

UDC 519.2:[004.9]+631.46

**Yu. O. Sotnykov, D. V. Gavva, S. V. Resnik, K. B. Novosad**  
Kharkiv National Agrarian University named after V. V. Dokuchayev  
e-mail: novosad-konst@rambler.ru; SerHEY021@yandex.ua

## COMPUTER TECHNOLOGIES OF EXPERIMENTAL DATA PREPARATION IN SOIL SCIENCE

*The main stages of data computer preparation obtained as a result of agronomic experiment for statistical processing are considered. The examples of the efficient use of computer programs while checking the primary experimental data characterizing the evolution of the typical chernozems are given.*

*Mathematical processing and analysis of information obtained in the research process is one of the most important and difficult components of any scientific experiment. The use of computer technologies in this process is not only reasonable, but obligatory. This greatly simplifies and accelerates data processing and it also makes obtained conclusions more informative and reasonable.*

*Therefore, the aim of the publication was to describe in a simplified and accessible form of basic methods and techniques of the preparatory stage of computer processing of experimental data of an agronomic experiment. The object of the research was selected a data array characterizing the evolution of chernozems typical of the South Eastern part of the Forest Steppe region of Ukraine. The research methods were to analyze the effectiveness of the instruments of experimental data computer processing at the preparatory stage of the research.*

*The proposed method of experimental data organization as tables "object-feature" allows conducting the analysis using the absolute majority of specialized statistical computer programs. But today a significant number of researchers as the basic data analysis tool unfortunately use only spreadsheets.*

*Spreadsheets are good for accumulation, data preservation, intermediate conversion, preliminary statistical estimation. However, the final statistical analysis is advisable to carry out with the help of specialized programs. Therefore, the choice of a specialized statistical software package is quite important. The article presents the selection criteria of such programs.*

*Based on our own practical experience, we can testify: pre-properly prepares data using any spreadsheets as well as primary data can be quickly analyzed in a specialized (including and commercial) program for the evaluation period.*

*The importance of preliminary classification of the agronomic experiment factors is considered, the examples of classification of non-parametric signs of the conducted experiment in the soil science and their influence on the results are given.*

*The order and the algorithm of data verification for legal compliance with the normal distribution in accordance with the international standards of carrying out experiments in agronomy are cited.*

*The preparatory stage of the experimental data analysis using the given algorithm will significantly speed up the process of data analysis based on computer technologies, improve the quality, clarity and validity of the received conclusions.*

**Keywords:** table "object-feature". Statgraphics Centurion, measurement scales, soils classification, extreme values, normal distribution.

УДК 519.2:[004.9] +631.46

**Ю. А. Сотников, Д. В. Гавва, С. В. Рєзнік, К. Б. Новосад**

*Харьковский национальный аграрный университет имени В. В. Докучаева  
e-mail: novosad-konst@rambler.ru; SeRHEY021@yandex.ua*

### **КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ ПОДГОТОВКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ В ПОЧВОВЕДЕНИИ**

*Рассмотрено основные этапы компьютерной подготовки данных, полученных в результате проведения агрономического эксперимента для статистической обработки. Приведены примеры эффективного использования компьютерных программ в ходе проверки первичных экспериментальных данных, которые характеризуют эволюцию черноземов.*

**Ключевые слова:** *таблица «объект-признак», Statgraphics Centurion, шкалы измерения, классификация почв, экспериментальные значения, нормальное распределение.*

УДК 519.2:[004.9] +631.46

**Ю. О. Сотников, Д. В. Гавва, С. В. Рєзнік, К. Б. Новосад**

*Харківський національний аграрний університет імені В. В. Докучаєва  
e-mail: novosad-konst@rambler.ru; SeRHEY021@yandex.ua*

### **КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ ПІДГОТОВКИ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ У ҐРУНТОЗНАВСТВІ**

*Розглянуто основні етапи комп'ютерної підготовки даних, отриманих у результаті проведення агрономічного експерименту для статистичного обробітку. Наведено приклади ефективного використання комп'ютерних програм під час перевірки первинних експериментальних даних, що характеризують еволюцію чорноземів типових.*

**Ключові слова:** *таблиця «об'єкт-ознака», Statgraphics Centurion, шкали вимірювання, класифікація ґрунтів, екстремальні значення, нормальний розподіл.*

**Вступ.** Математична обробка та аналіз інформації, що отримана в процесі дослідження – одна з найбільш відповідальних та складних складових будь-якого наукового експерименту. Використання в цьому процесі комп'ютерних технологій є не тільки доцільним, але й обов'язковим. Це значно спрощує та пришвидшує процес обробки даних, а отримані висновки дозволяє зробити більш інформативними та обґрунтованими. Слід розуміти, що очікувана ефективність використання будь-яких комп'ютерних технологій може бути досягнута лише при чіткому дотриманні жорстких правил організації даних та наявності в дослідника достатніх навичок користування обраними ним комп'ютерними програмами та методами.

На жаль, на практиці часто цим правилам не приділяється своєчасної,

належної уваги, що в кращому випадку, призводить до значного збільшення часу розрахунків, невиправданого їх ускладнення та недостатній інформативності отриманих результатів. У найгіршому випадку дослідник отримує помилкові результати і робить на їх основі хибні висновки. Частіше за все, причинами цього є недостатня теоретична підготовка та спроба реалізації в процесі обробки даних архаїчних алгоритмів «докомп'ютерної доби». До сьогодні, значна кількість учених агрономів, аналізуючи експериментальні дані діє як у прислів'ї – «*іти на комара з дрючком, а на вовка з швайкою*».

