

АГРОМЕТЕОРОЛОГІЯ

УДК 551.501:631.55

І.П. Криворучко

ОБҐРУНТУВАННЯ СПОСОБУ АНАЛІТИЧНОГО РОЗРАХУНКУ СТАТИСТИЧНОГО МАКСИМУМУ ВРОЖАЙНОСТІ

Проаналізовано поведження рядів урожайності озимої пшениці та ярого ячменю. Проведено порівняльний аналіз двох способів знаходження статистичного максимуму врожайності: графічного по ймовірнісній клітчатці та розрахункового для нормального розподілу. Подано обґрунтування близькості цих способів його знаходження.

Ключові слова: ряди урожайності, статистичний максимум урожайності, господарський максимум урожайності, нормальний закон розподілу, реперний рік, трендова складова.

Вступ

Часто для оцінки стану, властивостей чи прогнозування процесів, явищ виникає необхідність знаходження таких екстремальних значень, які можуть вважатися статистичними максимумами. Під статистичним максимумом урожайності будемо розуміти потенційну урожайність певної культури за деякий фіксований інтервал часу. Вона визначається як внутрішніми біологічними (генетичними) властивостями рослини, так і особливостями території (грунт, рельєф) та клімату. Але через сортове (генетичне) розмаїття посівів у межах господарства, мінливість технології вирощування та їх динаміку в часі осереднене по площі значення статистичного максимуму вмщує тренд урожайності і змінюється з року в рік. Прийнятий у агрометеорології розрахунок значень щорічного статистичного максимуму врожайності для певної території, що дістав назву господарського максимуму, здійснюється за методом, описаним у [1]. Але графічний характер способу знаходження цієї величини робить неможливим пряме його застосування в автоматизованих системах розрахунку врожайності сільськогосподарських культур.

Порівняльний аналіз способів знаходження статистичного максимуму врожайності

На прикладі ярого ячменю в [1] наведено розрахунок статистичного максимуму врожайності у Вінницькій області за даними останніх 25 років. Натуральний ряд урожайності розміщується в зростаючому порядку від першого до останнього номера незалежно від року. Для кожного номера цього ряду обраховується забезпеченість ($P, \%$) отримання заданої врожайності [2]:

$$P = 100 \cdot m / (n + 1), \quad (1)$$

де P – забезпеченість урожайності з даним порядковим номером; m – порядковий номер урожайності в зростаючому ряді; n – об'єм вибірки.

Отримані дані наносяться на бланк клітчатки ймовірностей, де по осі абсцис відкладається забезпеченість P в одиницях значень нормального розподілу, по осі ординат – урожайність Y у рівномірному масштабі. Потім проводиться пряма лінія так, щоб нанесені точки за візуальною оцінкою розміщувалися якомога ближче до неї. Вибравши правильне розміщення лінії, її екстраполюють до обрізу клітчатки. Статистичний максимум урожайності U_c знімають з цього графіка для забезпеченості 99,9 % ($U_c = 4,06$ т/га). До недоліку такої методики можна віднести наявність певного елемента суб'єктивізму, який полягає в індивідуальному (суб'єктивному) виборі розміщення лінії на координатній площині, а отже й в отриманні остаточного результату. Крім того, даний метод у наведеному вигляді неможливо прямо використати у автоматизованих системах розрахунку врожайності сільськогосподарських культур.

Ймовірність того, що можливе значення x , яке належить нормальному закону розподілу і буде знаходитися в межах від $-\infty$ до x_0 , визначається за виразом:

$$F(x_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_0} e^{-\frac{(t-\alpha)^2}{2\sigma^2}} \cdot dt, \quad (2)$$

де верхня межа інтегрування x_0 представляє собою відхилення величини x від α в одиницях σ [3]. Через симетрію нормального розподілу, можемо розглядати інтеграл лише для додатних значень змінної:

$$F(x_0) = \frac{1}{\sqrt{2\pi}} \int_0^{x_0} e^{-Z^2} \cdot dZ, \quad (3)$$

$$\text{де } Z = \frac{t - \alpha}{\sqrt{2} \cdot \sigma}.$$

Цей, так званий інтеграл Лапласа, має обраховані табличні значення. Задавши необхідне значення ймовірності, можемо з таблиці визначити відхилення в одиницях σ , яке забезпечує цю умову. Наприклад, значенню $F(t) = 0,999$ відповідає $t = 3,09$. Це означає, що за умови належності значення випадкової величини до цього ж розподілу воно буде менше за $Y_{\text{сер}} + 3,09 \cdot \sigma$ з ймовірністю $P = 0,999$. Це значення можемо прийняти за статистичний максимум цього розподілу Y_C , тобто:

$$Y_C = Y_{\text{сер}} + 3,09 \cdot \sigma \quad (4)$$

Фактично для визначення статистичного максимуму в традиційному графічному і чисельному за виразом (4) методах використовується та ж сама кумулятивна крива нормального розподілу (шкала абсцис для клітчатки ймовірностей). Але для останнього необхідності у побудові графіка немає, оскільки існує числовий зв'язок між кумулятивною (накопиченою) ймовірністю і величиною змінної, вираженою в одиницях стандартного відхилення. Оцінимо статистичний максимум урожайності за другим методом з використанням тих же даних, що і у прикладі [1].

Середнє значення урожайності та стандартне відхилення, отримані Дмитренко В.П. [1], відповідно складають 1,59 та 0,74 т/га. Звідси згідно з (4) $Y_C = 3,88$ т/га, що менше ніж на 4,5 % відрізняється від значення, отриманого в першому – графічному способі розрахунку.

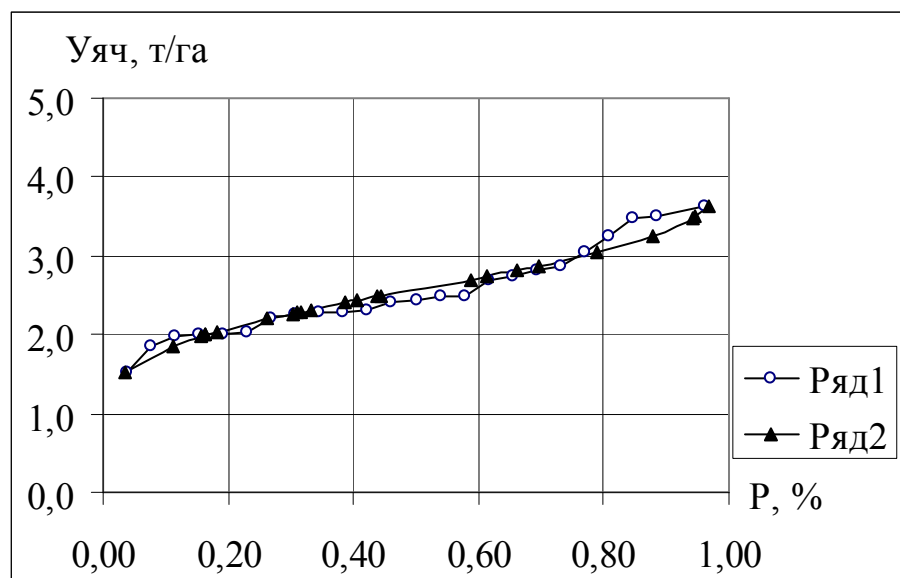


Рис. 1. Розподіл урожайності U ярого ячменю у Вінницькій області за період 1981-2005 рр. за даними забезпеченості P за: ряд 1 – рівнянням (1), ряд 2 – рівнянням (2)

Подібність двох способів розрахунку статистичного максимуму проявляється із співставлення інтегральних кривих за даними забезпеченості, розрахованими за (1) та (2). Графіки розподілу урожайності від забезпеченості нелінійні через використання даних за рівняннями (1) та (2) і лінійної осі абсцис, як видно з рис. 1.

Вочевидь, використання як аргументу цих залежностей значень нелінійної шкали, яка б точно відповідала (2), лінеаризувало б ці залежності (цей метод застосовується для клітчатки ймовірностей). Для цього пропонується скористатися зворотним перетворенням нормального розподілу для розрахованих значень забезпеченості за двома способами. Чисельно таке перетворення відповідає відносному відхиленню в одиницях σ від нульового середнього значення. У цьому випадку шкала осі абсцис співпадатиме з відповідною шкалою на клітчатці ймовірностей.

На рис. 2 – графік забезпеченості врожайності $U_{\text{ячм}}$ (ярий ячмінь, Вінницька область, 1981-2005 рр.), причому значення забезпеченості розраховувалось як значення зворотного нормального перетворення від кумуляти за (1). Як видно з рисунка, цей графік залежності дещо відрізняється від прямої лінії, що свідчить про відмінність значень забезпеченості за (1) від щорічних коливань урожайності та за нормальним законом розподілу, про що говорилося вище.

