

6. Моисейченко В.Ф., Трифонова М.Ф., Завирюха А.Х. Основы научных исследований в агрономии. Москва: Колос, 1996. 336 с.

7. Пакудин В.З. Оценка экологической пластичности сортов. Генетический анализ количественных и качественных признаков с помощью математико-статистических методов. 1973. С. 40–43.

8. Потанин В.Г., Алейников А.Ф., Степочкин П.И. Новый подход к оценке экологической пластичности сортов растений. Вавиловский журнал генетики и селекции. 2014. том 18. № 3. С. 548–552.

9. Склярова Н.П., Жарова В.А. Характеристика новых сортов картофеля по параметрам пластичности и стабильности. Селекция и семеноводство. 1989. № 2. С. 18–23.

10. Солонечний П.М. Гомеостатичність та селекційна цінність сучасних сортів ячменю ярого. Селекція і насінництво. 2013. № 103. С. 36.

УДК 004.4'2:631.526.3

ОСОБЛИВОСТІ СХОВИЩА ДАНИХ ТА ОБРОБЛЕННЯ РЕЗУЛЬТАТІВ КВАЛІФІКАЦІЙНОЇ ЕКСПЕРТИЗИ СОРТІВ РОСЛИН

Орленко Н.С. – к. е. н., доцент, старший науковий співробітник,

Український інститут експертизи сортів рослин

Карпич М.К. – науковий співробітник,

Український інститут експертизи сортів рослин

Коховська І.В. – науковий співробітник,

Український інститут експертизи сортів рослин

У статті проаналізовано вплив погодних умов на результати кваліфікаційної експертизи сортів рослин погодних умов у пунктах дослідження Українського інституту експертизи сортів рослин. Використано методи розвідувального та дисперсійного аналізу. Обґрунтовано важливість збереження інформації щодо природно-кліматичних умов відповідно до фенологічних стадій росту рослин у сховищі даних інформаційної системи.

Ключові слова: кваліфікаційна експертиза сортів рослин, придатність сортів до поширення, статистичні методи в селекції, дисперсійний аналіз, сховище даних, інтелектуальний аналіз даних, IBM SPSS Statistics.

Орленко Н.С., Карпич М.К., Коховская И.В. Особенности хранилищ данных и обработки результатов квалификационной экспертизы сортов растений

В статье проведен анализ влияния погодных условий на результаты квалификационной экспертизы сортов растений в пунктах исследования Украинского института экспертизы сортов растений. Используются методы разведочного и дисперсионного анализа. Обоснована важность сохранения информации о природно-климатических условиях согласно фенологических стадий роста растений в хранилище данных информационной системы.

Ключевые слова: квалификационная экспертиза сортов растений, статистические методы в селекции, дисперсионный анализ, хранилище данных, интеллектуальный анализ данных, IBM SPSS Statistics.

Orlenko N.S., Karpych M.K., Kokhovska I.V. Specific features of data warehouses and processing of data on plant variety qualification examination

The article analyzes the influence of weather conditions on the results of the qualification examination of plant varieties under the weather conditions at the Plant Variety Testing Divisions of the Ukrainian Institute for Plant Variety Examination. Exploratory data analysis (EDA) and ANOVA have been used. The importance of saving information about natural and climatic conditions in accordance with the phenological plant growth stages in the data warehouse is proved.

Key words: qualification examination of plant varieties, statistical methods in selection, ANOVA, data warehouse, data mining, IBM SPSS Statistics.

Постановка проблеми. Формування національних сортових ресурсів є одним із головних завдань аграрної політики України. Зростання потреб сільського господарства у забезпеченні високопродуктивними сортовими ресурсами рослин ставить особливі вимоги до кваліфікаційної експертизи сортів, методів та засобів оброблення даних її результатів.

Одним із важливих завдань планування та аналізу результатів експертизи є застосування засобів інтелектуального аналізу даних, що є невід'ємною частиною процесу оброблення та інтерпретації результатів спостережень і дослідів, які здійснює Український інститут експертизи сортів рослин у 24 пунктах дослідження, що розташовані в поліссі, лісо-степній зоні та степу. Підґрунтям для такого аналізу має служити сховище даних інформаційної системи УІЕСР.

