

УДК 004(92)+ 681:3.06(075.8)

Ю.О. Сотников, В.М. Сирий

Харківський національний аграрний університет ім. В. В. Докучаєва

ОСОБЛИВОСТІ ВИКОРИСТАННЯ ЕЛЕКТРОННИХ ТАБЛИЦЬ В АГРОНОМІЧНИХ ДОСЛІДЖЕННЯХ

Постановка проблеми. Зазвичай сенс проведення наукового експерименту в аграрній сфері полягає в побудові узагальнюючих висновків на базі обмеженої кількості первинних даних польового дослідження. Базові навички роботи з електронними таблицями (ЕТ) принципово дозволяють виконати їх обробку. Але проблема виникає тоді, коли дослідник стикається з необхідністю ефективно опрацювати великі масиви, поглибити аналіз у середовищі інших, професійних програм.

Аналіз останніх досліджень і публікацій. ЕТ, насамперед, створені для автоматизації обробки структурованих даних. Побудова таблиць та

система управління даними в них базується на стандартних алгоритмах роботи, реалізованих в абсолютній більшості програм [1]. Але інструменти ЕТ для аналізу даних, на жаль, не охоплюють повною мірою увесь спектр задач їх статистичної обробки [2], ускладнюють технології управління та обміну даних, а також їх експорт до іншого програмного забезпечення. В [3-7] вказується на недостатні можливості ЕТ для аналізу даних та низький рівень довіри до отриманих результатів.

Проблемою є і обмежені уявлення пересічних користувачів про ефективні можливості сучасних ЕТ, зокрема, попередньої підготовки даних для обробки іншими, професіональними пакетами.

Формулювання цілей статті. Автори статті мають на меті дати узагальнені рекомендації доцільності використання MS Excel в аналізі даних агрономічних досліджень, сформулювати основні правила створення масивів даних, їх трансформації, та перевірки валідності, розглянути ефективні технології аналізу даних в MS Excel, зокрема при використанні інструментів *Перевірка даних*, *Фільтр*, *Проміжні підсумки та Зведені таблиці*. Більш детальні методичні рекомендації авторів викладені в роботі [8].

Виклад основного матеріалу досліджень. На наш погляд, ЕТ доцільно використовувати у випадках: створення великих масивів даних для подальшого їх використання в спеціалізованих програмах статистичного моделювання; проведення валідації та верифікації отриманих даних; розрахунку інтегрованих показників та перекодування даних; проведення їх попереднього порівняльного аналізу; візуального визначення існування істотної різниці між варіантами.

В роботі обговорюються найважливіші аспекти застосування ЕТ у наведених випадках.

Створення масиву (списку) даних є першим етапом обробки інформації, який повинен відповідати наступним вимогам: в таблиці не повинно бути пустих рядків, колонок, формул та об'єднаних клітинок; кожне поле у першій комірці повинне мати заголовок (мітку); поля не містять пустих записів та груп даних, що повторюються.

Для ефективного заповнення даними комірок таблиці доцільно застосовувати стандартні *функції маркера автозаповнювання*. Для роботи з великими таблицями можна скористатися закріпленням рядків і колонок за мітками даних або інструментом *Форма*.

Приклад трансформації даних у вигляді таблиці (списку) наведений на рис. 1. Отримана база даних розмірністю $n \times m$, де n – загальна кількість об'єктів дослідження, m – кількість ознак (властивостей). Обов'язковим є рядок міток, для зручності додано стовпець з індексами об'єктів.

Особливості форматування отриманої таблиці та порядок розміщення колонок і рядків не мають принципового значення. При збільшенні кількості об'єктів або ознак таблиця даних масштабується.

Створення нових змінних (ознак) та перекодування. В процесі аналізу часто виникає необхідність створити нові розрахункові змінні (наприклад, нормалізувати первинні дані шляхом логарифмування), представити первинні данні в вигляді рангової шкали. Більшість статистичних процедур аналізу даних працює з числовими ознаками. Тому, актуальним є автоматичне перекодування текстових змінних в чисельні. Ці процедури просто реалізуються за допомогою стандартних функцій Excel.

