

# ПОРІВНЯЛЬНИЙ АНАЛІЗ ПАКЕТІВ EXCEL, MAPLE, MATLAB ПРИ ВИКОРИСТАННІ ЇХ ПІД ЧАС СТАТИСТИЧНОЇ ОБРОБКИ ДАНИХ

Т. М. Крохмаль<sup>1а</sup>, О. М. Нікітенко<sup>2б</sup>

<sup>1</sup> Харків, Спеціалізована школа з поглибленим вивченням англійської мови № 63

<sup>2</sup> Харків, Харківський національний університет радіоелектроніки

<sup>а</sup> sch63@kharkivosvita.net.ua

<sup>б</sup> nikon@kture.kharkov.ua

Математична статистика – розділ математики, в якому на основі дослідних даних вивчаються ймовірнісні закономірності масових явищ. Обробки даних, що здійснюється методами математичної статистики, потребують всі галузі досліджень: медицина, біологія, соціологія, математика, фізика, педагогіка тощо. До найважливіших розділів математичної статистики відносять:

- статистичні ряди розподілу;
- оцінка параметрів розподілу;
- закони розподілу вибіркового характеру;
- перевірка статистичних гіпотез;
- дисперсійний, кореляційно-регресійний, коваріаційний аналіз;
- факторний та кластерний аналіз тощо.

Тут розглядається лише один з перелічених розділів математичної статистики – оцінка параметрів розподілу, до яких відносяться такі параметри як математичне сподівання випадкової величини, її дисперсія, середньоквадратичне відхилення, асиметрія, ексцес та гістограма.

Статистичні розрахунки без допомоги ЕОМ є складними й потребують використання багатьох таблиць функцій та квантилів стандартних розподілів. Це не сприяє тому, щоб відчувати елемент новизни в матеріалі, який вивчається, змінити задовільно умови задач тощо. Використання ж спеціалізованих математичних пакетів під час навчання вимагає досить високого рівня підготовки з математичної статистики.

Більшість з існуючих математичних пакетів надають можливість користувачам оперувати з випадковими величинами, в тому числі й пакети, що набули широкої популярності: Excel, Maple, Matlab.

Статистика в цих пакетах має свою розвинену систему команд для обслуговування прикладних задач. Команди для статистичних робіт призначені тим категоріям користувачів, котрі потребують середовища, яке дозволяє легко переходити від однієї математичної спеціалізації до іншої, не витрачаючи зайвого часу на трансформацію даних й опануван-

ня різноманітних програмних засобів у вигляді набору команд для аналізу даних з обчисленням різноманітних середніх та квантилів, графічного зображення даних у вигляді гістограм та графіків, а також для обробки даних [1].

Метою цієї статті є порівняння результатів статистичних обчислень та побудови гістограм, що здійснено за допомогою згаданих пакетів.

Проілюструємо це, здійснивши обробку вибірки, обсяг якої складає 80 значень (табл. 1), за допомогою пакетів Excel, Maple, Matlab. Результати обробки вибірки, наведеної в табл. 1, подано в табл. 2.

Таблиця 1

**Вибірка**

13,39	13,46	13,26	13,59	13,54	13,42	13,53	13,5	13,52	13,36	13,57	13,31
13,42	13,53	13,33	13,36	13,37	13,45	13,57	13,37	13,39	13,34	13,33	13,26
13,38	13,55	13,43	13,44	13,31	13,32	13,58	13,3	13,62	13,34	13,64	13,56
13,53	13,29	13,5	13,34	13,37	13,44	13,66	13,5	13,4	13,28	13,43	13,4
13,51	13,24	13,44	13,33	13,33	13,58	13,43	13,4	13,23	13,48	13,49	13,26
13,3	13,34	13,53	13,25	13,54	13,5	13,42	13,28	13,45	13,4	13,55	13,47
13,4	13,54	13,48	13,28	13,32	13,36	13,38	13,31				

Таблиця 2

**Результати обробки вибірки**

	Вручну	Excel	Maple	Matlab
Середнє	13,42	13,42	13,42	13,42
Дисперсія вибірки	0,01136203	0,0113620	0,011362	0,0114
Стандартне відхилення	0,10659280	0,1065928	0,106593	0,1066
Асиметричність	0,194202	0,2017028	0,196660	0,1979
Ексцес	2,0440198	-0,884131	2,069893	2,0961

Як випливає з результатів обчислень, всі пакети подають однакові результати для математичного сподівання (середнього), дисперсії та середньоквадратичного відхилення.

