

СТВОРЕННЯ Й АПРОБАЦІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ СТАТИСТИЧНОГО МОДЕЛЮВАННЯ ВІРОГІДНОСТІ РЕЗУЛЬТАТІВ АГРОНОМІЧНИХ ЕКСПЕРИМЕНТІВ

П. Гнатів, д. біол. н.

ORCID ID: 0000-0003-2519-3235

О. Литвин, к. с.-г. н.

ORCID ID: 0000-0003-3966-9222

В. Іванюк, к. с.-г. н.

ORCID ID: 0000-0002-6885-9212

Н. Лагуш, к. с.-г. н.

ORCID ID: 0000-0002-1467-5805

В. Шестак, аспірант

ORCID ID: 0000-0003-4288-5794

Б. Коцюба, аспірант

ORCID ID: 0000-0001-7839-487X

Львівський національний університет природокористування

<https://doi.org/10.31734/agronomy2022.26.157>

Гнатів П., Литвин О., Іванюк В., Лагуш Н., Шестак В., Коцюба Б. Створення й апробація програмного забезпечення статистичного моделювання вірогідності результатів агрономічних експериментів

Існує широкий вибір математичних моделей статистики для агрономічних досліджень, які добре описані в літературі. Вони універсальні, проте досліднику раціональніше мати готову до застосування, зручну і швидку щодо отримання результату методику. Найпоширенішим є дисперсійний аналіз, який включений до пакетів *Statistics Calculators* чи *Functions of Excel*. Проте його використання у цих пакетах потребує додаткових кропітких розрахунків для отримання остаточного результату. Тому для оперативного застосування в математичному моделюванні вірогідності результатів експерименту реалізовано алгоритм дисперсійного аналізу однофакторного експерименту інтерпретованою об'єктно-орієнтованою мовою програмування високого рівня з чіткою динамічною типізацією – Python 3, із використанням середовища програмування *PyCharm Community Edition*. Для розроблення функціонального й інтерактивного інтерфейсу обрано бібліотеку *PyQt6* в комбінації з програмою для проєктування і дизайну інтерфейсів *Qt Designer v 5.9.6*. Для обчислення за математичними формулами використано бібліотеку *numpy*. Для роботи з табличними формами, переведенням у текстову форму, а також у формат *Excel.xlsx*, обрано бібліотеки *pandas* та *tabulate*. Так була створена робоча програма – *Dispersion.exe*. У ході апробації після занесення у програму вхідних показників урожаю за 2020–2021 роки вона їх опрацювала і формалізувала у підсумкову статистичну модель вірогідності припущення в експерименті у формі таблиці в середовищі *Microsoft Excel*. Таблиця містить усі кінцеві показники вірогідності результату експерименту з вивчення впливу добрив на врожайність озимого ячменю, розраховані на підставі відхилення нульової гіпотези варіації спостережень.

Ключові слова: дисперсія, ANOVA, Python, PyQt6, pandas, numpy, нульова гіпотеза, вірогідність.

Hnativ P., Lytvyn O., Ivaniuk V., Lahush N., Shestak V., Kotsiuba B. Creation and approvation of software for statistical modeling of the probability of agronomic experiment results

There is a wide choice of mathematical models of statistics for agronomic research, which are well described in the literature. They are universal and are often used, but it is more rational for a researcher to have a ready-to-use, convenient and fast method for obtaining the result. The most common is analysis of variance, which is included in the Statistics Calculators or Functions of Excel packages. However, its use in these packages requires additional painstaking calculations to obtain the final result. Therefore, for operational application in mathematical modeling of the probability of experimental results, the authors of the research implemented an algorithm for analysis of the variance of a one-factor experiment in the interpreted high-level object-oriented programming language with strict dynamic typing – Python 3 using PyCharm Community Edition programming environment. The PyQt6 library was chosen in combination with the Qt Designer v 5.9.6 interface design and development program to develop a functional and interactive interface. The numpy library was used to calculate mathematical formulas. The pandas and tabulate libraries were chosen to work with spreadsheets, text translations, and Excel.xlsx. This is how a working program called Dispersion.exe was created. During the approbation after entering the input indicators of the harvest for 2020–2021 in the program, it processed them and formalized them into a final statistical

model of the probability of assumptions in the experiment in the form of a table in Microsoft Excel. The table contains all the final indicators of the probability of the result of the experiment to study the effect of fertilizers on the yield of winter barley, calculated on the basis of the deviation of the null hypothesis of variation of observations.

