

УДК 004: 311.21: 338.43

Н. М. Самарець,

к. т. н., доцент, доцент кафедри інформаційних систем і технологій,  
Дніпропетровський державний агроуніверситет

Є. М. Харченко,

к. т. н., доцент, доцент кафедри інформаційних систем і технологій,  
Дніпропетровський державний агроуніверситет

Н. О. Чорна,

старший викладач кафедри інформаційних систем і технологій,  
Дніпропетровський державний агроуніверситет

## ВИКОРИСТАННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ У СТАТИСТИЧНОМУ АНАЛІЗІ ДАНИХ ДЛЯ АГРАРНИХ ПІДПРИЄМСТВ

N. Samarets,

Candidate of Engineering Sciences, Associate Professor,  
Associate Professor of Information Systems and Technology Department, Dnipropetrovs'k State Agrarian University

E. Kharchenko,

Candidate of Engineering Sciences, Associate Professor,  
Associate Professor of Information Systems and Technology Department, Dnipropetrovs'k State Agrarian University

N. Chorna,

Senior Lecturer of Information Systems and Technology Department, Dnipropetrovs'k State Agrarian University

### THE USE OF INFORMATION TECHNOLOGY IN THE STATISTICAL ANALYSIS OF DATA FOR AGRICULTURAL ENTERPRISES

Розглянуто сучасні інформаційні технології для статистичного аналізу експериментальних даних, сформульовано методологічні аспекти застосування деяких методів статистичної обробки та наведено приклади їх використання у практиці агропромислових підприємств.

The article considers the modern information technologies for statistical analysis of experimental data, formulates methodological aspects of use of certain methods of statistical analysis and gives examples of its use in operations of agricultural enterprises.

*Ключові слова: методи статистичного аналізу, інформаційні технології, агропромислові підприємства, перевірка гіпотез, дисперсійний аналіз, кореляція, регресія.*

*Key words: statistical analysis methods, information technology, agro-industrial enterprises, hypothesis testing, analysis of variance, correlation, regression.*

#### ПОСТАНОВКА ПРОБЛЕМИ

Ефективність та якість управління такою складною економічною системою, якою є сучасне агропромислове підприємство, суттєво зростають при застосуванні математичної статистики та комп'ютерної техніки як інструментів дослідницької роботи. Стохастична природа факторів сільськогосподарського виробництва (вплив на результати виробництва погодних факторів, масових захворювань тварин у періоди епідемій, враження сільськогосподарських культур комахами та грибовими захворюваннями тощо) створює в сільсько-

му господарстві деяку невизначеність у динаміці розвитку та вимагає застосування дієвих інструментів і методів статистичного аналізу експериментальних даних.

Використання специфічних методів математичної статистики та новітніх інформаційних технологій дає змогу систематизації, обробки і аналізу даних спостережуваних явищ з метою встановлення притаманних для них статистичних закономірностей та застосування останніх у практичній діяльності сільськогосподарських підприємств. Статистичні моделі використовують для діагнос-

тики стану об'єктів дослідження, при вивченні причинно-наслідкового механізму формування варіації та динаміки агротехнологічних явищ і процесів, у моніторингу кон'юнктури ринку, при прогнозуванні та прийнятті рішень.

Найважливіші розділи математичної статистики такі: статистичні ряди розподілу, оцінка параметрів розподілу, закони розподілу вибіркового характеру, перевірка статистичних гіпотез, дисперсійний, кореляційно-регресійний, коваріаційний аналіз. Останнім часом знаходять використання методи багатомірного статистичного аналізу — факторний аналіз, метод головних компонентів, кластерний аналіз тощо [4, с. 7]. Очевидно, що поширеному впровадженню методів математичної статистики для розв'язання різного роду господарських задач аграрних підприємств суттєво сприяє впровадження комп'ютерної техніки та новітніх інформаційних технологій.

### АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Дослідженнями проблеми статистичної обробки експериментальних даних займалися такі науковці, як С.А. Айвазян, Л.Н. Большее, В.П. Боровиков, Б.В. Гнеденко, С.В. Дронов, А.М. Єріна, Н.В. Смирнов, Г.А. Кимбл, А.И. Кобзарь, Н.Ш. Кремер, А.Т. Опря, А.И. Орлов, І.М. Яглом та ін.