Щодо коефіцієнтів асиметрії та ексцесу, то жоден результат не збігається.

Аналіз результатів обчислень показав, що збіг між цими обчисленнями відсутній через різне визначення коефіцієнтів асиметрії та ексцесу в наведених пакетах.

Теоретично коефіцієнт асиметрії, який характеризує несиметричність графіка функції розподілу і визначається як  $\frac{m_3}{\sigma^3}$ , де  $m_3$  – центральний емпіричний момент третього порядку, що визначається як

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 ;$$

$n$  – обсяг вибірки;

$x_i$  – елемент вибірки;

$\bar{x}$  – вибіркове середнє, яке визначається як

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i ;$$

$\sigma$  – підправлене середнє квадратичне або стандартне відхилення випадкової величини, яке визначається як

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} .$$

В пакеті Excel коефіцієнт асиметрії обчислюється за виразом

$$\frac{n^2}{(n-1)(n-2)} \frac{m_3}{\sigma^3} .$$

В системі комп'ютерної математики Maple коефіцієнт асиметрії обчислюється за виразом  $\frac{n}{(n-1)} \frac{m_3}{\sigma^3}$ .

В системі комп'ютерної математики Matlab коефіцієнт асиметрії збігається з теоретичним.

Теоретично коефіцієнт ексцесу, який характеризує сплющеність кривої розподілу та протяжність спадів, і визначається як  $\frac{m_4}{\sigma^4} - 3$ , де  $m_4$  – центральний емпіричний момент четвертого порядку, який визначається як  $m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4$ ;  $-3$  враховує той факт, що коефіцієнт ексцесу для нормального закону розподілу випадкових величин дорівнює 3.

Коефіцієнт ексцесу в пакеті Excel обчислюється за виразом

$$\frac{n^2(n+1)}{(n-1)(n-2)(n-3)} \frac{m_4}{\sigma^4} - \frac{3(n-1)^2}{(n-2)(n-3)} .$$

В системі комп'ютерної математики Maple коефіцієнт ексцесу обчислюється за виразом  $\frac{n}{(n-1)} \frac{m_4}{\sigma^4}$ .

В системі комп'ютерної математики Matlab коефіцієнт ексцесу обчислюється як теоретичний без урахування поправки на нормальний закон розподілу  $\frac{m_4}{\sigma^4}$ .

Для візуалізації відмінностей обчислення коефіцієнтів асиметрії та ексцесу їх наведено на рис. 1.

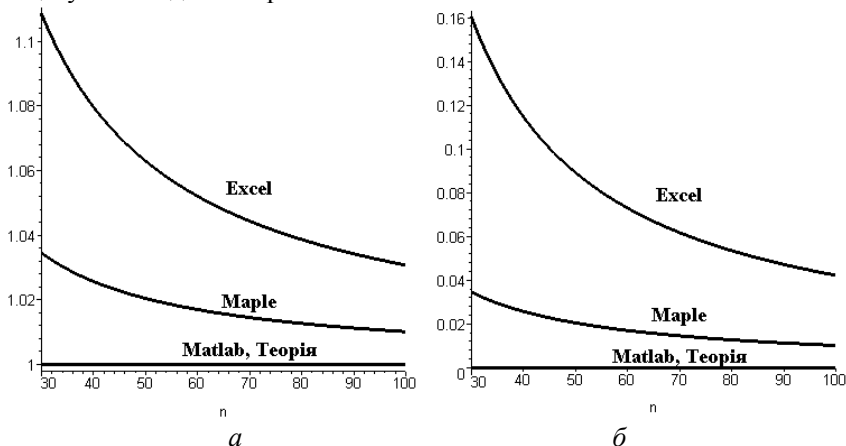


Рис. 1. Відмінності обчислення коефіцієнтів  $a$  – коефіцієнт асиметрії;  $b$  – коефіцієнт ексцесу

Результати побудови гістограми для цієї вибірки наведено на рис. 2.

З цього рисунку видно, що гістограми, які побудовані вручну та за допомогою систем комп'ютерної математики Maple та Matlab, є однаковими, а побудована за допомогою пакету Excel, має багато відмінностей.

Щоб з'ясувати причини такої розбіжності, проаналізуємо межі інтервалів на які поділено варіаційний ряд, що утворено з вибірки.