Тому **метою цієї роботи** є опис у спрощеній та доступній формі основних методів та прийомів підготовчого етапу комп'ютерної обробки експериментальних даних агрономічного експерименту.

**Об'єктом дослідження** було обрано масив даних, що характеризують еволюцію чорноземів типових південно-східного Лісостепу України.

**Методика досліджень** полягала в аналізі ефективності інструментів комп'ютерної обробки експериментальних даних на підготовчому етапі дослідження.

**Результати досліджень. Організація процесу накопичення та перетворення даних.** В абсолютній більшості випадків під час проведення наукових експериментів в агрономії об'єктом дослідження є складні біологічні системи (грунти, рослини, агробіоценози тощо). Такі об'єкти відрізняються між собою значною кількістю властивостей (ознак). Тому виникає гостра необхідність правильної організації отриманого в процесі експерименту ознакового простору. Отримуємо ситуацію, коли кожен досліджуваний об'єкт постає у вигляді послідовності значень відповідних ознак. Такий масив даних повинен оформлюватися у вигляді таблиці «об'єкт-ознака», рядки якої відповідають об'єктам дослідження (наприклад – зразкам ґрунту), а кожен стовпчик – векторам окремої ознаки (наприклад – глибини відбору зразку, вмісту гумусу і т.д.). У результаті отримуємо таблицю даних, яку можна обробляти за допомогою комп'ютерних програм (рис. 1).

*Створення такого масиву даних* є першим й найголовнішим етапом обробки отриманої інформації. Вона повинна проводитися за визначеними, обов'язковими до виконання, правилами:

1) У таблиці не повинно бути пустих рядків, стовпчиків, об'єднаних клітинок;

2) кожен стовпчик обов'язково повинен мати заголовок (мітку), що знаходяться в першому рядку даних. Доцільно за мітку використовувати повну назву ознаки (властивості) об'єктів, що досліджується, необхідною мовою;

3) кожне поле повинно мати значення в кожному рядку. Стовпці не повинні мати груп даних, що повторюються;

4) використання формул за рядками таблиці неприпустиме.

Експериментальні дані, організовані за таким алгоритмом можна аналізувати за допомогою абсолютної більшості спеціалізованих статистичних комп'ютерних програм. Зауважимо, що до сьогодні значна кількість дослідників, як основного інструмента аналізу даних, на жаль, використовує саме електронні таблиці (MS Excel, Open Calc, Google Spreadsheets та т.п.) (Бондарчук, 2010;

Мінько, 2015). На наш погляд, подібні програми, насправді, є доволі простим, практично ідеальним інструментом для створення великих масивів первинних експериментальних даних.

**1. Приклад організації експериментальних даних у вигляді таблиці «об'єкт-ознака» (фрагмент)**

№ зразка	Стационар	Варіанти	Глибина	Гумус	Ca2+	Mg2+	Na+	K+
1	Роганський	Дуб	0-5	8,6	36,55	5,77	0,32	0,85
2	Роганський	Дуб	5-20	6,8	35,15	5,73	0,26	0,82
3	Роганський	Дуб	20-40	5,1	34,46	5,68	0,26	0,55
4	Роганський	Береза	0-5	6,8	35,29	5,76	0,33	0,74
5	Роганський	Береза	5-20	4,7	34,58	5,78	0,26	0,68
6	Роганський	Береза	20-40	3,3	34,17	5,57	0,26	0,48
7	Роганський	Сосна	0-5	5,4	32,74	6,42	0,37	0,57
8	Роганський	Сосна	5-20	4,2	32,95	6,34	0,37	0,54
9	Роганський	Сосна	20-40	3,0	32,87	6,38	0,37	0,47
10	Роганський	Смерека	0-5	5,5	32,40	6,33	0,37	0,73
11	Роганський	Смерека	5-20	4,1	32,43	6,54	0,37	0,60
12	Роганський	Смерека	20-40	2,8	32,11	6,67	0,37	0,48
13	Роганський	Переліг кошений	0-5	5,9	38,47	4,74	0,37	0,53
14	Роганський	Переліг кошений	5-20	5,0	37,54	4,59	0,32	0,49
15	Роганський	Переліг кошений	20-40	3,4	36,82	4,41	0,27	0,40
16	Роганський	Переліг	0-5	6,2	39,26	4,26	0,37	0,45
17	Роганський	Переліг	5-20	5,9	38,10	4,21	0,37	0,33
18	Роганський	Переліг	20-40	3,5	37,88	4,15	0,37	0,34
19	Роганський	Рілля	0-5	5,6	33,96	5,65	0,37	0,55
20	Роганський	Рілля	5-20	5,4	33,40	5,47	0,37	0,49
21	Роганський	Рілля	20-40	4,1	33,41	5,55	0,37	0,34
22	Скрипаївський	Дуб	0-5	8,9	28,80	6,15	0,21	0,60
23	Скрипаївський	Дуб	5-20	3,6	27,56	6,04	0,22	0,43
24	Скрипаївський	Дуб	20-40	2,8	25,26	6,19	0,21	0,34
25	Михайлівський	Цілина абсолютна	0-5	10,1	35,10	5,70	0,21	1,72
26	Михайлівський	Цілина абсолютна	5-20	8,5	35,20	5,80	0,18	1,36
27	Михайлівський	Цілина абсолютна	20-40	6,4	34,60	5,60	0,25	0,73
28	Михайлівський	Цілина косима	0-5	9,6	34,40	5,55	0,20	1,75
29	Михайлівський	Цілина косима	5-20	7,9	34,60	5,50	0,22	1,29
30	Михайлівський	Цілина косима	20-40	6,6	34,20	5,44	0,21	0,67
31	Михайлівський	Лісосмуга	0-5	8,7	34,00	6,30	0,24	1,83
32	Михайлівський	Лісосмуга	5-20	7,5	33,30	6,12	0,23	1,41
33	Михайлівський	Лісосмуга	20-40	6,2	32,60	5,90	0,25	1,16
34	Михайлівський	Переліг	0-5	9,1	34,90	5,60	0,21	1,61
35	Михайлівський	Переліг	5-20	7,7	33,70	5,30	0,23	1,12
36	Михайлівський	Переліг	20-40	6,2	33,50	5,20	0,19	0,57
37	Михайлівський	Рілля	0-5	5,9	32,70	4,08	0,23	1,16
38	Михайлівський	Рілля	5-20	5,5	32,50	4,02	0,26	0,86
39	Михайлівський	Рілля	20-40	4,9	32,10	3,80	0,25	0,63

