Міністерство освіти і науки України
Одеський національний університет імені І. І. Мечникова
Біологічний факультет

Т. Г. Алєксєєва

ГЕНЕТИКО-СТАТИСТИЧНИЙ АНАЛІЗ КІЛЬКІСНИХ ОЗНАК КЛІТИН

Методичні вказівки до великого спеціального практикуму
для студентів IV курсу денної форми навчання та V курсу заочної форми
навчання напряму 6.040102 – «біологія»

Одеса
ОНУ
2013

УДК 575/576:311.001.8(075.8)
ББК 28.041.12я73
А471

Рекомендовано до друку
Вченою радою біологічного
факультету ОНУ імені
І. І Мечникова
Протокол № 8 від 2.04.2013

Рецензенти:

Ю. М. Олійник -

кандидат біологічних наук, доцент кафедри зоології;

Л. М. Карпов -

доктор біологічних наук, професор, завідувач кафедри фізіології людини і тварин.

Т. Г. АЛЄКСЄЄВА. ГЕНЕТИКО-СТАТИСТИЧНИЙ АНАЛІЗ КІЛЬКІСНИХ ОЗНАК КЛІТИН

Методичні вказівки до великого спеціального практикуму для студентів IV курсу денної форми навчання та V курсу заочної форми навчання напряму 6.040102 – «біологія»

© Т. Г. Алєксєєва, 2013-10-23
© Одеський національний університет імені І. І. Мечникова, 2013

Вступ

Як відомо, генетичний аналіз є основним методом генетики. За допомогою різних методичних напрямів генетичного аналізу вирішують велику кількість завдань сучасної генетики – аналізують природу спадкових змін, структуру і функції генів, визначають локалізацію генів у геномі і складають генетичні карти, досліджують генотипи як окремих особин, так і груп особин і генетичну структуру популяцій. У поле зору генетичного аналізу потрапляють як якісні, так і кількісні ознаки; дослідження останніх все частіше є предметом уваги зі сторони особливого напряму генетичної науки – статистичної генетики.

Вивчення мінливості і спадковості кількісних ознак вимагають як теоретичних знань основ генетичного аналізу, так і досить складної математичної обробки за спеціальними методиками. Ознайомлення з цими методиками і є метою даного розділу Великого спеціального практикуму.

ТЕМА 1. Організація та особливості проведення генетико-статистичного аналізу. *1.1. Планування та умови проведення генетико-статистичного дослідження*

Перед початком роботи треба звернути особливу увагу на те, що генетичний аналіз кількісних ознак, є, по суті, аналізом генетичних явищ зі статистичної точки зору. Таким чином, план експерименту повинен бути статистично продуманим, обґрунтованим. Потрібно завжди пам’ятати, що план досліду зумовлює і метод математичного аналізу результатів.

Дослід, який закладено для аналізу будь-якої кількісної ознаки, повинен обов’язково мати певну кількість повторностей, не менш ніж 4. Повторностями називають однойменні елементарні одиниці контрольного чи дослідного варіанту (наприклад, окремі ділянки у польовому досліді, чашки Петрі у лабораторному досліді і т.ін.).

При постановці дослідів потрібно дотримуватися єдності усіх умов, окрім одного – того, що досліджують, тобто зберігати принцип єдиної відмінності. Також при плануванні польового досліду (закладається при необхідності отримання рослинного матеріалу) необхідно закладати його рендомізованим методом, тобто методом випадкового розміщення ділянок для посіву.

Важливим є питання обсягу вибірки. При плануванні числа об’єктів потрібно керуватися «золотим правилом» - менше не можна, а більше не потрібно. Об’єктів у вибірці має бути досить багато, щоб за ними сформувати правильне уявлення про генеральну сукупність. Оптимальний обсяг вибірки можна оцінити за формулою: $n = \frac{t^2 s^2}{\Delta^2}$, де t - нормоване відхилення, значення якого вираховується за очікуваною ймовірністю

безпомилкового прогнозу; s - стандартне відхилення – визначається приблизно за розмахом варіювання ознаки (R=6s); Δ – бажана точність (задається дослідником).

Не менш важливою є умова репрезентативності вибірки. З метою отримати репрезентативну, тобто таку, яка об'єктивно відображає стан досліджуваної ознаки у генеральній сукупності, необхідно провадити відбір проб за такою методикою, яка б усувала можливість прояву систематичних похибок. Згідно з сучасною теорією вибіркового методу, рендомізований відбір усуває зміщення оцінок, значно покращує якість інформації та дозволяє експериментатору використовувати статистичні методи обробки даних.

Допускаючи певне спрощення, іноді отриману вибірку умовно розглядають як популяцію певного виду чи сорту, що дозволяє застосувати до неї методи статистичного аналізу.

Вибір дослідного матеріалу є визначальним для застосування тих чи інших методів статистичного аналізу. Наприклад, при роботі з віддаленими гібридами неможливо застосовувати методи визначення кількості генів, які відповідають за ту чи іншу ознаку, а при дослідженні прояву і успадковування генів у гаплоїдних тканинах неможливо використовувати діалельний аналіз і таке інше.

Висунення і перевірка гіпотез – основний метод у застосуванні математичної статистики. Статистичні методи чи критерії перевірки гіпотез є основою прийняття тих чи інших рішень при певній, зумовленій випадковою варіацією досліджуваних явищ, невизначеності. Практично перевірка гіпотез часто зводиться до виявлення наявності чи відсутності різниці між фактичними і теоретично очікуваними спостереженнями. Таким чином, гіпотезу, яка стверджує про випадковість відмінностей між вибірками, називають нульовою гіпотезою.

Прийняття нульової гіпотези означає, що дані спостережень не протирічать припущенню про відсутність різниці між фактичними і гіпотетичними (чи між двома рядами фактичних) розподілень. Відкидання нульової гіпотези означає, що емпіричні дані з нею несумісні, а вірною є інша, альтернативна гіпотеза, яку перевіряють у даному дослідженні.

1.2. Підготовка матеріалу для цитофотометрії і біометрії. Аналіз постійних мікропрепаратів пиляків пшениці, жита та пшенично-житніх гібридів, а також різних сортів пшениці та їх гібридів. Отримання цифрових мікрофотографій клітин придатної до аналізу якості.

Студентам пропонують постійні мікропрепарати пиляків наступних рослин – різних сортів пшениці, жита, а також пшенично-житніх гібридів і гібридів від міжсортового схрещування озимої пшениці та міжсортового схрещування ячменю.

Усі мікропрепарати були виготовлені у лабораторних умовах кафедри генетики та молекулярної біології. Зафіксований за методами Навашина та Карнуа рослинний матеріал доводили до парафіну за загальноприйнятою методикою [Паушева, 1988]. Для різки на мікротомі готували парафінові блоки методом закапування. Зрізи товщиною 10 мкм отримували на санному мікротомі, наклеювали їх на предметні скельця розчином альбуміну, що був виготовлений з білка курячого яйця або за допомогою дистиляту. Скельця зі зрізами залишали у термостаті при температурі 38 ^{o}C на декілька діб для висушування, після чого депарафінували і піддавали цитохімічним реакціям.

Відібрані для подальшої роботи постійні мікротомні препарати забарвлювали із застосуванням цитохімічних реакцій:

– на сумарні білки – з бромфеноловим синім по Мезіа [Паушева, 1988];
– на нуклеїнові кислоти – метиловим зеленим – піроніном по Тревану і Шарроку [Паламарчук, Веселова, 1965].

Повний перелік мікропрепаратів, які використовують на заняттях з ВСП, наведено у Додатку.

На рис. 1 наведено зразок мікрофотографії пиляка злаків.

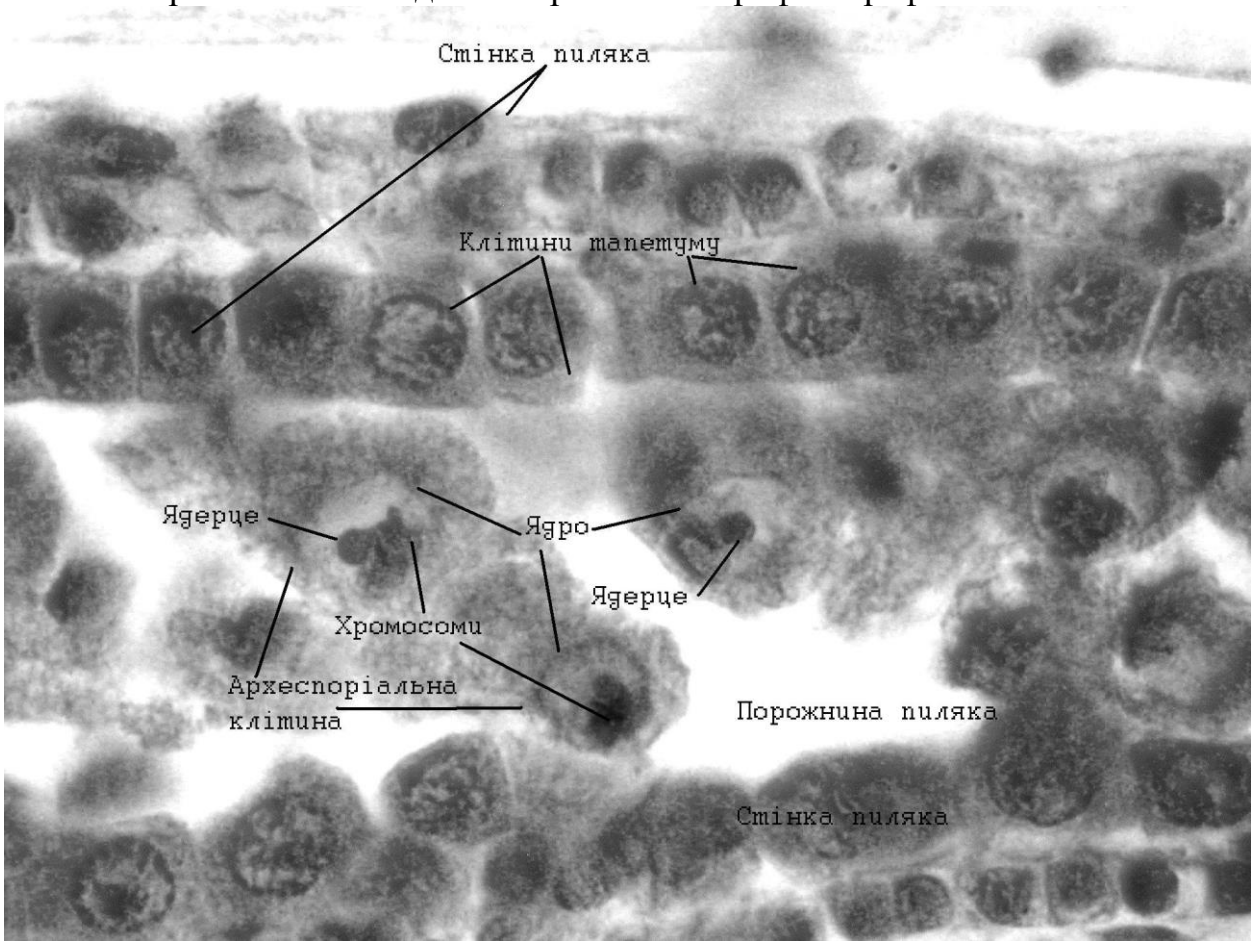

Рис. 1. Мікрофотографія пиляка пшениці сорту пшениці сорту Миронівська 808, зроблена при об’єктиві 40× та окулярі 15×.

Першим завданням студентів є знаходження клітин генеративних структур на мікропрепараті та отримання чітких репрезентативних мікрофотографій, придатних до подальшого аналізу.

