

УДК 651.92:001.817

ОСОБЛИВОСТІ ВИКОРИСТАННЯ ПРОГРАМ EXCEL І STATISTICA ДЛЯ ДИСПЕРСІЙНОГО АНАЛІЗУ ОДНОФАКТОРНИХ ЕКСПЕРИМЕНТІВ З ПОВТОРЕННЯМ

Бакарджисв Р. О., канд. техн. наук, доцент

Таврійський державний агротехнологічний університет Міністерства аграрної політики та продовольства України

Тел./факс: (0619) 42-24-11

Комарова І. Б., канд. с.-госп. наук

Інститут олійних культур Національної академії аграрних наук України

Тел./факс: (061) 259-77-80

Кисельов О. В., канд. техн. наук, ст. наук. співроб.

Інститут механізації тваринництва Національної академії аграрних наук України

Тел./факс: (061) 289-81-44

Розглянуто особливості використання програми Excel та пакету прикладних програм Statistica для проведення дисперсійного аналізу однофакторних експериментів з повторенням дослідів. На прикладах показано порядок його виконання при однаковому і різній кількості повторень експериментів. Представлено порівняння отриманих результатів і трудомісткість при використанні різних способів дисперсійного аналізу.

Ключові слова: дисперсійний аналіз, однофакторний експеримент, повторення дослідів, програма Excel, пакет прикладних програм Statistica.

Постановка проблеми. Дисперсійний аналіз – статистичний метод виявлення дії факторів, що впливають на властивості об'єкта, який вивчається. Його метою є отримання загальної оцінки наявності відмінності порівнюваних варіантів за результатами досліджуваних показників. Для цієї мети використовується критерій Р. А. Фішера (F-розподіл Фішера), який представляє собою відношення дисперсії s_v^2 величини досліджуваного показника порівнюваних варіантів (вибірок) до дисперсії s_z^2 неконтрольованих в експерименті джерел мінливості – дисперсії залишку (похибки). Подальша оцінка виконується зіставленням отриманого розрахункового значення з табличним, визначеним для відповідних ступенів вільності цих дисперсій – для варіантів v_v і для похибки v_z .

При цьому нульовою гіпотезою служить припущення: усі вибіркові середні є оцінками однієї генеральної середньої, і, отже, відмінності між ними несуттєві. Якщо дисперсії є випадковими із загальної генеральної вибірки, то фактичне значення критерію Фішера F_f не виходить за відповідні межі і не перевищує критичне теоретичне значення критерію F_T , який характеризує рівень можливих випадкових значень критерію Фішера, очікуваний при відповідних ступенях вільності. За умови $F_f < F_T$ нульова гіпотеза $H_0: d = 0$ не відкидається, і це свідчить про те, що між усіма вибірковими середніми немає істотних статистичних відмінностей, і на цьому перевірка закінчується; якщо генеральні параметри порівняльних груп різні, то $F_f \geq F_T$ і нульова гіпотеза відкидається [1, 2].

В однофакторному досліді з повтореннями варіювання показників за повтореннями обумовлено розходженням дій досліджуваних факторів, та інших причин випадкового характеру.

Аналіз останніх досліджень і публікацій. Доволі часто при аналізі технологій вирощування і селекційно-генетичних дослідженнях виникає потреба виконати аналіз однофакторного експерименту з повторенням. Використовувана в Інституті олійних культур НААН для цієї мети програма Dad.exe, розроблена ще в 1992 р. для використання з операційною системою DOS хоч і розраховує більшість статистичних оцінок вибірки, проте, вимагає ручного введення даних, кількість яких може сягати декілька сотень, а то й тисяч, і які вже, як правило, представлені первинним матеріалом, оснований на таблицях програмного забезпечення Microsoft Excel [3, 4], створюваних на основі первинної інформації.

Мета дослідження. У даній роботі розглядаються особливості проведення дисперсійного аналізу однофакторного експерименту з повторенням при використанні програм Excel і Statistica [5].