У цих таблицях можуть бути повністю відсутні розрахункові показники, а тому не виникає необхідності володіння навичками введення формул та проведення складних статистичних розрахунків. Якщо ж у період проведення польового експерименту можна забезпечити постійний доступ до мережі інтернет (наприклад, у період обліку врожаю) доцільно реєструвати дані безпосередньо в безкоштовних електронних таблицях типу Google Spreadsheet, Zohosheets і под. та зберігати результати у хмарному сховищі даних. Це дозволить значно підвищити ефективність, оперативність та якість обробітку даних, забезпечить високий рівень їх збереження.

Але будь-яка, без винятку ЕТ має досить обмежені можливості аналізу даних. В основному, це найпростіші інструменти статистичного аналізу. Цих інструментів не вистачить для проведення глибокого, професійного аналізу отриманих експериментальних даних, а ступінь довіри до результатів аналізу отриманих за допомогою ЕТ достатньо низький (Воробйов, 2009). Так, для реалізація будь-якого методу аналізу в ЕТ необхідно самостійно розробити відповідний алгоритм рішення та реалізувати його практично вручну, за допомогою знайдених формул. Оскільки ймовірність помилки у ході розв'язання задачі таким способом дуже велика, результат обов'язково необхідно перевіряти.

Під час використання спеціалізованих програм, навпаки, необхідно лише визначитися зі змінними, що аналізуються та обрати потрібний користувачу метод. Тому ймовірність помилки та витрати часу на обчислення за такого підходу – мінімальні.

ЕТ добре підходить для накопичення, збереження даних, проміжного перетворення, попередніх статистичних прикидок. Однак остаточний статистичний аналіз необхідно робити в спеціалізованих програмах (Мазур, 2014). Тому вибір спеціалізованого пакету статистичних програм є досить важливим.

**1. Вибір пакету прикладних програм.** У виборі слід звертати особливу увагу на такі фактори:

- Ліцензійне програмне забезпечення професіонального рівня потребує витрат коштів. Наприклад, JMP SAS – від \$23, SPSS – від \$73, Statgraphics Centurion XVII – \$30 (за 1 ліцензію). Але якщо існує будь-яка можливість використання саме комерційних програм — користуємося саме ними.

*Виходячи з власного практичного досвіду, можемо свідчити: попередньо правильно підготовлені за допомогою будь-яких електронних таблиць та первинні дані можуть бути досить швидко проаналізовані в спеціалізованій (зокрема, і комерційній) програмі за ознайомчий період.*

- Наявність в обраній програмі необхідних інструментів аналізу. Особливу увагу приділяємо методам багатомірної статистики, необхідних для дослідження складних, зокрема біологічних систем. Так, комерційна програма SPSS має в своєму арсеналі розвинений інструментарій багатомірного аналізу, а її практично повний безкоштовний «близнюк» – SPSS, ні.

- Простота і зручність користування, зрозумілість отриманих результатів. Значна кількість як безкоштовних (R), так і комерційних (Mathematica, Mathcad) потребує від користувача специфічних навичок і знань, як у користуванні програмою, так і в осмисленні результатів аналізу.

**2. Встановлення пріоритетів.** Перед початком дослідження необхідно навести отримані первинні дані відповідно до мети дослідження – класифікувати об’єкти дослідження за допомогою номінальних та порядкових вимірювальних шкал. У першу чергу це стосується якісних ознак, наприклад, назв стаціонарів, варіантів, глибин відбору зразків. Це дозволить розрізнити, ідентифікувати об’єкти та привласнити їм окремі числові значення (Верес, 2009).

