

UDC 004(92)+ 681:3.06(075.8)

Yu. O. Sotnykov, Cand. Sci. (Econ.), Assistant Professor

D. V. Gavva, Cand. Sci. (Agric.), Assistant Professor

Kharkiv National Agrarian University named after V. V. Dokuchaev

Kharkiv, e-mail: yesaul.halt@knau.kharkov.ua

SYSTEMATIZATION DATA OF THE AGRONOMIC EXPERIMENT USING SPREADSHEETS

The possibility of using Excel spreadsheets for primary systematization of data obtained as a result of the agronomic experiment is considered. For realizing the systemic entry in the science-related documents, it is necessary to carry out the procedure for re-assigning the data to the signatures of the values, including the data stratification. The purpose of such a systematization is to make the studied system understandable and convenient, simplified, more appropriate for the purpose of the study.

For this purpose, the collected data to convert a measuring scale. Measurement Scale - a sign rules for granting studied numerical value. The type of scale chosen principally zalezhateme choice of methods confirm the hypothesis of the study.

The number of possible benchmarking criteria influences the choice of type of scale. This means that the criterion can serve as an alternative to any of its sign, the value of which can be detected as one of the possible scales. After such criteria comparable to investigate found, it is possible to compare experiment variations.

The purpose of the research is to describe in a simplified and accessible form the basic methods and methods of systematization of the data of the agronomic experiment. The research methodology was to analyze the effectiveness of the tools of systematization of experimental data at the preparatory stage of the study.

The main purpose of this publication is to demonstrate the Excel spreadsheet processor to organize data in different types of scales.

The problem of data systematization is a cornerstone in the solution of the ABC-XYZ problem in market-making and logistics, VEN analysis-in the pharmaceutical industry, defined structures and quality of the crop in vegetable-growing, in the course of forest tax works, etc.

The practical use of these data systematization tools will greatly accelerate the analysis process, improve the quality, and make the conclusions understood.

Effective recomputation and data scaling with auto filters and formulas =IF(), =VLOOKUP(), =CONCATENATE(). Creating tables of crosstabulation using pivot tables.

Keywords: *data stratification, Spreadsheet, scales of measurements, data transformation.*

УДК 004(92)+ 681:3.06(075.8)

Ю. А. Сотников, канд. екон. наук, доцент**Д. В. Гавва, канд. с.-х. наук, доцент**

*Харьковский национальный аграрный университет им. В. В. Докучаева
г. Харьков, e-mail: yesaul.halt@knau.kharkov.ua*

СИСТЕМАТИЗАЦИЯ ДАННЫХ АГРОНОМИЧЕСКОГО ЭКСПЕРИМЕНТА С ПОМОЩЬЮ ЭЛЕКТРОННЫХ ТАБЛИЦ

Рассмотрены возможности использования электронных таблиц Excel для первичной систематизации данных, полученных в результате проведения агрономического эксперимента. Продемонстрировано эффективное перекодирование и шкалирование данных с помощью автофильтров и формул =IF(), =VLOOKUP(), =CONCATENATE(). Создание таблиц кростабуляции с использованием сводных таблиц.

Ключевые слова: систематизация данных, электронная таблица, сводная таблица, шкалы измерений, преобразование данных.

УДК 004(92)+ 681:3.06(075.8)

Ю. О. Сотников, канд. екон. наук**Д. В. Гавва, канд. с.-г. наук, доцент**

*Харківський національний аграрний університет ім. В. В. Докучаєва
м. Харків, e-mail: yesaul.halt@knau.kharkov.ua*

СИСТЕМАТИЗАЦІЯ ДАНИХ АГРОНОМІЧНОГО ЕКСПЕРИМЕНТУ ЗА ДОПОМОГОЮ ЕЛЕКТРОННИХ ТАБЛИЦЬ

Розглянуто можливості використання електронних таблиць Excel для первинної систематизації даних, отриманих у результаті проведення агрономічного експерименту. Продемонстровано ефективні способи перекодування та шкалювання даних за допомогою автофільтрів та формул =IF(), =VLOOKUP(), =CONCATENATE(). Створення та аналіз таблиць кростабуляції за допомогою реманенту «Зведені таблиці».

