

ЗАСТОСУВАННЯ ВЕЛИЧИНИ P – ЗНАЧЕННЯ P – VALUE ПРИ ПЕРЕВІРЦІ СТАТИСТИЧНИХ ГІПОТЕЗ

В статті розглянуто та проаналізовано важливість знаходження і застосування аналітиками P – value (P – значення) при перевірці статистичних гіпотез, адже після завершення перевірки нульової гіпотези аналітик має впевнитися, що результати перевірки повинні мати економічний сенс.

Постановка проблеми. При роз'язанні економічних та технологічних проблем, які характеризуються будь-яким середнім показником: середній обсяг реалізації продукції, середній прибуток підприємства, середня продуктивність праці, середній вміст нітратів у продукції, середня вологість сировини тощо, зустрічаються задачі на перевірку достовірності гіпотези про те, наскільки значущою (за заданим рівнем значущості α) є різниця між припущеним (гіпотетичним значенням) μ_0 та середньою вибірковою x_b .

Маючи тільки вибіркові значення, можна лише оцінити, але не визначити точно значення параметрів сукупності чи моделі. Ці оцінки є випадковими і змінюються від вибірки до вибірки. Тому важливо не тільки знати середні оцінки параметрів, знайдені на підставі вибірових даних, але й розуміти міру їх надійності та випадкового розсіювання, яке обумовлено випадковістю процесу формування вибірки. Для цього висуваємо гіпотезу (припущення) про значення генеральної середньої чи про порівняння дисперсій двох нормально розподілених сукупностей, тощо [1].

Нульова гіпотеза вважається вірною, поки не доведено, що вона є помилковою. Наведемо три способи визначення нульової H_0 та альтернативної H_1 гіпотез:

- 1) $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$.
- 2) $H_0: \mu \geq \mu_0$, $H_1: \mu < \mu_0$.

- 3) $H_0: \mu \leq \mu_0$, $H_1: \mu > \mu_0$.

Необхідно звернути увагу, що умова рівності завжди включається в нульову гіпотезу. Крім того, вибір комбінації має прямий вплив на область відхилення нульової гіпотези.

Вибір між трьома варіантами - це задача аналітика.

Аналітик самостійно може обирати рівень значущості для випробування. Але потрібно розуміти, що обираючи 1-відсотковий рівень значущості, в результаті перевірки нульової гіпотези рівно 1% всіх можливих спостережень призведуть до її відхилення. Наприклад, із 100 обчислених за вибірковими даними спостережуваних значень критерію невідхиленню нульової гіпотези будуть сприяти рівно 99, і лише 1 – відхиленню. Отже, якщо нульова гіпотеза відхиляється навіть на цьому рівні значущості, то статистичні докази за відхилення нульової гіпотези вважаються достатньо вагомими.

Аналогічно, для 10% рівня значущості, 10 із 100 обчислених спостережуваних значень критерію будуть підтверджувати, що нульова гіпотеза є хибною, тоді, як рівно 90 інших спостережуваних значень критерію будуть сприяти невідхиленню нульової гіпотези.

Таким чином, якщо нульова гіпотеза не була відхилена з 1% рівнем значущості, але відхиляється з 10% рівнем значущості, це пояснюється лише збільшенням рівня значущості (або збільшенням ймовірності відмови від нульової гіпотези з 1% до 10%). Отже, при збільшенні рівня значущості відхиляти нульову гіпотезу стає легше. Тому, якщо нульова гіпотеза відхиляється з меншими рівнями значущості, докази помилковості нульової гіпотези вважаються більш вагомими.

Тетяна Зінкевич,
канд. ек. наук,
доцент кафедри
фінансів
підприємств
ДВНЗ «Київський
національний
економічний
університет
імені Вадима
Гетьмана»

Валентина Лісовська,
канд. фіз. – мат.
наук, доц., заст.
зав. кафедри
вищої математики
ДВНЗ «Київський
національний
економічний
університет імені
Вадима Гетьмана»

Варвара Стасюк,
канд. пед. наук,
доц. кафедри
вищої математики
ДВНЗ «Київський
національний
економічний
університет імені
Вадима Гетьмана»

РИЗИК – МЕНЕДЖМЕНТ

Після завершення перевірки нульової гіпотези, аналітик має впевнитися, що результати перевірки мали економічний сенс. Незалежно від відхилення або невідхилення нульової гіпотези за результатами її статистичної перевірки, якщо отримані висновки не мають економічного сенсу, то аналітику може знадобитися проведення додаткових випробувань, можливо вже із іншою нульовою гіпотезою.

