

4c

УДК 519.233.3:004.415.53

ВИКОРИСТАННЯ ПЕРЕВІРКИ СТАТИСТИЧНИХ ГІПОТЕЗ В ІНФОРМАЦІЙНО-ТЕХНІЧНІЙ СФЕРІ. А/В ТЕСТУВАННЯ ТА ДОЦІЛЬНІСТЬ ЙОГО ЗАСТОСУВАННЯ

І. М. Лях, Ю. Ю. Білак, Л. Я. Данько-Товтин, В. В. Станищевський
ДВНЗ "Ужгородський національний університет"
пл. Народна, 3, Ужгород, 88000, Україна

Досліджено математичний апарат, який використовується для перевірки статистичних гіпотез та А/В тестування, а також здійснено порівняння доступних сервісів для проведення такого типу тестування.

Ключові слова: гіпотеза, тестування, сервіс, вибірка, програмне забезпечення.

Постановка проблеми. Сьогодні цифрова аналітика є невід'ємною частиною життєвого циклу програмного забезпечення, призначеного для використання у всесвітній мережі. Вся інформація про поведінку користувачів, зібрана за допомогою різних програмних продуктів та сервісів, дає змогу впроваджувати нові ідеї з апеляцією до статистично обґрунтованих даних.

Використання математичного апарату та статистично достовірних відомостей сприяє зменшенню ризику, дає можливість будувати процес розроблення програмного забезпечення, мінімізувати фінансові та людські ресурси, необхідні для впровадження, розвитку та ітеративно-прогресивної підтримки бізнес-ідей та програмних продуктів.

Перевірка та спростування статистичних гіпотез та А/В тестування є потужним інструментом, який керує онлайн бізнесом та веб-розробкою, проте за умови неправильного трактування чи недостатніх результатів може зашкодити.

Мета статті — дослідження математичного апарату, який використовується для перевірки статистичних гіпотез та А/В тестування, аналіз та порівняння доступних сервісів для проведення тестування, перевірка доцільності активного впровадження маркетингових досліджень такого типу в Україні. Використання математичних засобів для визначення кількості відвідувачів сайту, достатнього для отримання якісних даних для підтвердження чи спростування статистичних гіпотез щодо створення, зміни та розвитку веб-ресурсів в Україні.

Виклад основного матеріалу дослідження. Підготовка та аналіз результатів А/В тестів передбачає використання таких термінів та понять із галузі математичної статистики:

Статистична гіпотеза — наукова гіпотеза, яку можна перевірити на основі спостереження за процесом, який моделюється із використанням набору випадкових змінних [1].

Випробування статистичної гіпотези — це метод статистичного доведення, який використовують для перевірки статистичної гіпотези.

Статистичний критерій — строге математичне правило, за яким приймається або відкидається певна статистична гіпотеза [2].

Статистична значущість, р-рівень або р-значення результату у статистиці є оцінкою впевненості в його «істинності» (у розумінні «репрезентативності вибірки»).

У результаті перевірки статистичної гіпотези є можливість допустити помилки першого та другого роду.

Припустимо, що задано статистичний критерій:

$$f: R^n \rightarrow \{H_0, H_1\} \quad (1)$$

що зіставляє у кожній реалізації вибірки $X = x$ як одну із гіпотез, що маємо. Тоді можливі чотири ситуації:

1. Розподіл P^n вибірки X відповідає гіпотезі H_0 , і вона точно визначена статистичним критерієм, тобто виконується рівність:

$$f(x) = H_0 \quad (2)$$

2. Розподіл P^n вибірки X відповідає гіпотезі H_0 , але вона неправильно знехтувана статистичним критерієм, тобто виконується рівність:

$$f(x) = H_1 \quad (3)$$

3. Розподіл P^n вибірки X відповідає гіпотезі H_1 , і вона точно визначена статистичним критерієм, тобто $f(x) = H_1$:

$$f(x) = H_1 \quad (4)$$

4. Розподіл P^n вибірки X відповідає гіпотезі H_1 , але вона неправильно знехтувана статистичним критерієм, тобто $f(x) = H_0$:

$$f(x) = H_0 \quad (5)$$

У другому та четвертому випадках вважають, що відбулася статистична помилка, а називають її похибкою першого та другого роду відповідно (табл. 1).

Таблиця 1

Похибки при перевірці статистичних гіпотез

		Правильна гіпотеза	
		H_0	H_1
Результат застосування критерію	H_0	H_0 правильно прийнята	H_1 неправильно прийнята (похибка другого роду)
	H_1	H_0 неправильно знехтувана (похибка першого роду)	H_1 правильно знехтувана

Для отримання достовірних результатів та прийняття правильного рішення після проведення тесту, необхідно, щоб він досягнув певного розміру вибірки.

Розмір вибірки — це важливе питання в маркетингових тестах, адже воно суттєво впливає на достовірність результатів.

Без досягнення правильного та значущого розміру вибірки будь-який тест є просто здогадкою. Більший розмір вибірки збільшує впевненість у тому, що будь-яка зміна показника є справжньою, а не випадковістю [3].

Наявні різні методи обчислення розміру вибірки для різного типу показників, проте всі вони базуються на такій рівності:

$$N = \left[\frac{2\sigma}{\varepsilon} \right]^2 \quad (6)$$

де σ — стандартне відхилення випадкової величини; E — очікувана зміна показника в експерименті.