Матеріали та обладнання:
Біологічні мікроскопи
Цифрова фотокамера OLYMPUS C-370 ZOOM

В залежності від кількості студентів групу поділяють на підгрупи, що досліджують певний матеріал, на зразок:

1) 1 студент – препарати пиляків пшениці Миронівська 808,
   2 студент – препарати жита Харківське 60
   3 студент – препарати гібрида F_1 Миронівська 808 × Харківське 60

2) 1 студент – препарати пиляків пшениці Безоста 1
   2 студент – препарати пшениці Янтарна
   3 студент – препарати гібрида F_1 Безоста 1 × Янтарна
   4 студент – препарати гібрида F_2 Безоста 1 × Янтарна
   5 студент – препарати гібрида F_3 Безоста 1 × Янтарна

3) 1 студент – препарати пиляків ячменю Екзотик
   2 студент – препарати ячменю Незалежний
   3 студент – препарати гібрида F_1 Екзотик × Незалежний
   4 студент – препарати гібрида F_2 Екзотик × Незалежний

Кожен зі студентів у таких підгрупах буде працювати зі своїм матеріалом, досліджуючи і вимірюючи клітини чоловічої генеративної сфери певного злаку. У другій частині ВСП передбачено засвоєння методів статистичної варіаційної обробки первинних даних, отриманих студентами для кожної близькоспорідненої групи рослин. Вподовж цього часу кожен студент буде працювати з даними всього комплексу своєї групи.

ТЕМА 2. Освоєння комп'ютерної цитофотометрії. Робота з комп'ютерною цитофотометричною програмою PhotoM 1.21 (© А. Черниговский, 2001; ІЕФБ ім. І. М. Сеченова РАН). Отримання масиву первинних даних по наступним кількісним ознакам клітин: каріометричним (об'єм ядра та ядерця), цитохімічним (вміст РНК та загального білку у ядерці та цитоплазмі).

Матеріали та обладнання
Персональні ЕОМ
Комп'ютерна програма PhotoM 1.21 (© А. Черниговский, 2001; ІЕФБ ім. І. М. Сеченова РАН)
Комп'ютерна програма Excel Microsoft Office

Цей розділ ВСП повинен проводитися у комп'ютерному класі факультету. Студенти працюють з комп'ютерним емулятором -

цитофотометричною програмою PhotoM 1.21 (© А. Черниговский, 2001; ІЕФБ ім. І. М. Сеченова РАН).

Програма PhotoM 1.21 призначена для цитофотометрії. За її допомогою можна визначити лінійні розміри клітин і компонентів клітин, автоматично розрахувати площину об'єктів, визначити їх оптичну щільність. Оптична щільність розраховується як середній середній десятичний логарифм відношення яскравості точки фону до яскравості точки об'єкту (екстинкція) на фотографії постійного мікропрепарата.

Потрібно пам'ятати, що у загальному вигляді принципи цитофотометрії не відрізняються від дуже поширеного методу фотометрії, але за його застосування потрібно враховувати наступні обставини.

Речовини та забарвлені комплекси у клітині не утворюють гомогенну фазу, як це має місце у розчинах за звичайної фотометрії. Найчастіше вони локалізовані в окремих структурах, а іноді займають лише частину цієї структури, наприклад периферію. Крім того, ці речовини або комплекси в клітині можуть бути в повній мірі агреговані. Дуже важливим за проведення цитофотометрії є товщина шару, що пропускає світло. Й нарешті, особливістю цього методу є те, що результати не мають абсолютних значень – їх не можна вивести як одиниці ваги речовини на одну клітину. Найчастіше використовують умовні одиниці, такі як відсотки поглинання світла, екстинкція тощо. Тим не менш, незважаючи на такі обмеження, цитофотометрія дозволяє отримувати об'єктивні дані стосовно вмісту різноманітних речовин у компонентах клітин.

В ході заняття студенти повинні перенести власні, отримані на попередніх заняттях мікрофотографії клітин пиляків відповідних рослин. Уважно продивившись фотографії та визначивши клітини, придатні для подальшого аналізу, студенти запускають програму PhotoM.

На рис. 2. зображено скріншот даної програми, на якому зазначено інструменти, якими потрібно користуватися під час виконання завдань теми 2.

Програма має бути відкалібрована таким чином, щоби лінійні розміри клітин можна було отримати у мікрометрах. Цього досягають, використовуючи мікрофотографії шкали об'єкт-мікрометру. Об'єкт-мікрометр – це металева пластина за розмірами подібна до предметного скельця, у середині якої є скляна вставка. По центру вставки нанесена лінійка довжиною 1 мм, поділена на 100 частин. Розмір одного поділку дорівнює 10 мкм. Визначення збільшення мікроскопу та калібрування програми PhotoM визначається вимірянням на фотознімку відстані між штрихами сфотографованого зображення об'єкт-мікрометру.

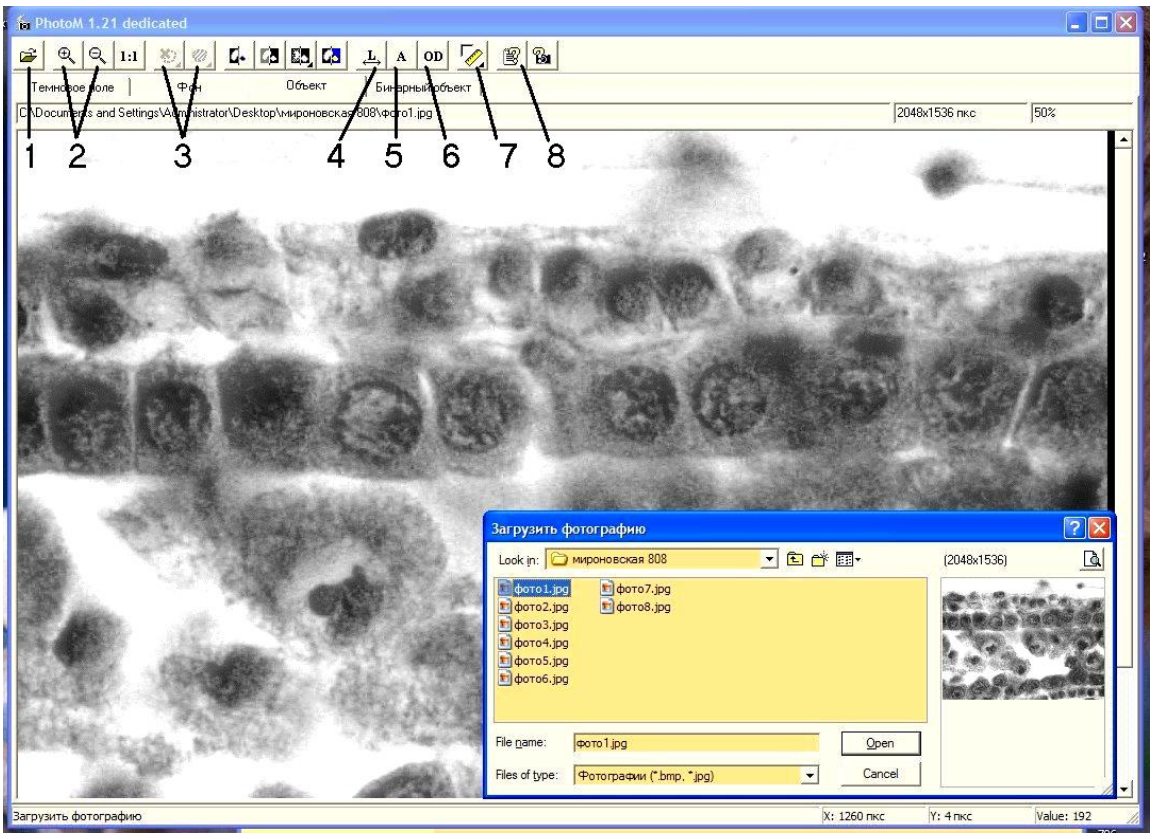

Рис. 2. Скріншот програми PhotoM 1.21.
(© А. Черниговский, 2001). Цифрами позначено інструменти:
1 – загрузка фотографії; 2 – збільшення та зменшення зображення; 3 – інструменти виділення області на фотографії та зняття виділення; 4 – виміряння відстані;
5 – виміряння площини; 6 – виміряння оптичної щільності, 7 – вікно калібровки,
8 – керівництво користувача

Завдання 2

Виміряти показники кількісних цитометричних ознак клітин генеративних структур досліджуваних варіантів злаків.

За проведення цієї роботи студенти повинні бути дуже уважними. Так як метою роботи є отримання масиву кількісних показників, які всебічно описують певну клітину, для того, щоби результати були вірогідними і вірно характеризували матеріал, необхідно чітко знати, до якої саме клітини відносяться певні вимірювання.

За допомогою інструментів програми студенти отримують лінійні показники кількісних каріометричних ознак клітин досліджуваних рослин, а саме – велику і маленьку вісі ядра та ядерця кожної клітини. У режимі виміру розмірів курсор комп'ютера приймає форму «хреста». При кліку лівою кнопкою у обраній точці фотографії в цю точку установлюється хрестоподібний маркер. Якщо після цього перемістити мишу у іншу точку і клікнути ще раз, то буде виміряна відстань між цими двома маркерами. На рис. 3 наведено скріншоти програми з вимірами ядра (а) і ядерця (б) археспоріальної клітини пшениці Миронівська 808

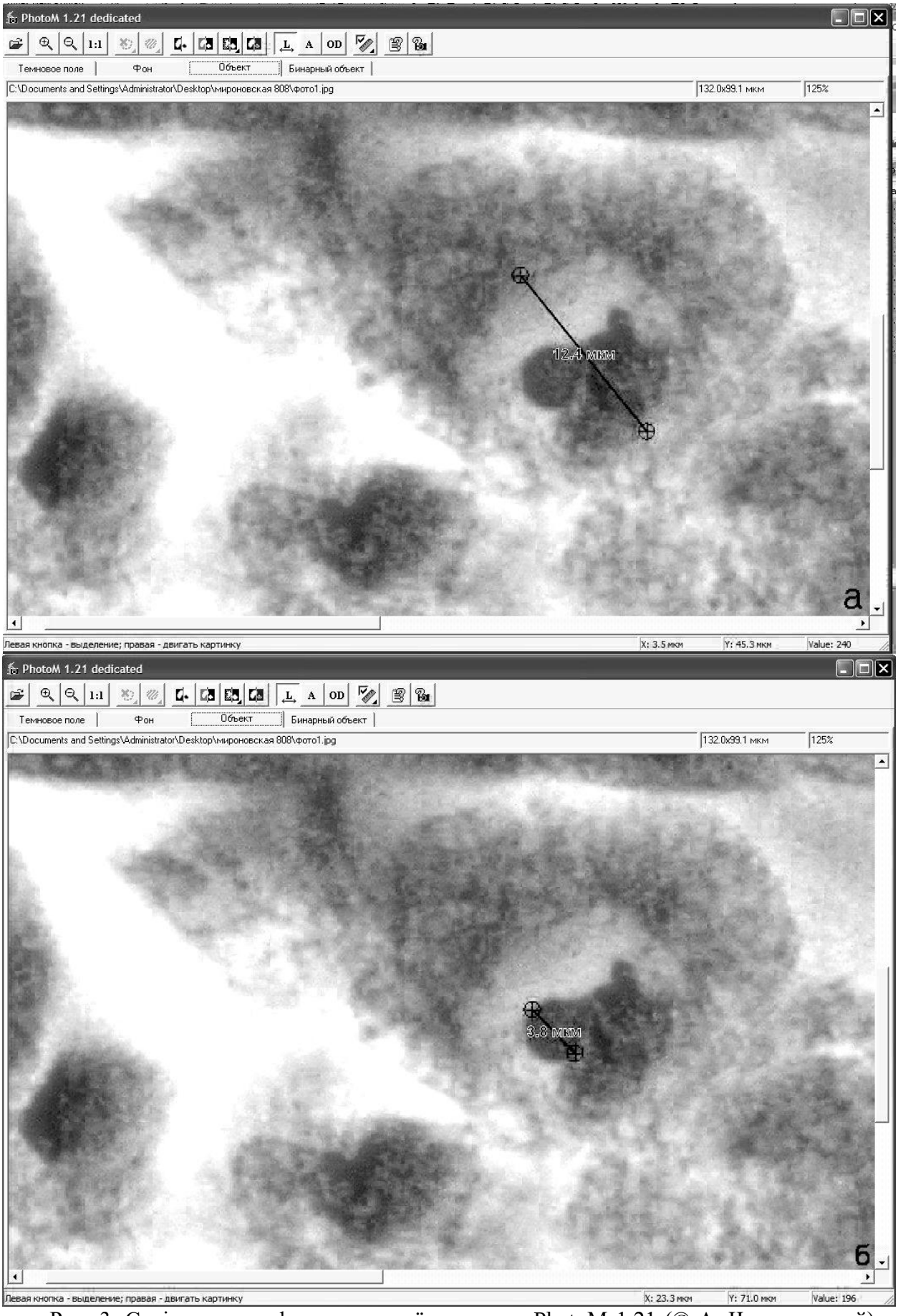

Рис. 3. Скріншот цитофотометричної програми PhotoM 1.21 (© А. Черниговский) – визначення лінійних розмірів компонентів клітини – ядра (а) та ядерця (б)

Ці показники потрібно занести у відповідні графи таблиці первинних вимірів. При цьому треба пам’ятати, що першою завжди записують більшу вісь, меншу – другою.