Ключові слова: систематизація даних, електронна таблиця, зведена таблиця, шкали вимірювання, перетворення даних.

Вступ. Після організації ознакового простору даних, отриманих у результаті проведення наукового експерименту та процедури їх очищення (Сотников Ю. О., 2015), настає потреба в їх осмисленні та первинній систематизації. Слушною є думка Умберто Еко що, «Деяка система починає діяти, наділяючи змістом щось таке, що спочатку цього змісту не мало, при

цьому деякі елементи цього «щось» зводяться в ранг означаючих. Але поки системи немає і поки це щось не наділене кодом, у ньому можливо безліч комбінацій, які знайдуть сенс тільки ПІЗНІШЕ, після накладання будь-якої системи» (Еко У., 2006).

Для реалізації системного підходу в наукових дослідженнях необхідно провести процедуру перевпорядкування даних відповідно до означених цілей, висунутих гіпотез та поглядів (*data stratification*). Мета такого перевпорядкування (систематизації) – зробити досліджувану систему зрозумілою та зручною, спрощеною, більш адекватною меті дослідження (Клир Дж., 1990, Джеффри Ч., 1980).

З цією метою зібрані дані потрібно перетворити у вимірювальні шкали. *Шкала вимірювання – це правила надання досліджуваній ознаці чисельного значення.* Від типу обраної шкали принципово залежатиме вибір методів підтвердження гіпотез дослідження. На вибір типу шкал впливає кількість можливих альтернатив (критеріїв) порівняння. Це означає, що критерієм якості альтернативи може слугувати будь-яка її ознака, значення якої можна зафіксувати в одній з можливих шкал. Після того такий критерій, зіставний з метою дослідження, буде знайдений, виявляється можливість порівняння варіантів експерименту.

На складність задач порівняння впливає кількість критеріїв. Багатокритеріальність реальних задач пов'язана не тільки з множиною цілей експерименту, але і з тим, що одну мету практично неможливо означити одним критерієм. З одного боку, багатокритеріальність є способом підвищення адекватності опису мети дослідження, а з іншого, – збільшує складність вирішення задачі. Тому потрібно, за можливістю мінімізувати кількість використовуваних критеріїв за максимального обґрунтування мети. Для упорядкування критеріїв у системному аналізі використовують три основних атрибути вимірювальних шкал, які визначають приналежність шкали до тієї чи іншої категорії вимірів, що формується у вигляді шкал.

- *упорядкованість* даних означає, що один пункт шкали відповідає вимірюваній властивості, більше, менше або дорівнює іншого пункту;

- *інтервальність* даних означає, що інтервал між будь-якими парою чисел, що відповідають вимірюваним властивостям, більший, менший або дорівнює інтервалу між іншою парою чисел;

- *точка відліку* означає, що набір чисел, які відповідають вимірюваним властивостями, має точку відліку, визначену за нуль, що відповідає повній відсутності властивості, що вимірюється.

Виходячи з викладеного вище, візьмемо собі за правило: *одна гіпотеза – одна властивість – одна шкала* (Hedden, Heather, 2010).

Розглянемо наступні найбільш уживані типи шкал: найменувань, порядкова, інтервалів, відносин, різниць, абсолютна.

Шкала найменувань використовується для віднесення одиниці

спостереження до однієї з груп. Кожна група має своє унікальне позначення, що не збігається ні з яким іншим. Змінні, представлені в шкалі найменувань, називаються номінальними. Арифметичні операції (додавання – віднімання, множення – ділення) і операція порівняння або відношення (більше або менше) не мають рації у використанні до даних шкали найменувань. Призначення цієї шкали полягає в проведенні якісної класифікації (віднесення до групи) без виконання будь-яких арифметичних операцій та операцій порівняння. Характеристикою є абсолютне число спостережень або їх відсоток входження в групу. Статистика центральної тенденції – мода. Як коди можна використовувати цифри або будь-який інший набір символів.

Порядкова шкала призначена для об'єднання одиниць спостереження в групи. Для даних цієї шкали можливі операції порівняння або відношення (більше, менше), але не мають сенсу арифметичні операції. Позначення класів (груп) для зручності використання присвоюються таким чином, щоб можна було виділити порядок класів. Статистикою центральної тенденції для групи даних порядковій шкали є мода та медіана. У цій шкалі даних припустимими операціями є операції порівняння або відношення. Змінні, представлені в порядковій шкали, іменуються порядковими або ранговими. Такі статистичні характеристики, як середнє значення в групі, помилка середнього, дисперсія, не мають сенсу для такого типу даних. Для порядкової шкали застосовуються методи, засновані на понятті рангів та належать до непараметричних статистик.

