

## Порівняльний аналіз статистичних програмних продуктів для кваліфікаційної експертизи сортів рослин на придатність до поширення

Н. В. Лещук\*, К. М. Мажуга, Н. С. Орленко\*\*,  
Є. М. Стариченко, Є. А. Шкапенко

Український інститут експертизи сортів рослин, вул. Генерала Родимцева, 15, м. Київ, 03041, Україна,  
e-mail: nadiya1511@ukr.net\*, n.s.orlenko@gmail.com\*\*

**Мета.** Визначити статистичні методи та інструментальні засоби (пакети прикладних програм) для створення системи підтримки прийняття рішення (СППР) кваліфікаційної експертизи сортів рослин на придатність до поширення (ПСП) у розрізі завдань оброблення даних. Обґрунтувати вибір програмних засобів оброблення статистичних даних польових та лабораторних досліджень, що входять до складу кваліфікаційної експертизи на ПСП. **Методи.** Аналітичний, який ґрунтується на порівнянні методів описової та багатовимірної статистики й засобів інтелектуального аналізу даних, отриманих під час кваліфікаційної експертизи на ПСП. Порівняльний аналіз програмних засобів оброблення статистичних даних для підготовки пропозицій щодо кінцевого рішення за заявкою на сорт рослин. **Результати.** Проведено декомпозицію завдань, що входять до складу системи підтримки прийняття рішень з кваліфікаційної експертизи сортів-кандидатів на ПСП. Проведено порівняння статистичного пакета SPSS, пакета аналізу, що входить до складу MS Excel, та мови програмування R за критеріями: зручність інтерфейсу, функціональність, якість представлення результатів обчислення, наочність графічної інформації, вартість програмного засобу. Обидва пакети широко застосовують у світі для статистичного оброблення даних, вони мають аналогічний склад функцій для обчислення статистик. **Висновки.** Виокремлено завдання ПСП, що рекомендовано вирішувати із застосуванням досліджуваних засобів. Як інструментальний засіб рекомендовано використовувати мову програмування R. Основною перевагою R порівняно з пакетом IBM SPSS Statistics є те, що R є програмним продуктом з відкритим кодом, який вільно поширюється.

**Ключові слова:** система підтримки прийняття рішень, кваліфікаційна експертиза, придатність сортів до поширення, умовний стандарт, статистичні методи в селекції, дисперсійний аналіз, кластерний аналіз, R, IBM SPSS Statistics.

### Вступ

Створення, поширення та комерційний обіг нових сортів рослин є одним із значущих чинників забезпечення продовольчої безпеки України. Тому формування національних сортових ресурсів є досить актуаль-

ним завданням. Це зумовлює актуальність створення системи підтримки прийняття рішень з проведення кваліфікаційної експертизи сортів, аналізу найбільш придатних методів та інструментальних засобів оброблення даних кваліфікаційної експертизи.

Одним з важливих завдань планування та аналізу даних експертизи є застосування статистичних методів і пакетів оброблення результатів спостережень і дослідів та їх інтерпретація. Встановлено інструментальні засоби, що є найбільш придатними для застосування та включення до системи підтримки прийняття рішення (СППР ПСП). Результати дослідження проілюстровано на прикладі статистичного оброблення даних польових та лабораторних досліджень сортів-кандидатів сої культурної (*Glicine max* (L.) Merrill).

Nadiya Leschuk  
<http://orcid.org/0000-0001-6025-3702>  
Kostiantyn Mazhuha  
<http://orcid.org/0000-0002-1434-8687>  
Nataliia Orlenko  
<http://orcid.org/0000-0003-0494-206>  
Yevhen Starychenko  
<http://orcid.org/0000-0001-8608-5268>  
Yevhenia Shkapenko  
<http://orcid.org/0000-0002-8600-1543>

*Мета досліджень* – визначити та обґрунтувати набір методів статистичного аналізу та інструментальних засобів (пакетів прикладних програм) для створення СППР кваліфікаційної експертизи на придатність сортів для поширення. Визначити найдоцільніші методи статистичного аналізу для оброблення даних кваліфікаційної експертизи та обґрунтувати вибір програмних засобів оброблення статистичних даних польових та лабораторних досліджень, що входять до складу кваліфікаційної експертизи на придатність сортів для поширення.