Аналіз останніх досліджень і публікацій. Кваліфікаційна експертиза сортів рослин проводиться відповідно до законодавчих актів [1–4] та включає комплекс польових та лабораторних досліджень, що визначені законодавством та методиками. У цьому контексті вітчизняними вченими було напрацьовано методики проведення експертизи сортів рослин [5–6]. Інструментарій та методика створення сховищ та вітрин даних описана в чисельних міжнародних та вітчизняних публікаціях, зокрема для сфери агрономії [7–10]. Методи та засоби інтелектуального аналізу, які висвітлені в роботах ряду вітчизняних та зарубіжних вчених [12–15], проаналізовано на придатність для застосування в процесі оброблення результатних даних кваліфікаційної експертизи сортів рослин.

Постановка завдання. Метою статті є аналіз особливостей сховища даних для збереження результатів кваліфікаційної експертизи та методів інтелектуального аналізу та інструментальних засобів (пакетів прикладних програм) для створення системи підтримки прийняття рішення для фахівців, які проводять кваліфікаційну експертизу на придатність сортів для поширення.

Авторами статті визначено такі завдання в рамках цієї роботи: проведення розвідуваного статистичного аналізу результатів кваліфікаційної експертизи, виявлення маргінальних значень результатів експертизи, виявлення підходів до збереження як розрахункових даних, так і текстової інтерпретації отриманих статистичних результатів у сховищі даних інформаційної системи УІЕСР.

Аналіз здійснювали під час оброблення даних кваліфікаційної експертизи сортів-кандидатів за період 2010–2017 рр. для 19 ботанічних таксонів, які обов'язково проходять експертизу в пунктах дослідження. Для ілюстрації є обрано дані результатів експертизи пшениці м'якої озимої та сої культурної. Для ретельного ознайомлення з вихідним статистичним матеріалом і з'ясування можливості застосування відповідних статистичних методів для його оброблення в технологію комп'ютерного оброблення даних було включено етап, що має назву «розвідувальний». Для оцінювання якості даних використано коробчасту діаграму, методи описової статистики та дисперсійного аналізу. Розраховано показники: середня врожайність, коригуючий фактор, сума квадратів розсіювань (загального, повторень, варіантів, залишку), розсіювання, фактичне значення критерію Фішера, найменша істотна різниця, порівняна похибка дослідів, значення меж довірчого інтервалу відповідно до загальноприйнятих методів.

Розрахунки було проведено з використанням тестової версії статистичного пакету IBM SPSS Statistics 22 (trial version).

Виклад основного матеріалу дослідження. Кваліфікаційна експертиза з польових та лабораторних досліджень здійснюється відповідно до вимог національного законодавства, міжнародних науково-методичних та методологічних вимог. Із метою виявлення найзагальніших закономірностей та тенденцій, законів розподілу величин проведено розвідувальний аналіз даних, результати яких представлено у графічному та

табличному варіанті нижче. Основними цілями розвідкового аналізу є максимальне «проникнення» в дані, виявлення основних структур, вибір найвагоміших змінних, виявлення відхилень та аномалій, перевірка основних гіпотез (припущень), тобто особливостей результатів кваліфікаційної експертизи, що можуть вплинути на модель сховища даних. На рисунку 1 подано коробчасту діаграму результатів розвіданого аналізу даних урожайності пшениці м'якої озимої сортів за період 2010–2017 рр.