Пастка при перевірці гіпотези полягає в тому, що вона відповідає лише на запитання: чи буде нульова гіпотеза відхилена на даному рівні значущості.

Наприклад, перевірка нульової гіпотези на основі даних конкретної вибірки може призвести до її відхилення з 10% рівнем значущості, але з 5% рівнем значущості вона вже відхилитися не буде. Ми дійсно не знаємо, в якій точці між 10% і 5% рівнями значущості нульова гіпотеза перетворилась з відхиленої на не відхилену. Ця точка називається р-значенням (або р-value).

Основний матеріал дослідження. Р-значення є найменшим значенням рівня значущості, при якому нульова гіпотеза може бути відхилена для даної вибірки.

Наприклад, припустимо, що спостережуване значення u^* (інакше його ще називають статистикою вибірки) знаходиться на відстані 1,96 стандартної помилки від припущеного значення параметра генеральної сукупності при односторонній перевірці гіпотези. Нагадаємо, що відстань в 1,96 стандартної помилки від припущеного значення параметра генеральної сукупності до спостережуваного значення критерію при односторонній перевірці гіпотези залишає критичну область в 2,5%. Це слідує з того, що критичні значення, які розділяють області відхилення та невідхилення нульової гіпотези, визначаються за формулами знаходження критичних точок для двосторонньої критичної області:

$$\Phi(z_{kp}) = \frac{1-\alpha}{2}, \text{ де } \alpha - \text{рівень значущості.}$$

Для $\alpha = 5\%$ маємо $\Phi(z_{kp}) = 0,475$, тоді за таблицями значень функції Лапласа, знайдемо $z_{kp} = 1,96$. Отже, -1,96 та 1,96 – критичні значення, що розділя-

ють область відхилення та невідхилення нульової гіпотези.

Тоді, саме 2,5% і буде р-значенням. Якщо рівень значущості обрати меншим, ніж 2,5%, то зробити висновки про відхилення нульової гіпотези не буде можливості, оскільки спостережуване значення критерію буде належати області невідхилення нульової гіпотези. Інакше, для рівня значущості більшого 2,5%, нульова гіпотеза буде відхилена (тобто спостережуване значення критерію буде належати критичній області). Легко бачити, що чим далі u^* – спостережуване значення критерію даної вибірки лежить від припущеного в нульовій гіпотезі значення параметра ознаки генеральної сукупності, тим меншим буде р-значення, і тому зменшується ймовірність отримання висновку про відхилення нульової гіпотези. Але, чим менше р-значення, тим вагомішими вважаються аргументи за відхилення нульової гіпотези на користь альтернативної.

Відомо два способи перевірки гіпотези про середнє значення генеральної сукупності у випадку відомого стандартного відхилення генеральної сукупності:

- 1) з використанням методу побудови області відхилення;
- 2) з використанням р - значення (метод р - value approach).

Для вирішення конкретних виробничих проблем значну допомогу менеджеру може надати якісне проведення перевірки статистичних гіпотез, коли нульова гіпотеза являє собою *status quo* (існуюче положення речей). Часто буває, що висновок, отриманий в результаті перевірки гіпотези, приймає форму деякого напрямку дій. Наприклад, якщо існують докази зростання або скорочення величини певного параметра, може бути прийнято рішення про новий план управління запасами, або про випуск нового виду продукції, перехід на кращі ліки для лікування хвороби тощо.

Щоб проілюструвати цей процес, розглянемо приклад нової системи розрахунків у супермаркеті.

Менеджер супермаркету обдумує впровадження нової системи розрахунків для клієнтів, які використо-

вують на касі кредитні електронні картки для розрахунків. Після всебічного фінансового аналізу він визначив, що нова система буде рентабельною, тільки якщо середнє значення всіх одноразових електронних платежів за місяць буде перевищувати 150 грн. Випадкова вибірка 300 одноразових електронних платежів за місяць показала, що середній платіж у вибірці становив 156 грн. Менеджер знає, що електронні платежі є приблизно нормально розподіленими зі стандартним відхиленням у 57 грн. Чи можна, виходячи з цих даних, зробити висновок, що впровадження нової системи розрахунків буде рентабельним?

