

Н.В. Лещук, кандидат
сільськогосподарських наук
М.М. Таганцова, аспірант
О.А. Стадніченко
Український інститут експертизи
сортів рослин

УДК 633.15:631.524.32

Методичні аспекти застосування гістограми і варіаційної кривої морфологічних ознак гібридів кукурудзи (*Zea mays* L.)

Методично обґрунтовано ідентифікацію (метод – морфологічний опис) якісних, кількісних і псевдоякісних ознак гібридів кукурудзи (*Zea mays* L.). Розкрито методичні аспекти практичного застосування гістограми та варіаційної кривої морфологічних ознак фенотипу за тесту на однорідність. Показано зв'язок кількісних ознак качана з господарсько-цінними показниками продуктивності рослин. Об'єктом досліджень були фізіологічні та біохімічні процеси формування продуктивності рослин у відповідні фенологічні фази росту і розвитку, які супроводжувалися відповідними спостереженнями та обліками вегетативних та генеративних органів кукурудзи. Отримані результати польових досліджень математично-статистично опрацьовані з метою пошуку граничних меж (min і max) довірчих інтервалів та підтверджено достовірність результатів досліджень для подальшої побудови гістограми та варіаційної кривої.

Ключові слова:

гібрид, ідентифікація, гістограма, варіаційний ряд, відхилення, варіаційна крива, ознака, продуктивність.

Постановка проблеми. Ідентифікація гібридів кукурудзи методом морфологічного опису якісних, кількісних і псевдоякісних ознак дає можливість отримати морфологічну кодову формулу фенотипу та встановити кореляцію з господарсько-цінними показниками продуктивних органів рослин. Саме кількісні ознаки качанів кукурудзи, зокрема: довжина качана, діаметр, кількість рядів зерен потребують удосконалення пошуку граничних меж (min і max) довірчих інтервалів й обґрунтування практичного застосування гістограми та варіаційної кривої для морфологічних ознак під час отримання результатів із ідентифікації гібридів або батьківських компонентів [1].

Мета і завдання досліджень – методичне обґрунтування морфологічного опису якісних, кількісних і псевдоякісних ознак під час ідентифікації гібридів кукурудзи, розкриття методичних аспектів практичного застосування гістограми та варіаційної кривої морфологічних

ознак фенотипу за тесту на однорідність, а також показати кореляційний зв'язок кількісних ознак фенотипу з господарсько-цінними показниками продуктивності гетерозисних гібридів кукурудзи.

Методи і методика досліджень.

Комплекс польових і лабораторних досліджень з ідентифікації гібридів та господарської оцінки продуктивності рослин кукурудзи проводили впродовж 2010–2011 рр. на дослідному полі Полтавської державної сортодослідної станції. Ґрунтово-кліматичні умови зони вирощування сприятливі для росту і розвитку рослин. Дослідження проводили відповідно до Методики проведення експертизи сортів кукурудзи звичайної *Zea mays* L. на відмінність, однорідність і стабільність та Методики на придатність до поширення. Застосовували польовий, лабораторний та математично-статистичний методи експериментального дослідження. Об'єкт досліджень – фізіологічні та біохімічні процеси формування продуктивності рослин у відповідні

фенологічні фази росту і розвитку, які супроводжувалися відповідними спостереженнями та обліками вегетативних та генеративних органів кукурудзи. Отримані результати польових досліджень підлягали математично-статистичному опрацюванню з метою пошуку граничних меж (min і max) довірчих інтервалів та підтвердження достовірності результатів досліджень для подальшої побудови гістограми та варіаційної кривої.

Результати досліджень. Ідентифікацію вегетативних і генеративних органів гібридів кукурудзи проводили методом морфологічного опису. На прикладі качана кукурудзи змодельємо визначення кодів проявлення кількісних морфологічних ознак генеративних органів кукурудзи.