### Матеріали та методика досліджень

Статистичні методи та інструментальні засоби порівнювали під час оброблення даних кваліфікаційної експертизи сортів-кандидатів на ПСП за період 2014–2016 рр. відповідно до методики [1].

Територіальна неоднорідність дослідних ділянок за ґрунтово-кліматичними умовами року дослідження у відповідній зоні вирощування створює передумови для обґрунтування однорідності та достовірності отриманих даних з використанням методів описової статистики та дисперсійного аналізу [2]. При цьому розраховують такі показники: середня врожайність, коригуючий фактор, сума квадратів розсіювань (загального, повторень, варіантів, залишку), розсіювання, фактичне значення критерію Фішера, найменша істотна різниця (НІР), відносна похибка дослідів, значення меж довірчого інтервалу відповідно до загальноприйнятих методів.

Кластеризація є логічним продовженням процедури класифікації, розв'язання завдань класифікації мають виявити групи сортів рослин, найбільш подібні між собою, що охоплюють як сорти-кандидати для включення до Реєстру сортів рослин України поточного року, так і сорти, внесені до Реєстру за останні п'ять років відповідно до Положення про Державний реєстр сортів рослин, придатних для поширення в Україні.

Розрахунки було проведено з використанням пакета аналізу, що входить до складу електронної таблиці Excel, мови програмування R та тестової версії статистичного пакета IBM SPSS Statistics 22 (trial version).

IBM SPSS Statistics – програмний пакет, який використовують для логічного статистичного аналізу великих обсягів даних [4–6]. Нині цей програмний продукт належить компанії IBM.

R є безкоштовним програмним середовищем з відкритим кодом, що перебуває у віль-

ному доступі. Середовище R містить широку гаму статистичних методів, функцій та графічних інструментів [7, 8]. На відміну від більшості комерційних статистичних програм, R не має кнопочково-віконного інтерфейсу, користувач має знати необхідні для роботи функції та синтаксис мови програмування. Для зручнішої роботи користувача з R розроблено графічні інтерфейси (GUI).

### Результати досліджень

Відповідно до методики проектування інформаційних систем (ІС), різновидом якої є система підтримки прийняття рішень (СППР), було проведено декомпозицію комплексу завдань ПСП та проаналізовано придатні для застосування статистичні методи та алгоритми моделювання. Дані аналізу наведено в таблиці 1.

Таблиця 1

Аналіз методів та алгоритмів, що є придатними для оброблення даних

| Назва завдання   | Метод або алгоритм         |
|--|----------------------------|
| Облік та аналіз господарсько-цінних ознак сортів рослин  | Методи описової статистики |
| Аналіз на однорідність результатів дослідження в межах природно-кліматичної зони   | Дисперсійний аналіз        |
| Облік та аналіз даних лабораторних досліджень  | Методи описової статистики |
| Аналіз на однорідність результатів дослідження за роки дослідження   | Дисперсійний аналіз        |
| Класифікація сортів-кандидатів з урахуванням умовного стандарту для показника врожайності та аналіз відповідності сортів-кандидатів за критерієм «середня врожайність», напрям використання із розрахуванням довірчого інтервалу | Кластерний аналіз          |

Оцінювання на придатність для MS Excel, IBM SPSS Statistics та R було здійснено під час виконання розрахунків із розв'язання трьох типів описаних вище завдань за такими критеріями: якість інтерфейсу користувача (usability) [9]; функціональність; якість представлення результатів обчислення; наочність графічної інформації; вартість програмного засобу.

Результати розрахунків (приклад типових звітів однофакторної дисперсійної моделі з повтореннями) наведено в таблицях 2–3.

Після знаходження істотної різниці здійснюють апостеріорне тестування коефіцієнтів для аналізу різниці між ознаками (урожайності залежно від метеорологічних умов поточного року в районі розміщення пунктів випробування, в яких проведено дослідів).