Цей приклад розглядає сукупність одноразових електронних платежів у супермаркеті. Для перевірки, чи буде рентабельним впровадження нової системи, ми ставимо наступне запитання: середнє значення по вибірці 156 є достатньо великим у порівнянні з 150?

Розглянемо два способи отримати відповідь на це питання.

Перший спосіб – це метод побудови області відхилення або метод перевірки статистичної гіпотези. Цей метод використовується при статистичних розрахунках, які здійснюються вручну.

Другий спосіб – метод дослідження з допомогою p - значення (інакше його називають p - value approach). Цей метод використовується, в основному, для розрахунків з використанням комп'ютера та статистичного програмного забезпечення.

Перший спосіб. 1. Визначаємо нульову та альтернативну гіпотези.

Для того, щоб зробити висновок, що система буде рентабельною, менеджеру необхідно показати, що середнє значення всіх одноразових електронних платежів за місяць буде перевищувати 150 грн. Якщо середнє значення всіх одноразових електронних платежів за місяць буде менше, або дорівнювати 150 грн., система не буде рентабельною. Отже, нульову та альтернативну гіпотези можна сформулювати так:

$H_0: \mu = 150$ – нова система розрахунків не встановлюється, оскільки вона не рентабельна;

$H_1: \mu > 150$ – нова система розрахунків встановлюється, оскільки вона є рентабельною.

На перший погляд здається, що раціонально було б відхилити нульову гіпотезу на користь альтернативної, якщо середнє значення по вибірці значно більше, ніж 150. Якщо, наприклад, обчислене середнє значення по вибірці дорівнювало би 400, то висновок про помилковість нульової гіпотези був би абсолютно очевидний. З іншого боку, значення \bar{X}_B , яке близьке до 150, наприклад, 151, не дозволяє нам відхилити нульову гіпотезу, оскільки цілком можливо, отримати середнє значення по вибірці 151, якщо середнє по сукупності, з якої була здійснена вибірка становить 150. На жаль, висновок про значення середньої генеральної сукупності не завжди є настільки очевидним. У прикладі, який розглядаємо, середнє значення по вибірці було обчислене і становить 156. Така оцінка, очевидно, не досить віддалена, і не досить близька до 150. Для висновку про значення середньої генеральної сукупності, встановлюємо область відхилення - діапазон значень критичної області. Якщо статистичні дані, що перевіряються (спостережуване значення u^*) потрапляє до цього діапазону, робимо висновок про можливість відхилення нульової гіпотези на користь альтернативної.

2. Оскільки середнє квадратичне відхилення генеральної сукупності відоме, статистичний критерій обираємо за формулою:

$$Z = \frac{\bar{X}_e - \mu_0}{\sigma(\bar{X}_e)} \quad , \quad \text{де } \sigma(\bar{X}_e) = \frac{\sigma}{\sqrt{n}}$$

3. Обчислюємо спостережуване значення статистичного критерію, як відношення різниці між середнім вибірки та припущеним в нульовій гіпотезі значенням середнього генеральної сукупності до стандартного відхилення вибіркової середньої:

$$z^* = \frac{\bar{X}_e - \mu_0}{\sigma(\bar{X}_e)} = \frac{156 - 150}{57/\sqrt{300}} \approx 1,82$$

4. Обираємо рівень значущості, критичну область (правостороння, лівостороння або двостороння), критичну точку.

Оскільки рівень значущості не задано за умовою задачі, обираємо його самостійно, нехай – 5%.

Цей рівень значущості дає нам правосторонню область відхилення нульової гіпотези $Z > z_{кр}$. За формулами

$$\Phi(z_{кр}) = \frac{1 - 2\alpha}{2}, \quad z_{кр} = \Phi^{-1}\left(\frac{1 - 2\alpha}{2}\right) =$$

знаходимо критичну точку: $\Phi(z_{кр}) = 0,45$, звідки $z_{кр} = 1,645$.

5. Оскільки спостережуване значення статистичного критерію $z^* = 1,82$ є більшим ніж критичне значення 1,645 та потрапляє до області відхилення нульової гіпотези, робимо висновок, що існує достатньо статистичних доказів на користь альтернативної гіпотези про те, що середнє значення всіх одноразових електронних платежів за місяць перевищує 150 грн.

Підкреслимо, що перевірка була проведена для 5-ти відсоткового рівня значущості. Відхилення нульової гіпотези дає підставу зробити висновок про доцільність встановлення нової системи розрахунків.