Установлено, що ідентифікація якісних і кількісних ознак вегетативних і генеративних органів рослин кукурудзи, значення кодів виявлення яких формують варіаційний ряд, представлений у вигляді діаграми,

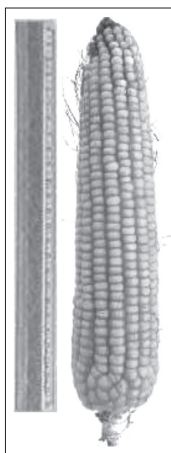


Рис. 1. Качан кукурудзи за довжиною, см.

на якій різна величина частот зображена стовпчиками різної висоти у вигляді гістограм можна представляти будь-які дані як кількісні, так і якісні, як дискретні, так і неперервні.

Варіаційна крива репрезентована лише неперервними (мірними) даними. Зображення варіаційного ряду у вигляді кривої,

ординати пропорційні його частотам. Саме ця крива найчастіше аналізується під час математично-статистичної обробки результатів. Вона є основою для розгляду питання щодо оптимального розподілу. Розглянемо випадок, коли маємо генеральну вибірку з 40 рослин на дослідній ділянці (кукурудзи). Кількісні ознаки спостерігаємо на 20 маркованих рослинах, рис. 2 [2].

Маємо результати вимірювань довжини качана, см:

a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀
19,5	19,5	20,0	18,5	20,5	18,5	20,0	19,0	19,0	20,0
a ₁₁	a ₁₂	a ₁₃	a ₁₄	a ₁₅	a ₁₆	a ₁₇	a ₁₈	a ₁₉	a ₂₀
19,5	19,0	19,0	20,0	19,0	19,5	20,5	18,5	20,5	19,5

Головною вимогою до вибірки є її достовірне відображення генеральної сукупності. Досягнути це можна випадковим відбором для формування вибірки. Так, кількісні параметри довжини качана заносимо у програму «Варіаційний ряд», записуючи вимір за виміром a₁ – , a₂₀. Суть ранжування полягає в розміщенні варіантів у строгому порядку зростання або спадання, рис. 3.

У нашому випадку будемо мати: 18,5; 18,5; 18,5; 19,0; 19,0; 19,0; 19,0; 19,0; 19,5; 19,5; 19,5; 19,5; 19,5; 20,0; 20,0; 20,0; 20,0; 20,5; 20,5; 20,5.

Але для зручності опрацювання запишемо наші результати за частотами:

вимір a _i	18,5	19,0	19,5	20,0	20,5
частота m _i	3	5	5	4	3
m _i a _i	55,5	95,0	97,5	80,0	61,5

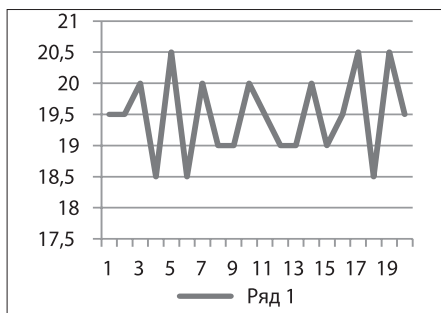


Рис. 2. Варіаційна крива кількісних значень довжини качана

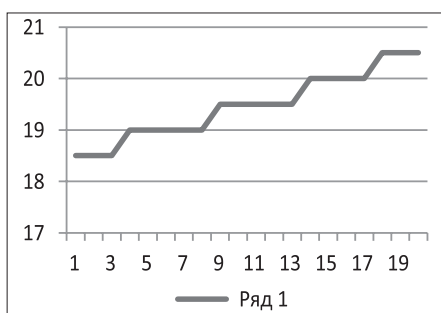


Рис. 3. Варіаційна крива ранжированого ряду

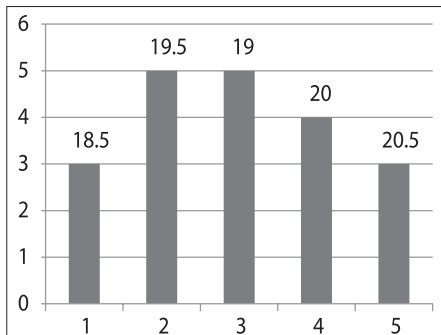


Рис. 4. Гістограма вибірки за частотами.