Таблиця 2

Тест внутрішньосуб'єктних ефектів (Test of Within-Subjects Effects)

| Measure (Вимір): Harverst (Урожайність) |                    |  |                              |                                 |                      |                                      |
|---|--------------------|--|------------------------------|---------------------------------|----------------------|--------------------------------------|
| Source (Джерело)                        |                    | Type III Sum of Squares<br>Сума квадратів типу III | df<br>Число ступенів свободи | Mean Square<br>Середній квадрат | F<br>Критерій Фішера | Sig<br>Ймовірність нульової гіпотези |
| Factor 1                                | Sphericity Assumed | 1,12   | 3                            | 0,37                            | 3,99                 | 0,14                                 |
|   | Greenhouse-Geisser | 1,12   | 1,00                         | 1,12                            | 3,99                 | 0,29                                 |
|   | Lower-bound        | 1,12   | 1,00                         | 1,12                            | 3,99                 | 0,29                                 |
| Factor 1 * Point                        | Sphericity Assumed | 14,41  | 9                            | 1,60                            | 17,15                | 0,02                                 |
|   | Greenhouse-Geisser | 14,41  | 3,00                         | 4,81                            | 17,15                | 0,17                                 |
|   | Lower-bound        | 14,41  | 3,00                         | 4,81                            | 17,15                | 0,17                                 |
| Error (factor 1)                        | Sphericity Assumed | 0,28   | 3                            | 0,09                            | –                    | –                                    |
|   | Greenhouse-Geisser | 0,28   | 1,00                         | 0,28                            | –                    | –                                    |
|   | Lower-bound        | 0,28   | 1,00                         | 0,28                            | –                    | –                                    |

Таблиця 3

Результатні показники за пунктами досліджень

| Measure (Вимір): Harverst (Урожайність) |                       |                        |                                  |                    |  |                            |  |
|---|-----------------------|------------------------|----------------------------------|--------------------|--|----------------------------|--|
| LSD (Найменша істотна різниця)          |                       |                        |                                  |                    |  |                            |  |
| (I) Пункт дослідження                   | (J) Пункт дослідження | (I-J) Різниця середніх | Std. Error<br>Стандартна похибка | Sig<br>Ймовірність | 95% Confidence Interval<br>95% Довірчий інтервал |                            |  |
|   |                       |                        |                                  |                    | Lower Bound<br>Нижня межа                        | Upper Bound<br>Верхня межа |  |
| 9                                       | 62                    | 0,66                   | 1,21                             | 0,682              | -14,77   | 16,09                      |  |
|   | 114                   | -1,43                  | 1,08                             | 0,412              | -15,15   | 12,29                      |  |
|   | 174                   | 3,20                   | 1,30                             | 0,246              | -13,34   | 19,75                      |  |
| 62                                      | 9                     | -0,66                  | 1,21                             | 0,682              | -16,09   | 14,77                      |  |
|   | 114                   | -2,09                  | 1,01                             | 0,286              | -14,89   | 10,70                      |  |
|   | 174                   | 2,54                   | 1,24                             | 0,289              | -13,24   | 18,32                      |  |
| 114                                     | 9                     | 1,43                   | 1,08                             | 0,412              | -12,30   | 15,15                      |  |
|   | 62                    | 2,09                   | 1,01                             | 0,286              | -10,70   | 14,88                      |  |
|   | 174                   | 4,63                   | 1,11                             | 0,150              | -9,48  | 18,75                      |  |
| 174                                     | 9                     | -3,20                  | 1,30                             | 0,246              | -19,75   | 13,33                      |  |
|   | 62                    | -2,54                  | 1,24                             | 0,289              | -18,32   | 13,23                      |  |
|   | 114                   | -4,63                  | 1,11                             | 0,150              | -18,75   | 9,48                       |  |

Результати тестування за критерієм найменшої істотної різниці (багаторазовий t-тест без альфа-кореляції) наведено в таблиці 3.

Типові графіки однофакторної дисперсійної моделі з повтореннями наведено на рисунках 1–2.

Діалогове вікно в однофакторній моделі та результати розрахунків у пакеті MS Excel наведено на рисунках 3–4.

На рисунку 5 наведено приклад імпортованих даних у середовищі графічного інтерфейсу R Studio.