Абсолютна більшість статистичних моделей так чи інакше пов’язані з класифікацією (Шитиков, 2005). В основі будь-якої моделі лежить замисел людини навчити комп’ютер «відрізнити одне від іншого», тобто за значенням прогнозованого показника-відгуку, явно або неявно оцінити деяку *категоріальність* об’єкта, процесу або явища, що вивчається. В кінцевому рахунку – отримати розподіл об’єктів дослідження за класами. Після отримання результатів такого моделювання приходить час діагностики – порівняння об’єктів дослідження між собою за визначеними відмінними ознаками класів (правилами дискримінації). Розподіл об’єктів дослідження за такими класами дозволяє зробити інформаційні таблиці більш компактними, оскільки кількість класів, що виділяються, завжди менша за кількість первинних об’єктів. У результаті отримуємо лаконічне, наочне й корисне представлення даних у просторі істотно меншої розмірності. За допомогою різноманітних методів шкалювання досягається необхідний ступінь формалізації статистичних моделей, достатній для комп’ютерної реалізації тих чи інших методів аналізу даних.

Наприклад, виходячи з даних, наведених у табл. 1 номінальні якісні дані повинні бути перекодовані в чисельні за порядковою шкалою та додані до таблиці: Михайлівський стаціонар – 1, Роганський – 2 тощо. Із кваліфікацією варіантів справа виглядає дещо складніше.

Так, за класифікацією Д.Г.Тихоненка (Тихоненко, 2009) ґрунти поділяються на природні, орні (агрогенні), техногенні (рекультивовані), урбаногенні (ґрунти населених пунктів). Останній тип у нашому випадку відсутній. Цілину можна відносити до 1 типу, ріллю до 2, усі інші до 3 типу ґрунтів. Таким чином, замість 8 варіантів досліді отримуюмо 3 типу ґрунтів 0. Але за дещо іншою модифікованою класифікацією постагрогенні ґрунти можна розділити також на перелоги та лісові штучні ценози (Герасимова, 2010). Таким чином, отримуємо вже чотири класи агроценозів (табл. 2).

Використання різних варіантів шкалювання дозволяє об’єднувати об’єкти «у такі групи, відносно яких можна висунути найбільшу кількість загальних ознак, і притому важливіших, ніж ті, які можна затверджувати відносно до всіх інших груп, за якими можна було б розподілити ці об’єкти. Таким чином, об’єкти можна було б розподіляти на підставі таких властивостей, які є спільними для багатьох інших або, принаймні, відповідають їх спільним ознакам. Із цих останніх потрібно вибирати такі, які були б найнадійнішими і найбільш безпосередніми ознаками, саме тоді ж мали такі властивості, на які треба зосередити увагу (Милль, 2011).

**3. Перевірка якості даних.** Безпосередньо перед початком проведення аналізу необхідно обов’язково провести перевірку отриманих даних з метою виявлення можливих помилок. Дуже часто під час перевірки даних, користувач

керується лише особистою інтуїцією, практичним досвідом та власною самовпевненістю. У кінцевому рахунку це значно збільшує час проведення аналізу, негативно впливає на якість отриманих результатів та їх осмислення, нерідко призводить до помилкових висновків.

## 2. Приклад кодування варіантів дослідів

Стационар	Код стаціо-нару	Варіант дослідів	Код варіанту	Класифікація (варіант I)	Кодування (I вар.)	Класифікація (варіант II)	Кодування (II вар.)
Михайлівський	1	Цілина абсолютна	1	Природні	1	Природні	1
Роганський	2	Цілина кошена	2				
Скрипаївський	3	Рілля	3	Агрогенні	2	Агрогенні	2
Глибина відбору зразків	Код глибини	Переліг	4	Пост-агрогенні	3	«Чисті» пост-агрогенні (переліги)	3
0-5		1	Переліг кошений				
		Берега	6				
5-20	2	Дуб	7				
		Сосна	8				
20-40	3	Смерека	9				
		Лісосмуга	10	Агрогенні, під лісонасадженнями		4	

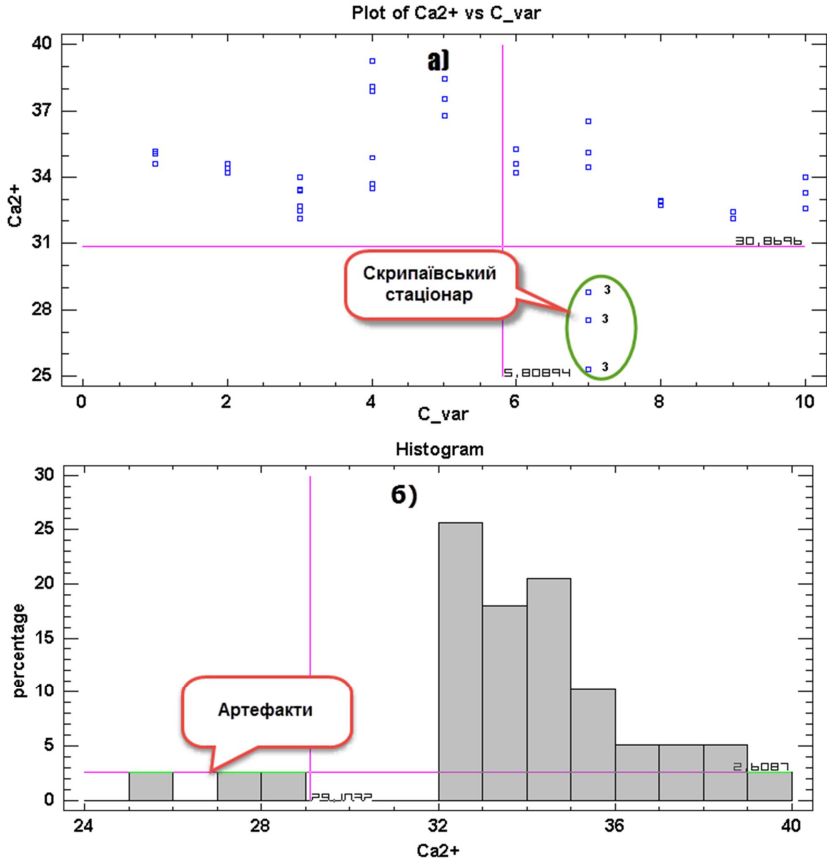
На сьогоднішній день існує декілька ефективних інструментів комп'ютерної перевірки даних, що базуються на методах математичної статистики. Розглянемо деякі з них:

✓ **Помилки вводу.** Під час використання спеціалізованих статистичних програм можливість допущення грубих помилок вводу даних зведена до мінімуму. Так, програма Statgraphics просто не дозволить ввести в стовпчик дані, що не відповідає типу даних, що був призначений раніше (наприклад, замість числової змінної 9,6 ввести текстову 9,6 або 9.6).

✓ **Помилки виміру (артефакти).** Найбільші складнощі виникають при перевірці числових даних. Дані такого типу перевіряються за розмахом, коли для кожної числової ознаки задаються верхні та нижні межі. Величину, що буде за цими межами (викид, артефакт, outliers) розглядають окремо. Причини появи таких значень можуть бути різні, але всі вони на початку дослідження повинні бути виявлені. Залежно від причини приймається рішення або про їх видалення, або про виправлення, або вилучення для подальшого окремого аналізу. Для ідентифікації таких аномальних значень використовують графіки розсіювання (X-Y plot) або гістограми. Результати такої перевірки для показника  $Ca^{2+}$  виглядатимуть так (рис. 1).

Рис. 1 свідчить, що графік розсіювання (а) дозволяє визначити сам факт наявності й кількість об'єктів, значення ознак яких максимально відхиляється від середніх, а також приблизні значення цієї ознаки та приналежність їх до відповідних варіантів дослідів. Практично той самий результат отримуємо і за

допомогою побудови гістограми розподілу, але без визначення варіантів досліджу. Обидва варіанти мають один суттєвий недолік – дослідник повинен самостійно вирахувати положення артефактів у таблиці первинних даних.



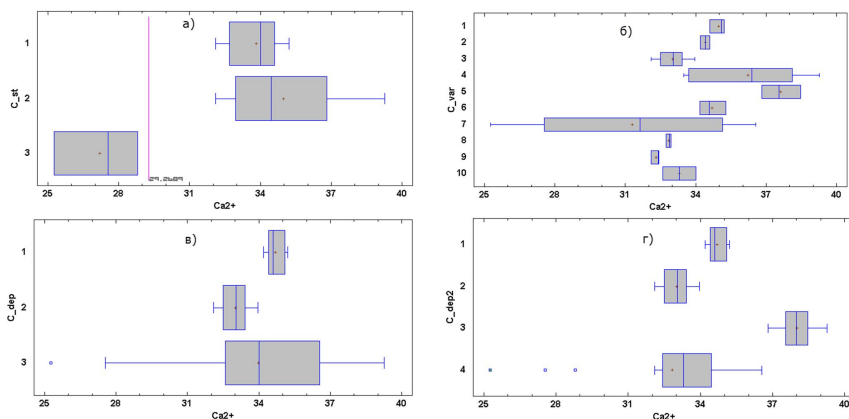
**Рис. 1.** Визначення артефактів вмісту обмінних катіонів кальцію графічним способом за допомогою: а) графіку X-Y б) гістограми розподілу

У Міжнародних Положеннях із провадження експериментальних досліджень в агрономії (Document, 2003) для виявлення артефактів рекомендують використовувати також коробчастий графік (box plot) (рис. 2), де для кожної групи даних рисується коробок з перемичкою. У межах кожного ящика знаходиться 75 % спостережень, позначками відмічаються середнє та медіану вибірки. Довжина вусів дорівнює, 1,5 довжини ящика. Показники, що знаходяться за межами вусів – артефакти, причину появи яких треба дослідити. Додатково за

*ISSN 2225-8701. Bulletin of Kharkiv National Agrarian University. 2015. № 1*



допомогою таких ящиків можна оцінити мінливість та симетрію розподілу. Серед недоліків такого графіка відмітимо, що його інформативність та обгрунтованість напряму залежить від якості обраної шкали розподілу. Так, виходячи з рис. 2а можна зробити висновок лише про значно менший обсяг  $\text{Ca}^{2+}$  Скрипаївського стаціонару, з рис. 2б – про незначну провосторонню асиметрію та значну мінливість цього показника у варіанті 7 (дуб). Асиметрію можна оцінити за зміщенням медіани в будь-який бік ящика та різній довжині вусів. Рис. 2в та 2г діагностують відповідно 1 та 3 артефакти. Знайти їх у таблиці первинних даних необхідно самостійно.