### ПОСТАНОВКА ЗАВДАННЯ

Метою статті є аналіз сучасних інформаційних технологій та основних методів статистичного аналізу експериментальних даних для одержання стійких кількісних характеристик параметрів причинно-наслідкових зв'язків явищ з метою здійснення різноманітних сільськогосподарських розрахунків, прогнозування діяльності агропромислових підприємств та прийняття науково обгрунтованих управлінських рішень.

### ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ ДОСЛІДЖЕННЯ

Програмне забезпечення для статистичних досліджень.

Основні загальні вимоги до програмного забезпечення для статистичного аналізу експериментальних даних наступні:

- наявність зручних засобів для роботи з вихідними даними;
- розрахунок статистичних характеристик;
- підтримка методів побудови моделей взаємозв'язків;
- підтримка методів оцінки адекватності моделей;
- реалізація методів аналізу і моделювання часових рядів;
- реалізація методів прогнозування;
- реалізація статистичних критеріїв;
- забезпечення можливості створення і збереження сценарію дослідження;
- візуалізація проміжних і кінцевих результатів дослідження.

Програмне забезпечення, яке застосовується в статистичних дослідженнях, можна розділити на наступні групи:

1) програми, що реалізують технологію електронних таблиць MS Excel, OpenOffice.org Calc та ін.;

2) статистичні пакети загального призначення: SPSS, STATISTICA, STATGRAPHICS та ін.;

3) програми, орієнтовані на рішення економетричних задач: Econometric Views, STADIA, Matrixer та ін.;

4) спеціалізовані статистичні пакети, призначені для рішення обмеженого кола завдань: EBPI-STA, МЕЗОЗАВР, ОЛІМП, Forecast Expert та ін.;

5) математичні пакети загального призначення Mathcad, Matlab, Maple, Mathematica та ін.

Основними перевагами програми для роботи з електронними таблицями MS Excel пакету Microsoft Office є доступність, наявність великого методичного забезпечення з прикладами використання та узгодження з більшістю статистичних пакетів прикладних програм, багаті засоби форматування та відображення даних. Широкому застосуванню MS Excel сприяє його здатність швидко та ефективно проводити статистичний аналіз та обробку даних, будувати графіки та діаграми, проводити різноманітні обчислення з використанням апарата статистичних, інженерних, фінансових функцій та формул, досліджувати вплив різних факторів на дані та ін.

Електронні таблиці MS Excel є, на наш погляд, найбільш зручним інструментом для статистичного аналізу експериментальних даних агропромислових підприємств. Авторами [1; 5] розроблено комплекс прикладних програм на базі MS Excel для розрахунків статистичних показників у сільськогосподарському виробництві. За їх допомогою можна отримати показники та їх графічне зображення для генеральної та вибіркової сукупностей, коефіцієнти кореляції та регресії, обчислити критерій  $\chi^2$ -квадрат, провести дисперсійний аналіз однофакторних та багатфакторних комплексів для багаточисельних вибірових сукупностей.

Статистичні методи перевірки гіпотез.

Завдання перевірки статистичних гіпотез постає в різних сферах сільськогосподарської практичної діяльності. Так, внаслідок наявності елементу випадковості при постановці дослідження досить часто виникає питання про розходження груп спостережень [5, с. 110]. Його вирішення зводиться до порівняння статистичних характеристик, що оцінюють параметри законів розподілу, тобто до перевірки статистичних гіпотез. Статистична гіпотеза — це наукове припущення про статистичні закони розподілу розглянутих випадкових величин, що може бути перевірене на основі вибірки. Майже завжди задача полягає в перевірці нульової гіпотези  $H_0$  — про відсутність реального розходження між фактичними і теоретичними показниками. Якщо розходження близькі до нуля чи знаходяться в області припустимих значень, то

Таблиця 1. Очікувані кількості комах за групами

Група комах	КІЛЬКІСТЬ КОМАХ				Σ	%
	загиблі		здорові			
	f	F	f	F		
Дослідна	28	19,2(F <sub>1</sub> )	2	10,8(F <sub>2</sub> )	30	60
Контрольна	4	12,8(F <sub>3</sub> )	16	7,2(F <sub>4</sub> )	20	40
Σ	32	32	18	18	50	100
%	64	-	36	-	-	100