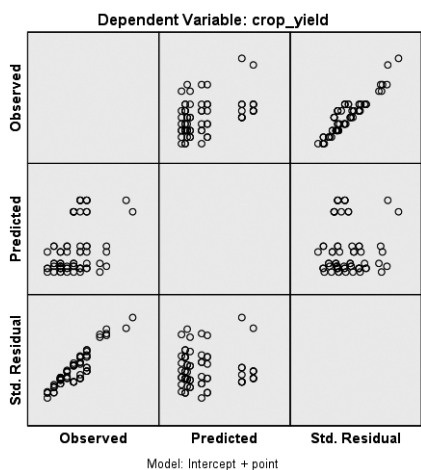


Рис. 1. Графік оцінювання врожайності сої культурної

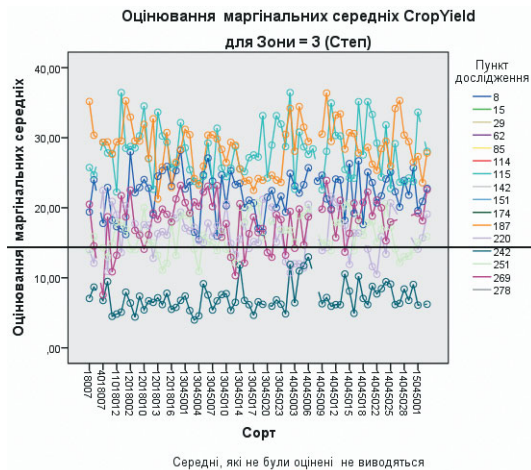


Рис. 2. Графік маргінальних середніх

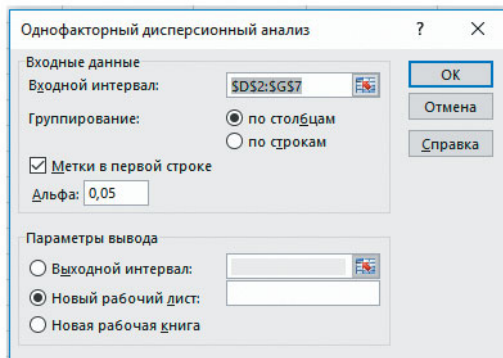


Рис. 3. Діалогове вікно під час внесення даних однофакторної моделі в пакеті MS Excel

|    | A                   | B         | C          | D           | E           | F           | G           |
|----|---------------------|-----------|------------|-------------|-------------|-------------|-------------|
| 3  | Підсумкові значення |           |            |             |             |             |             |
| 4  | Групи               | Кількість | Сума       | Середнє     | Дисперсія   |             |             |
| 5  | Сорт 18007          | 5         | 69,5418601 | 13,90837202 | 3,465053871 |             |             |
| 6  | Сорт 3018009        | 5         | 69,2301465 | 13,8460293  | 12,09187293 |             |             |
| 7  | Сорт 4018007        | 5         | 66,1166863 | 13,22333727 | 12,78687633 |             |             |
| 8  | Сорт 4018012        | 5         | 68,7020065 | 13,7404013  | 10,75815095 |             |             |
| 9  |                     |           |            |             |             |             |             |
| 10 |                     |           |            |             |             |             |             |
| 11 | Дисперсійний аналіз |           |            |             |             |             |             |
| 12 | Джерело варіації    | SS        | df         | MS          | F           | P-Значення  | F критерій  |
| 13 | Між групами         | 1,459539  | 3          | 0,486513019 | 0,049768666 | 0,984774391 | 3,238871517 |
| 14 | Всередині груп      | 156,4078  | 16         | 9,77548852  |             |             |             |
| 15 |                     |           |            |             |             |             |             |
| 16 | Всього              | 157,8674  | 19         |             |             |             |             |