Key words: variance, ANOVA, Python, MainWindow, DispOutput, null hypothesis, probability.

Постановка проблеми. Досвід використання персональних комп'ютерів (ПК) та програмного забезпечення (ПЗ) підтверджує, що практично для кожної науково-дослідницької задачі є можливість обрати оптимальний варіант комбінації *hardware-software*. Можна використовувати різне ПЗ та різні версії того самого ПЗ з різними системними вимогами на різноманітних апаратних платформах. Крім того, як розвиток *software*, так і розвиток *hardware*, відбувається взаємопов'язано і невинно [7]. Отже, з обранням для своїх досліджень певної системи засобів проблеми не виникне.

Для досягнення результативності ПЗ мають забезпечувати: можливість внесення доповнень, високий рівень сегментації, набір пропонованих методик аналізу і сервісних функцій, розвинені засоби налаштувань, зручні функції імпорту даних, а також налаштування на галузеву специфіку [4].

Python – високорівнева мова програмування загального призначення, орієнтована на підвищення продуктивності розробника і читання коду. Синтаксис ядра *Python* мінімалістичний. Водночас стандартна бібліотека передбачає безліч корисних функцій. *Python*, як і більшість бібліотек, безкоштовна й перебуває у відкритому доступі. Фреймворк – інфраструктура програмних рішень, що полегшує розробку складних систем.

В агрономічних дослідженнях науковець має справу з великою кількістю оцифрованих даних, які повинні підтвердити робочу гіпотезу експерименту, обґрунтувати виявлені закономірності та зв'язки явищ, з'ясовані в його процесі, або ж довести зворотнє – безперспективність висунутих припущень. Для цього потрібні

математичні моделі варіаційної статистики, широкий набір яких добре описаний у методичній літературі [1; 2; 5; 8].

Аналіз останніх досліджень і публікацій. На практиці інколи складно швидко і якісно опрацювати цифрові дані, особливо коли вони погано систематизовані, або ж цілком нові й складні для вибору статистичної моделі достовірності експерименту чи занесення їх у ПК. Виникають складнощі з виводом кінцевих – цільових результатів у прийнятній формі, як, наприклад, у пакетах *Statistics Calculators* [10] чи Функції *Excel* [6].

Постановка завдання. Наше завдання – показати ефективність та зручність простої програми *Dispersion_win7.exe*, написаної нами для дисперсійного аналізу на ПК агрономічних даних, що містять не менше двох і не більше 20 варіантів змінних у трьох–шести повтореннях спостережень.

Виклад основного матеріалу. Для створення алгоритму дисперсійного аналізу використали розробки Р. Фішера (1925) [8] в інтерпретації Б. Доспехова (1985) [1], Е. Ермантраут та ін. [2], Р. Майбороди і О. Сугакова [3], «Функції Excel (за категоріями)» [6], *Statistics Calculators*. Повний програмний код програми запозичений за посиланням на сайті *GitHub* [9].

Цифрові матеріали взято з результатів обліку врожаю озимого ячменю, вирощеного у 18-ти різних варіантах удобрення на ділянках із триразовою повторністю упродовж 2020–2021 років (табл. 1).

Таблиця 1

Вхідні цифрові результати експерименту для опрацювання

Варіант	2020 р.			2021 р.		
	Повторення			Повторення		
	I	II	III	I	II	III
1	42,7	44,6	43,8	47,1	45,4	44,1
2	43,4	44,2	42,2	45,5	44,2	43,0
3	61,8	58,5	60,6	55,9	56,6	55,3
·	·	·	·	·	·	·
·	·	·	·	·	·	·
·	·	·	·	·	·	·
17	75,2	71,4	72,2	73,1	71,7	73,6
18	72,2	73,8	72,1	73,2	75,2	77,3

Для впровадження у практичних дослідженнях аспірантів і магістрів методики здійснення дисперсійного аналізу за допомогою ПК спроектували програмне забезпечення інтерпретованою об'єктно-орієнтованою мовою програмування високого рівня з чіткою динамічною типізацією – *Python 3*, із використанням середовища програмування *PyCharm Community Edition*.