Для отримання більш-менш вірогідних результатів потрібно наміряти не менш ніж по п’ять клітин від п’яти різних рослин. Зразок заповнення таблиці отриманими даними про лінійні розміри ядра та ядерця клітин досліджуваних злаків наведено у табл. 1.

Таблиця 1

Лінійні розміри (мкм) археспоріальних клітин пиляків пшениці сорту Безоста 1

Номер повторності	Номер клітини	Ядро		Ядерце	
		більша вісь *a*	менша вісь *b*	більша вісь *a*	більша вісь *b*
1	1	12,4	10,2	3,8	3,6
	2	13,1	12,6	5,2	5,2
	3	12,0	11,1	5,2	4,6
	4	…	…	…	…
	5	…	…	…	…
2	1	…	…	…	…
…	…	…	…	…	…
5	5	…	…	…	…

За допомогою наступної формули

$$V = \frac{\pi}{6} \cdot a \cdot b^2,$$

де a – велика, b – маленька вісі,

визначають об’єм ядра та ядерця для кожної клітини окремо.

Для прискорення розрахунків зручно використовувати табличний процесор *Excel*. На рис. 4. наведено приклад розрахунків об’єму ядра.

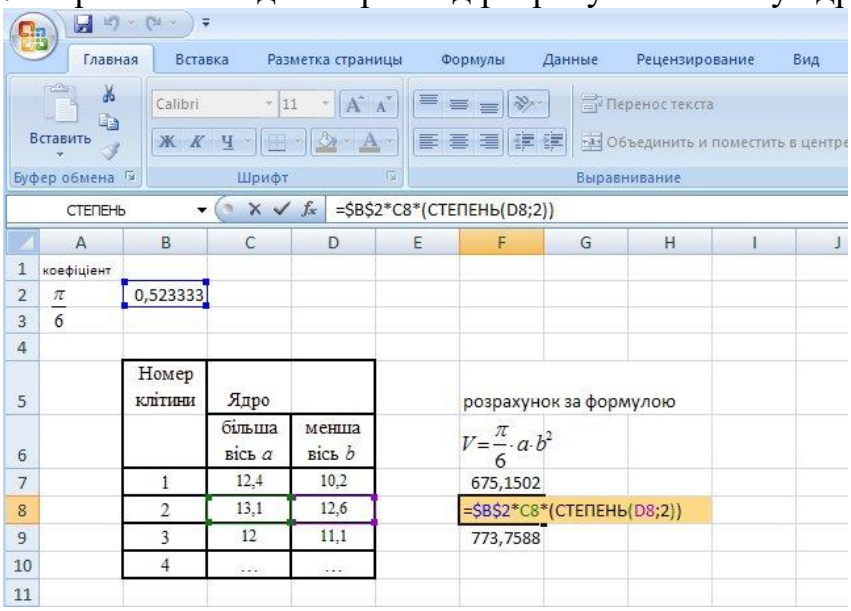

Рис. 4. Створення формул у *Excel*. Формули в *Excel* задають за допомогою посилань на комірки, що містять цифрові дані, і до яких застосовують арифметичні оператори; для задання коефіцієнту $\frac{\pi}{6}$ (значення якого не змінюється) використовують символ $.

Заповнюють наступну таблицю первинних данних, в яку вносять власне вже розраховані показники об’єму ядер та ядерець (табл. 2).

Таблиця 2

Об’єми ядра та ядерця (мкм^3) археспоріальних клітин пиляків пшениці сорту Безоста 1

Номер препарату (рослини)	Номер клітини	Ядро	Ядерце
1	1	675,15	25,77
	2	1088,41	73,58
	3	773,76	57,58
	4	…	…
	5	…	…
2	1	…	…
…	…	…	…
5	5	…	…

Для тих же клітин потрібно проаналізувати показники цитохімічних ознак – вмісту у компонентах клітин загальної кількості білку або РНК (в залежності від барвника, яким було забарвлено постійні мікропрепарати).

За допомогою інструментів програми PhotoM 1.21 вираховують оптичну щільність компонентів клітин як середній десятичний логарифм відношення яскравості точки фону до яскравості точки об’єкту на фотографії постійного мікропрепарату.

Важливо пам’ятати, що для розрахунку оптичної щільності використовують лише чорно-білі зображення. Тому отримані цифрові мікрофотографії заздалегідь повинні бути переведені у чорно-білий стан за допомогою будь-якого графічного редактора. Необхідною умовою є збереження единого шляху обробки усіх зображень в одному досліді, для того, щоб уникнути системних помилок, які можуть мати місце при недодержанні цього принципу.

Для розрахунку оптичної щільності потрібно вибрати режим врахування фону по середньому значенню у певній ділянці фотографії «Фон из штриховки на фотографии» та вибір області розрахунку об’єкта по певній його частині «Область расчета по выделенным областям объекта». Після вибору необхідних режимів натискання кнопки «Розрахунок» у ділянці виводу результатів з’являються загальні дані про усі задіяні об’єкти. На рис. 5 наведено скриншот програми, під час розрахунку оптичної щільності ядерця та цитоплазми.

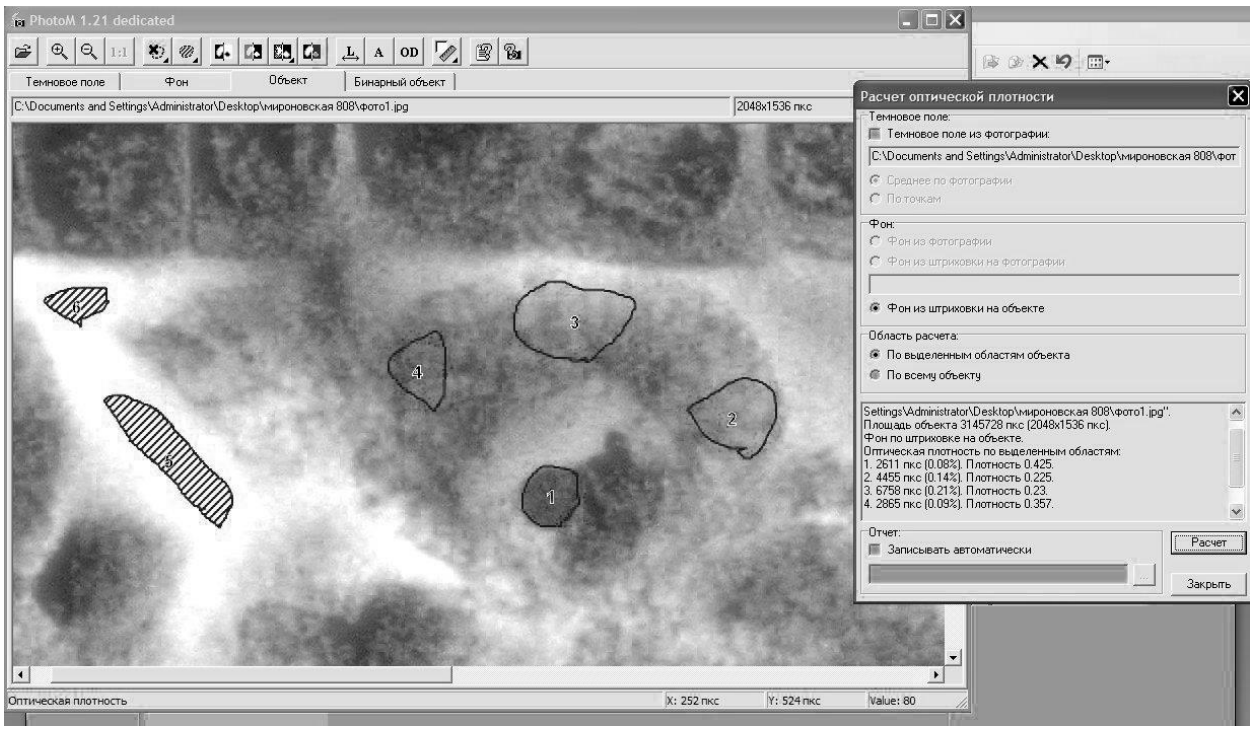

Рис. 5. Скріншот цитофотометричної програми PhotoM 1.21 (© А. Черниговский) – визначення оптичної щільності ядерця (маркер 1) та цитоплазми (маркери 2,3,4)

Таким чином заповнюють таблицю первинних даних про показники цитохімічних ознак досліджуваних клітин (Зразок заповнення наведено у таблиці 3).

Таблиця 3

Вміст загального білку у компонентах клітин археспорію пшениці сорту Миронівська 808

Номер препарату (рослини)	Номер клітини	Вміст речовини у ядерці	Вміст речовини у цитоплазмі			
			Точка 1	Точка 2	Точка 3	Середне арифметичне для цитоплазми
1	1	0,425	0,225	0,230	0,357	0,270
	2	0,223	0,195	0,189	0,224	0,203
	3	0,185	0,278	0,222	0,213	0,238
	4	…	…	…	…	…
	5	…	…	…	…	…
2	1	…	…	…	…	…
…	…	…	…	…	…	…
5	5	…	…	…	…	…

Таким чином, отримують масив даних, які характеризують кількісні ознаки кожної клітини. Подальша робота з ними буде значно полегшена систематизацією і представленням вихідних даних у вигляді графічного матеріалу.

ТЕМА 3. Організація масиву первинних даних для аналізу.

3.1. Побудова варіаційних кривих. Визначення статистичних параметрів. Мінливість ознаки. Коєфіцієнт мінливості.

Матеріали та обладнання
Персональні ЕОМ
Комп'ютерна програма Excel Microsoft Office

Як відомо, кількісна ознака в переважній більшості кодується не одним, а багатьма генами, взаємодія яких може бути дуже різноманітною. Мінливість, яка є наслідком розщеплення генів кількісних ознак, називають полігенною. Оскільки це розщеплення дуже складне і залежить не лише від кількості полігенів, але й від їх взаємодії між собою та іншими генами, то спадковість і мінливість кількісних ознак досліджуються за допомогою спеціальних статистичних методів.

Кожна з досліджуваних кількісних ознак може мати у різних особин різний ступінь прояву, тому кажуть, що ознака варіює, а властивість ознак – відрізнятися одна від одної навіть у однорідних сукупностях називається мінливістю чи варіюванням. Варіювання ознак у різних особин є наслідком варіювання як самих генів, так і взаємодії їх з різним генетичним оточенням у кожному індивідуальному випадку, а також взаємодією генів з мінливими умовами середовища. Таким чином, індивідуальний генотип особини зазнає впливу різних поєднань зовнішніх умов, які не завжди піддаються обліку і можуть мати стохастичний, тобто випадковий характер.

Для аналізу кількісних ознак їх організують у **варіаційний ряд** – подвійний ряд чисел, що показує, як значення ознаки пов'язані з їх повторюваністю. Весь розмах мінливості розбивають на рівні інтервали – класи, а потім визначають частоту для кожного класу.

Для побудови інтервального варіаційного ряду потрібно враховувати наступне:

1 Визначити розмах мінливості Для цього знайти найбільшу та найменшу дати сукупності – x_{max} і x_{min} . Різниця $R = x_{max} - x_{min}$ називається розмахом мінливості.

2. Визначити приблизне число класів, на яке буде розбита сукупність.