Об'єкт дослідження. Для ілюстрації виконання дисперсійного аналізу однофакторного досліді розглянемо приклад, наведений в [6, С. 30–35]. Значення залежної змінної (варіанту) визначалося в чотирьох повторних вимірах (ділянках) I, II, III і IV.

Для цього створюється вихідна таблиця, розташована в масиві даних A2:E8 (рис. 1), де значення варіант показника при повторенні замірів подано в масиві даних B4:E8.

При аналізі однофакторного експерименту з повторенням слід застосовувати інструмент Двухфакторный дисперсионный анализ без повторений, прийнявши одним фактором залежну змінну (варіант), а другим — повторення вимірів I, II, III й IV.

Натиснувши Данные (1) (рис. 1), потім Анализ данных (2) потрапляємо на панель Анализ данных, де слід вибрати інструмент

«Двухфакторный дисперсионный анализ без повторений» (3) і натисненням кнопки ОК (4) вивести панель «Двухфакторный дисперсионный анализ без повторений», на якій задати значення «Альфа: 0.05» (5) – імовірність найменш сприятливого типу похибки, виведення назв факторів ☒ «Метки» (6), указується виведення даних на ☒ «Новый рабочий лист» (7), задається інтервал «\$A\$3:\$E\$8» (8) і натиснути ОК (9).

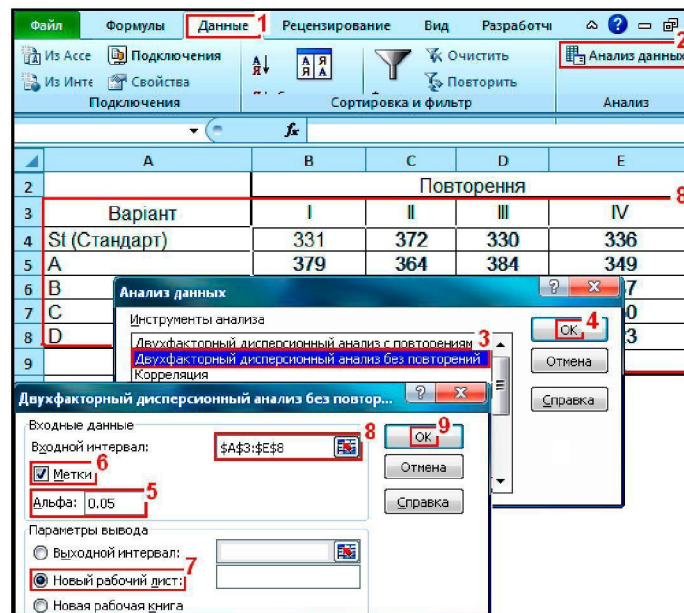


Рисунок 1 – Порядок застосування інструменту «Двухфакторный дисперсионный анализ без повторений» для аналізу однофакторного експерименту з повтореннями

Наслідок цих дій – лист виведення результатів у дещо редагованому вигляді представлений в таблиці 1.

Дані листа складаються як би із трьох таблиць, у верхній, яка є підсумковою представляє собою статистичну характеристику варіантів (масив даних A2:E7), у колонках наведено статистичні показники варіантів (сортрозразків) – кількість повторень, сума, середнє й дисперсія.

У нижній частині листа подана таблиця дисперсійного аналізу, яка розміщена в масиві даних A15:G19, у її стовбцях наведено: SS – суму квадратів відхилень, df – кількість ступенів вільності, MS – середній квадрат ефекту, розрахункове F і F – критичне значення критерію Фішера, p – рівень значущості критерію (похибка статистичного висновку).

Рядки таблиці показують саме яких джерел варіювання стосуються ці дані, тобто рядків (варіантів), колонок (років), похибки досліду (міжгрупового варіювання) та суми всіх перелічених (корегуючого чинника, поправки).

Отриманий результат наведено в масиві даних E16:E17.