Сума m_i дорівнює кількості нашої вибірки – 20. Гістограма вибірки за частотами у даному випадку буде мати такий вигляд, (див. рис. 4).

Для побудови варіаційної кривої гістограми потрібно математично опрацювати кількісні параметри довжини качана.

Середня арифметична – основний показник середньої якості досліджуваних об'єктів, обчислюється за формулою:

$$a = \frac{\sum_{i=1}^n a_i}{n}$$

Якщо вибірку розписати за частотами m_i, то середнє арифметичне можна обрахувати більш зручнішою формулою, де n = 20, k = 5, а середня арифметична:

$$a = (a_1 m_1 + a_2 m_2 + \dots + a_k m_k) / 20 = (55,5 + 95 + 97,5 + 80 + 61,5) / 20 = 19,48.$$

Другим параметром, який характеризує сукупність, є **середнє квадратичне відхилення (σ)**. Цей показник має виключне значення в математичній статистиці.

Він використовується як абсолютна міра різноманітності і, до того ж, його покладено в основу майже всіх характеристик мінливості, розподілу, кореляції, регресії, дисперсійного аналізу тощо. Основна формула, за якою визначається середнє квадратичне відхилення:

$$\sigma = \sqrt{\frac{\sum (a_i - a)^2}{v}} = \sqrt{\frac{\sum (a_i - a)^2}{n-1}},$$

(вибірка не згрупована по частотах)

$$\sigma = \sqrt{\frac{\sum (a_i - a)^2 m_i}{n-1}},$$

(згрупована)

Де σ (сигма) – **середнє квадратичне відхилення** (назва цього символу перейшла й на сам показник – середнє квадратичне відхилення часто називають просто сигмою); (a_i – a) – **центральне відхилення**, тобто різниця між кожною варіансою та середньою арифметичною; v – число степенів свободи; при розрахунках сигми число степенів свободи дорівнює кількості варіанс без одної, тобто v = n – 1, табл. 1 [2, 3].

Так як сума відхилень буде рівна нулю або наближатись до нуля, то відхилення варто піднести до квадрату та підрахувати суму квадратів відхилень. Для нашого прикладу:

$$\sum_i (a_i - a)^2 * m_i = 8.23,$$

$$\sigma = \sqrt{\frac{8.23}{19}} = \pm 0.65$$

Таблиця 1

Статистичні показники варіаційного ряду для довжини качана

m_i	a_i	$a_i m_i$	$(a_i - a)$	$(a_i - a) * m_i$	$(a_i - a)^2$	$(a_i - a)^2 * m_i$
3	18,5	55,5	-0,98	-2,94	0,96	2,88
5	19,0	95,0	-0,48	-2,4	0,23	1,15
5	19,5	97,5	0,02	0,1	0	0
4	20,0	80,0	0,52	2,08	0,27	1,08
3	20,5	61,5	1,02	3,06	1,04	3,12
Середня арифметична $a = 19,48$		Сума всіх відхилень дорівнює нулю або дуже близька до нуля		Сума квадратів $\sum_i (a_i - a)^2 * m_i = 8.23$		

Це пояснюється тим, що за обчислення сигми є одне обмеження величини ознаки у об'єктів, що вивчаються. Сигма вираховується для певної групи, яка має певну середню арифметичну. Тому різноманітність елементів, що утворюють середнє квадратичне відхилення, обмежена цією однією умовою, отже, число степенів свободи буде дорівнювати кількості варіанс без одної. Сигму майже ніколи не обчислюють за вищенаведеною формулою, це дуже незручно, адже треба спочатку обчислити середню арифметичну і лише згодом – сигму. Тому було запропоновано низку інших формул, які значно спрощують обчислення:

$$\sigma = \pm \sqrt{\frac{\sum a_i^2 - \frac{1}{n}(\sum a_i)^2}{n-1}}$$

де a – середня арифметична, n – число вивчених об'єктів,

$$\sum_{i=1}^n -$$

знак підсумовування (зазначено, що підсумовуються всі варіанси від $i = 1$ (позначка знизу) до $i = n$ (позначка зверху)); a_i – окремий вимір.