Рис. 4. Результатні дані однофакторної моделі в пакеті MS Excel

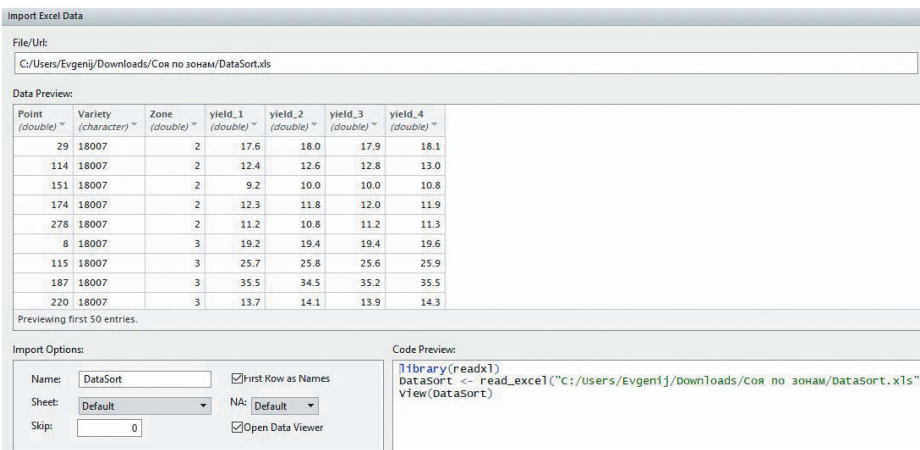


Рис. 5. Приклад графічного інтерфейсу R

Type III Repeated Measures MANOVA Tests:

Term: (Intercept)  
Sum of squares and products for the hypothesis:

(Intercept) 63100.6  
Багатовимірні критерії (для Intercept)

Multivariate Tests: (Intercept)

|                  | Df | test     | stat     | approx | F | num | Df         | den | Df | Pr(>F) |
|------------------|----|----------|----------|--------|---|-----|------------|-----|----|--------|
| Pillai           | 1  | 0.862632 | 81.63658 |        | 1 | 13  | 5.7604e-07 |     |    | ***    |
| Wilks            | 1  | 0.137368 | 81.63658 |        | 1 | 13  | 5.7604e-07 |     |    | ***    |
| Hotelling-Lawley | 1  | 6.279737 | 81.63658 |        | 1 | 13  | 5.7604e-07 |     |    | ***    |
| Roy              | 1  | 6.279737 | 81.63658 |        | 1 | 13  | 5.7604e-07 |     |    | ***    |

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Рис. 6. Результати дисперсійного аналізу MANOVA Tests

```

Term: Crop_yield
Sum of squares and products for the hypothesis:
      Crop_yield1  Crop_yield2  Crop_yield3
Crop_yield1  0.19993460  0.23726678  0.04811704
Crop_yield2  0.23726678  0.28156971  0.05710155
Crop_yield3  0.04811704  0.05710155  0.01158003
Багатовимірні критерії (для Crop_yield)
Multivariate Tests: Crop_yield
      Df test stat approx F num Df den Df Pr(>F)
Pillai      1 0.0411320 0.1572871    3  11 0.92279
Wilks      1 0.9588680 0.1572871    3  11 0.92279
Hotelling-Lawley 1 0.0428965 0.1572871    3  11 0.92279
Roy        1 0.0428965 0.1572871    3  11 0.92279
    
```

Рис. 7. Результати MANOVA за пунктами дослідження

Таблиця 4

**Аналіз інструментальних методів та алгоритмів під час проведення кваліфікаційної експертизи щодо придатності сортів для поширення**