Таблиця 1 – Результат дисперсійного аналізу однофакторного експерименту з повтореннями інструментом
«Двухфакторный дисперсионный анализ без повторений»

	A	B	C	D	E	F	G
1	Двухфакторный дисперсионный анализ без повторений						
2	ИТОГИ	Счет	Сумма	Среднее	Дисперсия		
3	St (Стандарт)	4	1369	342.3	400.25		
4	A	4	1476	369.0	250.00		
5	B	4	1582	395.5	187.67		
6	C	4	1657	414.3	632.92		
7	D	4	1828	457.0	1355.33		
8							
9	I	5	1934	386.8	1407.20		
10	II	5	2045	409.0	3147.50		
11	III	5	1988	397.6	2533.30		
12	IV	5	1945	389.0	2322.50		
13							
14	Дисперсионный анализ						
15	Источник вариации	SS	df	MS	F	P-Значение	F критическое
16	Строки	30686.3	4	7671.575	13.235030	0.000235	3.259167
17	Столбцы	1522.8	3	507.600	0.875713	0.480787	3.490295
18	Погрешность	6955.7	12	579.642			
19	Итого	39164.8	19				

Для визначення НІР (так як кількість повторень для всіх варіантів однакова) знизу до неї додаються рядки розрахунку похибки різниці середніх варіантів та критерію Стюдента з рівнем значущості α , прийнятим у відповідності з поз. (5) на рисунку 1.

Для виконання дисперсійного аналізу з додатковим використанням пакету прикладних програм Statistica також використовується таблиця, представлена на рисунку 1. Скопіюємо її масив даних A2:E8 в таблицю файлу ППП Statistica і перейменуємо назви колонок відповідно до її заголовку (рис. 2).

У меню **Statistics/Анализ** (1) використовуємо модуль **ANOVA/Дисперсионный анализ (ДА)** (2) і на панелі процедур і установок **General ANOVA/MANOVA:/Общий ДА:** вибираємо опцію **Repeated measures ANOVA/Повторные измерения ДА** (3) і **Quick specs dialog/Диалог** (4) та натискаємо ОК (5).

Натисканням кнопки **Variables/Переменные** (6) на вкладці **ANOVA/MANOVA Repeated Measures ANOVA: Повторные измерения ДА:** відкривається панель **Select dependent variables and optional categorical predictors (factors)/Выберите зависимые переменные и категориальные предикторы:**, де відмічаємо залежні змінні (7) і регресори (8) та натиснувши ОК (9) повертаємося на панель процедур і установок

ANOVA/MANOVA Repeated Measures ANOVA: Повторные измерения ДА:

При завданні факторів потрібно вказати, що залежні змінні містять значення одного фактору вимірів, який має чотири рівні (тобто повторення представити як повторність), інакше програма буде розглядати повторення I–IV як фактори й виконає багатомірний дисперсійний аналіз.