Значення 2σ є важливим виміром для відділення значущої маркетингової зміни від звичайних коливань. Згідно зі статистикою, якщо інших змін немає, близько 95% результатів тесту потрапляють в проміжок $\pm 2\sigma$ від середнього. А от збільшення чи зменшення за межі 2σ сигналізує про зміну, яку можна ідентифікувати [4].

Найточнішою та найзрозумілішою рівністю для розрахунку розміру вибірки є метод розрахунку для простих (бінарних) метрик.

До таких метрик належать:

1. CTR (від англ. Click Through Rate) — кількість переходів від однієї сторінки до іншої [5].

2. Коефіцієнт проходження опитувань.

3. Коефіцієнт перегляду електронних листів.

Щоб обчислити розмір вибірки для наведених та інших бінарних метрик, використовують рівність:

$$N = \frac{4 \cdot (t_{\frac{\alpha}{2}} + t_{\beta})^2 \cdot p \cdot (1 - p)}{E^2} \quad (7)$$

де $t_{\frac{\alpha}{2}}$ — наближене значення перцентилю нормального розподілу згідно із заданим рівнем значущості α . Для $\alpha = 95\%$, $t_{\frac{\alpha}{2}} = 1.96$; t_{β} — значення перцентилю ймовірності допустити помилку другого роду (β). Для $\beta = 0.2$, $t_{\beta} = 0.841$; 4 — значення, отримане піднесенням до квадрата об'єднаного стандартного відхилення; p — поточне значення метрики.

Якщо головною метрикою, яку перевірятимуть в експерименті, є безперервні дані (середня вартість покупки, середній рівень продажу [6]), то розмір вибірки розраховується так:

$$N = \frac{4 \cdot (t_{\frac{\alpha}{2}} + t_{\beta})^2 \cdot \sigma^2}{E^2} \quad (8)$$

де $t_{\frac{\alpha}{2}}$ та t_{β} — значення, аналогічні таким в (7); σ — стандартне відхилення індивідуальних значень.

Процес обчислення розміру вибірки для безперервних даних полягає у виконанні таких кроків [7]:

1. Визначення ключової метрики.
2. Визначення стандартного відхилення:
 - 2.1. Отримання індивідуальних даних за певний період часу.
 - 2.2. Розрахунок стандартного відхилення σ .
 - 2.3. Відкидання надто великих чи замалих значень вимірюваної величини (більше чи менше).
 - 2.4. Повторне обчислення стандартного відхилення (σ) для тих значень, які залишились після попереднього кроку.
3. Розрахунок розміру вибірки.

Таблиця 2

Порівняння наявних на ринку продуктів

	Optimizely Sample Size Calculator	Survey Monkey Sample Size Calculator	Survey System Sample Size Calculator
Бінарні метрики	+	+	+
Постійні метрики (грошові значення)	-	-	-
Редагування статистичної значущості	+	-	+ (фіксовані значення – 95 та 99%)
Редагування бажаної зміни метрики	+	-	-
Редагування потужності критерію	-	-	-

Висновки. Проведено аналіз та порівняно наявних сервісів (табл. 2.) для створення та проведення А/В тестів. Подано результати тестування.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Кендалл М. Статистические выводы и связи / М. Кендалл, А. Стьюарт. — М.: Наука, 1973.
2. Аністратенко В. О. Математичне планування експериментів в АПК / В. О. Аністратенко, В. Г. Федоров. — К.: Вища школа, 1993. — 374с.
3. Руденко В. М. Математична статистика / В. М. Руденко. — К.: Центр учбової літератури, 2012. — 304 с.
4. Christian, Brian (2000-02-27). «The A/B Test: Inside the Technology That's Changing the Rules of Business | Wired Business». Wired.com. Retrieved 2014-03-18.
5. «A/B Split Testing | Multivariate Testing | Case Studies». Visual Website Optimizer. Retrieved 2011-07-10.
6. Kohavi, Ron; Longbotham, Roger (2015). «Online Controlled Experiments and A/B Tests». In Sammut, Claude; Webb, Geoff. Encyclopedia of Machine Learning and Data Mining (PDF). Springer. pp. to appear.
7. Kohavi, Ron; Longbotham, Roger; Sommerfield, Dan; Henne, Randal M. (2009). «Controlled experiments on the web: survey and practical guide» (PDF). Data Mining and Knowledge Discovery (Berlin: Springer) 18 (1): 140–181. doi:10.1007/s10618-008-0114-1. ISSN 1384-5810.

USE OF VERIFICATION OF STATISTICAL HYPOTHESES
IN INFORMATION TECHNOLOGICAL SPHERE.
A/B TESTING AND ITS APPLICATION FEASIBILITY

I. M. Liakh, Y. Y. Bilak, L. Y. Dan'ko-Tovtyn, V. V. Stanysheskyi
SHEE "Uzhgorod National University"
3, Narodna Square, Uzhgorod, 88000, Ukraine

The article deals with mathematical tools used for testing statistical hypotheses and A/B testing, and the comparison of available services for this kind of testing has been made.

Keywords: hypothesis, testing, service, selection, software.

Стаття надійшла до редакції 28.07.2015.