нульова гіпотеза приймається. Якщо ж розходження виявляються в критичній для даного статистичного критерію області, неможливі при даній гіпотезі і тому несумісні з нею, то H<sub>0</sub> відкидається. Прийняття гіпотези означає, що дані не суперечать припущенню про відсутність розходжень між фактичними і теоретичними чи двома рядами фактичних розподілів. Відкидання гіпотези означає, що емпіричні дані несумісні з H<sub>0</sub> та вірна інша, альтернативна гіпотеза. Справедливість H<sub>0</sub> перевіряється обчисленням статистичних критеріїв перевірки для визначеного рівня значимості.

Критерій хі-квадрат (χ<sup>2</sup>) застосовують, коли необхідно встановити відповідність двох порівнюваних рядів розподілу — емпіричного і теоретичного чи двох емпіричних. Він представляє суму квадратів відхилень емпіричних частот f від теоретичних F, віднесену до теоретичних частот:

$$\chi^2 = \sum \frac{(f - F)^2}{F}$$

Цей показник використовується при вивченні якісних ознак для оцінки відповідності емпіричних даних теоретичній передумові, нульовій гіпотезі H<sub>0</sub>. Гіпотеза відкидається, якщо  $\chi^2_{\text{факт}} > \chi^2_{\text{теор}}$ , і не відкидається, якщо  $\chi^2_{\text{факт}} < \chi^2_{\text{теор}}$ ; коли фактичні та теоретичні очікувані частоти збігаються,  $\chi^2 = 0$ . Гранічні значення χ<sup>2</sup>, при яких H<sub>0</sub> приймається, приведені в довідкових таблицях. Величина χ<sup>2</sup> залежить від числа ступенів волі. У найбільш типових випадках число ступенів волі

Таблиця 2. Розподілення очікуваної кількості комах за групами

загиблих в основній групі	$F_1 = \frac{60 \cdot 32}{100} = 19,2;$
здорових у дослідній групі	$F_2 = \frac{60 \cdot 18}{100} = 10,8;$
загиблих у контрольній групі	$F_3 = \frac{40 \cdot 32}{100} = 12,8;$
здорових у контрольній групі	$F_4 = \frac{40 \cdot 18}{100} = 7,2.$

Таблиця 3. Розходження між фактичними та очікуваними кількостями комах у групах

	f-F <sub>i</sub>	(f-F <sub>i</sub> ) <sup>2</sup>	$\frac{(f - F_i)^2}{F_i}$
f-F <sub>1</sub>	8,8	77,44	4,04
f-F <sub>2</sub>	-8,8	77,44	7,17
f-F <sub>3</sub>	-8,8	77,44	6,05
f-F <sub>4</sub>	8,8	77,44	10,75
		χ <sup>2</sup> =	28,0

визначається за формулою (c - 1) · (K - 1), де c — число рядків, K — число стовпчиків в аналітичній таблиці.

Приклад із агрономії. З 50 жуків бурячного довгоносика було оброблено ділором 30 комах (дослідна група), а 20 не оброблялися (контрольна

	A	B	C	D	E	F	G	H	I	J
1	<b>Вычисление и использование критерия Хи-квадрат.</b>									
2	Условие задачи:	Из 50 жуков свекловичного долгоносика было обработано дилором 30 (опытные насекомые), а 20 не обрабатывались (контроль). В опытной группе погибло 28 долгоносиков, а в контрольной - 4. Доказывают ли результаты опыта токсическое действие дилора, или гибель жуков зависит от								
3										
4			f	F	f	F	сумма	процент		
5		опытная	28	19,2	2	10,8	30	60		
6		контрольная	4	12,8	16	7,2	20	40		
7		Сумма	32	32	18	18	50	100		
8			f·Fi	(f-F) <sup>2</sup>	(f-F) <sup>2</sup> /Fi					
9			8,8	77,44	4,033333333					
10			-8,8	77,44	7,17037037					
11			-8,8	77,44	6,05					
12			8,8	77,44	10,75555556					
13				χ <sup>2</sup> =	28,00925926					
14			При v = (c - 1) · (K - 1) = (2 - 1) · (2 - 1) = 1							
15			Теоретическое значение ( по таблице приложения ) χ <sup>2</sup> = 3,84							
16			Вывод:							
17			Наблюдается существенное увеличение гибели обыкновенного свекловичного долгоносика при обработке жуков дилором ( χ <sup>2</sup> <sub>факт</sub> > χ <sup>2</sup> <sub>теор</sub> ) и нулевая гипотеза о независимости гибели вредителя от дилора отвергается							
18										
19										
20										