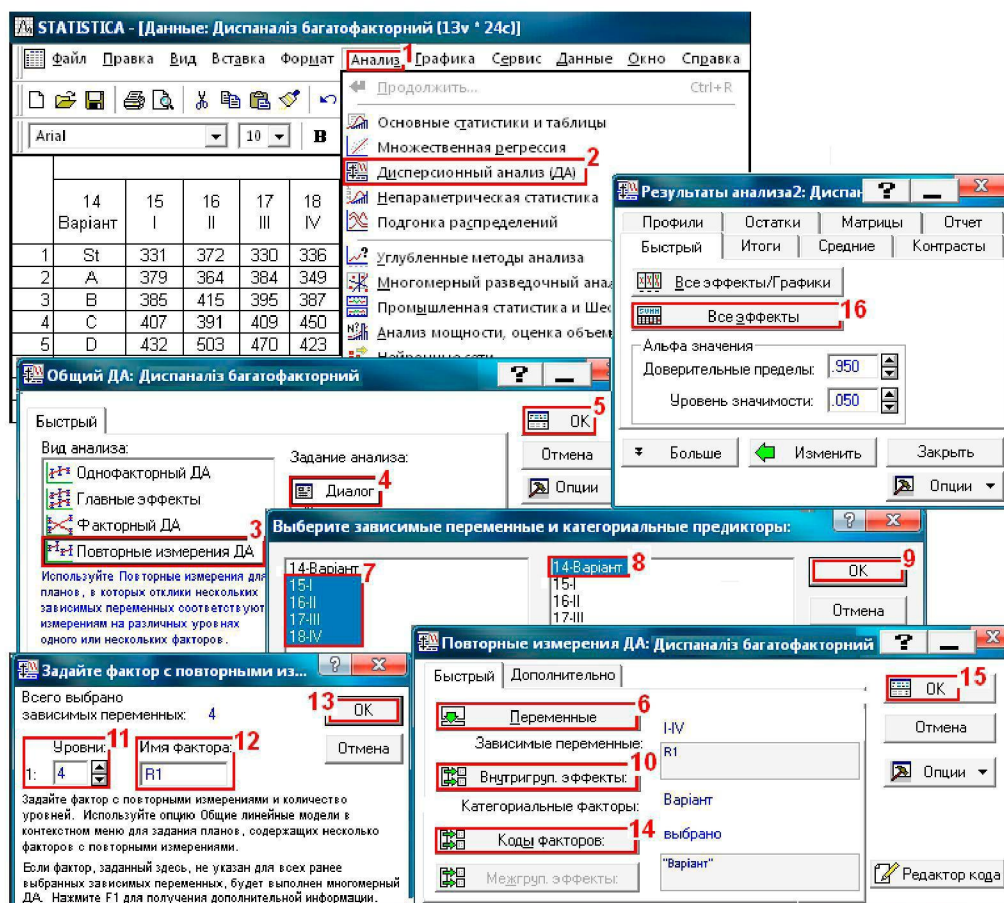


Рисунок 2 – Розрахунок дисперсійного аналізу однофакторного досліджу

Для того щоб визначити фактор, який нас цікавить, слід натиснути кнопку

Within effects: Внутригруп. эффекты (10) панели

ANOVA/MANOVA Repeated Measures ANOVA: Задать фактор с повторными из... і зада-
ти 4 рівні (11) вихідного фактора R1 (згідно кількості повторності) (12). Після натис-
кання ОК (13) знову повертаємося на панель

ANOVA/MANOVA Repeated Measures ANOVA: Повторные измерения ДА:

Коди Factor codes: Коды факторов (14), що визначають рівні міжгрупових факторів, задавати не потрібно, так як програма сама використовує в якості кодів міжгрупових факторів усі значення вхідних змінних. Далі натиснувши ОК

(15) переходимо на панель ANOVA Results 1: / Результати аналізу 2:, де задіявши кнопку All effects / Все эффекты (16) маємо таблицю 2 з результатами дисперсного аналізу.

Таблиця 2 – Результати дисперсійного аналізу

Эффект	Дисперсионный анализ с повторными измерениями Сигма-ограниченная параметризация (Диспанализ) Декомпозиция гипотезы				
	SS	Степени свободы	MS	F	p
Св. член	3129987,20	1	3129987,20		
Вариант	30686,30	4	7671,57		
Ошибка		0			
R1	1522,80	3	507,60		
R1*Вариант	6955,70	12	579,64		
Ошибка		0			

Копіюємо її із заголовками із ППП Statistica таблиці 2 у масив даних таблиці 3 B10:G16.

Таблиця 3 – Дисперсійний аналіз з використанням ППП Statistica

	A	B	C	D	E	F	G	
2	Варіант	Повторення				F12=E12/\$E\$15 F14=E14/\$E\$15		
3		I	II	III	IV			
4		St (Стандарт)	331	372	330			336
5		A	379	364	384			349
6		B	385	415	395			387
7		C	407	391	409			450
8		D	432	503	470			423
9								
10	Дисперсія		SS	Степени свободы	MS	F	p	
11	Корегуючий фактор, C	Св. член	3129987,20	1	3129987,20			
12	Варіантів, Cv	"Var1"	30686,30	4	7671,58	13,235	0,000235	
13		Ошибка		0				
14	Повторення, Cp	R1	1522,80	3	507,60	0,87571	0,480787	
15	Залишок (похибка), Cz	R1**"Var1"	6955,70	12	579,64			
16		Ошибка		0				

Для представлення результатів у масиві даних G3:J8 створюється таблиця, де поваріантно подано середнє значення показника, критерій Стюдента та його рівень значущості, пов'язані залежностями із двома першими таблицями.