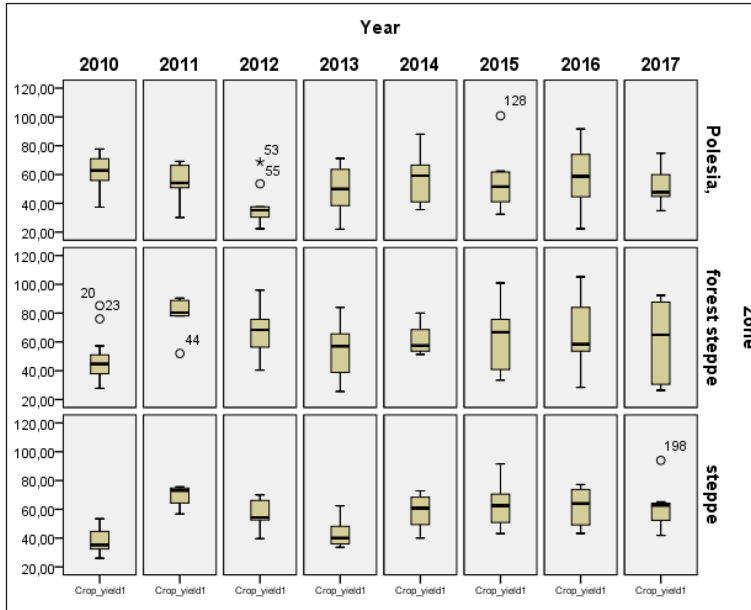


Рис. 1. Коробчаста діаграма результатів розвіданого аналізу даних урожайності пшениці м'якої озимої сортів за період 2010–2017 рр.

Як видно з рисунку 1, маргінальні значення в 2010 р. було отримано в пунктах дослідження № 20, 23 у лісо-степній зоні, в 2011 р. в пункті дослідження № 44 у лісо-степній зоні, в 2012 р. в пунктах дослідження № 53 та № 55 в поліссі, в 2016 р. в пункті дослідження № 128 в поліссі, в 2017 р. в пункті 198 степної зони.

Результати аналітичного аналізу урожайності (OLAP аналізу) наведено у таблицях 1 і 2.

Таблиця 1

Результати аналізу урожайності пшениці м'якої озимої в степовій зоні, 2014 р.

| | Mean | Std. Deviation | Variance | Minimum | Maximum |
|-------------|-------|----------------|----------|---------|---------|
| Crap_yield1 | 58,45 | 12,9190 | 166,903 | 40,00 | 72,80 |
| Crap_yield2 | 58,37 | 13,3209 | 177,446 | 39,60 | 72,00 |
| Crap_yield3 | 57,22 | 12,7564 | 162,726 | 40,00 | 71,20 |
| Crap_yield4 | 56,80 | 13,7830 | 189,973 | 36,80 | 72,00 |

У 2014 р. маргінальні значення урожайності в степній зоні відсутні. Досліди проводились у повтореннях. Середнє значення урожайності пшениці м'якої озимої

на дослідних ділянках становило 58,45 ц/ га, 58,37 ц/ га, 57,22 ц/ га та 56,80 ц/ га. Дисперсія становила 166,903, 177,446, 162,726 та 189,973.

Таблиця 2
Результати аналізу урожайності пшениці м'якої озимої
у лісо-степовій зоні в 2017 р.

| | Mean | Std. Deviation | Variance | Minimum | Maximum |
|-------------|-------|----------------|----------|---------|---------|
| Crop_yield1 | 61,76 | 25,4677 | 648,605 | 26,40 | 92,40 |
| Crop_yield2 | 62,39 | 25,2751 | 638,832 | 27,60 | 91,20 |
| Crop_yield3 | 62,50 | 25,4063 | 645,480 | 27,20 | 93,20 |
| Crop_yield4 | 62,70 | 25,52290 | 651,420 | 26,80 | 93,60 |

У 2017 р. в лісо-степовій зоні маргінальних значень також не було, але спостерігалась значно вища дисперсія, ніж у випадку з результатами 2014 р., яка становила 648,605, 638,832, 645,480 та 651,420. Мінімальне значення урожайності пшениці м'якої озимої на дослідних ділянках становило 26,40 ц/га, 27,60 ц/га, 27,20 ц/га та 26,80 ц/га. А максимальне значення, відповідно, – 92,40 ц/га, 91,20 ц/га, 93,20 ц/га та 93,60 ц/га.

Також було проведено дослідження урожайності сої культурної (*Glicine max* (L.) Merrill) із застосуванням методу дисперсійного аналізу. Результати розрахунків (приклад типових звітів однофакторної дисперсійної моделі з повтореннями) наведено у таблицях 3–4.