Рис. 1. Електронна таблиця умови задачі та результатів обчислень

група). У дослідній групі загинуло 28 довгоносиків, а в контрольній — 4. Необхідно з'ясувати, чи доводять результати досвіду токсичну дію ділора, або загибель жуків залежить від випадкових причин.

У таблицях 1, 2 наведено обчислені очікувані кількості комах по групах, де  $f$  і  $F$  — відповідно фактично спостережена та теоретично очікувана кількості. При  $v = (c - 1) \cdot (K - 1) = (2 - 1) \cdot (2 - 1) = 1$  теоретичне значення  $x^2 = 3,84$ .

Розходження між фактичними та очікуваними кількостями комах у групах наведено у таблиці 3.

На основі проведених обчислень можна зробити висновок, що спостерігається істотне збільшення загибелі звичайного бурячного довгоносика при обробці жуків ділором ( $x^2_{\text{факт}} > x^2_{\text{теор}}$ ), і нульова гіпотеза про незалежність загибелі шкідника від ділора відкидається.

Для рішення задачі застосовано електронні таблиці Excel Microsoft Office. На рис. 1 наведена умова задачі, розміщення даних та результати обчислень.

#### Дисперсійний аналіз.

Головне призначення дисперсійного аналізу — статистично виявити вплив різних факторів на мінливість ознаки, що вивчається. Для аграрних підприємств дисперсійний аналіз можна використовувати при оцінці ефективності препаратів, які застосовуються для боротьби з різноманітними хворобами тварин, при визначенні частки і вірогідності впливу факторів зовнішнього середовища на досліджувану ознаку [5, с. 117]. Його застосовують для порівняння декількох сортів однієї культури або декількох порід домашніх тварин за будь-якою із кількісних чи якісних ознак; встановлення реакції рослин або тварин на той або інший зовнішній вплив (добрива, харчі, пестициди і т. ін.). При цьому можливо як вивчення впливу кожного фактора окремо, так і їхнього спільного впливу.

При підборі даних для дисперсійного аналізу необхідно виконувати наступні вимоги:

- добір і вибірку необхідно проводити за принципом випадковості;
- вибірка повинна об'єктивно відображати генеральну сукупність, частиною якої вона є;
- вибірки можуть бути нечисленними і численними.

Мінливість ознак рослин і тварин визначається впливом численних факторів. Ознаки, що змінюються під впливом тих чи інших причин, називаються результативними. Фактори, які впливають на результативну ознаку, позначаються буквами латинського алфавіту ( $A, B, C$ ), а ознаки, що враховуються. —  $x, y, z$ . Фактори в досліді випробуються серійно, тобто мають градації. Градація фактора — це ступінь його впливу на результативну ознаку. Наприклад, різні норми витрати пестицидів, норми випуску, трихограми та ін., тобто градації — це відособлені один від одного гру-

пи (класи). Вони позначаються тими ж буквами, що і фактори: наприклад, градації фактора  $A$  позначаються через  $A_1, A_2, A_3$  і т.д.

Фактори, що впливають на ступінь варіювання результативної ознаки, поділяються на організовані та випадкові. Організовані (чи регульовані, систематичні) фактори викликаються дією досліджуваного в експерименті фактора, якому в досліді надається декілька значень (градацій). Наприклад, кілька норм витрат пестициду і т.д. Випадкові фактори (причини) обумовлюються природним варіюванням всіх ознак біологічних об'єктів у природі. Це неорганізовані, неконтрольовані в даному досвіді фактори. До випадкових факторів відноситься, наприклад, мінливість, обумовлена погрішностями вимірів при постановці досвіду.