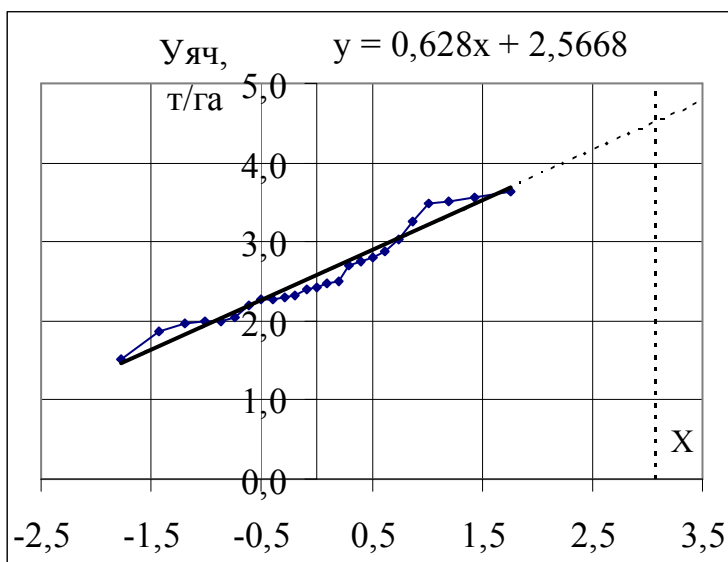


Рис. 2. Графічне представлення розрахунку статистичного максимуму урожайності ярого ячменю у Вінницькій області за період 1981-2005 рр. за даними забезпеченості за рівнянням (1)

X – відносне відхилення врожайності від середньої в одиницях σ

Оскільки кумулята за (2) знаходилася за рівнянням нормального розподілу, графік залежності $Y_{\text{ячм}}$ від відносного відхилення X (в одиницях σ) є прямою лінією (рис. 3). На рисунках наведені відповідні зображеним графікам рівняння з урахуванням тренду лінійного типу, розраховані в Excel з використанням методу найменших квадратів. Крім того, вертикальна пунктирна лінія з абсцисою $x = 3,09$ вказує на положення значення відносного відхилення, яке відповідає забезпеченості $P = 0,999$.

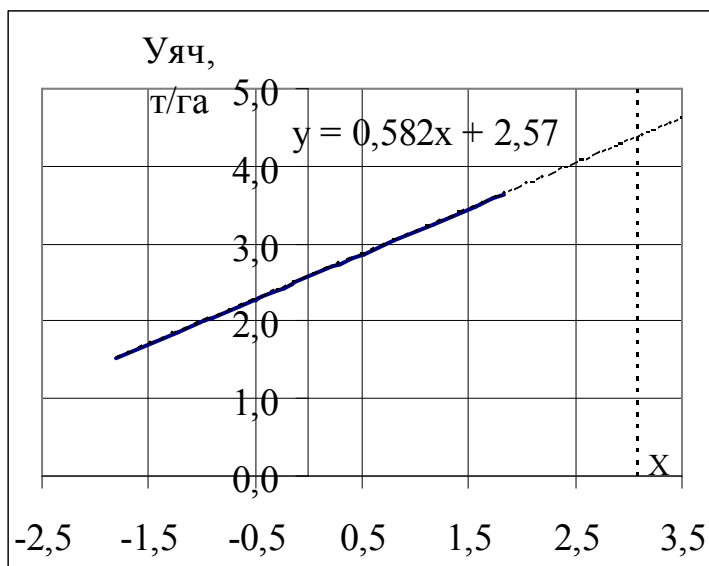


Рис. 3. Графічне представлення розрахунку статистичного максимуму врожайності ярого ячменю у Вінницькій області за період 1981-2005 рр. за даними забезпеченості за рівнянням (2)

Вочевидь, точки перетину ліній з урахуванням тренду і ліній, які відповідають забезпеченості $P = 0,999$, згідно з методикою, дають нам очікувані значення статистичного максимуму.