Наведені вище дві шкали (найменувань та порядкові) призначені для опису якісних змінних. Опис кількісних змінних здійснюється в метричних шкалах, представлених інтервально та шкалою відносин.

Інтервальна шкала призначена не тільки для кількісного порівняння, але і для оцінки рівня відмінностей між одиницями спостереження, вона дозволяє як впорядковувати дані, так і чисельно відображати відмінності між ними. За даними, представленими в цій шкалі, припустимі операції додавання і віднімання, які дозволяють оцінити, «на скільки» дані різняться. Але дуже сумнівним є результат порівняння «у скільки разів відрізняються спостерігаються дані». Таким чином, інтервальні шкали припускають операцію додавання і не завжди мають властивість, що описується операцією множення.

У шкалі відношень з даними допустимі всі логічні і арифметичні операції, що не змінює структуру шкали. Окрім всіх властивостей інтервальної шкали, у шкалі відношень з'являється можливість розрахувати, у скільки разів одне значення більше або менше від іншого. Усі арифметичні операції з даними цієї шкали мають сенс. Статистики - середнє, медіана і мода - в метричній шкалі є мірою центральної тенденції.

Шкали відношень й інтервальна різняться між собою наявністю абсолютного початку координат, не надаючи принципового впливу на види статистичних показників. За інформативністю шкали можна розташувати в порядку зростання у такий спосіб: «найслабша» – шкала найменувань,

«міцніша» – порядкова шкала, найбільш «міцна» – метрична шкала (інтервальна і відношень). Перевести дані перекодуванням можна тільки в одному напрямку інформативності від більш «міцної» шкали до менш «міцної» із втратою певної міри інформативності.

Наведені теоретичні основи класифікації даних необхідні, по перше, завжди існує потреба зрозуміти й оцінити адекватність результатів досліджень. По-друге, тому що перед дослідником, який не володіє математичними методами обробки даних, може постати завдання зібрати дані і сформулювати мету їх обробки для спеціалістів, здатних використовувати математичні методи.

Метою досліджень є опис у спрощеній та доступній формі основних методів та прийомів систематизації даних агрономічного експерименту.

Об'єктом дослідження було обрано масив даних, який отримано у ході лабораторних досліджень чорноземів типових глибоких важкосуглинкових на лесах у межах південно-східного Лісостепу України.

Методика досліджень полягала в аналізі ефективності інструментів систематизації експериментальних даних на підготовчому етапі дослідження.

Результати та обговорення. Основним завданням цієї публікації є демонстрація табличного процесора Excel для систематизації даних у вигляді різних типів шкал. Проблема систематизації даних є наріжною у вирішенні задачі ABC-XYZ-аналізу в маркетингу та логістиці, VEN-аналізі – у фармацевтиці, визначені структури та якості врожаю в овочівництві, під час проведення робіт з лісової таксації, ґрунтознавчо-картографічних завдань тощо. Автоматизація цього процесу за допомогою ET Excel потребує відповідного рівня знань та навичок (Минько О. О., 2015). У ET Excel існує декілька можливостей перетворення даних.

1. *За допомогою автофільтрів.* Найбільш простий та швидкий, але найменш гнучкий, непрофесійний.

Організуємо ознаковий простір у формі таблиці «об'єкт-ознака». Попередньо дані не сортуємо. Створюємо окремий стовпчик для необхідного критерію. У нашому випадку Код_глиб – код глибин. Припустимо, нам необхідно об'єднати дані вологості у двох шарах ґрунту (опад та 0-5 – шар 1; 5-20 та 20-40 – шар 2) (рис. 1).