Дисперсійний аналіз — це розчленування загальної суми квадратів відхилень і загального числа ступенів волі на частини компоненти, що відповідають структурі експерименту, оцінка значимості дії і взаємодії досліджуваних факторів по  $F$ -критерію. Він дозволяє одночасно обробляти дані декількох вибірок (варіантів), що складають єдиний статистичний комплекс, оформлений у вигляді спеціальної робочої таблиці. Дисперсійний комплекс — це сукупність градацій з обмірюваними значеннями результативної ознаки. При вивченні одного фактора комплекс називається однофакторним, а декількох з їхньою дією і взаємодією — багатфакторним. Дисперсійний аналіз розроблений і введений у практику сільськогосподарських і біологічних досліджень англійським ученим Р.Фішером, що відкрив закон розподілу відносини середніх квадратів (дисперсій):

$$F = \frac{\sigma_1^2}{\sigma_2^2},$$

де  $\sigma_1^2$  — середній квадрат вибірових середніх;  
 $\sigma_2^2$  — середній квадрат об'єктів.

При обробці однофакторних статистичних комплексів, що складаються з декількох незалежних вибірок ( $r$  варіантів), загальна мінливість результативної ознаки, вимірювана загальною сумою квадратів  $C_y$ , розчленовується на два компоненти: варіювання між вибірками (варіантами)  $C_x$  і в середині вибірок  $C_z$ . У загальній формі мінливість ознаки буде подана виразом  $C_y = C_x + C_z$ . Варіація між вибірками (варіантами) є частиною загальної дисперсії, обумовленою дією досліджуваних факторів. Дисперсія в середині вибірок характеризує випадкове варіювання досліджуваної ознаки, тобто помилку досвіду. Загальне число ступенів волі ( $N - 1$ ) також розчленовується на дві частини: ступеня волі для варіантів ( $r - 1$ ) та для випадкового варіювання ( $N - r$ ).

Приклад із ветеринарії. Необхідно вивчити дію різних доз біостимулятора на життєздатність поросят (визначається за приростом живої маси в

Таблиця 4. Градації дисперсійного комплексу

Показник	Градації					Підсумкові дані
	1	2	3	4	5	Число градацій $r = 5$
Варіанти (дані) $x$	2	3	4	8	4	$H_{\Sigma} = \frac{(\sum \sum x)^2}{N} = 556,96$ $\bar{X} = \frac{\sum \sum x}{N} = 4,72$
	1	6	7	6	6	
	3	4	5	9	5	
	1	3	9	7	4	
	2	2	6	8	3	
$n_i$	5	5	5	5	5	$N = 25$
$\sum x$	9	18	31	38	22	$\sum \sum x = 118$
$H_i = (\sum x)^2 / n_i$	16,2	64,8	192,2	288,8	96,8	$\sum H_i = 658,8$
$\sum x^2$	19	74	207	294	102	$\sum \sum x^2 = 696$

Таблиця 5. Дисперсії та варіанси дисперсійного комплексу

Розрахунок дисперсій та варіанс		Підсумкові дані	
Факторіальна дисперсія $C_x = \sum H_i - H_{\Sigma}$	$= 658,80 - 556,96 = 101,84$	Показник сили впливу $\eta_x^2 = \frac{C_x}{C_y}$	$= \frac{101,84}{139,04} = 0,732$
Випадкова дисперсія $C_z = \sum \sum x^2 - \sum H_i$	$= 696 - 658,8 = 37,2$	Критерій вірогідності $F = \frac{\sigma_x^2}{\sigma_z^2}$	$= \frac{25,46}{1,86} = 13,69$
Загальна дисперсія $C_y = \sum \sum x^2 - H_{\Sigma}$	$= 696 - 556,96 = 139,04$	$\nu_1 = r - 1$	$= 5 - 1 = 4$
Факторіальна варіанса $\sigma_x^2 = \frac{C_x}{r - 1}$	$= \frac{101,84}{5 - 1} = 25,46$	$\nu_2 = N - r$	$= 25 - 5 = 20$
Випадкова варіанса $\sigma_z^2 = \frac{C_z}{N - r}$	$= \frac{37,2}{25 - 5} = 1,86$	$F_{st}$	2,9; 4,4; 7,1

кілограмах за період дослідження). Для цього сформовано 5 груп по 5 голів у групі, у яких випробувалися різні дози препарату.