Таблиця 3
Результати дисперсійного аналізу

| Source (Джерело) | | Type III Sum of Squares | df. | Mean Square | F | Slg. |
|---------------------------|--------------------|-------------------------|--------|-------------|--------|-------|
| factor1 | Sphericity Assumed | 1,117 | 3 | 0,372 | 3,991 | 0,143 |
| | Greenhouse-Geisser | 1,117 | 1,000 | 1,117 | 3,991 | 0,295 |
| | Huynh-Feldt | 1,117 | . | . | . | . |
| | Lower-bound | 1,117 | 1,000 | 1,117 | 3,991 | 0,295 |
| factor1 * Variety | Sphericity Assumed | 70,927 | 162 | 0,438 | 4,691 | 0,113 |
| | Greenhouse-Geisser | 70,927 | 54,000 | 1,313 | 4,691 | 0,354 |
| | Huynh-Feldt | 70,927 | . | . | . | . |
| | Lower-bound | 70,927 | 54,000 | 1,313 | 4,691 | 0,354 |
| factor1 * Point | Sphericity Assumed | 14,414 | 9 | 1,602 | 17,159 | 0,020 |
| | Greenhouse-Geisser | 14,414 | 3,000 | 4,805 | 17,159 | 0,175 |
| | Huynh-Feldt | 14,414 | . | . | . | . |
| | Lower-bound | 14,414 | 3,000 | 4,805 | 17,159 | 0,175 |
| factor1 * Variety * Point | Sphericity Assumed | 18,645 | 36 | ,518 | 5,549 | 0,091 |
| | Greenhouse-Geisser | 18,645 | 12,000 | 1,554 | 5,549 | 0,321 |
| | Huynh-Feldt | 18,645 | . | . | . | . |
| | Lower-bound | 18,645 | 12,000 | 1,554 | 5,549 | 0,321 |
| Ошибка (factor1) | Sphericity Assumed | 0,280 | 3 | 0,093 | | |
| | Greenhouse-Geisser | 0,280 | 1,000 | 0,280 | | |
| | Huynh-Feldt | 0,280 | . | . | | |
| | Lower-bound | 0,280 | 1,000 | 0,280 | | |

Оскільки дисперсійний аналіз не вказує групи, які відрізняються одна від іншої, а лише зазначає наявність таких розбіжностей, то після знаходження суттєвої різниці виконуються апостеріорні тести коефіцієнтів для аналізу різниці між ознаками (урожайності залежно від метеорологічних умов поточного року) з району розташування пунктів випробування, в яких проведено дослід. Результати тесту за критерієм найменшої істотної різниці (багаторазовий t-тест без альфа-кореляції) наведено у таблиці 4.

Таблиця 4

Результати тесту за критерієм найменшої істотної різниці

| (I) Пункт дослідження | (J) Пункт дослідження | x (I-J) | Std. Error | Slg | 95% Confidence Interval | |
|-----------------------|-----------------------|---------|------------|-------|-------------------------|-------------|
| | | | | | Lower Bound | Upper Bound |
| 9 | 62 | 0,6620 | 1,21454 | 0,682 | -14,7701 | 16,0942 |
| | 114 | -1,4301 | 1,08005 | 0,412 | -15,1534 | 12,2933 |
| | 174 | 3,2058 | 1,30212 | 0,246 | -13,3392 | 19,7508 |
| 62 | 9 | -,6620 | 1,21454 | 0,682 | -16,0942 | 14,7701 |
| | 114 | -2,0921 | 1,00693 | 0,286 | -14,8864 | 10,7022 |
| | 174 | 2,5437 | 1,24214 | 0,289 | -13,2392 | 18,3267 |
| 114 | 9 | 1,4301 | 1,08005 | 0,412 | -12,2933 | 15,1534 |
| | 62 | 2,0921 | 1,00693 | 0,286 | -10,7022 | 14,8864 |
| | 174 | 4,6358 | 1,11101 | 0,150 | -9,4808 | 18,7525 |
| 174 | 9 | -3,2058 | 1,30212 | 0,246 | -19,7508 | 13,3392 |
| | 62 | -2,5437 | 1,24214 | 0,289 | -18,3267 | 13,2392 |
| | 114 | -4,6358 | 1,11101 | 0,150 | -18,7525 | 9,4808 |