Для розробки функціонального й інтерактивного інтерфейсу обрали бібліотеку *PyQt6* у комбінації з програмою для проектування і дизайну інтерфейсів *Qt Designer v 5.9.6*. Для обчислення за математичними формулами використовували бібліотеку *numpy*. Для роботи з табличними формами, переведенням у текстову форму, а також у формат *Excel.xlsx*, скористалися бібліотеками *pandas* та *tabulate*. Основний скелет програми побудований із двох класів. Перший клас описує поведінку інтерфейсу і його взаємодію із вхідними даними користувача:

```
class MainWindow (QMainWindow,
Ui_MainWindow):
    def __init__(self, parent=None, *args,
**kwargs):
        def updateTable(self):
        def updateTableSize(self):
        def saveConfigs(self):
        def loadConfigs(self):
        def getMatrix(self):
        def writeCells(self, data: list):
        def exportToExcel(self):
        def showResults(self):
```

Другий клас описує математичні операції із введеними даними експерименту і виведення підсумків дисперсійного аналізу у прийнятній для користувача формі:

```
class DispOutput:
    def __init__(self, X):
    def roundVals(self, n):
    def toMarkdown(self) -> str:
    def toExcel(self):
```

Основні обчислення проводяться при ініціалізації класу *DispOutput*:

```
X: np.ndarray = X
l: int = np.shape(X)[0] # Число варіантів
n: int = np.shape(X)[1] # Число спостережень
N: int = np.size(X) # Загальна кількість спостережень
V: np.ndarray = np.sum(X, axis=1) # Суми
avg: float = np.average(X)
C: float = pow(np.sum(X), 2) / N
CY: float = np.sum(np.square(X)) - C
CV: float = np.sum(np.square(V)) / n - C
```

```
CZ: float = CY - CV
s2v: float = CV / (l - 1) # Середній квадрат варіантів
s2: float = CZ / (N - l) # Середній квадрат помилки
v: float = 100 * math.sqrt(s2) / avg # Коефіцієнт варіації, %
sx: float = math.sqrt(s2 / n) # Помилка досліджу
sd: float = math.sqrt(2 * s2 / n) # Помилка різниці середніх
sd_percent: float = 100 * sd / avg # Відносна помилка різниці середніх
Ff: float = s2v / s2
F05: float = ft.f05_distr(n, N - l)
t05: float = ft.t_crit(0.95, N - l)

HCP05: float = t05 * sd
HCP05_percent: float = (HCP05 * 100) / avg
```

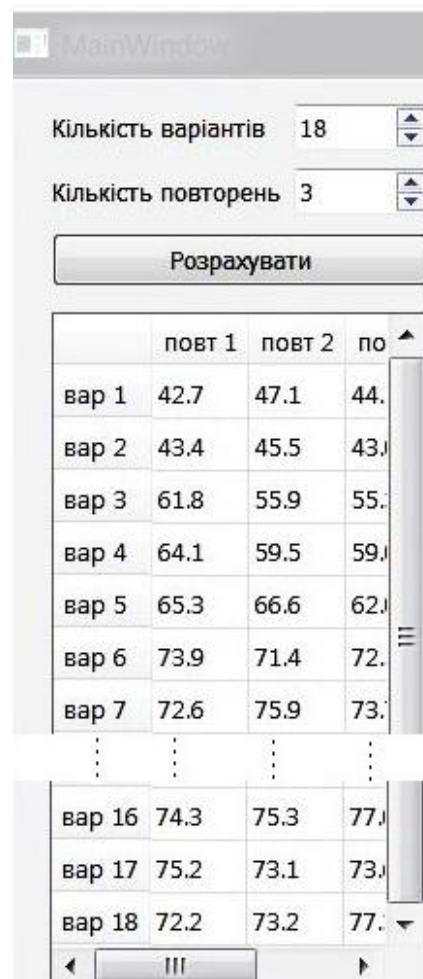


Рис. 1. Робоче вікно програми *Dispersion_win7.exe* для запису цифрових даних експерименту за кожен рік, якщо рік не є фактором

Із повним робочим кодом програми можна ознайомитися за посиланням на сайті *GitHub* – *software development platform* (<https://github.com/dimbaida/variance-analysis>) [9].