У масиві даних F12 і F14 заноситься розрахункове значення критеріїв Фішера, у масиві даних G12 і G14 – рівень їх значущості, порядок визначення яких наведено на рисунку 2.

Наявна повна відповідність результатів таблиць 2 і 1, проте менш трудомістким є дисперсійний аналіз з використанням інструменту «Двухфакторный дисперсионный анализ без повторений» програми Excel.

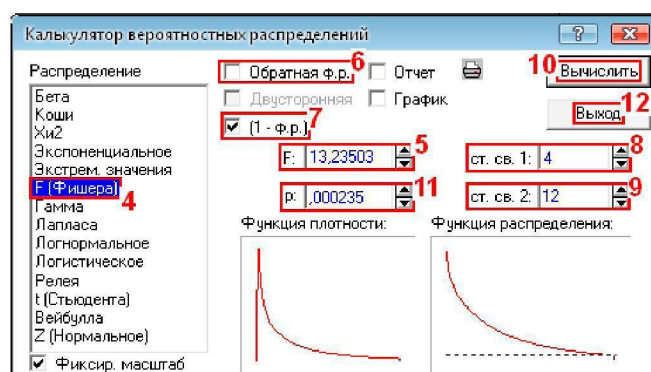


Рисунок 2 – Визначення рівня значущості p критерію Фішера за допомогою «Калькулятора вероятностных распределений ППП Statistica»

Застосування для дисперсійного аналізу однофакторного дослід з повтореннями інструменту аналізу програми Microsoft Excel «Дисперсионный анализ» **Однофакторный дисперсионный анализ** представляє результати дисперсійного аналізу відповідно усього фактору, тобто коли повторення чи варіанти розглядаються як одне ціле і показники дисперсії не розкладаються за цими факторами.

Проте, якщо виконувати дисперсійний аналіз масиву даних почергово групуванням за варіантами (рядками) і за повтореннями (стовбцями) можна отримати відповідні дані за якими розрахувати потрібне значення. Хоч цей метод досить громіздкий, проте він дає змогу проводити дисперсійний аналіз при відсутності деяких даних масиву.

Для прикладу розглянемо аналіз масиву B4:E8 таблиці 3 зі штучно вилученими значеннями в масивах даних C5, D5 і E6, представлених в таблиці 4.

Таблица 4 – Пропущені дані при повторенні експерименту

	A	B	C	D	E
1		Повторення			
2	Варіант	I	II	III	IV
3	St (Стандарт)	331	372	330	336
4	A	379	364	384	349
5	B	385			387
6	C	407	391	409	
7	D	432	503	470	423

Виконавши почергово **Однофакторный дисперсионный анализ** масиву даних A3:E7 (табл. 4) при групуванні **по строкам** і масиву даних B2:E7 при групуванні **по столбцам**, отримуємо відповідні листи виведення результатів. З даних на-

ведених на них формуємо таблицю (табл. 5) аналогічну за формою таблиці 1, вставляючи підсумкові і дисперсійні таблиці отриманих листів виведення результатів.

Таблиця 5 – Результат дисперсійного аналізу з використанням інструменту «Однофакторный дисперсионный анализ»