Таблиця 4

Число класів варіаційного ряду залежно від числа спостережень [Атраментова, Утевська, 2008]

Число спостережень n	Число класів К
25 – 40	5 – 6
40 – 60	6 – 8
60 – 100	7 – 10
100 – 200	8 – 10
> 200	10 - 15

3. Віднайти значення класового інтервалу за формулою, де $\lambda = \frac{x_{max} - x_{min}}{K}$, де λ – класовий інтервал, x_{max} і x_{min} - максимальна і мінімальна дати сукупності, К – число класів, на які буде розбита сукупність. При виборі меж класових інтервалів треба керуватися міркуваннями зручності. Усі інтервали повинні буди рівновеликими і не повинні перекриватися.

Отриманий інтервальний ряд зручно представити у вигляді **гістограми**.

Табличний процесор *Excel* є потужним і зручним інструментом первинного статистичного аналізу. За його допомогою зручно представляти дані у графічному вигляді. На рис. 6 наведено побудову варіаційного ряду для значень оптичної щільності ядерець археспоріальних клітин жита Харківське 60.

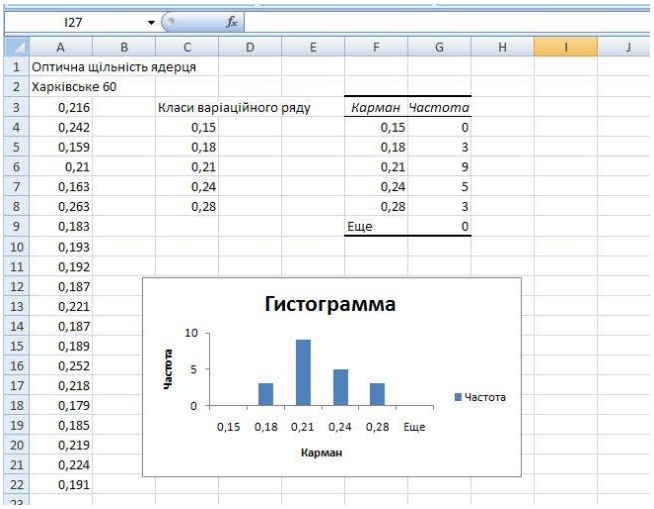

Рис. 6. Побудова гістограми засобами *Excel*. Для цього потрібно, спираючись на межі різноманітності даних експерименту (оптична щільність ядерця жита Харківське 60) визначити класи варіаційного ряду. Після цього необхідно активувати настройку «Анализ данных», вибравши пункт «Гистограмма», задати діапазон первинних данних та класів варіаційного ряду «Карман».

Варіаційні ряди та їх графіки дають наочне уявлення про варіювання ознак, але вони не є достатніми для повного опису об'єктів. Для досягнення цієї мети використовують особливі логічно та теоретично обґрунтовані числові показники, що носять назву статистичні характеристики. Для докладного ознайомлення з математичним підґрунтям необхідно звернутися до підручників, що присвячені питанням біологічної статистики, наприклад «Біометрія» Г. Ф. Лакина чи «Статистичні методи в біології» Л. О. Атраментової та О. М. Утевської.

Таким чином, для всебічного аналізу отриманої сукупності потрібно обчислити наступні статистичні показники – характеристики положення і показники варіації.

Усі ці показники також можуть бути обчислені за допомогою *Excel.*

Характеристика положення відображає рівень розвитку ознаки у певній сукупності, або центральні тенденції – це **середня арифметична**. Формула середньої арифметичної – сума всіх дат, що поділена на їх кількість

$$\bar{x} = \frac{\sum x}{n},$$

де x – окремі дати, а n – кількість дат. Похибка середньої арифметичної - $s_{\bar{x}} = \frac{s}{\sqrt{n}}$, де s – стандартне відхилення.

Інформацію про різноманіття ознак дають показники варіації. Вони характеризують розсіювання дат навколо центру розподілу і таким чином описують мінливість досліджуваних ознак.

Це розмах варіації, дисперсія, стандартне відхилення та коєфіцієнт варіації.

Межі та розмах варіації R_{Lim}

Максимальне і мінімальне значення ознаки в групі називаються межами варіації. Різницю між максимальною і мінімальною датами виражає розмах $R_{Lim} = x_{max} - x_{min}$.

Стандартне (середньоквадратичне) відхилення (у різних посібниках позначається як s чи σ) є найбільш зручною мірою різноманітності. Розраховується за формулою:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}},$$

де x – окрема дата, $\bar{x}$ - середне арифметичне, n – кількість дат. Статистична похибка стандартного відхилення $s_s = \frac{s}{\sqrt{2n}}$, где s – стандартне відхилення, n- обсяг вибірки.

Для розрахунків за допомогою калькулятора зручно використовувати робочу формулу $s(\sigma) = \sqrt{\frac{\sum x_i - \frac{(\sum x_i)^2}{n}}{n-1}}$ [Рокицкий, 1973].

Для автоматичного розрахунку середньоквадратичного відхилення у *Excel* існує функція СТАНДОТКЛОН(*массив даних*).

Дисперсія – один з найважливіших показників різноманітності використовується для порівняльної оцінки однойменних величин

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

Статистична похибка дисперсії

$$s_{s^2} = \frac{s^2}{\sqrt{2n}},$$

де s^2 – дисперсія, n- обсяг вибірки.

В *Excel* дисперсія вибірки розраховується за допомогою функції ДИСПР(*массив даних*)

Коефіцієнт мінливості – зручний показник, значення якого вимірюють у відсотках, що дає додаткову можливість порівнювати мінливість ознак, які мають різні одиниці виміру.

$Cv = \frac{s}{x} 100\%$, де s – стандартне відхилення, $\bar{x}$ - середне арифметичне.

Статистична похибка коефіцієнту мінливості

$$s_{Cv} = \sqrt{\frac{Cv^2}{2n}},$$

де Cv – коефіцієнт варіації, n - обсяг вибірки.

Мінливість прийнято вважати незначною, якщо коефіцієнт мінливості не перевищує 10 %, середньою, якщо вона в межах від 10 до 20 %, і значною, якщо коефіцієнт вищий ніж 20%. Застосовуючи коефіцієнт мінливості у якості характеристики варіювання ознаки, необхідно враховувати одиниці розмірності досліджуваної ознаки: лінійні чи вагові (об'ємні). Показано, що при лінійному виміренні ознаки коефіцієнт мінливості буде приблизно у три рази менший, ніж за кубічного виміру тієї ж ознаки [Лакин, 1990].

Завдання 3.1

Завданням студентів є аналіз мінливості кількісних ознак клітин досліджуваних злаків. На прикладі отриманих первинних даних потрібно побудувати варіаційний ряд і гістограму. За наведеними формулами необхідно розрахувати статистичні показники положення (середню арифметичну) і варіації (розмах варіації, стандартне відхилення, дисперсію та коефіцієнт мінливості).

Після проведення розрахунків потрібно написати звіт – охарактеризувати мінливість кількісних ознак досліджених клітин (пиляків батьківських сортів злаків і гібридів) як самих по собі, так і у порівнянні між собою.

3.2. Параметричні методи дослідження. Формулювання мети дослідження. Перевірка нормальності розподілу.

Для використання потужних методів дослідження що використовують для аналізу кількісних ознак у певній вибірці, необхідно з'ясувати, як розподіляється ознака у генеральній сукупності. Ці питання докладно розібрані у значній кількості статистичних посібників [Рокицкий, 1973; Лакін, 1990; Атраментова, Утевська, 2008 та багато інших], з якими студенти повинні бути добре ознайомлені.

Нормальний розподіл є класичним розподілом математичної статистики. Він характеризує безперевні випадкові величини, значення яких визначаються безліччю одночасно діючих незалежних факторів. За нормальним законом розподіляються більшість біологічних ознак. Тим не менш, потрібно підкреслити, що певні ознаки (події) можуть бути описані іншими розподілами. Наприклад, коли мова йде про несумісні події – народження дівчинки чи хлопчика, передається чи ні нащадкам певний рецесивний алель – такі події підкоряються біноміальному розподілу. Біноміальний розподіл застосовують у генетиці для аналізу співвідношення статі або генотипів у родинах однакового розміру [Атраментова, Утевська, 2008[.

Ймовірність дуже рідких подій (народження близнюків, виникнення мутацій тощо) – підлягає розподілу Пуассона (який є окремим випадком біноміального розподілу) Характерною ознакою розподілу Пуассона є рівність значень дисперсії s^2 та середньої арифметичної. Такий асиметричний розподіл має вигляд літери J. Існують і інші асиметричні розподіли, наприклад розподіл Максвелла [Лакин, 1990].

Загальноприйнятою є думка, що біологічні кількісні ознаки розподіляються у відповідності до нормального закону. Нормальний розподіл характеризує безперервні випадкові величини, значення яких визначаються безліччю діючих незалежних факторів. При $n \to \infty$ розподіл стає безперервним, а полігон розподілу перетворюється на плавну лінію – нормальну варіаційну криву, або криву Гаусса (рис. 6).

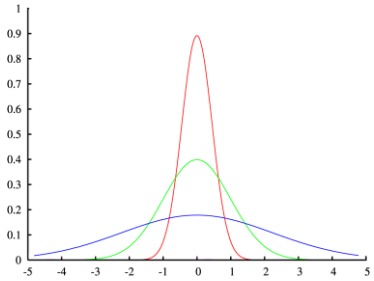

Рис. 6. Загальний вигляд кривих Гаусса.

Для вибору адекватного статистичного методу необхідно провести так звану перевірку на нормальність. Якщо вибірка взята з генеральної сукупності, де досліджувана ознака розподіляється у відповідності з нормальним законом, використовують параметричні методи – порівняння середніх арифметичних за допомогою критерію Ст'юдента, порівняння дисперсій за допомогою критерію Фішера, параметричні кореляційний і дисперсійний аналізи.

Якщо вибірка походить з генеральної сукупності, у якій розподіл ознак не відповідає нормальному закону, то використовують непараметричні методи статистичного аналізу.

Існує декілька способів перевірки на нормальність - вибір способу визначається, головним чином, обсягом вибірки. Якщо обсяг вибірки досить значний (до 50 дат, більш, ніж 50 дат) для неї розраховують показники ассиметрії та ексцесу. У випадку великих вибірок (більш, ніж 50 дат) можна використати і критерій χ^2. Відповідні формули легко знайти у підручниках з варіаційної статистики (наприклад, Атраментова, Утевська, 2008; Лакин, 1990).

У випадку нечисленної вибірки перевірку на нормальність проводять за допомогою критерію Шапіро-Уілка за наступною схемою [Атраментова, Утевська, 2008]:

1. Дати ранжують у порядку зростання $x_1, x_2, \ldots . x_{n-1}, x_n$
2. Обчислюють величину ***b*** за формулою:

$b = a_1(x_n - x_1) + a_2(x_{n-1} - x_2) + \ldots .$. Для цього знаходять різницю між датами що однаково відстоять від центру розподілу. Якщо ряд нараховує непарну кількість дат, то дата, що знаходиться у центрі $X_{центр}$, виключається із розрахунків.

Значення коефіцієнтів ***a*** знаходять у таблицях-додатках статистичних посібників [Атраментова, Утевська, 2008]. Для зручності розрахунків коефіцієнти для певної кількості дат наведено нижче.

Таблиця 5

Коєфіцієнти ***a*** для критерію Шапіро-Уілка за перевірки на нормальність (n – обсяг вибірки) [Атраментова, Утевська, 2008].

n	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
20	0,4734	0,3211	0,2565	0,2085	0,1686	0,1334	0,1013	0,0711	0,0422	0,0140					
25	0,4450	0,3069	0,2543	0,2148	0,1822	0,1539	0,1283	0,1046	0,0823	0,0610	0,0403	0,0200	0,0000		
30	0,4254	0,2944	0,2487	0,2148	0,1870	0,1630	0,1415	0,1219	0,1036	0,0862	0,0697	0,0537	0,0381	0,0227	0,0076

Повна таблиця значень коефіцієнтів *a* наведена у Додатку (табл. 1)

За допомогою критерію W тестують альтернативну гипотезу H_A – розподіл не є нормальними

3. Критерій Шапіро-Уілка розраховують за формулою:

$$W = \frac{b^2}{(n-1)s^2},$$

де s – стандартне відхилення.