На основі результатів дослідження побудовано дисперсійний комплекс, наведений у

Таблиця 6. Початкові дані для визначення показників зв'язку між кількісними ознаками

x	5,8	8,3	6,0	9,8	6,2	7,4	7,2	8,6	7,7	8,0
y	10,0	11,6	9,5	13,0	9,6	11,0	10,1	12,2	10,5	13,3

Таблиця 7. Результати роботи інструменту "Регрессія"

ВЫВОД  
ИТОГОВ

Регрессионная статистика						
Множественный R	0,855976598					
R-квадрат	0,732695936					
Нормированный R-квадрат	0,699282928					
Стандартная ошибка	0,760204831					
Наблюдения	10					
Дисперсионный анализ						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	1	12,67270891	12,67270891	21,92846382	0,001576184	
Остаток	8	4,623291086	0,577911386			
Итого	9	17,296				
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	4,034387187	1,523661936	2,647823044	0,029354786	0,520816464	7,54795791
x	0,939415042	0,200610382	4,682783768	0,001576184	0,476806671	1,402023412

Таблиця 8. Експериментальні дані залежності вмісту сухої речовини вівса від зовнішніх факторів

№ досліду	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	У
1	14,3	10,8	44,0	50	15,78
2	17,3	10,7	44,5	35	15,33
3	16,6	9,3	40,0	43	16,07
4	11,9	9,0	34,5	43	17,78
5	15,3	11,0	43,0	26	18,23
6	18,2	9,9	43,5	30	16,23

Таблиця 9. Вид рівнянь регресії та ступінь апроксимації

Зовнішні фактори	Регресійна модель	Ступінь апроксимації
температура повітря	$Y = 8095 - 2155,8x + 213,91x^2 - 9,3578x^3 + 0,1524x^4$	$R^2 = 0,96$
кількість годин соячного сівва	$Y = 64653 - 26031x + 3926,8x^2 - 262,98x^3 + 6,597x^4$	$R^2 = 0,98$
температура ґрунту	$Y = 3436,9 - 260,82x + 6,59x^2 - 0,0554x^3$	$R^2 = 0,79$
вологість повітря	$Y = 136,55 - 9,64x + 0,2515x^2 - 0,0021x^3$	$R^2 = 0,77$

табл. 4, 5. Він має таку структуру: обсяг комплексу  $N = 25$ , число градацій  $r = 5$ , число варіант у градації  $n = 5$ .

На основі проведених розрахунків зроблено наступний висновок: із усіх факторів, що впливали на приріст живої маси порослят, частка впливу нового рецепта біостимулятора складає 73,2%.

Визначення показників зв'язку між кількісними ознаками.

Більшість явищ природи, які є об'єктами наукових досліджень, характеризуються кореляційними зв'язками, коли одному значенню аргументу відповідають кілька значень функції. Вони можуть бути прямими і зворотними, слабкими і сильними, лінійними і криволінійними, простими і множинними. Для вивчення тісноти зв'язків і виду залежностей між явищами, що спостерігаються при досліджуванні, застосовується кореляційно-регресійний аналіз. Вивчення взаємозв'язків кореляційного типу має особливо велике значення при аналізі явищ, які складаються під впливом значної кількості факторів. Такими явищами і є сільськогосподарські виробничі процеси — складні багатоступеневі системи, на які впливають різноманітні чинники.

За допомогою регресійного аналізу по експериментальним даним встановлюють форму залежності між вхідними факторами  $X_1, X_2, \dots, X_n$  та вхідним фактором  $Y$  у вигляді:

$$Y = F(X_1, X_2, \dots, X_n),$$

де  $n$  — кількість вхідних факторів.

Коефіцієнт кореляції  $r$  визначає тісноту зв'язку між досліджуваними змінними для лінійної регресії, а важливим показником якості регресійної моделі є множинний коефіцієнт детермінації  $R$ , який показує, на скільки регресійна модель пояснює розсіювання експериментальних даних.

Приклад із ветеринарії. Одержати та оцінити рівняння парної лінійної регресії  $\hat{y}_x = a_0 + a_1x$  між числом еритроцитів у млн/мкл ( $X$ ) і вмістом гемоглобіну в г/л ( $Y$ ) у крові овець, згідно з даними

вибіркової сукупності за допомогою інструменту "Регрессия", вбудованого в електронні таблиці MS Excel.