**Рис. 2. Варіанти коробчастих** графіків, в залежності від обраних шкал вимірювання.  
а) за стаціонарами; б) за варіантами дослідів; в) за типами ґрунту(I вар.);  
г) за типами ґрунту(II вар.)

У більшості спеціалізованих статистичних програм існують окремі інструменти для визначення артефактів. Так, у програмі Statgraphics Centurion XVII це *Outlier Identification*. Після вибору процедури користувачу необхідно лише обрати змінну, що перевіряється, ввести рівень довірчого інтервалу, відсоток цензурування даних, кількість крайніх значень (максимальних та мінімальних), що будуть відображатися на графіку Outlier Plot (Графік викидів) (рис. 3). Такий графік дозволяє візуально ідентифікувати екстремальні значення показника, що аналізується.

Аналіз графіка базується на правилі  $3\sigma$  – практично всі значення нормально розподіленої випадкової величини лежать в інтервалі  $[\bar{x} - 3\sigma; \bar{x} + 3\sigma]$  – відхилення від середнього значення не повинні перевищувати трьох стандартних відхилень. За графіком можна легко визначити що показник  $\text{Ca}^{2+}$  ґрунтів Скрипаївського стаціонару перевищує вказані межі, а такий саме показник ґрунтів під перелогом Роганського стаціонару має максимальні відхилення серед припустимих. Для полегшення ідентифікації екстремальних значень автоматично формується таблиця відсортованих відхилень (табл. 3).

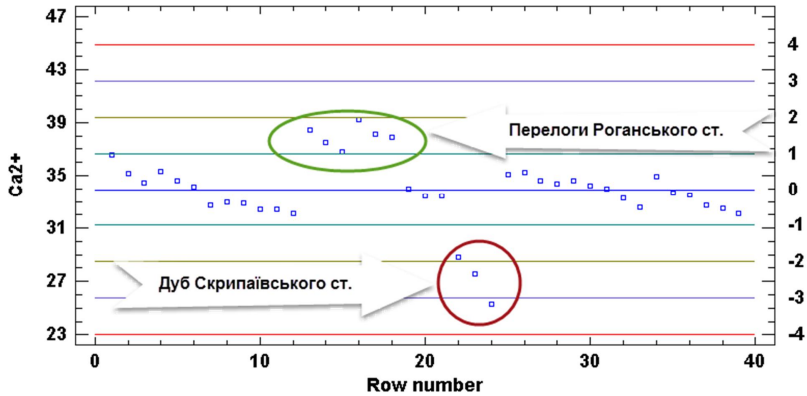


Рис. 3. Графік викидів показника  $Ca^{2+}$ , сформований програмою Statgraphics

Табл. 3. Екстремальні значення показника  $Ca^{2+}$

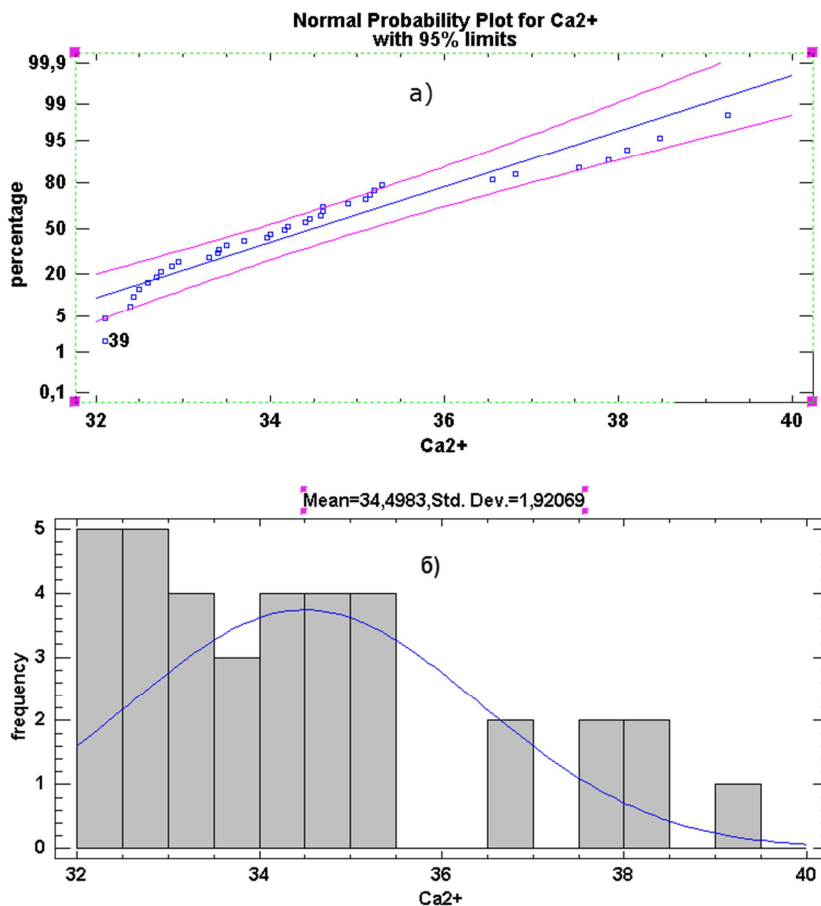
№ рядка	Значення ознаки	Коеф. відхилення
24	25,26	-3,18082
23	27,56	-2,33773
22	28,8	-1,88319
39	32,1	-0,673535
...		
14	37,54	1,32057
18	37,88	1,4452
17	38,1	1,52584
13	38,47	1,66147
16	39,26	1,95105

Аналіз даних таблиці дозволив на підготовчому етапі аналізу виявити грубу помилку: серія зразків ґрунту Скрипайвського стаціонару не належать до досліджуваних нами чорноземів звичайних (вони належать до сірих-опідзолених), та можуть бути обґрунтовано виключені з подальшого дослідження. Інші відхилення (перелogi Роганського стаціонару) не є критичними, але причини їх появи можуть бути визначені окремо. Потребують перевірки також дані по нижньому шару ріллі Михайлівського стаціонару.