Вплив густини стояння рослин та щільювання з підживленням на біометричні показники цибулі ріпчастого сорту "Золотистий"

Густина стояння рослин, тис./га	Рік	Маса, г			
		листя		цибулини	
		Період обліку			
		1	2	1	2
400	2011	33,0	27,0	28,5	47,0
	2012	43,0	19,2	16,0	25,0
	2013	18,0	20,0	17,0	69,0
	2014	27,8	42,3	20,9	63,8
	В середньому	30,5	27,1	20,6	51,2
600	2011	28,5	24,5	23,0	46,5
	2012	36,0	18,6	14,0	25,5
	2013	18,3	21,0	21,3	68,7
	2014	27,0	38,4	19,0	66,5
	В середньому	27,5	25,6	19,3	51,8

1 - Утворення цибулини

2 - Збирання врожаю

а) Презентаційна таблиця, непридатна для аналізу

№ запису	Густина стояння рослин, тис./га	Рік	Період обліку	Маса листя, г	Маса цибулини, г
1	400	2011	Утворення цибулини	33,0	28,5
2	400	2012	Утворення цибулини	43,0	16,0
3	400	2013	Утворення цибулини	18,0	17,0
4	400	2014	Утворення цибулини	27,8	20,9
5	400	2011	Збирання врожаю	27,0	47,0
6	400	2012	Збирання врожаю	19,2	25,0
7	400	2013	Збирання врожаю	20,0	69,0
8	400	2014	Збирання врожаю	42,3	63,8
9	600	2011	Утворення цибулини	28,5	23,0
10	600	2012	Утворення цибулини	36,0	14,0
11	600	2013	Утворення цибулини	18,3	21,3
12	600	2014	Утворення цибулини	27,0	19,0
13	600	2011	Збирання врожаю	24,5	46,5
14	600	2012	Збирання врожаю	18,6	25,5
15	600	2013	Збирання врожаю	21,0	68,7
16	600	2014	Збирання врожаю	38,4	66,5

б) Правильно трансформовані дані (таблиця "об'єкт – ознака")

Рис. 1. Трансформація даних для комп'ютерного аналізу

Перевірка валідності отриманих даних, та їх розподілу на «нормальність». Перед обробкою даних слід переконатися, що вони введені правильно і мають коректні значення. При введенні великих масивів практично неможливо уникнути грубих помилок. Найпоширенішими з них є ввід замість числових даних тексту, неможливих від’ємних значень, зайвих розрядів, неприпустимих символів. Часто джерелом помилок є дані, імпортовані з інших програм, сайтів або введені скануванням з паперових носіїв.

Для перевірки даних при вводі можна використати інструмент *Перевірка даних* у групі *Знаряддя даних* з вкладки *Дані*.

Перевірку уведених даних можна здійснити також за допомогою інструменту *Описова статистика* з надбудови *Пакет аналізу*, або використанням статистичних функцій. Для цього, на окремому аркуші для діапазонів даних, що перевіряються, застосовують формули обраних показників описової статистики. Для перевірки може бути обрано будь-який критерій, наприклад, "правило 3σ ". Під час перевірки даних доцільно провести попередню оцінку отриманого розподілу "на нормальність". Для візуалізації помилок ЕТ можна настроїти так, щоб неприпустимі значення обводилися контурами, які б після корегування даних автоматично приховувалися.

Нашвидкуруч перевірку даних можна виконати за допомогою технології фільтрації. Для цього слід скористатися кнопкою *Фільтр* у групі *Сортування й фільтр* з вкладки *Дані*. Якщо відсортувати дані від найбільшого до найменшого, пусті клітинки і максимальні показники опиняться на початку списку, мінімальні - в кінці. Після редакції аномальних даних можна повернутися до первинного порядку за допомогою поля з номерами записів. Проведення попереднього порівняльного аналізу. Найбільш розповсюдженим видом аналізу в агрономічних дослідженнях є різноманітні способи порівняння групових середніх. Отримання первинної таблиці групових середніх ознак за обраними варіантами для багатьох дослідників є проблемою. Для її подолання доцільно застосувати інструмент *Проміжні підсумки* з меню *Дані*. На рис. 2 наведена технологія розрахунку середніх значень показників маси листя та цибулини за фазами розвитку та варіантами посіву, за даними рис. 1.

1. Сортування даних за ключами "Період обліку", Густина стояння" та "Рік". Слід пам'ятати, що від коректно проведеного сортування залежить правильність отриманого результату. 2. Запит інструменту *Підсумки* з групи *Структура* на вкладці *Дані*. У полі *За кожної зміни* слід вказати дані "Густина стояння". У полі *Використовувати функцію* - обрати

функцію *Середнє*. У полі *Додати підсумки до* - вказати дані "Маса листя" та "Маса цибулини". Повтор п 2 з вибором у полі *За кожної зміни* даних поля "Період обліку". В результаті отримана бажана дворівнева таблиця групових середніх. (рис. 2, б). Для того, щоб скористатися її копією можна виділити лише видимі комірки таблиці за допомогою інструменту *Виділити* у групі *Редагування* вкладки *Головна*.