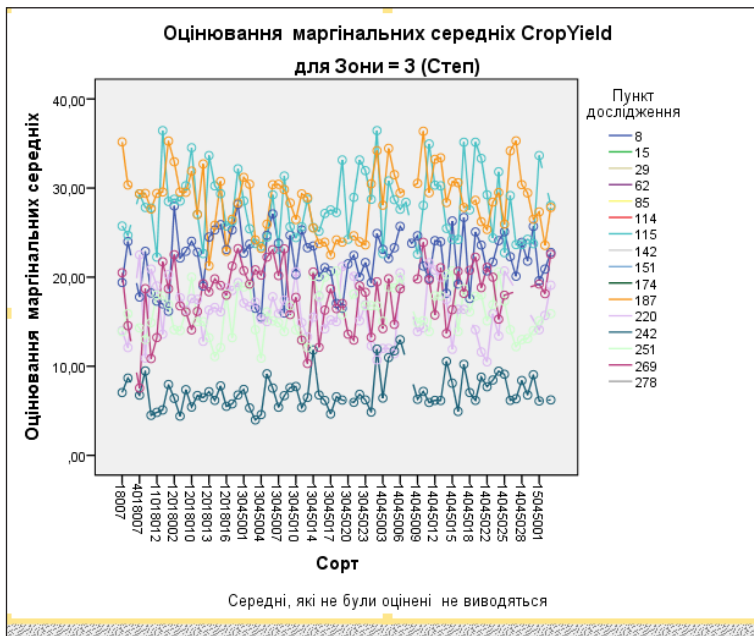
На рисунку 2 відображено графік оцінювання маргінальних значень врожайності сої культурної у розрізі сортів та пунктів дослідження в степній зоні.

Як під час дослідження пшениці м'якої озимої та і під час дослідження сої культурної були виявлені маргінальні значення, а дисперсія свідчить про велику розпорошеність значень, урожайність погодних умов та територіальну неоднорідність дослідних ділянок дисперсійного аналізу. Це пояснюється впливом природно-кліматичних умов. Зауважимо, що клімат України є надзвичайно чутливим до зміни глобального клімату. Зміни щорічної температури повітря характеризують динаміку відхилень від норми в різних агрокліматичних зонах України – Степу, Лісостепу, Поліссі.

Цей факт зумовлює потребу у використанні інноваційних засобів забезпечення накопичення даних щодо умов. Такими інноваційними засобами є програмні продукти VitalFields, Climate Basic, Climate Pro, Climate FieldView, ADAMA та Meteo {11}.

З метою аналізу впливу кліматичних змін на господарсько-цінні ознаки сортів рослин треба забезпечити збереження даних кліматичних умов відповідно до фенологічних фаз розвитку рослин у сховищі даних ІС УІЕСР.

Зауважимо, що сховище даних також використовується як джерело даних для проведення інтелектуального аналізу даних із використанням багатомірного статичного аналізу (дисперсійний та кластерний аналіз), узагальнення інформації, яка надходить із різних автоматизованих робочих місць, що входять до складу автоматизованої інформаційної системи УІЕСР.



*Рис. 2. Графік оцінювання маргінальних значень
врожайності сої культурної в степній зоні*

Оперативні дані кваліфікаційної експертизи сортів рослин зберігаються в АІС УІЕСР, які потім завантажуються у сховище даних (СД). Сховище даних АІС УІЕСР побудоване на основі клієнт-серверної архітектури. Перед завантаженням до СД дані інтегруються та агрегуються за сортами рослин, напрямками їх використання, ботанічними таксонами, природно-кліматичними зонами, роками проведення експертизи тощо. Ці дані є інваріантними у часі та організовані з мінімальною збитковістю інформації. Перед завантаженням у сховища даних інформація фільтрується, а також формується деяка підсумкова інформація (наприклад середня урожайність сорту, приведена до стандартної вологості, дисперсія, помилка середньої тощо). Модель сховища даних організовано відповідно до основних аспектів діяльності УІЕСР.