Алгоритм кодування наступний: Виділяємо таблицю. Активуємо інструмент «Дані-Автофільтр». Перший рядок перетворюється в «поля з незнадаючим списком». Активуємо список у стовпчику шар та виділяємо в ньому позиції що належить до критерію 1-2 верхніх шарів ґрунту (рис. 2). У першу клітинку стовпчика Код_глиб вводимо необхідну позначку коду – цифру 1 та копіюємо у видиму частину діапазону за допомогою маркеру заповнення (табл. 3). Результат відобразиться на екрані (рис. 4). Аналогічну операцію проводимо для шарів, що залишилися.

Наведений спосіб дуже простий, гарно працює за невеликого розміру вибірки та малій кількості градацій критерію. Не потребує від користувача

володіння вводом формул. Абсолютно негнучкий. За зміни меж критеріїв та збільшення обсягу вибірки неефективний. У таких випадках доцільно використовувати формулу =IF().

	В	Е	Ф	Н	Q
1	культура	Повторн	Шар	Вологіст	Код глиб
2	Закритий ґрунт	1	0-5	0,907	
3	Відкритий ґрунт	3	5-20	1,458	
4	Відкритий ґрунт	1	5-20	1,718	
5	Закритий ґрунт	3	0-5	2,165	
6	Відкритий ґрунт	2	5-20	2,255	
7	Закритий ґрунт	2	0-5	2,577	
8	Вереза	3	опад	4,077	
9	Дуб	2	0-5	4,285	
10	Відкритий ґрунт	1	0-5	4,425	
11	Відкритий ґрунт	3	0-5	4,616	

Рис. 1. Показники вологості зразків ґрунту під різними культурами

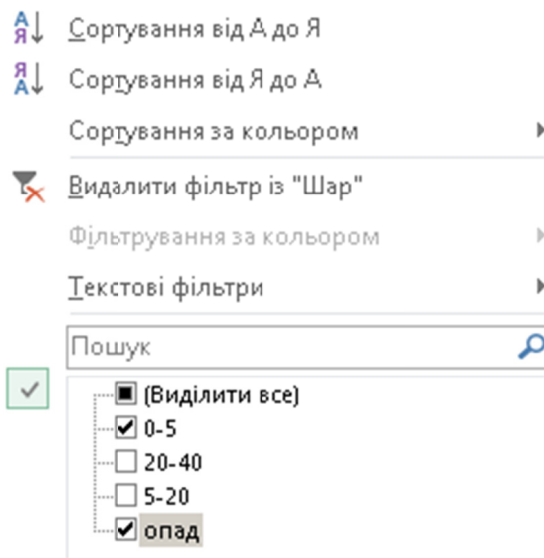


Рис. 2. Вікно автофільтру

Ф	Н	Q
Шар	Вологіст	Код глиб
0-5	11,125	1,00
опад	11,340	
опад	11,396	
0-5	12,097	
опад	12,116	

Рис. 3. Копіювання коду

Шар	Вологіст	Код глиб
0-5	11,125	1,00
20-40	11,248	
опад	11,340	1,00
опад	11,396	1,00
5-20	11,790	
0-5	12,097	1,00
опад	12,116	1,00

Рис. 4. Перший результат кодування

2. За допомогою формули =IF(). Як доповнення до первинного масиву даних на окремому аркуші Excel створюємо довідник відповідності критеріїв їх числовим кодам (рис. 5).

Далі, у стовпчик відповідного критерію вводимо формулу =IF(), що містить посилання на відповідні клітинки аркуша «Довідник». Цей спосіб значно спрощує процес перекодування даних, але потребує гарних навичок вводу формул. Незручний за великої кількості градацій критерію. Так, у нашому випадку, у всього чотирьох градаціях шарів ґрунту формула у клітинці Q2 матиме повний вигляд (рис. 6): =IF(F2=Довідник!\$A\$2;Довідник!\$B\$2;IF(F2=Довідник!\$A\$3;Довідник!\$B\$3;IF(F2=Довідник!\$A\$4;Довідник!\$B\$4;IF(F2=Довідник!\$A\$5;Довідник!\$B\$5;))))).

	A	B	C
1	Шар	Код	
2	Опад	1	
3	0-5	1	
4	5-20	2	
5	20-40	2	
6			

Рис. 5. Коды шарів ґрунту

	F	H	Q	R
1	Шар	Вологість,	Код_гл	
2	0-5	0,907	=IF(F2=Довідник	
3	5-20	1,458	2	
4	5-20	1,718	2	
5	0-5	2,165	1	
6	5-20	2,255	2	
7	0-5	2,577	1	
8	опад	4,077	1	

Рис. 6. Кодування за допомогою IF()

Щоб уникнути згаданих вище незручностей, рекомендуємо освоїти користування найбільш ефективною формулою Excel для сегментування великих масивів даних =VLOOKUP().