4. $W_{факт}$ порівнюють з критичним табличним значенням. При $W_{факт} < W_{табл}$, альтернативна гіпотеза залишається чинною. Якщо $W_{факт} > W_{табл}$, то альтернативна гіпотеза відхиляється і розподіл вважають нормальним

Таблиця 6

Критичні значення критерію Шапіро-Уілка для перевірки нормальності розподілу при рівнях значущості *p=0,05* та *p=0,01* (n – обсяг вибірки) [Атраментова, Утевська, 2008.]

n	$W_{0,01}$	$W_{0,05}$
20	0,868	0,905
25	0,888	0,918
30	0,900	0,927

Повна таблиця критичних значень наведена у Додатку (табл. 2)

Серед причин помилкової асиметрії розподілів можна вказати технічні помилки (недбале ставлення до чистоти експерименту, нерендомізований добір досліджуваних об’єктів у вибірку). Певний вклад у це можуть внести середовищні умови та зумовлена будь-якими причинами неоднорідність сукупності, а також внутрішньоалельна та міжалельна взаємодії генів.

Нижче наведено приклад визначення нормальності розподілу даних про оптичну щільність (вміст загального білку у ядерці) археспоріальних клітин жита Харківське 60.

Первинні дані вмісту білку				Ранжовані від меншого до більшого			
0,216	0,263	0,221	0,179	0,159	0,187	0,193	0,221
0,242	0,183	0,187	0,185	0,163	0,187	0,21	0,224
0,159	0,193	0,189	0,219	0,179	0,189	0,216	0,242
0,21	0,192	0,252	0,224	0,183	0,191	0,218	0,252
0,163	0,187	0,218	0,191	0,185	0,192	0,219	0,263

Після того, як дані було розташовано у порядку зростання, потрібно визначити значення b, використовуючи значення таблиці 5 (для кількості 20 дат):

b = 0,4734(0,263-0,159)+ 0,3211(0,252-0,163)+ …. + 0,0422(0,210-0,191)+0,014(0,193-0,192) = 0,119

Розраховуємо стандартне відхилення за формулою $s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$,

s = 0,028.

5. Тестуємо гіпотезу (H_A: розподіл не є нормальним за допомогою критерію W).

$$W_{факт} = \frac{0,119}{(20-1)0,028^2} = 0,947$$

Таким чином, $W_{факт}$ перевищує $W_{табл}$ (яке дорівнює 0,905) . Це значить, що альтернативна гіпотеза відхиляється і розподіл кількісної ознаки «вміст білку в ядерці» вважають нормальним.

Завдання 3.2

Здійснити перевірку на нормальність розподілу досліджуваних кількісних ознак клітин за допомогою критерію Шапіро-Уілка, обґрунтувати та пояснити отримані результати.

ТЕМА 4. Визначення кількості генів, що впливають на кількісну ознаку. Формули Касла і Райта, Серебровського. Умови та обмеження використання цих формул.

Матеріали та обладнання

Персональні ЕОМ

Комп'ютерна програма Excel Microsoft Office

У представників усіх популяцій спостерігається мінливість за великою кількістю генів, які впливають на різноманітні морфологічні ознаки. У тих випадках, коли доводиться оперувати показниками кількісних ознак, варіація яких неперервна (а до числа таких відносяться більшість ознак, у тому числі і господарсько-важливих), відмінності між особинами представляють собою ряд з усіма переходами, у якому немає можливості виділити окремі, чіткі групи.

Тому генетика кількісних ознак потребують особливих математичних підходів [Тоцький, 2008].

Аналіз генетичного підґрунтя кількісних ознак повинен враховувати дві обставини а) сильну залежність їх прояву від умов зовнішнього середовища і б) можливу детермінацію їх багатьма генами.

Розвиток будь-якої з таких ознак у кінцевому рахунку залежить від сукупної дії багатьох генів так званих полігенів. Однак кожний з цих генів не може бути виділений окремо. Тому для характеристики популяції за подібними ознаками не можна використовувати методики генетики якісних ознак – наприклад, визначення співвідношень генотипів, чи частот генів тощо. Единий шлях, зокрема, на першому етапі аналізу – використання звичайних статистичних параметрів – середньої арифметичної, стандартного (середньоквадратичного) відхилення, дисперсії та інших.

Для більш детального ознайомлення зі шляхами та методами статистичної характеристики популяцій за кількісними ознаками необхідно звернутися до монографії П. Ф. Рокицького «Введение в статистическую генетику».

Традиційно кількісні або варіюючи ознаки розглядаються як полігенні, а саме варіювання ознаки пояснюється дією багатьох генів і впливом різних умов середовища. У генетичному аналізі таких ознак допускається однакова сила дії кожного з полігенів і їх незалежне успадкування. На цьому допущенні складено і рекомендовано формули для обчислення можливої кількості генів, що обумовлюють ознаку.

Закономірності успадкування кількісних ознак:

1. Крива розподілу особин F_1 за досліджуваною ознакою зазвичай розташовується між кривими розподілу батьківських форм. Середня арифметична F_1 найчастіше займає проміжне положення між середніми арифметичними батьківських форм.
2. Середня арифметична F_2 приблизно дорівнює середній арифметичній F_1, але варіювання особин F_2 значно більше, ніж варіювання особин F_1
3. Криві розподілу особин із зворотних схрещувань (беккроссов) F_1 на кожного з батьків зрушені ближче до кривих розподілу тих з батьківських форм, з якими було проведено беккрос
4. Криві розподілу F_3 від особин F_2, узятих з різних частин варіаційного ряду F_2, розташовуються по різному в залежності від того, з якої частини ряду взяті особини F_2

Були запропоновані декілька теорій, які пояснюють успадковування полімерних чи множинних генів. Найбільш прийнята серед них – теорія полімерних генів. Її базові посилання такі:

1) гени адомінантні, тобто гетерозиготи по тому чи іншому полімерному гену займають суворо проміжне положення між двома гомозиготами;

2) за силою дії гени рівні одне одному;

3) дія декількох генів сумується адитивно;

4) чим більше пар генів приймають участь у розщепленні, тим менше у F_2 та F_3 доля особин, що фенотипово подібні до вихідними батьківськими формами;

5) частоти градацій кількісної ознаки серед нащадків відповідають коефіцієнтам розкладення біному Ньютона $(a + b)^n$, де n – кількість пар генів, за якими відбувається розщеплення.

Зрозуміло також, що окремі гени можуть мати різний ступень домінування, аж до зверхдомінування. Дії окремих полімерних генів можуть бути не рівні одне одному. Сумарна дія декількох генів може бути не лише простою адитивною, але й і ускладнюватися різними формами міжгенної взаємодії. Все це може відобразитися на розщепленні у F_2, навіть без урахування зовнішнього середовища.

Таким чином, питання вирахування навіть кількості генів, які відповідають за кількісну ознаку, стає досить заплутаним.

Для визначення кількості генів у таких випадках потрібно вдатися до певних спрощень та обмежень, а саме:

- Дія досліджуваних генів однакова і вони адомінантні;
- Схрещувані форми гомозиготні за цими генами;
- Усі гени, що посилюють ознаку, знаходяться у одного батька, а ті, що послаблюють, у другого

Тим не менш, суттєво, що невиконання окремих обмежуючих умов впливає на отриманий результат (на кількість генів, яку визначають), лише в сторону його збільшення. Таким чином, застосування формул дає змогу встановити мінімальну кількість генів, за якими розрізняються форми, що приймають участь у схрещуванні.

Вперше таку формулу для визначення кількості генів n (кількість алелів тоді 2n) запропонували Касл і Райт (цит. за [Рокицкий, 1974]):

$n = \frac{D^2}{8(s_{F_2}^{\ 2} - s_{F_1}^{\ 2})}$, где D – різниця між середніми батьківських порід чи ліній, s^2 –дисперсії F_1 та F_2.

В підручнику П. Ф. Рокицького, (як і у багатьох інших керівництвах з біологічної статистики), для позначення дисперсії використовується символ σ (сігма). Таким чином, $\sigma^2 = s^2$.

Для виключення впливу домінування А. С. Серебровським було запропоновано більш загальну формулу

$n = \frac{3D^2 - 4DD_1 + 4D_1^{\ 2}}{16(s_{F_2}^{\ 2} - s_{F_1}^{\ 2})}$, у якій додано величину D_1 – різниця між середньою величиною ознаки F_1 і ознаки у рецесивного типу.

Ще одна формула була запропонована Бертоном: $n = \frac{0{,}25(0{,}75 - h + h^2)D^2}{(s_{F_2}{}^2 - s_{F_1}{}^2)}$, де D – різниця між середніми батьківських форм, а $h = \frac{\bar{x}_{F_1} - \bar{x}_{P_1}}{D}$. Ця формула має ті ж обмеження, що і попередні формули [Рокицкий, 1974].

Таким чином, незважаючи на неточність цих методів у зв'язку з обмежуючими їх умовами, вони дають інформацію про ступінь генетичних відмінностей по даній кількісній ознаці між формами, що схрещуються

Перед тим, як розпочати виконання завдання, студенти повинні визначити, які саме із наданих груп та комбінацій схрещування придатні для визначення кількості генів, що кодують досліджувані кількісні ознаки.

Завдання 4

Завданням є визначення кількості генів, що відповідають за досліджувані кількісні ознаки трьома різними способами та порівняння отриманого результату.

ТЕМА 5. Оцінка ступеня фенотипового домінування. Перевірка вірогідності відмінностей між середніми досліджуваних ознак за допомогою критерію Ст'юдента. Оцінювання ступеня фенотипового домінування за формулою Peter - Frey з інтерпретацією по Beil, Atkins. Визначення прояву гетерозису.

Матеріали та обладнання

Персональні ЕОМ

Комп'ютерна програма Excel Microsoft Office

Результатом реалізації генотипу у конкретних умовах середовища є його фенотиповий прояв. З огляду на це, успадковування кількісних ознак може бути визначеним шляхом оцінки ступеня фенотипового домінування h_p в F_1 за формулою, запропонованою Петером та Фреєм [Petr, Frey, 1966].

Правомірне використання формули домінування Petr - Frey можливе тільки в тому випадку, коли статистично доведено суттєвість відмінностей між батьківськими формами за досліджуваною ознакою.

Перевірка вірогідності відмінностей середніх арифметичних досліджуваних ознак

Традиційно для перевірки вірогідності відмінностей між досліджуваними вибірками використовують **критерій Ст'юдента**, який дає можливість визначити вірогідність відмінностей за порівняння середніх арифметичних. Вибір формули для обчислення критерію Ст'юдента залежить від того, рівні чи ні дисперсії порівнюваних груп, залежні чи незалежні групи. В залежності від співвідношення (рівності) дисперсій використовують різні формули.

Корректне розуміння рівності дисперсій та незалежності (залежності) вибірок розглядається у статистичній літературі, наприклад у посібнику Л. О. Атраментової і М. О. Утевської «Статистичні методи в біології», 2008 чи підручнику О. І. Орлова «Прикладная статистика», 2004.

Для перевірки вірогідності нульової гіпотези замість критерію Ст'юдента часто доцільно використовувати **критерій Крамера-Уелча** [Орлов, 2004].

$$T = \frac{\sqrt{mn}(\bar{x} - \bar{y})}{\sqrt{ns_x^2 + ms_y^2}}$$

Нульова гіпотеза формулюється так: середні арифметичні генеральних сукупностей, з яких узяті вибірки, не розрізняються.

Додатковою перевагою цього критерію є те, що не потрібно визначати рівність дисперсій. Правило прийняття рішення для критерію Крамера-Уелча виглядає так:

Якщо $T_{факт} \leq t_{кр}$, то нульова гіпотеза про рівність генеральних середніх залишається чинною, різниця між вибірковими середніми вважається випадковою. Якщо $T_{факт} > t_{табл}$, то нульова гіпотеза відхиляється на прийнятому рівні значущості α, різниця між вибірковими середніми вважається вірогідною.

Показово, що критичні значення для критерію Крамера-Уелча залежать лише від рівня значущості α й виражаються через критичні значення критерію Стюдента наступним чином:

$T_{факт}(0,01) = t_{кр}(0,01;\infty) = 2,58$
$T_{факт}(0,05) = t_{кр}(0,05;\infty) = 1,96$
$T_{факт}(0,1) = t_{кр}(0,1;\infty) = 1,65$

У прикладній статистиці найбільш часто використовують рівень значущості α=0,05. Тоді значення модулю статистики *T* Крамера-Уелча потрібно порівнювати з гранічним значенням $\Phi(1-\frac{\alpha}{2}) = 1,96$ [Новиков, 2004; Орлов, 2004].

