

ОЦІНКА ОДНОРІДНОСТІ ВИБІРОК ВИПАДКОВИХ ЧИСЕЛ ПЛАТФОРМИ .NET

Оцінимо однорідність двох вибірок. Перша вибірка складається з 1000 рівномірно розподілених у діапазоні від 0 до 100 цілих випадкових чисел і згенерована з використанням класу Random із простору System платформи .Net (варіант 1). Друга вибірка має той самий розмір, вихідні дані та згенерована з використанням класу RandomNumberGenerator з простору System.Security.Cryptography (варіант 2).

У гіпотезі однорідності потрібно визначити, чи ці вибірки є однорідними, тобто отриманими з одного і того ж розподілу, або закон розподілу спостережень від вибірки до вибірки змінювався. Тобто гіпотезою однорідності є твердження $H_0 F_1(x) \equiv F_2(x)$, де $F_i(x)$ – функція розподілу i -ї вибірки.

Для оцінки однорідності двох вибірок використовувався критерій однорідності χ^2 -квадрат. Діапазон значень вибірок було розбито на 15 інтервалів. Для двох вибірок з однаковим розміром 1000 значення статистики χ^2 -квадрат дорівнювало 6,7. За рівня значимості $\alpha=0,05$ критична межа дорівнює 23,7.

Так як значення статистики χ^2 -квадрат менше критичної межі, то виконується згода гіпотези однорідності для двох даних вибірок.

На рис.1 наведено гістограму для розподілу значень χ^2 -квадрат для 1000 пар вибірок розміру 1000.

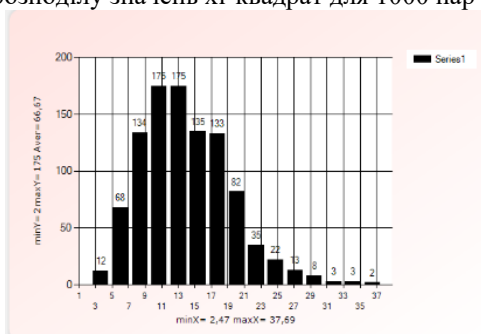


Рис. 1. Гістограма розподілу значень χ^2 -квадрат

Причому з 1000 одержаних значень χ^2 -квадрат кількість значень χ^2 -квадрат більших величини критичної межі 23,7 дорівнює 51. Тобто, з 1000 пар вибірок 51 пара не відповідає поняттю однорідності, а 949 – відповідають.

Оцінимо так само випадковість тих самих двох вибірок. Для оцінки випадковості вибірки використовувався критерій інверсій, статистикою даного критерію є число інверсій I у вибірці. Гіпотеза про випадковість не відхиляється, якщо $I_{\alpha/2} < I < I_{1-\alpha/2}$, де α – рівень значущості. Число інверсій I приймає цілі значення, безліч яких залежить від розміру вибірки n . Математичне очікування та дисперсія статистики I мають відповідно вигляд $E = n \cdot (n-1)/4$,

$D = (2 \cdot n^3 + 3 \cdot n^2 - 5 \cdot n)/72$. Нормалізована статистика має вигляд

$$I^* = (I - E)/D^{0.5} \text{ і за } n \geq 30 \text{ досить точно описується стандартним нормальним законом.}$$

Для двох вибірок з однаковим розміром 1000 значення нормалізованої статистики виявилось рівно відповідно 0,0474 та 0,0674. При рівні $\alpha = 0,05$ критична межа дорівнює $\pm 2,326$.

Оскільки значення нормалізованих статистик перебувають у інтервалі від $-2,326$ до $2,326$, то для двох даних вибірок виконується згода гіпотези про випадковість.

На рис. 2 наведена гістограма для розподілу значень нормалізованих статистик для 1000 вибірок розміру 1000 для варіанта 1. Причому з 1000 отриманих значень статистики I^* кількість 29 значень I^* не потрапили в інтервал $\pm 2,326$.



Рис. 2. Гістограма розподілу значень I^* для варіанта 1

Список використаних джерел

1. Огірко О. І., Галайко Н. В. Теорія ймовірностей та математична статистика: / О. І. Огірко, Н. В. Галайко. – Львів: ЛьвДУВС, 2017. – 292 с.
2. Шпігельгальтер Д. Мистецтво статистики. Прийняття аргументованих рішень на підставі даних: / Д. Шпігельгальтер. – К.: Країна мрій Київ, 2023. – 384 с.