| Назва інструментального засобу  |   |   |
|---|---|---|
| IBM SPSS Statistics   | Microsoft Excel (Пакет аналізу даних)   | R   |
| Якість інтерфейсу користувача (usability)   |   |   |
| Роботу можна виконувати з використанням RKward, RapidMiner, Deducer, R Commander, RExcel, Sage, але для побудови статистичної моделі застосовують консольне введення інформації | Діалогові вікна дають змогу формувати модель аналізу та обирати число повторень та виміру, за якими здійснюються розрахунки, формувати складові моделі, вибирати тип моделі, контрасти, апріорні критерії, тип графіків | Недоліком є те, що необхідно виконувати додаткові дії з транспонування масиву даних. Діалогова форма не дає користувачу змоги вибрати критерії для розрахунку |
| Оцінювання наявності функцій для реалізації завдань ПСП   |   |   |
| Описова статистика  | Описова статистика  | –   |
| Дисперсійний аналіз   | Дисперсійний аналіз   | –   |
| Кластерний аналіз   | –   | Кластерний аналіз   |
| Якість представлення результатів обчислення   |   |   |
| Аналіз багатфакторних критеріїв, перевірки внутрішніх групових ефектів, перевірки внутрішніх групових контрастів тощо   | Перелік результатних показників, що формує пакет, обмежено такими значеннями: сума квадратів, дисперсія, число ступенів свободи, середнє квадратичне відхилення, критерій Фішера, значущість                            | Аналіз з використанням багатфакторних критеріїв, перевірка внутрішніх групових ефектів і контрастів та факторів   |
| Наочність графічної інформації  |   |   |
| Є можливість редагувати формат та написи графіків у графічному редакторі  | Графіки можна сформувати стандартними засобами пакета   | Формат графіків задається програмним кодом і потребує додаткових знань та навичок з їх правильного представлення  |

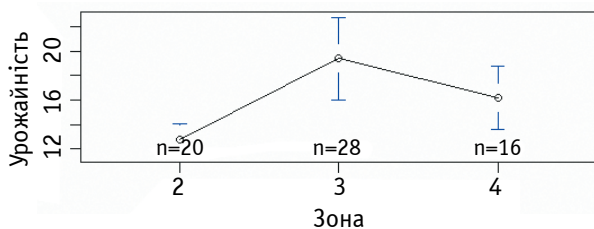


Рис. 8. Графік оцінювання середньої врожайності сої культурної по природно-кліматичних зонах

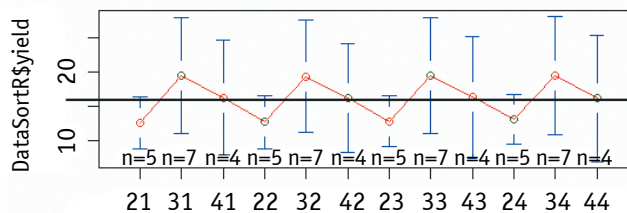


Рис. 9. Діаграма взаємодії для середньої врожайності сої культурної

Результати дисперсійного аналізу наведено на рисунках 6–7. Графіки, що побудовано з використанням програмного коду мови R для дисперсійної моделі, наведено

на рисунках 8–9. Підсумкові результати аналізу інструментальних засобів: IBM SPSS Statistics, MS Excel, R зведено в таблицю 4.

## Висновки

Проведено аналіз методів та алгоритмів, що є придатними для оброблення даних кваліфікаційної експертизи сортів рослин на ПСП.

Здійснено порівняльний аналіз інструментальних засобів, призначених для статистичного оброблення даних. Проведено статистичний аналіз на експериментальних даних польових досліджень сої культурної із застосуванням MS Excel, R, IBM SPSS Statistics, використовуючи критерії: якість інтерфейсу користувача, функціональність, якість представлення результатів обчислення, наочність графічної інформації.

Виявлено, що пакет прикладних програм IBM SPSS Statistics переважає за критеріями: якість інтерфейсу користувача, якість представлення результатів обчислення, пакет аналізу MS Excel та R. Для використання під час оброблення завдань ПСП рекомендовано застосовувати мову програмування R, оскільки цей інструментальний засіб програмування розповсюджується безкоштовно за ліцензією GNU General Public License.

## Використана література

1. Методика проведення кваліфікаційної експертизи сортів рослин на придатність до поширення в Україні. Загальна частина. Вінниця : ФОП Корзун Д. Ю., 2016. 120 с.
2. Горяинова Е. Р., Панков А. Р., Платонов Е. Н. Прикладные методы анализа статистических данных. Москва : Высшая школа экономики, 2012. 312 с.
3. Compton M. E. Statistical methods suitable for the analysis of plant tissue culture data. *Plant Cell. Tiss. Organ. Cult.* 1994. Vol. 37, Iss. 3. P. 217–242. doi: 10.1007/BF00042336
4. Bryman A., Cramer D. *Quantitative Data Analysis with IBM SPSS 17, 18 and 19: A Guide for Social Scientists*. New York : Routledge, 2011. 408 p.
5. Levesque R. *SPSS Programming and Data Management: A Guide for SPSS and SAS Users*. 4<sup>th</sup> ed. Chicago, IL : SPSS Inc, 2007. 540 p.
6. Бюль А., Цефель П. SPSS: искусство обработки. Анализ статистических данных и восстановление скрытых закономерностей. Санкт-Петербург : ДиаСофтЮП, 2002. 608 с.