3. За допомогою =VLOOKUP(). Може ефективно використовуватися як у побудові номінальної, так і, особливо, інтервальної шкали безперервних змінних, (наприклад, показник вологості ґрунту).

Для цього на окремому аркуші («Довідник») створити таблицю кодування де в крайньому лівому стовпчику вказувалися верхні межі необхідних інтервалів обов'язково в порядку зростання. Кількість інтервалів та їх межі можуть бути встановлені за обраним користувачем алгоритмом, який можна за необхідністю змінювати (у нашому прикладі межі обрано умовно). У правому стовпчику вказуються відповідні цим інтервалам числові або символні коди (рис. 7).

	H	R	S	T	U	V	W
1	Вологість	Код_вол.					
2	1,200	2,00					
3	1,458	2,00					
4	1,718	2,00					
5	2,165	3,00					
6	2,255	3,00					
7	2,577	3,00					
8	4,077	3,00					

D	E
Вологість	Код вол.
0,00	1,00
1,00	2,00
2,00	3,00
5,00	4,00

Довідник База_ Лк

=VLOOKUP(H8;Довідник!\$D\$2:\$E\$5;2)

Рис. 7. Використання формули Vlookup()

Розглянемо параметри функції =VLOOKUP(Н8;Довідник!\$D\$2:\$E\$5;2) (рис. 8).

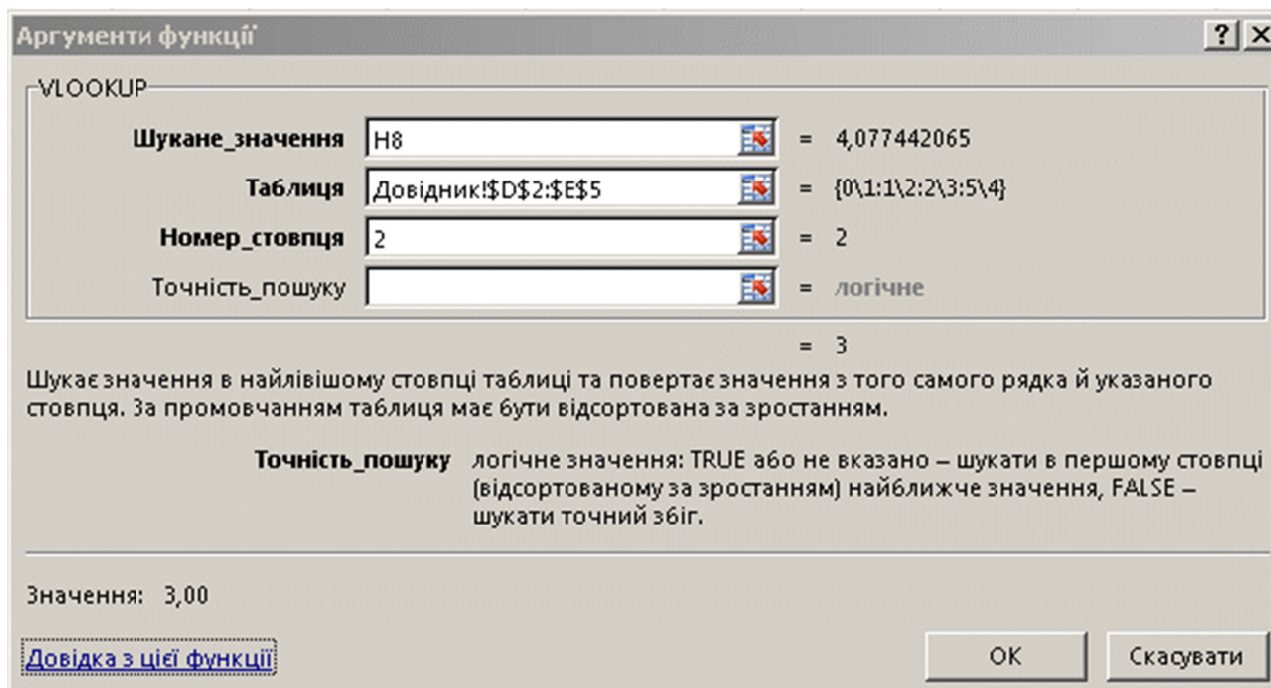


Рис. 8. Вікно «Аргументи функції» VLOOKUP

Н8 – відносне посилання на клітинку поточного аркуша Excel, у якій знаходиться *шукане значення*.