	A	B	C	D	E	F	G
1	Двухфакторный дисперсионный анализ без повторений						
2	ИТОГИ	Счет	Сумма	Среднее	Дисперсия		
3	St (Стандарт)	4	1369	342.3	400.25		
4	A	4	1476	369.0	250.00		
5	B	2	772	386.0	2.00	B16=B24	
6	C	3	1207	402.3	97.33	G16=G24	
7	D	4	1828	457.0	1355.33		
8							
9	I	5	1934	386.8	1407.20	B17=B30	
10	II	4	1630	407.5	4181.67	G17=G30	
11	III	4	1593	398.3	3374.92	B19=B26	
12	IV	4	1495	373.8	1546.25	C19=C26	
13							
14	Дисперсионный анализ						
15	Источник вариации	SS	df	MS	F	P-Значение	F критическое
16	Строки	29300.11	4	7325.028	14.146860	0.000171	3.259167
17	Столбцы	2576.23	3	858.74	0.338937	0.797559	3.410534
18	Погрешность	3637.19	12	303.099			
19	Итого	35513.53	16				
20	B18=B19-B16-B17 C18=C16*C17						
21	D18=B18/C18						
22	Дисперсионный анализ по строкам						
23	Источник вариации	SS	df	MS	F	P-Значение	F критическое
24	Между группами	29300.113	4	7325.028	14.146860	0.000171	3.259167
25	Внутри групп	6213.4167	12	517.785			
26	Итого	35513.53	16				
27							
28	Дисперсионный анализ по столбцам						
29	Источник вариации	SS	df	MS	F	P-Значение	F критическое
30	Между группами	2576.23	3	858.74	0.338937	0.797559	3.410534
31	Внутри групп	32937.3	13	2533.638			
32	Итого	35513.53	16				

Об'єднавши, як показано на полі таблиці 5, отримані у обох випадках дані в кінцеву таблицю (масив даних A15:G19) розрахунку дисперсій і виконавши додатково ряд визначень, наведених за представленими також на полі формулами, маємо результат визначення критерію Фішера F для варіантів і повторень (відповідно масиви даних E16 і E17).

Висновки. При виконанні дисперсійного аналізу однофакторних експериментів з повторенням дослідів з використанням програми Excel і ППП Statistica отримується повна відповідність результатів, проте менш трудомістким є дисперсійний аналіз з використанням інструменту «Двухфакторный дисперсионный анализ без повторений» програми Excel. У разі різній кількості повторень експериментів слід застосовувати інструмент «Однофакторный дисперсионный анализ»

програми Excel, аналізуючи досліджуваний масив почергово за досліджуваними факторами і за повтореннями.

Перелік посилань

1. *Доспехов Б. А.* Методика полевого опыта (с основами статистической обработки результатов исследований) / Доспехов Б. А. – М. : Агропромиздат, 1985. – 365 с.
2. *Литтл Т. М.* Сельскохозяйственное опытное дело. Планирование и анализ / Т. М. Литтл, Ф. Дж. Хиллз : пер. с англ. Б. Д. Кирюшкина; под ред. и с предисловием Д. В. Васильевой. – М. : Колос, 1981. – 320 с.
3. Використання Microsoft Excel 97 : Пер. з англ. / Брюс Холберг, Шеррі-Кінкоф, Біл Рей [та ін.] – К. : М. :СПб. :Видавн. дім «Вільямс», 1998. – 736 с.
4. Системний аналіз в селекції польових культур. Навчальний посібник. / [Літун П. П., Кириченко В. В., Петренкова В. П., Коломацька В. П.]. – Х. : Інст. Рослин, 2009, – 354 с.
5. *Боровиков В.* STATISTICA. Искусство анализа данных на компьютере : Для профессионалов. / В. Боровиков. – 2-е изд. (+CD). – СПб. : Питер, 2003. – 688 с. : ил.
6. *Ушкаренко И. А.* Планирование эксперимента и дисперсный анализ данных полевого опыта / И. А. Ушкаренко, И. Я. Скрипников. – К. – Одесса : Вища шк., 1988. – 120 с.

FEATURES OF THE USE OF THE PROGRAMS OF EXCEL AND STATISTICA FOR THE ANALYSIS OF VARIANCE OF EXPERIMENTS WITH ONE FACTOR WITH A REITERATION

Summary. Peculiarities of the use of the Excel program and package of applied programs Statistica for the analysis of variance of one-factorial experiments with the repetition of the experiments. The examples shows the order of its execution, with the same or a different number of replications of experiments. A comparison of the results obtained and time consuming when using different methods of analysis of variance.