Предметна організація даних у сховищі сприяє як спрощенню аналізу, так і підвищенню швидкості виконання аналітичних запитів. Дані, отримані з оперативних БД, накопичуються в сховищі у вигляді «історичних шарів», кожен з яких стосується конкретного періоду часу. Це дає змогу аналізувати тенденції в сортівивченні.

Висновки і пропозиції. У роботі було проведено статистичний аналіз результатів кваліфікаційної експертизи сортів рослин. Виявлено значний вплив на урожайність кліматичних умов, що склались у пунктах дослідження, методики проведення експертизи сортів рослин на відмінність, однорідність та стабільність та на придатність сортів для поширення, на структуру сховища даних.

З метою врахування впливу на урожайність рослин кліматичних умов рекомендовано до складу моделі сховища даних включити таблиці, які будуть містити інформацію щодо погодних умов. Як зовнішнє джерело даних щодо погодних умов має застосовуватись інформація, що отримана в результаті використання програмних

продуктів VitalFields, Climate Basic, Climate Pro, Climate FieldView, ADAMA або Meteo. Рекомендовано зберігати дані щодо фенологічних стадій розвитку рослин разом із даними щодо кліматичних умов кожного пункту дослідження. Обґрунтовано необхідність використання сховища даних як підґрунтя інтелектуального аналізу результатів кваліфікаційної експертизи,

Перспективами подальших досліджень є застосування експертної системи з метою автоматизованого формування рекомендацій експертного висновку про результати кваліфікаційної експертизи.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. Про науково-технічну інформацію: Закон України. URL: <http://zakon2.rada.gov.ua/laws/show/3322-12>.
2. Про охорону прав на сорти рослин: Закон України. URL: <http://zakon0.rada.gov.ua/laws/show/3116-12>.
3. Про інформацію: Закон України. URL: <http://zakon2.rada.gov.ua/laws/show/2657-12>.
4. Про доступ до публічної інформації: Закон України. URL: <http://zakon0.rada.gov.ua/laws/show/2939-17>.
5. Концепція формування національних сортових ресурсів на 2006–2011 роки № 302-р. URL: <http://www.agroperspectiva.com/>
6. Методика державного випробування сортів рослин на придатність до поширення в Україні. Офіційний бюлетень «Охорона прав на сорти рослин». К., 2003. № 1, ч. 3. С. 5.
7. Методика проведення кваліфікаційної експертизи сортів технічних та кормових культур. Офіційний бюлетень «Охорона прав на сорти рослин». К., 2003. № 3, ч. 3. С. 18–36.
8. Aditya Kumar Gupta. Multidimensional schema for agricultural Data Warehouse ISSN: 2319 - 1163. Volume: 2 Issue: 3. 245–253.
9. Rajni Jindal1, Shweta Taneja COMPARATIVE STUDY OF DATA WAREHOUSE DESIGN APPROACHES: A SURVEY. International Journal of Database Management Systems (IJDMs) Vol. 4, No. 1, February 2012 DOI: 10.5121/ijdms.2012.4104 33.
10. Mumick I.S. Maintenance of data cubes and summary tables in a warehouse / I.S. Mumick, D. Quass, B.S. Mumick. Stanford University, Database group, 2006. URL: <http://www.db.stanford.edu/pub/papers/cube.maint.ps>.
11. ADAMA Crosses \$3Bn Sales Mark: Reports Solid Business Momentum for Q4 & FY 2013". PRNewswire. 2014-03-19. URL: <https://www.adama.com/en/media/press-releases/adama-crosses-3bn-sales.html>.
12. Kempton R.A., Fox P.N. Statistical methods for plant variety evaluation Great Britain: Chapman and Hall, 1997. 188 p.
13. Ситник В.Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг) Київ: КНЕУ, 2007. 376 с.
14. Brendan O'Connor. Comparison of data analysis packages: R, Matlab, SciPy, Excel, SAS, SPSS, Stata. URL: <http://brenocon.com/blog/2009/02/comparison-of-data-analysispackages-r-matlab-scipy-excel-sas-spss-stata/> (21.11.2012).