УДК 004.4'2: 631.526.3

**Лещук Н. В.\***, **Мажуга К. М.**, **Орленко Н. С.\*\***, **Стариченко Е. М.**, **Шкапенко Е. А.** Сравнительный анализ статистических программных продуктов для квалификационной экспертизы сортов растений на пригодность к распространению // *Plant Varieties Studying and Protection*. 2017. Т. 13, № 4. С. 429–435. <https://doi.org/10.21498/2518-1017.13.4.2017.117757>

Український інститут експертизи сортів рослин, ул. Генерала Родимцева, 15, г. Київ, 03041, Україна, e-mail: nadiya1511@ukr.net\*, n.s.orklenko@gmail.com\*\*

**Цель.** Определить статистические методы и инструментальные средства (пакеты прикладных программ) для создания системы поддержки принятия решения (СППР) квалификационной экспертизы сортов на пригодность их к распространению (ПСП) в разрезе задач обработки данных. Обосновать выбор программных средств обработки статистических данных полевых и лабораторных исследований, которые входят в состав квалификационной

7. Baumer B. S., Kaplan D. T., Horton N. J. *Modern Data Science with R*. Boca Raton : Chapman & Hall/CRC, 2017. 556 p.
8. Чешкова А. Ф., Алейников А. Ф., Степочкин П. И. Применение графических возможностей программной среды R для анализа экспериментальных данных по селекции тритикале. *Вычислительные технологии*. 2016. Т. 21, спец. вып. 1. С. 104–115.
9. Кіріленко Ю., Кузнецова Ю., Соколова Є., Фролова Г. Методи оцінювання usability інтерфейсу користувача. *Вісн. Нац. ун-ту «Львів. політехніка»*. Серія: Комп'ютерні науки та інформаційні технології. 2013. № 751. С. 244–256.

## References

1. *Metodyka provedennia kvalifikatsiinoi ekspertyzy sortiv roslyn na prydatnist do poshyrennia v Ukraini. Zahalna chastyna* [Regulations on the procedure and the conduct of qualification tests for suitability of crop varieties for dissemination in Ukraine. General part]. (2016). Vinnytsia: FOP Korzun D. Yu. [in Ukrainian]
2. Goryainova, E. R., Pankov, A. R., & Platonov, E. N. (2012). *Prikladnye metody analiza statisticheskikh dannykh* [Applied methods of statistical data analysis]. Moscow: Vysshaya shkola ekonomiki. [in Russian]
3. Compton, M. E. (1994). Statistical methods suitable for the analysis of plant tissue culture data. *Plant Cell. Tiss. Organ. Cult.* Vol. 37, Iss. 3. P. 217–242. doi: 10.1007/BF00042336
4. Bryman, A., & Cramer, D. (2011). *Quantitative Data Analysis with IBM SPSS 17, 18 and 19: A Guide for Social Scientists*. New York: Routledge.
5. Levesque, R. (2007). *SPSS Programming and Data Management: A Guide for SPSS and SAS Users*. (4<sup>th</sup> ed.). Chicago, IL: SPSS Inc.
6. Byuyul, A., & Tsefel, P. (2002). *SPSS: iskusstvo obrabotki. Analiz statisticheskikh dannykh i vosstanovlenie skrytykh zakonmernostey* [SPSS: Art of Handling. Analysis of statistical data and restoration of hidden patterns]. St. Petersburg: DiaSoftYuP. [in Russian]
7. Baumer, B. S., Kaplan, D. T., & Horton, N. J. (2017). *Modern Data Science with R*. Boca Raton: Chapman & Hall/CRC.
8. Cheshkova, A. F., Aleynikov, A. F., & Stepochkin, P. I. (2016). Application of graphical capabilities of the R programming environment for analysis of experimental data on triticale breeding. *Vychislitel'nye tekhnologii* [Computational Technologies], 21(1), 104–115. [in Russian]
9. Kirilenko, Yu., Kuznetsova, Yu., Sokolova, Ye., & Frolova, H. (2013). Methods to estimate the usability of a user's interface. *Visnyk Natsionalnoho universytetu «Lvivska politehnika»*. Seriya: Kompiuterni nauky ta informatsiini tekhnologii [Bulletin of National University "Lvivska Politehnika". Ser.: Computer Engineering and Information Technologies], 751, 244–256. [in Ukrainian]