Застосування зведених таблиць (ЗТ). ЗТ є одним з універсальних та корисних інструментів будь-яких ЕТ. Особливо ефективні їх можливості при обробці великих масивів даних. Наведемо лише декілька корисних прикладів їх практичного застосування в агрономічних дослідженнях.

1	2	3	4	A	B	C	D	E	F
		№ запису	Густина стояння рослин, тис./га	Рік	Період обліку	Маса листя, г	Маса цибулини, г		
		1							
		2	5	400	2011	Збирання врожаю	27,0	47,0	
		3	6	400	2012	Збирання врожаю	19,2	25,0	
		4	7	400	2013	Збирання врожаю	20,0	69,0	
		5	8	400	2014	Збирання врожаю	42,3	63,8	
		6		400	Середнє		27,1	51,2	
		7	13	600	2011	Збирання врожаю	24,5	46,5	
		8	14	600	2012	Збирання врожаю	18,6	25,5	
		9	15	600	2013	Збирання врожаю	21,0	68,7	
		10	16	600	2014	Збирання врожаю	38,4	66,5	
		11		600	Середнє		25,6	51,8	
		12			Збирання врожаю Середнє		26,4	51,5	
		13	1	400	2011	Утворення цибулини	33,0	28,5	
		14	2	400	2012	Утворення цибулини	43,0	16,0	
		15	3	400	2013	Утворення цибулини	18,0	17,0	
		16	4	400	2014	Утворення цибулини	27,8	20,9	
		17		400	Середнє		30,5	20,6	
		18	9	600	2011	Утворення цибулини	28,5	23,0	
		19	10	600	2012	Утворення цибулини	36,0	14,0	
		20	11	600	2013	Утворення цибулини	18,3	21,3	
		21	12	600	2014	Утворення цибулини	27,0	19,0	
		22		600	Середнє		27,5	19,3	
		23			Утворення цибулини Середнє		29,0	20,0	
		24			Загальне середнє		27,7	35,7	
		25							

а) Підсумки I рівня

1	2	3	4	A	B	C	D	E	F
		№ запису	Густина стояння рослин, тис./га	Рік	Період обліку	Маса листя, г	Маса цибулини, г		
		1							
		6		400	Середнє		27,1	51,2	
		11		600	Середнє		25,6	51,8	
		12				Збирання врожаю Середнє	26,4	51,5	
		17		400	Середнє		30,5	20,6	
		22		600	Середнє		27,5	19,3	
		23				Утворення цибулини Середнє	29,0	20,0	
		24				Загальне середнє	27,7	35,7	
		25							

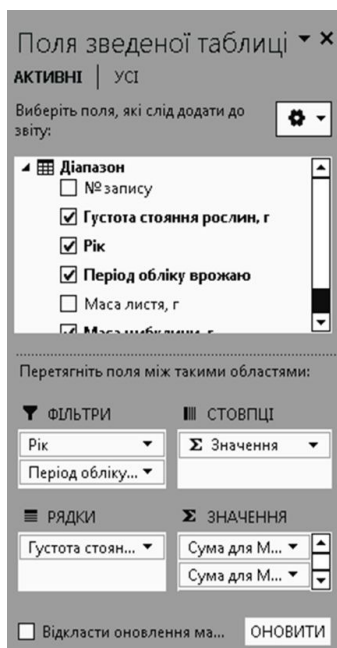
б) Підсумки II рівня

Рис. 2 Підготовка багаторівневих підсумків

Наприклад, потрібно розрахувати абсолютну та відносну різницю середніх показників маси цибулин між двома варіантами густини стояння рослин. Для цього можна застосувати наступну технологію.

1. Для створення первісного макету ЗТ на вкладці *Вставка* слід скористатися кнопкою *Зведена таблиця*. Макет ЗТ формується в діалоговому вікні *Поля зведеної таблиці* (рис. 3, а). У прикладі поле "Густина стояння" розташовується в області *Рядки*, а "Маса цибулини" в області *Значення* у відповідності до потрібної кількості результативних показників. В області *Фільтри* застосовуються поля "Рік" і "Період обліку".

2. Модифікація макету ЗТ здійснюється вибором з контекстного меню для обраних полів опцій *Параметри значення поля* та *Відобразити значення як*. У прикладі, в якості поля, обраний показник густоти стояння, в якості елемента (бази порівняння) – значення 400 (рис. 3, б).