Другий спосіб. Статистична процедура це тільки один з декількох факторів, який розглядається менеджером для прийняття рішення. Використовуючи перший спосіб розв'язку задачі, менеджер з'ясував, що існує достатньо статистичних доказів для висновку про те, що середнє значення всіх одноразових електронних платежів за місяць перевищує 150 грн. Проте, перед початком здійснення будь-яких дій, необхідно дослідити ще багато інших факторів, таких, як вартість, фізична можливість перебудови системи розрахунків, можливість помилки (у нашому випадку – помилки першого типу).

Для отримання повної користі від інформації, яка стала доступною у результаті перевірки статистичної гіпотези та прийняття більш обґрунтованого рішення, корисно визначити міру кількості статистичних фактів, які підтримують альтернативну гіпотезу. Таку міру та можливість оцінити найменшу ймовірність припущення помилки першого типу надає нам p – значення.

Ми відхилили нульову гіпотезу на 5-ти відсотковому рівні значущості.

Знайдемо p -значення – найменшу ймовірність припущення помилки першого типу, або найменший рівень значущості, на якому нульова гіпотеза може бути помилково відхилена на користь альтернативної. Оскільки мова йде про наявність помилки першого типу, вважаємо, що нульова гіпотеза H_0 : $\mu = 150$ справедлива. Тоді для нашого прикладу p -значення є ймовірністю спостереження середнього по вибірці щонайменше у розмірі 156, коли середнє по сукупності становить 150. Так,

$$\begin{aligned} p\text{-значення} &= P(\bar{X}_B > 156) = \\ &= P\left(\frac{\bar{X}_B - \mu_0}{\sigma/\sqrt{n}} > \frac{156 - 150}{57/\sqrt{300}}\right) = \\ &= P(Z > 1,82) = 1 - P(Z \leq 1,82) = \\ &= 1 - 0,9656 = 0,034. \end{aligned}$$

Таким чином, ймовірність спостерігати середнє значення щонайменше таким же великим як 156 у сукупності, де середнє складає 150, становить 0,0344, що є малим значенням. Інакше кажучи, ми спостерігали малоїмовірну подію – малоїмовірну настільки, що маємо серйозні сумніви, щоб зробити припущення про те, що нульова гіпотеза справедлива. Тобто, отримали вагому підставу відхилити нульову гіпотезу і підтримати альтернативну.

Інтерпретація p -значення. p -значення надає корисну інформацію, оскільки отримане значення є мірою статистичного доведення, яке підтримує альтернативну гіпотезу. Для розуміння цього пояснення у повній мірі, звернемося до табл.1, де вказано декілька можливих значень \bar{X}_B , відповідних їм спостережуваних даних z^* та p -значення для розглянутого прикладу.

Відзначимо, що чим ближче \bar{X}_B до гіпотетичного значення 150, тим більшим є її p -значення. Чим більша різниця між \bar{X}_B та 150, тим меншим є її p -значення.

Чим далі знаходиться \bar{X}_B від значення 150, тим легше зробити висновок про відхилення нульової гіпотези на користь альтернативної.

Таким чином, чим менше p -значення, тим більше статистичних фактів на користь підтримки альтернативної гіпотези.

Середнє вибіркове \bar{X}_B	Спостережуване значення z^* критерію Z: $z^* = \frac{\bar{X}_B - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X}_B - 150}{57/\sqrt{300}}$	p-значення
150	0	0,5000
152	0,19	0,4247
154	1,22	0,1117
156	1,82	0,0344
158	2,43	0,0078
160	3,04	0,00135

На рис 1 графічно зображено інформацію табл. 1.

У зв'язку з цим виникає питання, наскільки малим має бути р-значення для висновку про те, що альтернативна гіпотеза вірна? У загальному випадку відповідь залежить від кількості факторів, включаючи вартість помилок першого та другого типу. У розглянутому прикладі помилка першого типу трапляється, якщо менеджер прийме рішення на користь нової системи розрахунків, у випадку коли вона нерентабельна. Якщо вартість цієї помилки висока, намагаємось зменшити її ймовірність. У випадку з методом області відхилення, досягатимемо цього встановленням рівня значущості досить низьким, наприклад, 1%. Використовуючи метод дослідження р - значення, повинні наполягати на тому, щоб р-значення було досить малим, надаючи достатній доказ для висновку, що середнє значення всіх електронних ра-

хунків за місяць буде перевищувати 150 грн. для встановлення нової системи розрахунків.