Робоче вікно програми *Dispersion_win7.exe* для запису цифрових даних експерименту за кожен рік, якщо рік не є фактором, показано на рис. 1.

Результативне вікно обчислень у програмі *Dispersion.exe* із записом цифрових результатів показано на рис. 2.

Перенесені у формат середовища *Microsoft Excel* результати обчислень у програмі *Dispersion.exe* показані на рис. 3.

The screenshot shows the 'Form' window of the Dispersion.exe application. It contains two main tables and several text-based statistics.

Вхідні дані (Input Data):

Варіанти	1	2	3	К-ть спост.	Суми	Середні
1	42.70	47.10	44.10	3	133.90	44.63
2	43.40	45.50	43.00	3	131.90	43.97
3	61.80	55.90	55.30	3	173.00	57.67
⋮	⋮	⋮	⋮	⋮	⋮	⋮
17	75.20	73.10	73.60	3	221.90	73.97
18	72.20	73.20	77.30	3	222.70	74.23

Загальна кількість спостережень: 54
Загальна сума: 3611.6
Середнє по досліді: 66.88

Результати дисперсійного аналізу (Dispersion Analysis Results):

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	Fф	F05
Загальна	5476.32	53	--	--	--
Варіантів	5331.37	17	313.61	77.89	2.84
Залишок (помилки)	144.95	36	4.03	--	--

Критерій суттєвості: 77.89
Критерій F на 5%-му рівні значимості: 2.84
Помилка досліді: 1.16
Помилка різниці середніх: 1.64
Відносна помилка різниці середніх: 2.45%
Коефіцієнт варіації: 3.0%
НІР абсолютне: 3.32
НІР відносне: 4.97%

At the bottom of the window, there is a button labeled "Експортувати в Excel" (Export to Excel).

Рис. 2. Результативне вікно програми *Dispersion.exe* для запису цифрових результатів

	A	B	C	D	E	F	G	H
1	Варіанти	1	2	3	К-ть спост.	Суми	Середні	
2	1	42,7	47,1	44,1	3	133,9	44,6	
3	2	43,4	45,5	43,0	3	131,9	44,0	
4	3	61,8	55,9	55,3	3	173,0	57,7	
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
18	17	75,2	73,1	73,6	3	221,9	74,0	
19	18	72,2	73,2	77,3	3	222,7	74,2	
20	Загальна сума				54	3611,6	66,9	
21								
22								
23	Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	Fф	F05		
24	Загальна	5476,32	53	--	--	--		
25	Варіантів	5331,37	17	313,61	77,89	2,84		
26	Залишок (помилки)	144,95	36	4,03	--	--		
27								
28								
29	Критерій суттєвості				77,89			
30	Критерій F на 5%-му рівні значущості				2,84			
31	Помилка дослід, т/га				1,16			
32	Помилка різниці середніх, т/га				1,64			
33	Відносна помилка різниці середніх, %				2,45			
34	Коефіцієнт варіації, %				3,00			
35	НІР абсолютне, т/га				3,32			
36	НІР відносне, %				4,97			
37								

Рис. 3. Підсумкова статистична модель вірогідності припущення в експерименті з вивчення впливу добрив на врожайність озимого ячменю у 2021 році, розрахована на підставі відхилення нульової гіпотези варіації спостережень у форматі Microsoft Excel

Формалізований результат дисперсійного аналізу ANOVA представлених, як приклад, цифрових даних на предмет достовірності впливу факторів (варіантів) на врожайність озимого ячменю поданий у табл. 2.

Статистична модель достовірності експерименту свідчить, що нульова гіпотеза має бути відхилена, оскільки критерій значущості Фішера розрахунковий більший за критерій Фішера табличний на 5 % рівня значущості ($F_{\text{факт}} > F_{05}$). Помилка дослід 2020 року становить 0,92 ц/га, 2021 – 1,16 ц/га. Помилка різниці середніх відповідно становить 1,30 та 1,64 ц/га. Розрахунок відносної помилки показав, що у 2021 році дослід

був менш точним, ніж 2020 року. Відповідно найменша істотна різниця (НІР₀₅) між середніми за варіантами за абсолютним показником (ц/га) була більшою у 2021 році, що свідчить про меншу силу впливу експериментальних варіантів порівняно з дисперсією даних за повторностями.