а) Робота з макетом

	В	С	Д	Е
1	Рік	All		
	Період обліку врожаю	збору вро...		
		Середнє значення Маса цибулини, г	Середнє значення Маса цибулини, г	Середнє значення Маса цибулини, г
4	Позначки рядків			
5	400	51,275	100,00%	
6	600	51,8	0,525	101,02%
7				

Callouts in the image point to the following cells:

- Середнє**: points to the cell containing '51,275' in row 5, column C.
- Частка**: points to the cell containing '0,525' in row 6, column C.
- Зведена відмінність**: points to the cell containing '101,02%' in row 6, column D.
- Відмінність**: points to the cell containing '1,02%' in row 6, column E.

б) Модифікація зведеної таблиці

Рис. 3 Створення зведеної таблиці

Окремої уваги заслуговує можливість групувати дані ЗТ за датами. Потреба у використанні цієї можливості виникає, наприклад, при необхідності отримати зведені дані за окремі проміжки часу, накопичувальні підсумки, визначення структури показників за часом та інше.

Для розв'язання цієї задачі з контекстного меню для поля "Дата" обирається опція *Групувати* і потрібний термін часу (у прикладі 10 днів).

Далі, за описаним вище алгоритмом, розраховується маса цибулин із зростаючим підсумком (рис. 4).

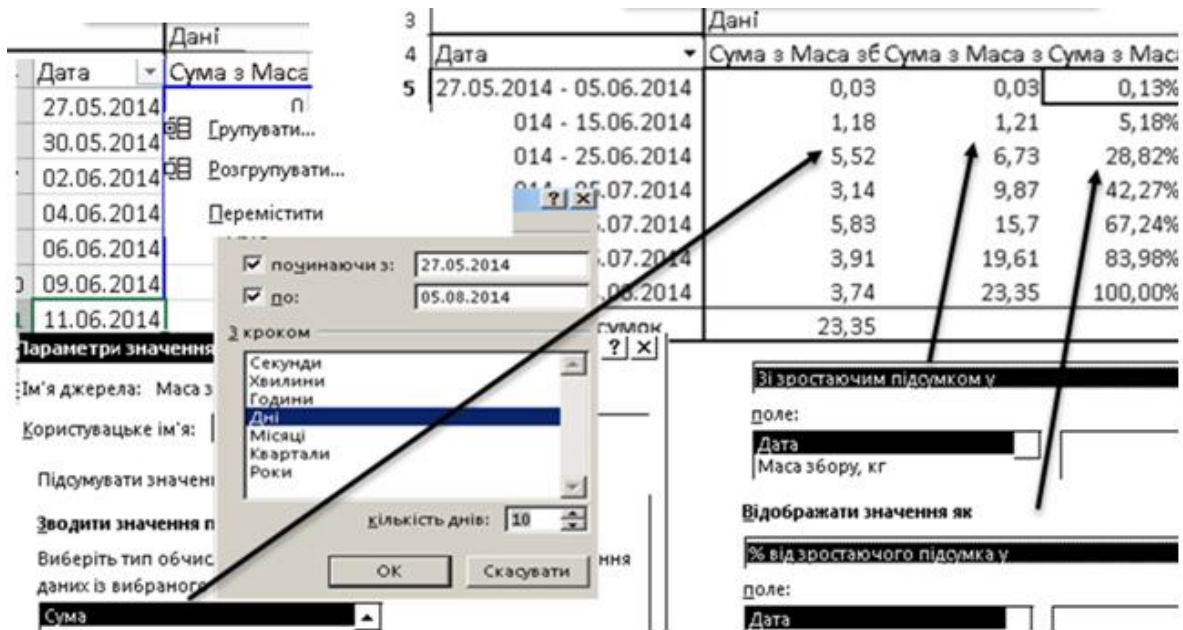


Рис. 4. Зведена таблиця для аналізу груп даних

Візуальне визначення існування істотної різниці між варіантами. За допомогою графічної інтерпретації даних ЗТ можна попередньо перевірити існування істотної різниці між варіантами, що досліджується (рис. 5). Цей спосіб не дає чисельної оцінки, але дозволяє швидко та наглядно оцінити саме існування різниці. В меню *Конструктор, Додати елемент діаграми* можна встановити потрібні *Планки похибок*.