Висновки з проведеного дослідження. Якщо р-значення менше 0,01, стверджується, що існує неспростовний доказ для прийняття альтернативної гіпотези як істинної. Вважається, що перевірка має високий рівень статистичної значущості.

Якщо р-значення знаходиться у межах від 0,01 до 0,05, існує твердий доказ того, що альтернативна гіпотеза істинна. Результат вважається статистично значущим.

Якщо р-значення знаходиться у межах від 0,05 до 0,10 – докази істинності альтернативної гіпотези вважаються непереконливими.

Якщо р-значення більше 5%, результат вважається статистично незначущим.

Якщо р-значення перевищує 0,10, робиться висновок, що не існує доказів

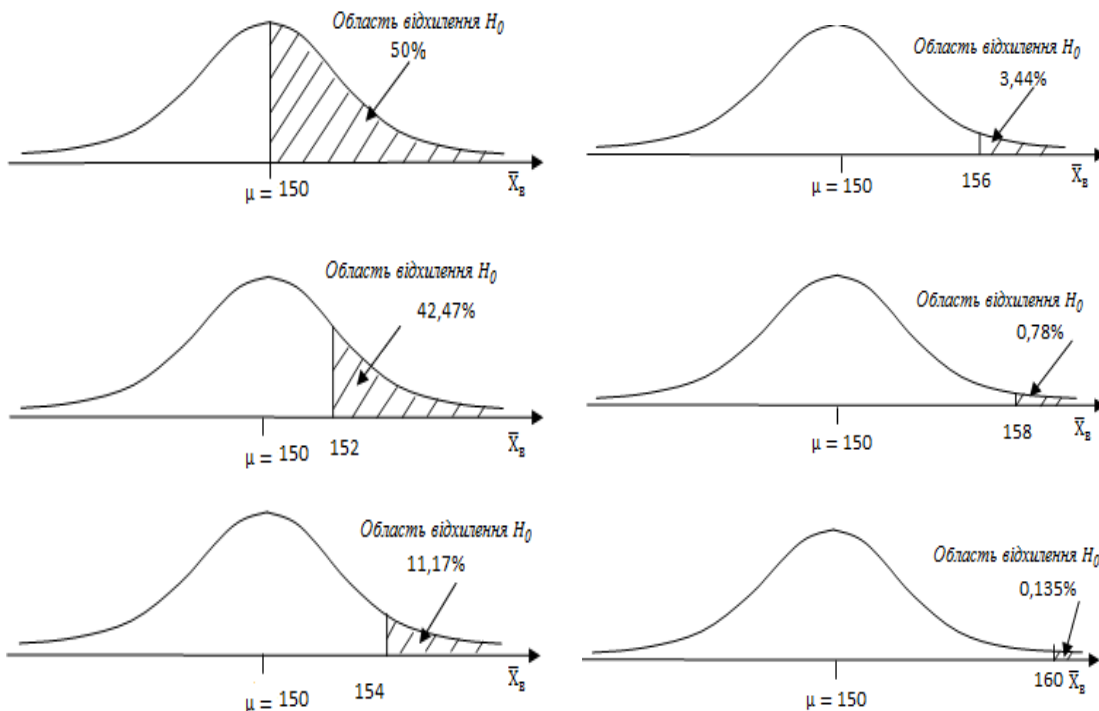


Рис. 1.

для визнання альтернативної гіпотези істинною.

Таким чином, досяжний рівень значущості p -значення – це найменша величина рівня значущості, при якому нульова гіпотеза відхиляється для даного спостережуваного значення u^* обраного критерію U .

ЛІТЕРАТУРА

1. *Зінькевич Т.О., Лісовська В.П.* Вчені записки. (ДВНЗ «Київський національний економічний університет ім. В. Гетьмана»), №14.

2. *Гаркавий В.К., Ярова В.В.*, Математична статистика: Навчальний посібник. – К.: ВД «Професіонал», 2004. – 384с.

3. *Пономаренко О.І., Перестюк М.О., Бурим В.М.* Сучасний економічний аналіз: У 2-х ч. Ч. I. Мікроекономіка: навч. посіб. – К.: Вища шк., 2004. – 262 с.

4. Gerbal Keller «Statistics for management and economics», South-Western GENGAGE Learning. – 889 s.