Рівняння з урахуванням лінійного тренду, наведені на полі графіків рис. 2 та рис. 3, становлять два розподіли нормального виду і є сумою постійної складової, яка відповідає середньому значенню ряду врожайностей (2,57 т/га) та добутку σ та X (σ для ряду урожайностей за нормальним розподілом становить 0,582 т/га), тобто: $Y_C = Y_{\text{сер}} + X \cdot \sigma$. Оскільки тренд для графіка на рис. 2 обрахований за даними забезпеченості за (1), коефіцієнт при X у рівнянні відрізняється від вказаного вище значення ($\sigma_1 = 0,628$ т/га). Тому знайдені значення Y_C за графічним методом $Y_{C.1} = 4,50$ т/га та за (4) $Y_{C.2} = 4,37$ т/га не співпадають.

Згідно з [2] точні розподіли екстремальних значень можуть бути виписані у вигляді функцій від вихідного розподілу та об'єму вибірки n . Для необмеженої випадкової величини максимум математичного очікування найбільшого \tilde{x}_n зростає зі збільшенням n і визначається за формулою:

$$\tilde{x}_n = \bar{x} + \frac{\sigma \cdot (n-1)}{\sqrt{2 \cdot n-1}}. \quad (5)$$

Наприклад, для ряду з числом членів $n = 25$ коефіцієнт при σ становить 3,428. За більш «строгу» оцінку статистичного максимуму урожайності можна взяти величину: $Y_{C.3} = X_{\text{сер}} + 3,43 \cdot \sigma$. Для порівняння в табл. 1 наведено три оцінки статистичного максимуму урожайності озимої пшениці для деяких областей України за період 1981-2005 рр. ($Y_{C.1}$ – статистичний максимум урожайності, визначений за першим, графічним способом; $Y_{C.2}$ – за розрахунковим згідно з (2); $Y_{C.3}$ – максимум математичного очікування найбільшого значення за (5)).

Таблиця 1

Оцінки статистичного максимуму урожайності (т/га) озимої пшениці на 2006 рік, одержаних трьома способами для деяких областей України

Параметр	Область						
	Вінницька	Київська	Харківська	Волинська	Черкаська	Чернігівська	Луганська
$Y_{C.1}$	5,7	5,8	5,9	5	6,25	4,6	5,3
$Y_{C.2}$	5,49	5,55	5,72	4,81	6,2	4,35	4,81
$Y_{C.3}$	5,73	5,78	6,01	5,02	6,48	4,56	5,07

Як відомо, господарський максимум урожайності розраховується для наступних сезонів з урахуванням тренду [4]:

$$Y_t = y_o + k \cdot \sigma_N + A(t - t_o). \quad (6)$$

Для року, наступного за взятим як реперний $(t - t_o) = 1$, за господарський максимум урожайності приймаємо екстраполяційне

значення статистичного максимуму, обраховане за формулою лінійного тренду згладжених значень за попередні 11 років. Оскільки головне завдання роботи – показати близькість двох описаних вище способів знаходження статистичного максимуму врожайності, нижче у табл. 2 наведено наслідки розрахунку Y_C та Y_G озимої пшениці за графічним та чисельним (2) способами для Вінницької області (2000 рік взято як реперний).

Таблиця 2

Порівняльний аналіз даних розрахунку господарського максимуму врожайності (т/га) озимої пшениці у Вінницькій області за двома способами

Рік	$Y_{C.1}$	A_C	$Y_{C.2}$	$Y_{C.2yT}$	$Y_{G.1}$	$Y_{G.2}$
2000	5,03	-0,146	4,86	5,03		
2001	5,03	-0,124	4,87	5,04	4,88	4,88
2002	4,89	-0,074	4,72	4,88	4,73	4,92
2003	4,78	-0,103	4,63	4,79	4,59	4,81
2004	4,64	-0,123	4,48	4,63	4,44	4,69
2005	4,54	-0,061	4,39	4,54	4,3	4,51
2006	4,56	0,022	4,40	4,56	4,15	4,48

Порівняння значень $Y_{C.1}$ та $Y_{C.2}$, наведених у таблицях і отриманих для інших рядів урожайності, показує систематичне зміщення $Y_{C.2}$ відносно $Y_{C.1}$ в область менших значень. Беручи до уваги рівність середніх значень урожайності, яка входить до кожного з Y_C , урівноважити ці величини можна відповідним вибором коефіцієнта при σ у формулі (4) для $Y_{C.2}$. На підставі узагальнення розрахованих значень $Y_{C.1}$ та $Y_{C.2}$ (всього 270 випадків) отримано середнє значення коефіцієнта при σ та його середнє квадратичне відхилення, які складають 3,35 та 0,011 відповідно, тобто замість формули (4) пропонується аналогічна:

$$Y_{C.2yT} = Y_{\text{сеп}} + 3,35 \cdot \sigma, \quad (7)$$

де $Y_{C.2yT}$ – уточнене значення статистичного максимуму урожайності.