Варто зазначити, що $\sigma = \pm 0.65$ має два знаки («+» та «-»). Це свідчить про те, що варіанти відхиляються від середньої арифметичної у межах двох діапазонів. Середнє квадратичне відхилення є показником мінливості ознаки.

За правилом 3σ майже всі варіанти мають вміщатись у інтервал: $[a - 3\sigma : a + 3\sigma]$, тобто якщо мінімальний варіант не нижче $a_{\min} \geq a - 3\sigma$, а максимальний

варіант $a_{\max} \leq a + 3\sigma$, то спостереження ведуться над однорідною сукупністю.

В нашому випадку маємо інтервал:

$$[19.48 - 3 * 0.65 : 19.48 + 3 * 0.65] \Leftrightarrow [17.53 : 21.43]$$

Основною перевагою середньоквадратичного відхилення є повна кількісна характеристика мінливості показника, що вивчається. Однак порівняти мінливість двох груп з різним середнім значенням ознаки, що вивчається та мінливість різних ознак з допомогою показника не можна. Ось тут на допомогу приходить наступний показник мінливості – коефіцієнт мінливості або варіації. Відношення середньоквадратичного відхилення до середньої арифметичної для даної вибірки, виражене у процентах, є коефіцієнтом варіації:

$$V = \frac{\sigma}{a} 100\%, \text{ у нашому прикладі } V = 3.4\%$$

Коефіцієнт мінливості дає можливість порівняти ступінь мінливості різних ознак. Чим коефіцієнт мінливості вище, тим вище загальна мінливість ознаки. Низькі коефіцієнти мінливості $V \leq 10\%$ вказують на генеральну однорідність сорту за даною ознакою, а високі коефіцієнти – про її неоднорідність.

Для побудови гістограми або варіаційного ряду:

1. Визначаємо розмір варіювання результатів кількісних параметрів довжини, тобто різницю між найбільшим і найменшим значенням ряду вимірів: $R = a_{\max} - a_{\min}$, $R = 20,5 - 18,5 = 2$.

R – розмір варіювання

a_{\max} – максимальне значення вибірки

a_{\min} – мінімальне значення вибірки
2. Визначаємо розмір довірчого інтервалу:

$$I = \frac{R}{K}, \text{ } I \text{ – розмір інтервалу}$$

$K = \sqrt{n}$ – число груп, у нашому прикладі $I = \frac{2}{5} = 0.4$

3. Визначаємо верхню та нижню межу кожної підгрупи, але в нашому прикладі в підгрупах містяться рівні за значенням виміри.

4. Готуємо макет таблиці згрупованого розподілу частот результатів вимірів. В першу колонку записують інтервал підгрупи, а в другій – число результатів вимірів, які входять в даний інтервал тобто частоту.

Інтервал	18,5–18,8	18,9–19,2	19,3–19,6	19,7–20,0	20,1–20,5
Частота	3	5	5	4	3

Межі, в яких з тією чи іншою вірогідністю міститься середня арифметична генеральної сукупності, називаються Межами довіри. А інтервал між цими межами – довірчим інтервалом. Виходячи з певних законів нормального розподілу, можна стверджувати, що генеральна середня арифметична не має відрізнятись від знайденої вибіркової середньої більше, ніж на величину максимально можливої похибки, яка визначається шляхом перемноження статистичної похибки з так званим критерієм Стьюдента (t). Значення цього критерію береться із спеціальних таблиць (див. табл. 2). У загальному вигляді максимальна погрішність оцінювання генеральної середньої (її позначають грецькою літерою Δ – дельта) може бути записана таким чином: $\Delta = \pm m_a$, а межі довірчого інтервалу для генеральної середньої арифметичної як $a \pm m_a$. Отже, для різних рівнів довіри це виглядатиме так:

Опрацьовані кількісні значення довжини за програмою «Варіаційний ряд»

Середнє значення	Дисперсія	Середнє квадратичне відхилення	Помилка середньої	Коефіцієнт варіації	Точність
$a = 19,48$	$D = 0,42$	$\sigma = \pm 0,65$	$\Delta = m_a \approx \pm 0,15$	$V = 3,4\%$	$m_V = \frac{\Delta}{a} = 0,8$

З вірогідністю 95 % –

$$a \pm t m_a = a \pm 1,96 m_a,$$

З вірогідністю 99 % –

$$a \pm t m_a = a \pm 2,58 m_a,$$

З вірогідністю 99,9 % –

$$a \pm t m_a = a \pm 3,30 m_a$$

де $\Delta = \pm m_a = \frac{\sigma}{\sqrt{n}}$ – похибка

середньої арифметичної, у нашому

прикладі

$$\Delta = \pm m_a = \frac{\sigma}{\sqrt{n}} = \frac{0,65}{4,47} \approx 0,15$$

Для більшості біологічних і екологічних досліджень у багатьох випадках визнається достатнім 95% рівнем довіри, при якому вважається за достатнє, якщо істотність висновків визнається в 95 випадках із 100.

У нашому випадку $19,48 \pm 1,96 \cdot 0,15$, тобто маємо межі довіри [19,19 : 19,77].

Опрацьовані результати за програмою «Варіаційний ряд» можемо подати таким чином, табл. 2.

Для пошуку коду прояву відповідної морфологічної ознаки «Качан за довжиною» використаємо середнє значення $a = 19,48$ та середнє квадратичне відхилення $\sigma = \pm 0,65$. Саме $a \pm \sigma$ дасть можливість знайти граничні межі \min та \max кількісних значень довжини качана, а в подальшому встановити код прояву відповідної ознаки. Адже встановлення коду та ступінь його прояву є обов'язковим під час проведення експертизи на однорідність. Алгоритм наших розрахунків варто подати таким чином:

$$a_{\min} = 19,48 - 0,65 = 18,83 \text{ см.}$$

$$a_{\max} = 19,48 + 0,65 = 20,13 \text{ см.}$$

Довжина качана кукурудзи знаходиться в граничних межах 18,83–20,13 см. Кількісні параметри довжини відносяться до коду проя-

ву – 5 (середня). Вибірка качанів для обстеження представлена качанами середньої довжини.

Висновки. Математично-статистичне опрацювання кількісних ознак гібридів кукурудзи за експертизи на однорідність методом ідентифікації «морфологічний опис» є обов'язковим для встановлення коду прояву ознак та підтвердження достовірності результатів спостережень.

Математичний вираз $a \pm \sigma$ дасть можливість знайти граничні межі \min та \max кількісних значень довжини качана та в подальшому встановити код прояву кількісної ознаки для кожного гібриду зокрема.

У вигляді гістограм можна представляти будь-які дані, як кількісні, так і якісні, як дискретні, так і неперервні; у вигляді варіаційної кривої можна репрезентувати лише неперервні (мірні) дані, які пропорційні частотам варіаційного ряду.

ВИКОРИСТАНА ЛІТЕРАТУРА

1. Методика проведення експертизи сортів зернових і зернобобових, лісогосподарських, декоративних і плодово-ягідних, лікарських і овочевих культур на відмінність, однорідність і стабільність (*Zea mays* L.) / В оф. бюл. Охорона прав на сорти рослин. – Київ: ПП «Видавництво «Фенікс», 2010. – Вип. 3, Ч 3. С. 37–71.
2. Доспехов, Б.А. Методика полевого опыта / Б.А. Доспехов. – М.: Агропромиздат, 1985. – 351 с.
3. Зайцев, Г.Н. Математическая статистика в экспериментальной ботанике / Г.Н. Зайцев. – М.: Наука, 1984. – 423 с.