На підставі наявності критерію НІР₀₅ розраховують різницю між варіантами удобрення та контрольним варіантом, який буде обрано за стартовий. Різниця між варіантами та контролем, яка буде меншою за НІР₀₅, вказує на відсутність достовірного впливу порівнюваного варіанта, а та, яка буде більшою, – на агрономічну ефективність пропонованого варіанта удобрення.

Базові формули та цифрові показники статистичної моделі вірогідності експерименту

Показник	Формула розрахунку	2020 р.	2021 р.
Середній квадрат варіантів	$s_p^2 = \frac{C_V}{(l_v - 1)}$	294,66	313,61
Середній квадрат помилки	$s^2 = \frac{C_E}{(n - 1)}$	2,53	4,03
Критерій значущості розрахунковий	F _{факт}	116,56	77,89
Критерій на 5%-му рівні значущості	F _{05табл}	2,84	2,84
Помилка досліду, ц/га	$s_x = \frac{s^2}{n}$	0,92	1,16
Помилка різниці середніх, ц/га	$s_d = \frac{s}{\sqrt{\frac{1}{n}}}$	1,30	1,64
Відносна помилка різниці середніх, %	$s_{d, \%} = \frac{s_x}{\bar{x}} \times 100$	1,96	2,45
НІР ₀₅ абс, ц/га	НІР ₀₅ = $t_{05} \times s_d$	2,63	3,32
НІР ₀₅ відн, %	НІР _{05, \%} = \frac{t_{05} \times s_d}{\bar{x}} \times 100}	3,97	4,97
Коефіцієнт варіації, %	$V = 100 \times \frac{\sqrt{s^2}}{\bar{x}}$	2,4	3,0

Висновки. Програма *Dispersion.exe* значно простіша у користуванні порівняно з пакетами *STATISTICA* та Функції *Excel*, оскільки є вузько-спеціалізованою прикладною програмою для прямого застосування. Результативні підсумкові форми програми *Dispersion.exe* завершені, читабельні і зрозумілі пересічному експериментатору.

Перспективне розширення можливостей програми *Dispersion.exe* для дво- і трифакторних експериментів в агрономії, над чим слід працювати в подальшому.

Бібліографічний список

1. Доспехов Б. А. Методика полевого опыта (с основами статистической обработки результатов исследований). 5-е изд., перераб. и доп. Москва: Агрпроомиздат, 1985. 351 с.
2. Ермантраут Е. Р., Присяжнюк О. І., Шевченко І. Л. Статистичний аналіз агрономічних дослідних даних в пакеті Statistica-6. Київ: Поліграф Консалтинг, 2007. 55 с.
3. Майборода Р. Є., Сугакова О. В. *Статистичний аналіз даних за допомогою пакету STATISTICA, 2022.* URL: <http://matphys.rpd.univ.kiev.ua/downloads/courses/mmatstat/StatAn.doc> (дата звернення: 24.05.2022).

4. Роїк М. В., Присяжнюк О. І., Денисюк В. О. Огляд програмних засобів статистичного аналізу даних. *Ефективна економіка.* 2017. № 7. URL: <http://www.economy.nayka.com.ua/?op=1&z=5676> (дата звернення: 24.05.2022).

5. Самарець Н. М., Харченко Є. М., Чорна Н. О. Використання інформаційних технологій у статистичному аналізі даних для аграрних підприємств. *АГРОСВІТ,* 2013. № 20. С. 14–20.

6. *Функції Excel (за категоріями).* URL: <https://support.office.com/uk-ua/article> (дата звернення: 24.05.2022).

7. 50 Years of Moore's Law. 2022. URL: <https://www.intel.com/content/www/us/en/silicon-innovations/moores-law-technology.html> (Accessed: 24 May 2022).

8. Fisher R. A. 1925. *Statistical methods for research workers* (Modified Sept. 2005.) URL: <http://psychclassics.yorku.ca/Fisher/Methods/> (Accessed: 24 May 2022).

9. Baidachnyi Dmytro. 2022. *GitHub – software development platform.* 2022. URL: <https://github.com/dimbaida/variance-analysis> (Accessed: 24 May 2022).

10. *Statistics Calculators: Anova Calculator.* URL: <https://calculator-online.net/anova-calculator/> (Copyrights 2022 Calculator-online.net) (Accessed: 24 May 2022).

Стаття надійшла 28.05.2022