4. На сьогоднішній день більшість методів, що використовується в аналізі даних агрономічних досліджень базується на припущенні про їх «приблизний нормальний розподіл». Тому, **перевірка даних на відповідність закону нормального розподілу** є обов'язковим.

Після вилучення з вибірки артефактів (об'єктів Скрипайвського стаціонару) та виправлення можливих помилок проведемо обов'язкову перевірку на нормальність. Традиційно така перевірка проводиться шляхом побудови графіків розподілу, побудованих за допомогою ППП Statgraphics Centurion XVII (рис. 4).

Графік 4а свідчить, що абсолютна більшість точок знаходиться в межах заданого (95 %) довірчого інтервалу, крім точки 39, що відповідає зразку «глибина 20-40 см»; рілля; Михайлівський Степ). Порівняння гістограми розподілу з теоретичною кривою нормального розподілу (4б) свідчить про приблизну відповідність показників  $\text{Ca}^{2+}$  закону нормального розподілу.



**Рис. 4. Відображення розподілу величини  $\text{Ca}^{2+}$  за допомогою:**  
**а) графіка нормальної ймовірності б) гістограми розподілу**

Отриманий тип гістограми розподілу – т.зв. «обривчаста гребінка з островом» та невеликою скошеністю в бік більших значень (що припустимо) досить розповсюджений в агрономічних дослідженнях. Така гістограма свідчить

про наступне: а) перемішані зразки різного типу, або помилки виміру – «острів»; б) виникає коли ширина стовпчика не кратна одиниці виміру (як у нашому випадку) або при помилці реєстрації – «острів»; в) при неможливості забезпечити достатню точність вимірів виникає «обрив» і відповідно, за рахунок округлення даних у більшу сторону виникає правостороння асиметрія розподілу.

#### 4. Показники описової статистики змінної $\text{Ca}^{2+}$

Average	34,4983
Median	34,185
Mode	34,6
Coeff. of variation	5,56748%
Std. skewness	2,10103
Std. kurtosis	0,160745

Той самий висновок підтверджується одночасно отриманими табличними даними результатів тестування, згенерованими в програмі Statgraphics. Нульова гіпотеза про відповідність досліджуваних величин закону нормального розподілу підтверджується відповідними тестами. У нашому випадку при рівні значущості  $\alpha = 0,05$ , для виборки, що перевіряється нульова гіпотеза (Null hypothesis), приймається ( $P\text{-value} = 0 < 0,05$ ). Висновок також підтверджується відповідними показниками описової статистики показника  $\text{Ca}^{2+}$  (табл.4):

- ✓ тотожність показників положення (моди, медіани, мередньої арифметичної);
- ✓ коефіцієнт варіації не перевищує 6 %;
- ✓ показник кривизни з правого боку виходить за межі  $\pm 2$ , а крутості знаходиться в цих межах.

Підтвердивши таким чином приналежність досліджуваних змінних до нормального закону розподілу, отримуємо обґрунтовану можливість використання абсолютної більшості методів статистичного моделювання.

**Висновки.** Розглянута послідовність підготовки первинних експериментальних даних статистичного обробітку з використанням комп'ютерних технологій. Надано: 1. Правила організації ознакового простору у вигляді таблиці «об'єкт-ознака». 2. Принципи вибору та використання необхідного програмного забезпечення. 3. Обґрунтування вибору класифікаційної шкали ознак, що досліджуються. 4. Наведено приклади комп'ютерної перевірки експериментальних даних. 5. Розглянуто порядок перевірки даних на відповідність закону нормального розподілу.

Проведення підготовчого етапу аналізу експериментальних даних за наведеним алгоритмом дозволить значно прискорити процес аналізу даних на основі комп'ютерних технологій, підвищити якість, зрозумілість та обґрунтованість отриманих висновків.

#### СПИСОК ЛІТЕРАТУРИ / REFERENCES

Бондарчук С. С. Основы практической биостатистики: учебное электронное издание / С. С. Бондарчук. – Томск: ТПУ, 2010. – [Електрон. ресурс]. – Режим доступу: <http://koi.tspu.ru/biostat/Foundations%20of%20practical%20biostatistics.pdf>.

Bondarchuk S. S., 2010, "Osnovu praktycheskoy byostatystyky. Uchebnoe elektronnoe yzdanye", Tomsk, ТРНУ, [Elektron. resurs], Rezhym dostupu: <http://koi.tspu.ru/biostat/Foundations%20of%20practical%20biostatistics.pdf>.

**Мінько** О. О. Сучасний аналіз даних в Excel для науковців / О. О. Мінько. – К.: Науково-учбовий центр прикладної інформатики НАН України, 2015. – 476 с.