Як приклад, наведено порівняння вмісту РНК у ядерцях археспоріальних клітин пшениці Безоста 1 і жита Харківське 60. Нехай обсяг вибірки пшениці $n_1 = 20$ клітин, $\bar{x}_1 = 0,166$, $s_1 = 0,039$, а для другої вибірки (жито) ці показники відповідно складуть $n_2 = 20$, $\bar{x}_2 = 0,177$, $s_2 = 0,020$.

Тоді величина модулю Крамера-Уелча дорівнює:

$$T = \frac{\sqrt{n_1 n_1}(\bar{x}_1 - \bar{x}_2)}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} = \frac{\sqrt{20 \times 20}(0,166 - 0,177)}{\sqrt{20 \times 0,039^2 + 20 \times 0,020^2}} = -0,24$$

Оскільки отримане значення за абсолютною величиною є меншим, ніж 1,96, то нульова гіпотеза залишається чинною та на рівні значущості 0,05 можна стверджувати, що відмінності між вибірками є випадковими.

Оцінка ступеню фенотипового домінування [Petr, Frey, 1966]

$$h_{P_{F_1}} = \frac{\overline{X}_{F_1} - \overline{X}_{MP}}{D_1},$$

де $\overline{X}_{F_1}$ - середнє значення ознаки у гібриду,

$\overline{X}_{MP}$ - середнє значення ознаки у обох батьків;

$D_1 = \overline{X}_{H_p} - \overline{X}_{MP}$ - різниця між середнім значенням ознаки у кращої батьківської форми і середнім значенням у обох батьківських форм.

Для визначення фенотипового домінування ознак у F_2 пшенично-житніх гібридів використовували формулу $h_{P_{F_2}} = \frac{2(\overline{X}_{F_2} - \overline{X}_{MP})}{D_2}$,

де $D_2 = \overline{X}_{H_p} - \overline{X}_{MP}$.

Характер успадковування і диференціації ступеней домінування оцінювали відповідно до градації, запропонованої [Beil, Atkins, 1965]:

h_p < -1 від'ємне наддомінування;
h_p = - 1 від'ємне повне домінування;
-1 < h_p < 0 часткове домінування «гіршої» батьківської форми;
h_p = 0 проміжне успадковування;
0 < h_p < 1 часткове домінування «кращої» батьківської форми;
h_p = 1 повне домінування «кращої» форми;
h_p > 1 наддомінування, справжній гетерозис.

Справжній гетерозис (переважання гібридом кращої батьківської форми за даною ознакою) розраховують за формулою:

$$ht(\%) = \frac{\overline{X}_{F_1} - \overline{X}_{Hp}}{\overline{X}_{Hp}} \cdot 100,$$

де $\overline{X}_{F_1}$ - середне значення ознаки у гібриду F_1, а $\overline{X}_{Hp}$ - середне значення ознаки у кращої батьківської форми.

<u>Завдання 5</u>

<u>За допомогою вказаних вище формул оцінити ступінь фенотипового домінування досліджуваних кількісних ознак у пшенично-житніх гібридів та міжсортових пшеничних і ячмінних гібридів першого покоління.</u>

ТЕМА 6. Дисперсійний аналіз. Створення дисперсійної решітки. Проведення дисперсійного аналізу за допомогою програми Excel.

Матеріали та обладнання

Персональні ЕОМ

Комп'ютерна програма Excel Microsoft Office

Найчастіше дисперсійний аналіз застосовують у випадках, коли потрібно оцінити силу впливу фактору на величину (середній рівень) ознаки; він є потужним інструментом аналізу кількісних ознак і необхідним підґрунтям для будь якого-генетичного дослідження у цій галузі. Цей метод широко використовується для планування експерименту і статистичної обробки його даних. В основі цього методу полягає закон розподілення відношень середніх квадратів (дисперсій):

$$F == \frac{s_1^2}{s_2^2} = \frac{середній\ квадрат\ вибіркових\ середніх}{середній\ квадрат\ об'єктів}$$

Загальним для усіх видів дисперсійного аналізу є порівняння декількох груп за різноманітністю об’єктів, що їх складають. Та ознака, яка підлягає дослідженню, називається результативною. Беруть до уваги, що причиною різноманітності цієї ознаки є певний фактор (чи фактори), який організують у дослідженні відповідно до прийнятих методів. У генетичних дослідженнях найчастіше фактором виступає стан організму (прояв тієї чи іншої кількісної ознаки).

Ступені дії фактору називаються градаціями фактору (різні генотипи). Наявність декількох градацій фактору – необхідна умова щодо проведення генетичного аналізу. Вважають, що різні градації фактору неоднаково впливають на досліджувану ознаку, тобто мають статистичний вплив на неї.

У дисперсійному аналізі вивчають три види статистичних впливів: факторіальний, випадковий і загальний.

Факторіальний вплив – вплив фактору, що є організованим і може бути проконтрольований у експерименті. Зрозуміло, що на ознаку діють не лише організовані фактори, але й безліч інших, які не враховуються у дослідженні. Ці фактори називають випадковими, а їх вплив – випадковим. Вони створюють загальний фон, на якому відбувається дослідження. Чим більше факторіальний вплив відрізняється від випадкового, тим статистично більш вірогідним є дія фактору. Сумарний вплив усіх організованих і неорганізованих факторів називається загальним впливом.

За допомогою однофакторного дисперсійного аналізу вивчають вплив одного фактору, за допомогою двофакторного – аналізують вплив двох факторів, а також вплив взаємодії між факторами.

При дисперсійному аналізі одночасно опрацьовують дані декількох вибірок (варіантів), що складають єдиний статистичний комплекс, оформлений у вигляді спеціальної робочої таблиці.

Для використання цього потужного інструменту потрібно витримати умову вірної побудови дисперсійного комплексу у відповідності з обраною моделлю експерименту. Відносно аналізу кількісних ознак це означає, що при закладці генетичного або селекційного досліду масив кожного з дослідних об’єктів (у даному випадку рослин) повинен поділятися на декілька повторностей (не менш ніж 4). Це необхідна умова; закладку повторностей

кожного варіанту здійснюють ще на етапі планування досліду. Для більш детального знайомства з основами підготовки експерименту та проведення дисперсійного аналізу необхідно звернутися до підручника Б. О. Доспехова «Методика полевого опыта».

Математичний сенс дисперсійного аналізу – розчленування загальної суми квадратів відхилень і загального числа ступенів свободи на частини – компоненти (які повинні відповідати структурі експерименту) і оцінка значущості дії та взаємодії досліджуваних факторів за F-критерієм.

Сучасні комп'ютерні статистичні програми провадять складні розрахунки дисперсійного аналізу автоматично, проте вкрай важливо розуміти сутність моделей розрахунку, покладених в основу автоматичних розрахунків та необхідність правильної побудови дисперсійних комплексів. Сучасний експеримент, у якому досліджуються кількісні ознаки, повинен закладатися з урахуванням відповідних моделей дисперсійного аналізу.

При роботі з однофакторними статистичними комплексами, які складаються з декількох незалежних вибірок, то загальна мінливість результативної ознаки, що вимірюється загальною сумою квадратів Cy поділяється на два компоненти: варіювання між вибірками (варіантами) Cv в всередині вибірок Cz. У загальній формі мінливість ознаки може буди представлена виразом Cy = Cv + Cz.

Варіювання між вибірками (Cv) – це та частина загальної дисперсії, яка зумовлена дією досліджуваних факторів, а дисперсія всередині вибірок (Cz) – характеризує випадкове варіювання, тобто похибку експерименту.

Загальне число ступенів свободи (N-1) також поділяється на частини: ступені свободи для повторень: n – l, ступені свободи для варіантів (l-1) і для випадкового варіювання (l-1)(n-1).

Якщо ж, як у випадку класичних генетичних експериментів, досліджують однофакторні сполучені статистичні комплекси, коли вибірки (варіанти) пов'язані якоюсь загальною, контрольованою умовою, наприклад, наявністю n організованих повторностей у досліді, тоді загальна сума квадратів розкладається на три частини: варіювання повторностей Cp, варіантів Cv і випадкове Cz. У таких випадках загальна мінливість та загальне число ступенів свободи можуть бути представлені виразами:

$$Cy = Cp + Cv + Cz$$

$$(N-1) = (n-1)+(l-1)+ (n-1)\ (l-1)$$

Статистичний аналіз даних провадять у три етапи [Доспехов, 1985].

1. Складають розрахункову таблицю, розташовуючи у ній вихідні дані по рядам і стовпцям, визначають суми і середні по варіантам і

повторенням, загальну суму і середнє значення ознаки по досліду (зразок у таблиці)

Таблиця 7

Розташування даних у таблиці дисперсійного комплексу

Варіанти	Вихідні дані X				Число спостережень n	Сума по варіантам (V)	Середні по варіантам
1	X_{11}	X_{12}	…	X_{1n}	n_1	V_1	$\bar{x}_1$
2	X_{21}	X_{22}	…	X_{2n}	n_2	V_2	$\bar{x}_2$
…	…	…	…	…	…	…	…
l	X_{l1}	X_{l2}	…	X_{ln}	n_l	V_l	$\bar{x}_l$
Суми по повторностям	P_1	P_2	…	P_n	$N = \Sigma n$	$\Sigma X = \Sigma V = \Sigma P$	$\bar{x} = \frac{\sum x}{N}$

2. Вираховують суми квадратів відхилень за формулами наступної таблиці і визначають фактичне значення критерію $F_ф$

Попередньо необхідно зробити розрахунок корректуючого фактору С за формулою:

$C = \bar{x}\ \Sigma X$ – коректуючий фактор, поправка

Таблиця 8

Формули для розрахунку сум квадратів відхилень, дисперсії та критерію $F_ф$

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_ф$	$F_т$
Загальна Cy	ΣX^2 - C	N-1	-	-	-
Повторності P	ΣX^2:*l* - C	n-1	-	-	-
Варіантів Cv	ΣV^2: n - C	*l*-1	$s^2_v = C_V / l - 1$	$s_v^2 : s^2$	за статистичними таблицями*
Залишок Cz (похибка)	$C_y - C_P - C_v$	(*l*-1)(n-1)	$s^2 = C_Z /$ (n-1)(*l* -1)	-	-

* теоретичне значення знаходять по статистичним таблицям, виходячи з *l*-1 ступеня свободи для дисперсії варіантів (числівник) та N- *l* ступенів для залишка (знаменник). При $F_ф > F_т$ у досліді є істотні відмінності по варіантам на 5% рівні значущості, і тоді нульова гіпотеза (усі відмінності між варіантами випадкові) відхиляється.

3. Оцінка істотності різниці між середніми

Необхідно розуміти, що критерій F лише встановлює факт наявності істотних відмінностей між середніми, але не вказує, між якими середніми ці відмінності мають місце. Тому, якщо загальна оцінка за цим критерієм встановлює наявність варіантів, які істотно відрізняються від решти, і нульова гіпотеза про рівність параметрів відхиляється, тоді необхідно визначити, які саме варіанти істотно розрізняються. Для цього визначають похибки досліду та НІР (найменша істотна різниця) за наступними формулами:

$s_{\bar{x}} = \sqrt{\frac{s^2}{n}}$ - похибка досліда;

$s_d = \sqrt{\frac{2s^2}{n}}$ - похибка різності середніх;

Визначення найменшої істотної різниці розраховують за формулою

$$НІР_{05}=t_{05}s_d$$

де значення критерію t_{05} беруть з таблиць додатків посібників по статистиці для N-*l* ступенів свободи дисперсії залишку (таблиця 3 Додатку). Для цього використовують значення рівня значущості та число ступенів свободи залишкової дисперсії. Індексами при НІР та t записані показники рівня значущості. Таким чином, ті різності між середніми, які є більшими, ніж $НІР_{05}$, є істотними з 5-% рівнем значущості. Потрібно пам'ятати, що 5%-ому рівню значущості відповідає 95%-вий рівень вірогідності, а 1%-ому – 99%-вий [Доспехов, 1985].