Результати обчислення меж інтервалів, що виконано за допомогою пакету Excel, наведено в таблиці 3.

Таблиця 3

Межі інтервалів за пакетом Excel

Bin	Frequency
13,23	1
13,28375	8
13,3375	12
13,39125	14
13,445	14
13,49875	7
13,5525	15
13,60625	6
More	3

Результати обчислення меж інтервалів, що здійснено за допомогою

інших пакетів, наведено в таблиці 4.

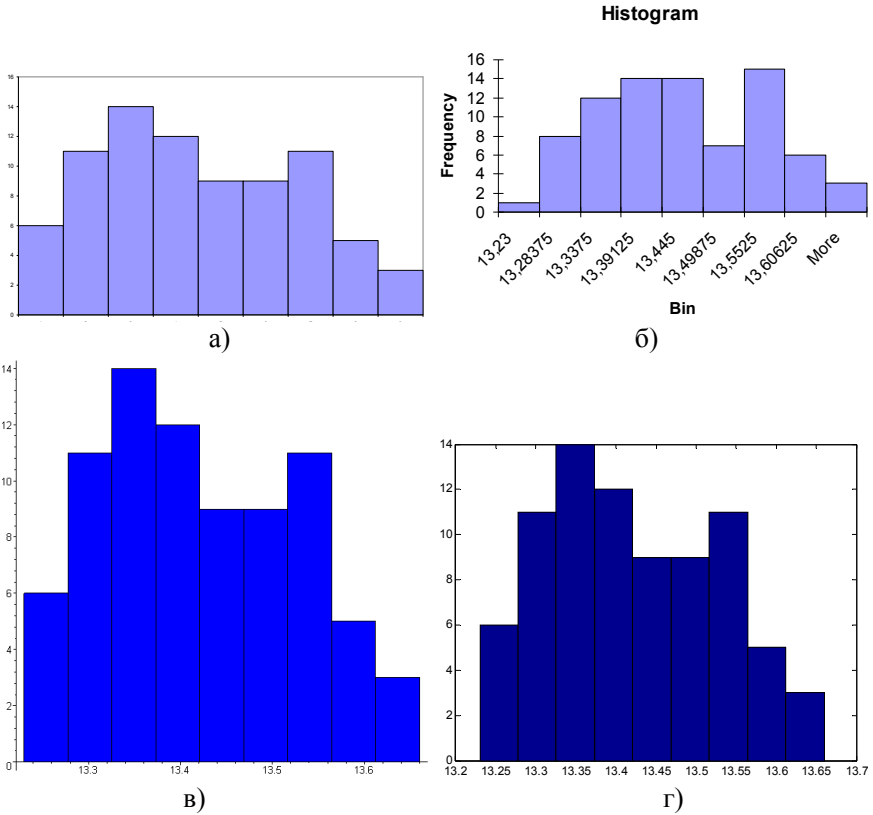


Рис. 2. Гістограми: а – вручну; б – Excel; в – Maple; г – Matlab

Таблиця 4

**Межі інтервалів за іншими обчисленнями**

Bin	Frequency
13.23 .. 13.27778	6
13.27778 .. 13.32556	11
13.32556 .. 13.37333	14
13.37333 .. 13.42111	12
13.42111 .. 13.46889	9
13.46889 .. 13.51667	9
13.51667 .. 13.56444	11
13.56444 .. 13.61222	5
13.61222 .. 13.66	3

З порівняння даних з таблиць 3 та 4 випливає, що в пакеті Excel межі інтервалів обчислюються з похибками, а це призводить до неправильного визначення кількості елементів, які потрапляють в ці інтервали. Отже, для того, щоб правильно побудувати гістограму за допомогою пакету Excel, попередньо необхідно обчислити межі інтервалів.

Таким чином, під час обчислення статистичних характеристик за допомогою комп'ютерних пакетів необхідно або здійснити попереднє порівняння результатів обчислень, що не завжди зручно, або з'ясувати за якими формулами відбуваються обчислення необхідних параметрів і вжити відповідних заходів для усунення можливих розбіжностей.

#### Список використаних джерел

1. Крохмаль Т. М. Пакет символічних обчислень Maple для вивчення теми «Закони розподілу випадкових величин» / Т. М. Крохмаль, О. М. Нікітенко // Теорія та методика навчання математики, фізики, інформатики : збірник наукових праць. Випуск ІХ. – Кривий Ріг : Видавничий відділ НМетАУ, 2011. – С. 82-87.