Min'ko O. O., 2009, "Suchasnyu analiz danykh v Excel dlya naukovtsiv", K., Naukovouchoyvouy tsestr prykladnoyi informatyky NAN Ukrayiny, 476 p.

**Воробійов** К. П. Формат сучасної журнальної публікації за результатами клінічного дослідження / К. П. Воробійов // Морфологія. – 2009. – № 4, Т. III. – С. 102-114.

Vorobyov K. P., 2009, "Format suchasnoyi zhurnal'noyi publikatsiyi za rezul'tatamy klinichnoho doslidzhennya", Morfolohiya, № 4, Т. III, P. 102-114.

**Мазур** Н. П. Статистичні критерії як інструмент аналізу отриманих даних моніторингового дослідження / Н. П. Мазур // Проблеми сучасної педагогічної освіти. Сер.: Педагогіка і психологія. – Ялта, 2014. – № 44(2). – С. 190-197. – [Електрон. ресурс]. – Режим доступу: [http://elibrary.kubg.edu.ua/7956/1/N\\_Mazur\\_PSP0\\_44%282%29\\_IS.pdf](http://elibrary.kubg.edu.ua/7956/1/N_Mazur_PSP0_44%282%29_IS.pdf).

Mazur N. P., 2014, "Statystychni kryteriyi yak instrument analizu otrymanykh danykh monitorynhovoho doslidzhennya", Problemy suchasnoyi pedahohichnoyi osvity. Ser.: Pedahohika i psykholohiya, Yalta, № 44(2), P. 190-197, [Elektron. resurs], Rezhym dostupu: [http://elibrary.kubg.edu.ua/7956/1/N\\_Mazur\\_PSP0\\_44%282%29\\_IS.pdf](http://elibrary.kubg.edu.ua/7956/1/N_Mazur_PSP0_44%282%29_IS.pdf).

**Верес** Ю. Шкалювання при розподілі обмежених ресурсів / Ю. Верес, А. Катренко // Вісн. Нац. ун-ту "Львів. політехніка". – 2009. – № 650. – С. 205-213.

Veres Yu. Katrenko A., 2009, "Shkalyuvannya pry rozpodili obmezenykh resursiv", Visn. Nats. un-tu "L'viv. politekhnika", № 650, P. 205-213.

**Шитиков** В. К. Методы синтетического картографирования территории (на примере эколого-информационной системы Volgabas) / В. К. Шитиков, Г. С. Розенберг, Н. В. Костина // Количественные методы экологии и гидробиологии: Сб. науч. тр., посвященный памяти А. И. Баканова/ Отв. ред. чл.-корр. РАН Г.С. Розенберг. – Тольятти: СамНЦ РАН, 2005. – С. 167-227. – [Електрон. ресурс]. – Режим доступу: <http://www.ievbras.ru/ecostat/Kiril/Article/A16/Volgabas.htm>.

Shytykov V. K., Rozenberh H. S., Kostyna N. V., 2005, "Metodu syntetycheskoho kartohrafyrovannyau terrytoryu (na prymerе ekoloho-ynformatsyonnoy systemy Volgabas)", Kolychestvennye metody ekolohyy u hydrobyolohyy (Sbornyk nauchnikhtrudov, posvyashchennyy ramyaty A. Y. Bakanova), Otv. red. chl.-korr. RAN H. S. Rozenberh, Tol'yatty, SamNTs RAN, P. 167-227, [Elektron. resurs], Rezhym dostupu: <http://www.ievbras.ru/ecostat/Kiril/Article/A16/Volgabas.htm>

**Тихоненко** Д. Г. Еволюція і класифікація агрогенних ґрунтів України / Д. Г. Тихоненко // Вісн. Чернів. ун-ту. Сер. Біологія. – Чернівці, 2012. – С. 96-100.

Tykhonenko D. G., 2012, "Evolutsiya i klasyfikatsiya ahrohennykh gruntiv Ukrayiny", Visn. Cherniv. un-tu. Ser. Biolohiya, Chernivtsi, P. 96-100.

**Герасимова** М. Классификация почв России: обсуждение на сайте в Интернете / М. Герасимова, С. Хохлов // Почвоведение (Eurasian Soil Science). – 2010. – № 12. – С. 1449–1459.

Herasyмова M. Khokhlov S., 2010, "Klasyfykatsyya pochv Rossyy: obsuzhdenye na sayte v Ynternete", Pochvovedenye (Eurasian Soil Science), № 12, P. 1449–1459.

**Милль** Джон Стюарт Система логики силлогистической и индуктивной: Изложение принципов доказательства в связи с методами научного исследования / Джон Стюарт Милль. – М.: ЛЕНАНД, 2011. – 832 с.

Myll' Dzhon Styuart, 2011, "Systema lohyky syllohystycheskoy u ynduktyvnoy: Yzlozhenye pryntsupov dokazatel'stva v svyazy s metodamy nauchnoho yssledovannya", M., LENAND, 832 p.

**Document** TGP/8, 2003, "Use of statistical procedures in distinctness, uniformity and stability testing", Associated Document to the General Introduction to the Examination of Distinctness, Uniformity and Stability and the Development of Harmonized Descriptions of New Varieties of Plants (document TG/1/3), International union for the protection of new varieties of plants, Geneva, 212 p.