Таблиця – ім'я області даних у першому стовпчику якої, проводиться пошук збігу значення з параметром Н8. Стовпець, за яким здійснюється пошук, завжди повинен бути першим. Область даних можна також задати, вказавши адресу лівої верхньої і правої нижньої клітинки. Наприклад: =VLOOKUP(Н8;Довідник!D2:E5;2).

Тут посилання на область даних відносне, тобто під час копіювання формули адреса області даних буде змінюватися, що призведе до помилок, тому посилання треба робити абсолютним: =VLOOKUP(Н8;Довідник!\$D\$2:\$E\$5;2)

Нагадаємо, відносним посиланням вважається то, що під час копіювання формули за стовпцями / рядках посилання буде змінюватися в тому ж напрямку. Щоб під час копіювання / формули в інші стовпці / рядки посилання на стовпець / рядок змінювалася її можна зробити абсолютною за одним / обома параметрами. Наприклад: \$Н\$8 – абсолютне посилання за стовпцем та рядком (абсолютне посилання на клітинку; \$А8 – абсолютне за стовпчиком, відносне за рядком; Н\$8 – відносне за стовпчиком, абсолютне за рядку; Н8 – відносне посилання на клітинку.

Позначку абсолютної адресації – \$ можна вставити у формулу вручну, або виділити в рядку формул, циклічно натискаючи F4, домогтися потрібного

результату. Відносно посилання на область даних виходить автоматично, якщо область даних вказується виділенням і знаходиться в тій самій книзі, де й наведена формула. Якщо область даних знаходиться в іншій книзі, то посилання на цю область автоматично проставляється у формі абсолютного посилання на виділену область.

Номер_стовпця. Обов'язковий параметр. *Номер стовпця* в аргументі таблиця, з якого функція має повернути відповідне значення. Якщо аргумент *номер_стовпця* має значення 1, функція повертає значення з першого стовпця таблиці; якщо аргумент *номер_стовпця* має значення 2, функція повертає значення із другого стовпця таблиці тощо.

Якщо аргумент *номер_стовпця*: менший за 1, функція VLOOKUP повертає значення помилки #VALUE!. більший за число стовпців у таблиці, функція VLOOKUP повертає значення помилки #REF!.

Точність_пошуку. Додатковий параметр. Логічне значення, що вказує, який саме збіг потрібно знайти функції VLOOKUP – точний чи приблизний.

Якщо аргумент *точність_пошуку* пропущений або має значення TRUE (істина), функція повертає точний чи приблизний збіг. Якщо точний збіг не знайдено, функція повертає наступне за величиною значення, менше від значення аргументу шукане_значення. Якщо аргумент *точність_пошуку* має значення FALSE (хибність), значення в першому стовпці таблиці сортувати не потрібно. Якщо аргумент *точність_пошуку* має значення FALSE (хибність), функція VLOOKUP знайде лише точний збіг. Якщо за аргументом шукане_значення функція знаходить у першому стовпці таблиці два збіги чи більше, використовується лише перше знайдене значення. Якщо точний збіг не знайдено, функція повертає значення помилки #N/A.