Господарський максимум $Y_{G.1}$ розраховувався згідно з [1] з фіксацією реперного року. Так, для даних табл. 2 за реперний рік взято 2000 р. (5,03 т/га), а A фіксоване і обчислене для того ж року (-0,146 т/га·рік). Господарський максимум $Y_{G.2}$ розраховувався за (6) з

урахуванням уточненої формули для статистичного максимуму (7) з «плаваючим» реперним роком. Тобто, для розрахунку господарського максимуму на поточний рік використовуються значення Y_{C,y_T} та тренду урожайності минулого року. Наприклад, розраховане $Y_{Г,2}$ для 2002 р. (4,92 т/га) є сумою $Y_{C,2y_T}$ (5,04 т/га) та A (-0,124 т/га·рік) за 2001 р. Значення $Y_{Г,1}$ та $Y_{Г,2}$ на 2000 р. не розраховувалися через відсутність у таблиці даних попереднього періоду.

Використання для розрахунку господарського максимуму першого екстраполяційного члена мінімізує можливий негативний вплив на $Y_{Г}$ трендової складової через вибір найменшого можливого значення множника $(t - t_0)$, рівного одиниці, для рівняння (6). З табл. 2 видно, що значення $Y_{Г,2}$ краще корелює зі значеннями $Y_{C,2y_T}$, в той час, як відхилення між $Y_{C,1}$ та $Y_{Г,1}$ досягає 0,41 т/га для 2006 року і залежить від вибору реперного року.

Висновки

1. Для знаходження господарського максимуму врожайності в автоматизованих системах прогнозування і оцінки стану сільськогосподарських культур пропонується використовувати розрахунковий метод за формулою (6) з $k = 3,35$ і множника $(t - t_0) = 1$, причому за реперний рік береться попередній рік і тренд A розраховується також для попереднього року.

2. Вибір величини коефіцієнта при σ у (6), рівного 3,35, зумовлено необхідністю збігу величин статистичного максимуму врожайності, знайденого графічним та розрахунковим способами.

3. Показано, що використання для розрахунку господарського максимуму урожайності першого екстраполяційного члена мінімізує можливий негативний вплив на $Y_{Г}$ трендової складової.

* *

1. Методические указания по составлению агрометеорологических прогнозов урожайности ярового ячменя по территории УССР, БССР и МССР. – М.: Гидрометеоздат, 1983. – 31 с.
2. Гумбель Э. Статистика экстремальных значений. – М.: Мир, 1965. – 452 с.
3. Математическая статистика / *Иванова В.М., Калинина В.Н., Нешумова Л.А., Решетникова И.О.* / Под ред. *Длина А.М.* – М.: Высш. школа, 1975. – 398 с.
4. *Дмитренко В.П.* Принципи і засоби визначення потенціалу врожаю

сілськогосподарських культур за еколого-географічними засадами // Наук. пр. УкрНДГМІ. – 2005. – Вип. 254. – С. 9-29.

*Український науково-дослідний
гідрометеорологічний інститут, Київ*

Криворучко И.П.

Обоснование способа аналитического расчета статистического максимума урожайности

Проанализировано поведение рядов урожайности озимой пшеницы и ярового ячменя. Проведен сравнительный анализ двух способов нахождения статистического максимума урожайности: графического по клетчатке вероятностей и расчетного для нормального распределения. Дается обоснование близости этих способов расчетов.

Ключевые слова: ряды урожайности, статистический максимум урожайности, хозяйственный максимум урожайности, нормальный закон распределения, реперный год, трендовая составляющая.

Kryvoruchko I.P.

The substantiation of the method analytical calculation statistical maximum of productivity

The behavior of productivity series of a winter wheat and summer barley are analysed. The comparative analysis of two ways of a finding of a statistical maximum of productivity was carried out: graphic on probability paper and calculated for normal distribution. The substantiation of affinity of these ways of determination is given.

Keywords: time-series of productivity, statistical maximum of productivity, economic maximum of productivity, normal law distribution, anchor year, trend component.