У табл. 6 наведено початкові дані для визначення показників зв'язку між кількісними ознаками, у табл. 7 — результати роботи інструменту "Регрессия" (Сервис → Анализ данных → Регрессия).

Рівняння парної лінійної регресії залежності між вмістом гемоглобіну ( $Y$ ) в г/л та числом еритроцитів ( $X$ ) у млн/мкл у крові овець має вид:

$$\hat{y}_x = 4,034 + 0,939x.$$

Проведені розрахунки свідчать, що при зміні числа еритроцитів у крові на 1 млн/мкл вміст гемоглобіну зростає в середньому на 0,939 г/л. Аналіз тісноти зв'язку між ознаками показує існування сильної додатної кореляції:  $r = 0,86$ . Коефіцієнт детермінації  $R^2 = 0,73$ , тобто 73% варіації  $Y$  пояснюється варіацією  $X$ , а 27% варіації  $Y$  пояснюється дією не включених до регресійної моделі факторів. Оцінка значущості коефіцієнтів регресії по  $t$ -розподілу Ст'юдента показує їх статистичну значущість на рівні значущості  $\alpha = 0,05$ :  $t_{a_0} = 2,65$ ;  $t_{a_1} = 4,68$ ;  $t_{8;0,05} = 2,31$ , таким чином  $t_{a_0} > t_{8;0,05}$ ;  $t_{a_1} > t_{8;0,05}$ . Оцінка значущості рівняння регресії в цілому за допомогою  $F$ -критерію Фішера дозволяє зробити висновок, що регресійна модель адекватна: розрахункове значення  $F = 21,93$ , критичне значення  $F_{1;8;0,05} = 5,32$ , отже,  $F > F_{1;8;0,05}$ .

Для коефіцієнтів регресії можна записати їх довірчі інтервали при рівні значущості  $\alpha = 0,05$ :

$$0,52 \leq a_0 \leq 7,55; \quad 0,48 \leq a_1 \leq 1,40.$$

За допомогою отриманого рівняння регресії розраховують прогнозне значення  $\hat{y}_x$  при будь-якому значенні  $x$ . Точковий прогноз одержуємо при підстановці прогнозних значень факторної ознаки в рівняння регресії, наприклад,  $\hat{y}_x(10) = 13,43$ .

Приклад із агрономії. Дослідити вплив факторів зовнішньої середовища на хімічний склад зеленої маси вівса протягом доби. У табл. 8 наведено експериментальні дані для визначення показників зв'язку між кількісними ознаками. Прийнято до уваги наступні зовнішні фактори погодних умов:  $X_1$  — температура повітря, °C;  $X_2$  — кількість годин сонячного сяйва;  $X_3$  — температура ґрунту, °C;  $X_4$  — вологість повітря, %;  $Y$  — вміст сухої речовини, %.

У результаті проведення нелінійного регресійного аналізу залежності хімічного складу вівса від окремих факторів погодних умов побудовано чотири поліноміальні моделі:

$$1) Y = A_0 + A_1 X_1 + A_2 X_1^2 + A_3 X_1^3 + A_4 X_1^4,$$

$$2) Y = A_0 + A_1 X_2 + A_2 X_2^2 + A_3 X_2^3 + A_4 X_2^4,$$

$$3) Y = A_0 + A_1 X_3 + A_2 X_3^2 + A_3 X_3^3,$$

$$4) Y = A_0 + A_1 X_4 + A_2 X_4^2 + A_3 X_4^3,$$

та отримано наступні результати:

1) збільшення порядку поліному не дає значного покращення ступеню апроксимації;

2) зменшення порядку поліному дає суттєве зниження ступеню апроксимації;

3) поліноми вище п'ятого порядку не поліпшують якості регресійної моделі з точки зору ступеня апроксимації.

Вид рівнянь регресії та ступінь апроксимації рівнянням регресії вихідних даних наведено у таблиці 9.