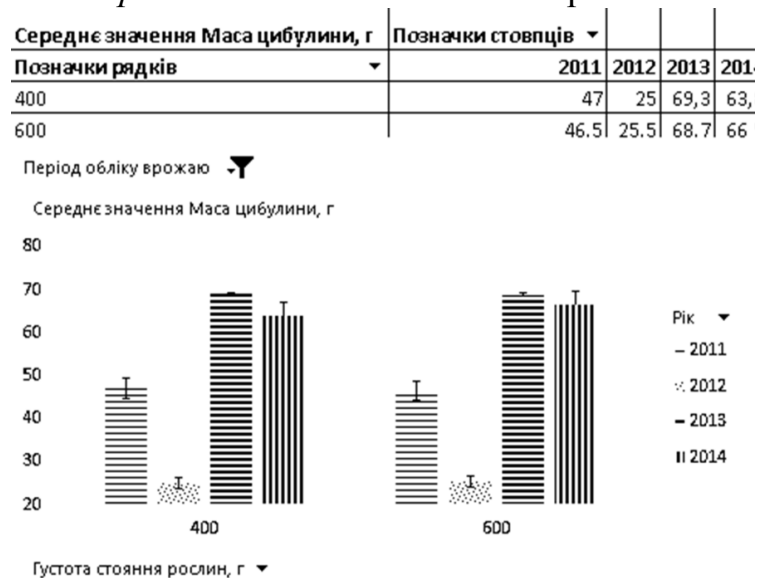


Рис. 5. Візуальна оцінка результатів дослідження

Висновки. Оволодіння технологіями, які рекомендовані в даній статті, дозволить значно підвищити ефективність процесів обробки та аналізу даних спеціалістами з агрономічних спеціальностей.

Матеріал статті може бути використаний пересічними користувачами для подальшого поглибленого вивчення можливостей електронних таблиць в аналізі даних польового досліджу.

Бібліографічний список. 1. Кирюшин Б.Д. и др. Основы научных исследований в агрономии /Б. Д. Кирюшин, Р. Р. Усманов, И.П. Васильев. - М.: Колосс, 2009. -398 с. 2. Остапенко Р.И. Математические основы психологии: Учеб.-метод. пособ.- Воронеж: ВГПУ. – 2010.- 76 с. 3. Адамов А.А. Теория вероятностей и математическая статистика. Прикладная статистика с использованием MS Excel: учеб. пособие / А.А. Адамов. – Пермь: Изд-во Перм. гос. техн.ун-та, 2008.- 174 с. 4. Саймон, Джинжер Анализ данных в Excel: наглядный курс создания отчетов, диаграмм и сводных таблиц. : - М.: Издательский дом "Вильямс", 2004. – 528 с. 5. Ивантер Э.В., Коросов А.В. Введение в количественную биологию. – Учеб. пособ. - Петрозаводск: Петрозаводский гос. ун-т, 2010. – 302 с. 6. Джелен Билл, Александер Майкл. Сводные таблицы в Microsoft Excel 2013. - М.: ООО "И.Д. Вильямс", 2013. –448 с. 9. 7. Мазур Н.П. Статистичні критерії як інструмент аналізу отриманих даних моніторингового дослідження. 8. Сотников. Ю.О., Сирий В.М. Информатика: Приклади використання електронних таблиць в агрономічних дослідженнях. [Електронний ресурс]: методичні вказівки - Харківський національний аграрний університет ім. В.В. Докучаєва, 2015. – Режим доступу: <http://knau.kharkov.ua/kafedra-kiber.html>

Сотников Ю.А., Сырый В.Н. Особенности использования электронных таблиц в агрономических исследованиях. Разработаны унифицированные рекомендации по применению эффективных технологий обработки данных агрономических исследований с помощью электронных таблиц. Рассмотрены возможности автоматизированной проверки данных и их предварительного анализа. Результаты могут быть использованы в среде специализированных программ.

Sotnikov Y.O., Syry V.M. Application of spreadsheet in agricultural researches. Spreadsheet use general problems in agricultural research data analysis and ways of their resolutions. Purpose: Provision to researches of uniform rules, guidelines, and effective technologies in creating and processing of the large data volumes in MS Excel (Windows, Microsoft Office) and its

analogues. Results: The article contains recommended options of the spreadsheet use expediency, general rules for data sets creation and primary information transformation, effective methods of automated data verification, data preliminary analysis and preparation for its use in specialized programs. Technology of the automatic cell filling marker function, window shapes, data validity verification, sorting, filtering, elements of descriptive statistics, built-in features and error visualization tools are proposed for the data sets creation. Efficient technologies of the group-averages tables using subtotals calculation are recommended for the preliminary comparative analysis. Preliminary result preparation methods for the use and further in-depth analysis with specialized software are shown. Particular attention is paid to discussing the possibilities in grouped tables application. Useful examples of their use in agricultural research are given. In particular, we show methods of the absolute and relative difference calculations, data grouping, cumulative conclusions receiving, parameters structure defining, and graphic interpretation of the grouped tables for the existence visual determination of the significant differences between the variants.