експертизи на ПСП. **Методы.** Аналитический, основанный на сравнении методов описательной и многомерной статистики и средств интеллектуального анализа данных, полученных при проведении квалификационной экспертизы на ПСП. Сравнительный анализ программных средств обработки статистических данных для подготовки предложений по окончательному решению по заявке на сорт растений. **Результаты.** Проведена декомпозиция задач

которые входят в состав системы поддержки принятия решений по квалификационной экспертизе сортов-кандидатов на ПСП. Проведено сравнение статистического пакета SPSS, пакета анализа, который входит в состав MS Excel и языка программирования R по критериям: удобство интерфейса, функциональность, качество представления результатов расчетов, наглядность графической информации, стоимость программного средства. Оба пакета широко применяются в мире для статистической обработки данных, имеют аналогичный состав функций для расчета статистик. **Выводы.** Выделены задачи ПСП,

которые рекомендуется решать с применением исследуемых средств. В качестве инструментального средства целесообразно использовать язык программирования R. Основным преимуществом R по сравнению с пакетом IBM SPSS Statistics является то, что R – программный продукт с открытым кодом, который свободно распространяется.

**Ключевые слова:** система поддержки принятия решений, квалификационная экспертиза, пригодность сортов к распространению, условный стандарт, статистические методы в селекции, дисперсионный анализ, кластерный анализ, R, IBM SPSS Statistics.

UDC 004.4'2: 631.526.3

**Leschuk, N. V.\***, **Mazhuha, K. M.**, **Orlenko, N. S.**, **Starychenko, Ye. M.**, & **Shkapenko, Y. A.** (2017). Comparative analysis of statistical software products for the qualifying examination of plant varieties suitable for dissemination. *Plant Varieties Studying and Protection*, 13(4), 429–435. <https://doi.org/10.21498/2518-1017.13.4.2017.117757>

*Ukrainian Institute for Plant Variety Examination, 15 Henerala Rodymtseva Str., Kyiv, 03041, Ukraine, e-mail: nadiya1511@ukr.net\*, n.s.orlenko@gmail.com\*\**

**Purpose.** To define statistical methods and tools (application packages) for creating the decision support system (DSS) for qualifying examination of plant varieties suitable for dissemination (VSD) in the context of data processing tasks. To substantiate the selection of software for processing statistical data relative to field and laboratory investigations that are included into the qualifying examination for VSD. **Methods.** Analytical one based on the comparison of methods of descriptive and multivariate statistics and tools of intellectual analysis of data obtained during qualifying examination for VSD. Comparative analysis of software tools for processing statistical data in order to prepare proposals for the final decision on plant variety application. Decomposition of tasks was carried out which were included into the decision support system for qualifying examination of varieties-candidates for VSD. **Results.** Statistical package

SPSS, analysis package included in MS Excel and program language R was compared for the following criteria: interface usability, functionality, quality of calculation result presentation, visibility of graphical information, software cost. The both packages were widely used in the world for statistical data processing, they have similar functions for statistics calculation. **Conclusion.** Tasks of VSD were separated and recommended to tackle using investigated tools. Program language R was a product recommended to use as a tool. The main advantage of R as compared to the package IBM SPSS Statistics is the fact that R is an open source software.

**Keywords:** decision support system, qualifying examination, varieties suitability for dissemination, conditional standard, statistical methods in breeding, dispersion analysis, cluster analysis, R, IBM SPSS Statistics.

Надійшла / Received 06.09.2017

Погоджено до друку / Accepted 09.11.2017