На основі проведених обчислень зроблено такі висновки. Найвища залежність вмісту сухої речовини вівса спостерігалась від температури повітря та кількості годин сонячного сяйва, менший вплив виявили температура ґрунту та вологість повітря. Нелінійний характер апроксимації ускладнює виконання прогнозу, оскільки потребує для цього виявлення сезонних компонент, лінії тренду та можливих циклічних коливань. Для виконання таких досліджень необхідно мати більший обсяг вибірки, а відповідно і дослідів.

### ВИСНОВКИ

Для систематизації, обробки і аналізу даних спостережуваних явищ з метою встановлення притаманних для них статистичних закономірностей необхідне використання специфічних методів математичної статистики та інформаційних технологій. Застосування багатого арсеналу методів статистичної обробки експериментальних даних та економіко-математичного моделювання є важливою складовою успішних досліджень та якісного аналізу в сільськогосподарській сфері. Так, зокрема, у практиці підприємств АПК доцільно використовувати наступні методи статистичного аналізу: перевірка статистичних гіпотез по критерію  $\chi^2$ -квадрат, дисперсійний аналіз, розрахунок коефіцієнтів кореляції та регресії, основні теоретичні положення та приклади, застосування яких наведено у представленій роботі.

Пріоритетним напрямом вдосконалення методологічних аспектів статистичної обробки даних

для агропромислових підприємств є широке застосування сучасних інформаційних технологій. Зручним та ефективним інструментом для статистичного аналізу експериментальних даних є програма для роботи з електронними таблицями MS Excel, за допомогою якої і були проведені розглянуті дослідження.

### Література:

1. Леснікова І.Ю. Основи роботи і вирішення задач сільського господарства в середовищі електронних таблиць Excel / І.Ю. Леснікова, Є.М. Харченко. — Дніпропетровськ: Пороги, 2002. — 147 с.

2. Дронов С.В. Многомерный статистический анализ: Учебное пособие / С.В. Дронов. — Барнаул: Изд-во Алт. гос. ун-та, 2003. — 213 с.

3. Єріна А.М. Статистичне моделювання та прогнозування: Навч. посіб. / А.М. Єріна. — К.: КНЕУ, 2001. — 170 с.

4. Опря А.Т. Статистика (модульний варіант з програмованою формою контролю знань): Навч. посібник / А.Т.Опря. — К.: Центр учбової літератури, 2012. — 448 с.

5. Терещенко М.В. Моделювання технологічних процесів у середовищі Microsoft Excel: Навч. посібник / М.В. Терещенко, Є.М. Харченко, В.М. Ковшов, І.Ю. Леснікова, В.О. Петренко, О.І. Гогенко, Ф.К. Клименко. — Дніпропетровськ: Пороги, 2005. — 266 с.

### References:

1. Lesnikova, I.Yu. and Kharchenko, Ye.M. (2002), *Osnovy roboty i vyrishennia zadach sil's'koho hospodarstva v seredovyschi elektronnykh tablyt' Excel* [Fundamentals and solving problems of agriculture in the environment spreadsheet Excel], Porohy, Dnipropetrovs'k, Ukraine.

2. Dronov, S.V. (2003), *Mnogomernyj statisticheskij analiz: Uchebnoe posobie* [Multivariate statistical analysis: Teach. manual], Publishing house Alt. State University, Barnaul, Russia.

3. Erina, A.M. (2001), *Statystychne modeliuвання ta prohnozuvannia: Navch. Posibnyk* [Statistical modeling and prediction: Teach. manual], KNEU, Kyiv, Ukraine.

4. Oprya, A.T. (2012), *Statystyka (modul'nyj variant z prohramovanoi formoiu kontroliu znan')*. Navch. posibnyk [Statistics (modular version with a programmable form of control knowledge)]. Teach. manual], Tsentr uchbovoi literatury, Kyiv, Ukraine.

5. Tereshchenko, M.V. Kharchenko, Ye.M. Kovshov, V.M. Lesnikova, I.Yu. Petrenko, V.O. Hohenko, O.I. and Klymenko, F.K. (2005), *Modeliuвання tekhnolohichnykh protsesiv u seredovyschi Microsoft Excel: Navch. posibnyk* [Modeling of processes in an environment Microsoft Excel: Teach. manual], Porohy, Dnipropetrovs'k, Ukraine.

*Стаття надійшла до редакції 27.09.2013 р.*