
m0<-U+V; R<-V+S
with_correction<-matrix(c(U, V, m0, T, S, m-m0, m-R, R, m), nrow=3,
                              ncol=3, byrow=FALSE)
dimnames(with_correction)<-list(c("# невідхилених Hi", "# відхилених Hi",
"Разом"), c("# справедливих Hi", "# несправедливих Hi", "Разом"))
print(with_correction)

df1[200,4]<-min(1,df1[200,1])
for (i in 199:1){df1[i,4]<-min(1,(m/i)*df1[i,1]*total,df1[i+1,4])}
colnames(df1)<-c("pvalue", "Hi", "color", "pvalue_modify")

library(ggplot2)
ggplot(df1, aes(x=1:200, y=pvalue_modify, color=Hi))+geom_point(size=1.5)+
  scale_x_continuous(name = "i", limits = c(0, 200))+
  scale_y_continuous(name = "p-value", limits = c(0, 1))+
  scale_color_manual(values = c("blue", "red"),
                     labels = c("true hypothesis", "false hypothesis"))+
  geom_hline(linetype="dotted",yintercept=0.05, color="darkred", size=1)+
  geom_text(aes(x=0, y=0.08), label=c(as.character(0.05)), color="black")+
  theme(legend.title = element_blank(),legend.position = c(0.84, 0.2))+
  ggtitle("P-values in ascending order")
```

Список рекомендованої літератури

1. F. Bretz, T. Hothorn, P. Westfall, *Multiple Comparisons Using R*, *CRC Press*, 2016.
2. Z. K. Šidák, *Rectangular Confidence Regions for the Means of Multivariate Normal Distributions*, *Journal of the American Statistical Association*, 62(318)(1967), 626-633
3. Y. Benjamini, Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, *Journal of the Royal Statistical Society*, 57(1) (1995), 289–300.
4. Bondarenko I.S., Kravchenko S.V. *On the frequentist approach to multivariate landing page testing* // *Visnyk DNU. Series: Modelling*. — Dnipro: DNU. — 2017. — Issue 9. — No. 8. — pp. 142–151.
5. Евгений Рябенко *Построение выводов по данным*: [Електрон. ресурс] // URL: https://www.coursera.org/learn/stats-for-data-analysis
6. Page Piccinini *A/B testing in R*: [Електрон. ресурс] // URL: https://www.datacamp.com/courses/ab-testing-in-r
7. Ryan Grossman *Customer Analytics & A/B Testing in Python*: [Електрон. ресурс]// URL: https://www.datacamp.com/courses/customer-analytics-ab-testing-in-python

Зміст

Вступ
Розділ 1. Множинна перевірка статистичних гіпотез
- 1.1. Математична постановка задачі
  - 1.1.1. Задача перевірки однієї статистичної гіпотези
  - 1.1.2. Задача множинної перевірки статистичних гіпотез
- 1.2. Метод Бонферроні
  - 1.2.1. Статистична процедура методу
  - 1.2.2. Модельний експеримент
- 1.3. Спадна процедура множинної перевірки гіпотез
  - 1.3.1. Статистична процедура методів
  - 1.3.2. Метод Шидака
  - 1.3.3. Метод Холма
  - 1.3.4. Модельний експеримент
- 1.4. Висхідна процедура множинної перевірки гіпотез
  - 1.4.1. Статистична процедура методів
  - 1.4.2. Метод Бенджаміні-Хохберга
  - 1.4.3. Метод Бенджаміні-Хохберга-Іекутіелі
  - 1.4.4. Модельний експеримент

Розділ 2. Мультиваріантне тестування
- 2.1. Термінологія інтернет-маркетингу
- 2.2. Постановка задачі
- 2.3. *Z*–критерій для двох часток
- 2.4. Число унікальних відвідувачів, необхідне для тестування
- 2.5. Реалізація A/B тестування
- 2.6. Реалізація мультиваріантного тестування

Додаток 1. Модельний експеримент
Додаток 2. Статистична процедура методу Бонферроні
Додаток 3. Статистична процедура методу Шидака
Додаток 4. Статистична процедура методу Холма
Додаток 5. Статистична процедура методу Бенджаміні-Хохберга
Додаток 6. Статистична процедура методу Бенджаміні-Хохберга-Іекутіелі