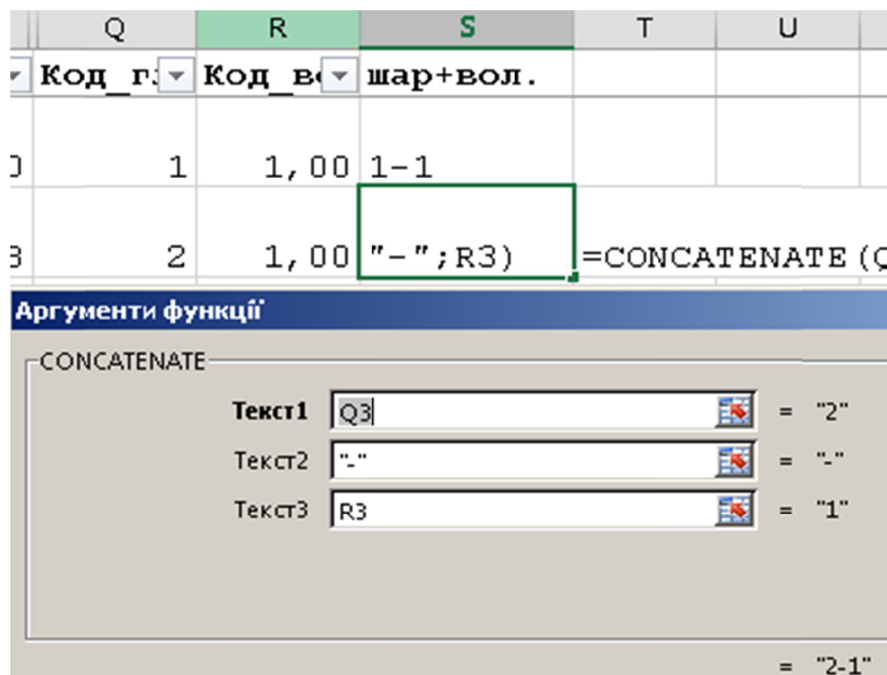


Рис. 9. Поєднання текстових полів за допомогою функції =CONCATENATE

Кваліфіковане використання функції =VLOOKUP() полягає у систематизуванні дослідницьких даних, і в подальшому, дуже швидкому трансформуванні їх за необхідним алгоритмом.

Так використовуючи реманент «зведена таблиця» можна, наприклад отримати перехресну таблицю за двома критеріями, або використовуючи функцію =CONCATENATE() об'єднати групи за обраними критеріями (рис. 9).

Висновки. Розглянуто приклади класифікації первинних експериментальних даних з використанням електронних таблиць. Наведено можливість використання автофільтрів під час перекодування даних; використання спеціалізованих формул =IF(), =VLOOKUP() та =CONCATENATE() для кодування ознак за необхідною шкалою; створено двомірну таблицю (таблиці крос-табуляції) за допомогою інструменту «зведені таблиці». Практичне використання цих інструментів систематизації даних дозволить значно прискорити процес аналізу, підвищити якість і зрозумілість отриманих висновків.

СПИСОК ЛІТЕРАТУРИ / REFERENCES

Сотников Ю. О. Комп'ютерні технології підготовки експериментальних даних в ґрунтознавстві / Ю. О. Сотников, Д. В. Гавва, С. В. Резнік, К. Б. Новосад // Вісник ХНАУ. Сер. «Ґрунтознавство, агрохімія, землеробство, лісове господарство, екологія ґрунтів». – 2015. – № 1. – С. 83–95.

Sotnikov Yu. O., Gavva D. V., Reznik S. V., Novosad K. B., 2015, "Computer technologies for the preparation of experimental data in soil science", Bulletin of the KhNAU. Ser "Soil science, agrochemistry, agriculture, forestry, ecology of soils", Kharkiv, KhNAU, № 1, pp. 83–95.

Эко У. Отсутствующая структура. Введение в семіологію / У. Эко. – СПб.: Симпозиум, 2006. – 544 с.

Eko U., 2006, "Missing structure. Introduction to Semiology", St. Petersburg, Symposium, 544 p.

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

Minko O. O., 2015, "Modern analysis of data in Excel for scientists", Kiev, Scientific and Educational Center of Applied Informatics of the National Academy of Sciences of Ukraine, 476 p.

Джеффри Ч. Биологическая номенклатура / Ч. Джеффри. – М.: Мир, 1980. – 176 с.

Jeffrey C., 1980, "Biological nomenclature", Moscow, Mir, 176 p.

Hedden, Heather, 2010, "The accidental taxonomist", Information today inc., Medford, New Jersey, 600 p.

Клир Дж. Системология. Автоматизация решения системных задач / Дж. Клир. – М.: Радио и связь, 1990. – 534 с.

Klir J., 1990, "Systemology. Automation of the solution of system tasks", Moscow, Radio and Communication, 534 p.