Нижче наведено розібраний приклад дисперсійного аналізу (рівномірні комплекси) кількісної ознаки «об'єм ядерця» однієї комбінації віддаленого схрещування – пшениці, жита та їх гібрида.

Дослід, результатом якого стали постійні мікропрепарати, що були використані для отримання мікрофотографій клітин, був закладений у п'яти повторностях. Номера повторностей вказані на етикетках мікропрепаратів. Таким чином, для заповнення первинної таблиці дисперсійного комплексу потрібно отримати середню арифметичну значення ознаки для кожної повторності окремо і внести її у таблицю у відповідну ячейку.

Таблиця 10

Об'єм ядерця у клітинах трьох досліджуваних злаків

Варіанти, генотипи	Повторності, мкм					Число спосте-режень	Сума по варіантам V	Середні по варіантам
Безоста 1	28,42	29,59	23,83	31,58	26,69	5	140,12	28,03
Харківське 60	19,39	17,48	16,46	18,63	19,99	5	91,95	18,39
F_1 Безоста 1 × Харківське 60	23,14	24,97	24,28	26,93	23,75	5	123,09	24,62
Суми по повторностям (P)	70,96	72,04	64,58	77,14	70,44	15	355,18	23,67

Визначаємо значення загальної вибірки N, який розраховують шяхом помноження кількості повторностей n=5 на кількість варіантів *l* = 3. Таким чином, N = 15.

Визначаємо коректуючий фактор $C = (\Sigma X)^2 : N = (335,18)^2:15 = 8410,55$

$Cy = \Sigma X^2 – C = (28,42^2 + 29,59^2 + … + 23,75^2) – 8410,55 = 289,92$

$C_P = Cp = \Sigma P^2/l – C = (70,962 + 72,042 + …70,442)/3 - 8410,55 = 26,76$

$Cv = \Sigma V^2/n - C = (140,12^2 + 91,95^2 + 123,09^2)/5 - 8410,55 = 238,66$

$Cz = Cy - Cp - Cv = 289,92 - 26,76 - 238,66 = 24,49$

Таблиця 11

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_ф$	$F_т$
Загальна Cy	289,92	14	-		
Повторностей C_P	26,76	4	-		
Варіантів Cv	238,66	2	119,33	38,99	4,46
Залишок (похибка)	24,49	8	3,06		

При $F_ф > F_т$ у досліді спостерігаються суттєві відмінності по варіантам на 5% рівні значущості, таким чином нульова гіпотеза – фактор генотипу не чинить статистичного впливу на ознаку – відкидається

Визначають похибки досліду і НІР:

$s_{\bar{x}} = \sqrt{\frac{3,06}{5}} = 0,78$ мкм - похибка досліду;

$s_d = \sqrt{\frac{2 \times 0,78}{5}} = 0,55$ мкм - похибка різності середніх;

Показник t_{05} за 8 ступенів свободи дорівнює 2,31, $t_{01} = 3,36$.

Таким чином, найменша істотна різниця для 5% рівня значущості $НІР_{05} = t_{05}s_d = 0,40$ мкм

Після проведення розрахунків можна зробити висновок про те, що відмінності у розмірах ядерець у клітинах пшениці, жита та пшенично-житнього гібриду F_1 першого покоління суттєві і зумовлені генотипом досліджуваних злаків. Показник $НІР_{05}$ дозволяє судити про вірогідність відмінностей між середніми значеннями ознаки усіх досліджуваних злаків. Після його розрахунку можна стверджувати, що об'єм ядерця пшениці вірогідно перевищують відповідні показники жита та пшенично-житнього гібриду першого покоління.

Також необхідно знати, що визначення різниці між середніми груп, що складають дисперсійний комплекс, можна вирахувати не лише за допомогою Нір, а ще й за допомогою методу Тьюкі (для рівновеликих груп) та методу Шеффе (для груп рівного або різного обсягу) [Атраментова, Утевська, 2008].

Завдання 6

Використовуючи отриманий раніше масив даних, провести дисперсійний аналіз прояву однієї з кількісних ознак клітин генеративних структур злаків для пари батьківських рослин і їх гібриду.

ТЕМА 7. Коєфіцієнт успадковуваності. Коєфіцієнт успадковуваності як показник частки генетичного варіювання у загальному фенотиповому варіюванні ознаки**.** Визначення коєфіцієнту успадковуваності/

Матеріали та обладнання
Персональні ЕОМ
Комп'ютерна програма Excel Microsoft Office

Як вже зазначалося, на кількісні ознаки (які мають неперервний ряд прояву) дуже сильно впливають фактори зовнішнього середовища. Такі модифікаційні зміни також неперервні і ще більше стирають розбіжності між окремими класами [Тоцький, 2008].

На практиці, тим не менш, постійно виникає потреба визначити, в якому ступені мінливість кількісних ознак залежить від генотипу, і в якому – від модифікацій. Для цього досліджують загальну фенотипову мінливість у популяції. Фенотипова мінливість популяції завжди може бути виявлена практично і виміряна дисперсією (варіансою у декількох виданнях, наприклад у підручнику П. Ф. Рокицького чи посібнику Г. К. Дремлюка, И. Ф. Герасименко).

Дуже детально принципи виділення компонентів фенотипової дисперсії біологічна та математична сутності окремих варіанс-дисперсій, які характеризують популяції, розібрані у підручнику П. Ф. Рокицького «Введение в статистическую генетику». У нашому випадку ми коротко розглянемо мінімум, необхідний для коректного визначення одного з найбільш важливих параметрів, що описують популяцію за кількісними ознаками – коефіцієнту успадковуваності.

Потрібно пам'ятати, що генетико-статистичне дослідження кількісних ознак провадиться лише для популяцій. У популяціях, що складаються з генетично різних особин (не клонів), загальна фенотипова варіанса популяції (σ_P^2) є результатом сумації відмінностей між генотипово відмінними особинами, з яких складається досліджувана популяція (ці відмінності вимірюються варіансою σ_G^2) і відмінностей, які викликані умовами зовнішнього середовища (σ_E^2, вимірюються варіансою в межах кожної групи особин, більш-менш подібних генотипово).

Таким чином, фенотипова варіанса розкладається на два компоненти – генотипову та середовищну варіанси. З точки зору генетичного аналізу, найсуттєвішим при дослідженні є визначення того, яку долю у загальній фенотиповій дисперсії ознаки займає дисперсія, що залежить від генотипових відмінностей у популяції, чи встановлення величини $\frac{\sigma_G^2}{\sigma_P^2}$. У найпростішому випадку вона дорівнює: $\frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}$. Ця величина, яку позначають як h^2, і є коефіцієнтом успадковуваності. Він є мірилом питомої ваги генотипової мінливості у загальній мінливості популяції [Рокицкий, 1974].

З огляду на те, що генотипова варіанса може складатися із трьох компонентів, доля яких у сумарній величині генотипової варіанси може бути різною, то прийнято розрізняти два види успадковуваності – у «вузькому» і «широкому» розумінні.

Успадковуваність у «широкому розумінні» - це вплив усіх спадкових факторів організму – може бути виражена як відношення генотипової варіанси до загальної фенотипової варіанси:

$$H^2 = \frac{\sigma_G^2}{\sigma_P^2} = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}$$

Коефіцієнт успадковуваності може мати значення від 0 до 1, або від 0 до 100% і відображує вклад адитивних генетичних факторів у досліджуваний фенотип.

Успадковуваність у «вузькому розумінні» - враховує генетичну різноманітність, що зумовлена адитивною спадковістю (адитивний чи сумарний ефект генів є сумою ефектів окремих алелів даного локусу).

Загальну генотипову варіансу (σ_G^2) розчленовують на адитивну генотипову варіансу (σ_A^2), домінантну генотипову варіансу (σ_D^2) і епістатичну генотипову варіансу (σ_1^2):

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_1^2$$

Таким чином, при визначенні успадковуваності у «вузькому розумінні» вираховують відношення величини генетичної адитивної дисперсії σ_A^2 до загальної дисперсії σ_P^2. Під σ_A^2 розуміють ту частину, що пов'язана з дією домінантних генів, які пригнічують прояв рецесивних алелів. Із σ_G^2 у даному випадку виключають ефекти, що зумовлені епістазом.

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2} = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_D^2 + \sigma_1^2 + \sigma_e^2}$$

Оскільки σ_A^2 завжди менша чи дорівнює σ_G^2, h^2 завжди буде менший чи дорівнювати H^2. Потрібно розуміти, що виділення адитивної генотипової варіанси із загальної генотипової не завжди можливо. Вирахування коєфіцієнта успадковуваності у «вузькому розумінні» потребує постановки спеціальних дослідів [Дремлюк, Герасименко, 1992].

Чим вищий показник коефіцієнту успадковуваності, тим більша доля фенотипової варіації досліджуваних варіантів зумовлена їх генотипом. Високі значення h^2 можуть вказувати на велику генотипову різноманітність вихідних батьківських форм, а також свідчити про те, що на дану ознаку мало впливає паратипова (середовищна) мінливість. Низькі коефіцієнти успадковуваності вказують на слабку генотипову різноманітність батьків чи (або) сильну залежність ознаки від зовнішніх умов. Таким чином, показник h^2 зменшується за інбридингу і довготривалого добору на відповідну ознаку, тому що і перший, і другий процеси призводять до генотипового вирівнювання нащадків. Оптимізація умов, за яких розвиваються організми, обумовлює підвищення коефіцієнта успадковуваності, бо сприяє більш повному вияву генетичних потенцій. У селекційній практиці цей коефіцієнт має істотне прогностичне значення – чим він вищий у досліджуваної групи особин, тим кращого результату слід очікувати від добору.

Для визначення коефіцієнта успадковуваності використовують кореляційний, регресійний і дисперсійний аналізи.

Застосування дисперсійного аналізу дозволяє розчленувати загальну фенотипову варіансу на генотипову і середовищну та вірно оцінити ступінь впливу генотипу і середовища на вираженість ознаки у нащадків.

Для розчленування загальної фенотипової варіанси на варіансу, що зумовлена генотиповими відмінностями варіантів і варіансу, що викликана умовами зовнішнього середовища і неврахованими факторами, прирівнюють середні квадрати до їх оцінюваних параметрів, внаслідок чого отримують:

$$\hat{\sigma}_G^{\,2} = \frac{(\hat{\sigma}_e^{\,2} + n\sigma_G^{\,2}) - \hat{\sigma}_e^{\,2}}{n},$$

звідки $\hat{\sigma}_P^{\,2} = \hat{\sigma}_G^{\,2} + \hat{\sigma}_e^{\,2}$ [Дремлюк, Герасименко, 1992].

У табл. 12 наведено приклад заповнення первинної таблиці дисперсійного аналізу, дані якої потім будуть використані для розрахунку коефіцієнта успадковуваності у широкому сенсі.

Таблиця 12

Первинна таблиця дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	Параметри, що оцінюють
Загальна Cy	ΣX^2 - C	N-1	-	
Повторностей C_P	ΣX^2:*l* - C	n-1	-	$\sigma_e^2 + a\sigma_b^2$
Варіантів Cv	ΣV^2: n - C	*l*-1	$s^2_v = C_V/l - 1$	$\sigma_e^2 + n\sigma_G^2$
Залишок (похибка)	$C_y - C_P$ - Cv	(*l*-1)(n-1)	$s^2 = C_Z/$ (n-1)(*l* -1)	σ_e^2

Коефіцієнт успадковуваності у широкому сенсі дорівнює:

$$H^2 = \frac{\hat{\sigma}_G^{\,2}}{\hat{\sigma}_G^{\,2} + \hat{\sigma}_e^{\,2}}.$$

Розібраний вище приклад дисперсійного аналізу можна використати для визначення коефіцієнту успадковуваності в «широкому розумінні»

Таблиця 13

Первинна таблиця дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	Параметри, що оцінюють
Загальна Cy	289,92	14	-	
Повторностей C_P	26,76	4	-	$\sigma_e^2 + a\sigma_b^2$
Варіантів Cv	238,66	2	119,33	$\sigma_e^2 + n\sigma_G^2$
Залишок (похибка)	24,49	8	3,06	σ_e^2

$\hat{\sigma}_e^2 = 3{,}06$

$\hat{\sigma}_e^2 + n\sigma_G^2 = 238{,}66$

Звідси: $\hat{\sigma}_G^2 = \frac{238{,}66 - 3{,}06}{5} = 47{,}12$

$\hat{\sigma}_P^2 = \hat{\sigma}_G^2 + \hat{\sigma}_e^2 = 47{,}12 + 3{,}06 = 50{,}18$

Коефіцієнт успадковуваності у широкому сенсі дорівнює:

$$H^2 = \frac{\hat{\sigma}_G^2}{\hat{\sigma}_G^2 + \hat{\sigma}_e^2} = \frac{47{,}12}{50{,}18} = 0{,}93$$

Ми отримали дуже високий показник успадковуваності ознаки «об'єм ядерця» для аналізованих злаків, який свідчить про те, що основна доля фенотипової варіації досліджуваних злаків зумовлена їх генотипом.

Як вже зазначалося, для вірогідного визначення коефіцієнтів успадковуваності кількісних ознак у «широкому» та «вузькому» розумінні потрібно дуже суворо підійти до питань організації дисперсійних комплексів та вибору досліджуваного матеріалу. Розроблені статистично обґрунтовані варіанти і схеми дисперсійного аналізу, які дозволяють визначити як обидва коефіцієнти успадковуваності, так і такі важливі для генетичної характеристики матеріалу і селекційної роботи з ним показники як загальна та специфічна комбінаційні здатності. Проте для виправданого їх використання експеримент повинен закладатися за певними схемами, зокрема ієрархічною схемою дисперсійного аналізу [Рокицкий, 1974; Турбин и др., 1974; Дремлюк, Герасименко, 1992].

З огляду на те, що коефіцієнт успадковуваності – параметр, який завжди відноситься до популяції, у якій можливі різноманітне поєднання зовнішніх умов, з одного боку, і різна ступінь генетичної мінливості, з іншого боку, то становиться зрозуміла певна його обмеженість. Потрібно мати на увазі, що успадковуваність – це не лише характеристика ознаки, але й і популяції і тих середовищ них умов, у яких знаходяться особини даної популяції. Так як h^2 вираховують на основі ряду величин, то зміна у будь-якій з них буде впливати на його значення. Таким чином, розраховані показники h^2 завжди відносяться до конкретних популяцій, особини яких знаходяться у певних же зовнішніх умовах.

Тим не менше, коефіцієнт успадковуваності відноситься до найважливіших параметрів за генетичного аналізу кількісних ознак у популяції.

Завдання 7

Умовно представити усю кількість досліджуваних клітин кожного зі злаків як популяцію і за допомогою дисперсійного аналізу встановити коєфіцієнт успадковуваності для цитометричних ознак у широкому сенсі.

Рекомендована література

1. Атраментова Л. О., Утєвська О. М. Статистичні методи в біології. Х: ХНУ імені В. Н. Каразіна, 2007. – 288 с.

2. Доспехов Б. А. Методика полевого опыта (с основами статистической обработки результатов исследований). - М.: «Колос», 1985. – 351 с.

3. Дремлюк Г. К., Герасименко В. Ф. Приемы анализа комбинационной способности и ЭВМ-программы для нерегулярних скрещиваний. – Москва ВО «Агропромиздат» 1991 – Одесса СГИ 1992. – 144 с.

4. Новиков Д. А. Статистические методы в педагогических исследованиях (типовые случаи). М.: МЗ-Пресс, 2004. – 67 с.

5. Орлов А. И. Прикладная статистика. Учебник. - М.: Экзамен, 2004. - 656 с.

6. Паламарчук И. А., Веселова Т. Д. Учебное пособие по ботанической гистохимии. – М.: Изд-во МГУ, 1965. – 108 с.

7. Паушева З. П. Практикум по цитологии растений. – М.: Агропромиздат, 1988. – 271 с.

8. Рокицкий П. Ф. Биологическая статистика. – Минск: Выш. школа, 1973. – 319 с.

9. Рокицкий П. Ф. Введение в статистическую генетику. – Минск: Выш. школа, 1974. – 452 с.

10. Тоцький В. М. Генетика. – Одеса: Астропринт, 2008. – 712 с.

11. Турбин Н. В., Хотылева Л.В., Тарутина Л. В. Диаллельный анализ в селекции растений. – Минск: Наука и техника, 1974. – 180 с.

12. Beil G. M., Atkins R. E. Inheritance of quantitative characters in grain sorghum // Jowa State. J. Sci. – 1965. – V. 39, № 3. – P. 345 - 358.

13. Petr F. C., Frey K. J. Genotypic correlations, dominance and heritability of quantitative characters in oats. // Crop Sci. – 1966. – V. 6, № 3. – P. 259 -262.

Додаток

Повний перелік постійних мікротомних препаратів пиляків злаків:
М'яка озима пшениця (2n = 42)
Безоста 1
Миронівська 808
Альбатрос одеський
Янтарна
Жито озиме (2n = 14)
Харківське 60
Ячмень озимий (2n = 14)
Екзотик
Незалежний
Пшенично-житні гібриди F_1 (2n = 28)
Безоста 1 × Харківське 60
Миронівська 808 × Харківське 60
Альбатрос одеський × Харківське 60
Міжсортовий гібрид пшениці F_1 (2n = 42)
Безоста 1 × Янтарна
Міжсортовий гібрид ячменю F_1 (2n = 14)
Екзотик × Незалежний
Міжсортовий гібрид пшениці F_2 (2n = 42)
Безоста 1 × Янтарна
Міжсортовий гібрид ячменю F_2 (2n = 14)
Екзотик × Незалежний
Міжсортовий гібрид пшениці F_3 (2n = 42)
F_3 Безоста 1 × Янтарна

Таблиця 1

Коєфіцієнти а для критерію Шапіро-Уілка за перевірки на нормальність (n – обсяг вибірки) *

n	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
2	0,7071														
3	0,7071	0,0000													
4	0,6872	0,1677													
5	0,6646	0,2413	0,0000												
6	0,6431	0,2806	0,0875												
7	0,6233	0,3031	0,1401	0, 0000											
8	0,6052	0,3164	0,1743	0,561											
9	0,5888	0,3244	0,1976	0,0947	0,0000										
10	0,5739	0,3291	0,2141	0,1224	0,0399										
11	0,5601	0,3315	0,2260	0,1429	0,0695	0,0000									
12	0,5475	0,3325	0,2347	0,1586	0,0922	0,0303									
13	0,5359	0,3325	0,2412	0,1707	0,1099	0,0539	0,0000								
14	0,5251	0,3318	0,2460	0,1802	0,1240	0,0727	0,0240								
15	0,5150	0,3306	0,2495	0,1878	0,1353	0,0880	0,433	0,0000							
16	0,5056	0,3290	0,2521	0,1939	0,1447	0,1005	0,0593	0,0196							
17	0,4968	0,3273	0,2540	0,1988	0,1524	0,1109	0,0725	0,0359	0,0000						
18	0,4886	0,3253	0,2553	0,2027	0,1587	0,1197	0,0837	0,0496	0,0163						
19	0,4808	0,3232	0,2561	0,2059	0,1641	0,1271	0,0932	0,0612	0,0303	0,0000					
20	0,4734	0,3211	0,2565	0,2085	0,1686	0,1334	0,1013	0,0711	0,0422	0,0140					
21	0,4643	0,3185	0,2578	0,2119	0,1736	0,1399	0,1092	0,0804	0,0530	0,0263	0,0000				
22	0,4590	0,3156	0,2571	0,2131	0,1764	0,1443	0,1150	0,0878	0,0618	0,0368	0,0122				
23	0,4542	0,3126	0,2563	0,2139	0,1787	0,1480	0,1201	0,0941	0,0969	0,0459	0,0228	0,0000			
24	0,4493	0,3098	0,2554	0,2145	0,1807	0,1512	0,1245	0,0997	0,0764	0,0539	0,0321	0,0107			
25	0,4450	0,3069	0,2543	0,2148	0,1822	0,1539	0,1283	0,1046	0,0823	0,0610	0,0403	0,0200	0,0000		
26	0,4407	0,3043	0,2533	0,2151	0,1836	0,1563	0,1316	0,1089	0,0876	0,672	0,0476	0,0284	0,0094		
27	0,4366	0,3018	0,2522	0,2152	0,1848	0,1584	0,1346	0,1128	0,0923	0,0728	0,0540	0,0358	0,0178	0,0000	
28	0,4328	0,29992	0,2510	0,2151	0,1857	0,1601	0,1372	0,1162	0,0965	0,0778	0,0598	0,0424	0,0253	0,0084	
29	0,4291	0,2968	0,2499	0,2150	0,1864	0,1616	0,1395	0,1192	0,1002	0,0822	0,0650	0,0483	0,0320	0,0159	0,0000
30	0,4254	0,2944	0,2487	0,2148	0,1870	0,1630	0,1415	0,1219	0,1036	0,0862	0,0697	0,0537	0,0381	0,0227	0,0076

*Атраментова, Утевська, 2008, табл. 6, стр. 259

Таблиця 2

Критичні значення критерію Шапіро-Уілка *W* для перевірки нормальності розподілу в нечисленних вибірках (n≤30) при рівнях значущості *p=0,05* та *p=0,01* (n – обсяг вибірки)*

n	$W_{0,01}$	$W_{0,05}$
3	0,753	0,767
4	0,687	0,748
5	0,868	0,762
6	0,713	0,788
7	0,730	0,803
8	0,749	0,818
9	0,764	0,829
10	0,781	0,842
11	0,792	0,850
12	0,805	0,859
13	0,814	0,866
14	0,825	0,874
15	0,835	0,881
16	0,844	0,887
17	0,851	0,892
18	0,858	0,897
19	0,863	0,901
20	0,868	0,905
21	0,873	0,908
22	0,878	0,911
23	0,881	0,914
24	0,884	0,916
25	0,888	0,918
26	0,891	0,920
27	0,894	0,923
28	0,896	0,924
29	0,898	0,926
30	0,900	0,927

*Атраментова, Утевська, 2008, табл. 7, стр. 260

Таблиця 3

Значення критерію t
при рівнях значущості *p=0,05, p=0,01та p= 0,001**

n	Рівень значущості		
	0,05	0,01	0,001
1	12,71	63,66	-
2	4,30	9,93	31,60
3	3,18	5,84	12,94
4	2,78	4,60	8,61
5	2,57	4,03	6,86
6	2,45	3,71	5,95
7	2,37	3,50	5,41
8	2,31	3,36	5,04
9	2,26	3,25	4,78
10	2,23	3,17	4,59
11	2,20	3,11	4,44
12	2,18	3,06	4,32
13	2,16	3,01	4,22
14	2,15	2,98	4,14
15	2,13	2,95	4,07
16	2,12	2,92	4,02
17	2,11	2,90	3,97
18	2,10	2,88	3,92
19	2,09	2,86	3,88
20	2,09	2,85	3,85
21	2,08	2,83	3,82
22	2,07	2,82	3,79
23	2,07	2,81	3,77
24	2,06	2,80	3,75
25	2,06	2,79	3,76
26	2,06	2,78	3,71
27	2,05	2,77	3,69
28	2,05	2,76	3,67
29	2,05	2,76	3,66
30	2,04	2,75	3,65
50	2,01	2,68	3,50
100	1,98	2,63	3,39
∞	1,96	2,58	3,29

*Доспехов, 1985, табл. 1, стр. 317

Учебное издание

Алєксєєва Т. Г.

ГЕНЕТИКО-СТАТИСТИЧНИЙ АНАЛІЗ КІЛЬКІСНИХ ОЗНАК КЛІТИН

Методичні вказівки до великого спеціального практикуму
для студентів IV курсу денної форми навчання та V курсу заочної форми навчання напряму 6.040102 – «біологія»

Печатается в авторской редакции

Підп. до друку . Формат
Гарн. Таймс. Умов.-друк. арк. . Тираж прим.
Зам. №

Видавець і виготовлювач
Одеський національний університет
імені І. І. Мечникова
Свідоцтво ДК № 4215 від 22.11.2011 р.

Україна, 65082, м. Одеса, вул. Єлісаветинська, 12
Тел. (048) 723 28 39. E-mail: druk@onu.edu.ua