

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
МИКОЛАЇВСЬКИЙ НАЦІОНАЛЬНИЙ АГРАРНИЙ УНІВЕРСИТЕТ
Факультет ТВППТСБ
Кафедра генетики, годівлі тварин та біотехнології

СПЕЦІАЛЬНІ ІНФОРМАЦІЙНІ СИСТЕМИ І ТЕХНОЛОГІЇ

**Методичні рекомендації щодо виконання лабораторно-практичних
занять для здобувачів ступеня вищої освіти ступеня доктора філософії зі
спеціальності 204 «ТВППТ»**



Миколаїв
2017

УДК 004 : 636.03
С71

Друкується за рішенням науково-методичної комісії факультету ТВППТСБ Миколаївського національного аграрного університету від 23.11.2017 р., протокол № 3.

Укладач:

С. С. Крамаренко – д-р біол. наук, доцент, професор кафедри генетики, годівлі тварин та біотехнології Миколаївського національного аграрного університету.

Ю. В. Волосюк – канд. техн. наук, доцент, завідувач кафедри інформаційних систем і технологій Миколаївського національного аграрного університету

Рецензенти:

О.В. Жуков – д-р біол. наук, доцент, професор кафедри зоології та екології Дніпропетровського національного університету імені Олеся Гончара;

Є.В. Баркарь – кандидат с.-г. наук, доцент, доцент кафедри генетики, годівлі тварин та біотехнології Миколаївського НАУ.

З М І С Т

Вступ	4
Тема 1. Інформаційно-обчислювальні системи в селекції	5
Тема 2. Сучасні ППП: їхнє призначення, структура і головні особливості функціонування	11
Тема 3. Основні автоматизовані інформаційні системи (АІС) в тваринництві	18
Тема 4. Основні завдання інформаційно-обчислювальних систем в селекції	25
Тема 5. Методика математико-статистичного аналізу імуногенетичних даних для великої рогатої худоби ..	33
Тема 6. Автоматизація аналізу варіаційного ряду	41
Тема 7. Автоматизація біометричного оброблення даних зоотехнічного обліку	46
Тема 8. Автоматизація оцінки бугаїв-плідників за нащадками	50
Тема 9. Автоматизація порівняльного аналізу в ряду “предки-нащадки”	53
Тема 10. Автоматизація розрахунків коефіцієнтів успадкування та повторюваності	57
Тема 11. Автоматизація розрахунків коефіцієнтів лінійної регресії	60
Тема 12. Автоматизація алгоритму автоматизованого дисперсійного аналізу	63
Тема 13. Автоматизація аналізу терміну господарського використання тварин	65
Тема 14. Моделювання селекційного процесу за допомогою ПЕОМ	69
Список рекомендованої літератури	72

ВСТУП

Організація великомасштабної селекції тварин молочних порід на базі використання досягнень популяційної генетики стала можливою завдяки застосуванню ПЕОМ (персональних електронних обчислювальних машин) для збирання, накопичення і обробки даних племінного обліку. Індексна оцінка племінних якостей тварин, моделювання та оптимізація програм великомасштабної селекції, оцінка бугаїв за якістю нащадків та ряд інших заходів великомасштабної селекції практично нездійсненні без використання ПЕОМ.

Застосування ПЕОМ сприяє підвищенню ефективності племінної роботи: підвищуються точність оцінки племінних якостей тварин, ефективність відбору, підбору, тощо. Але особливо великий потенціал різкого підвищення ефективності племінної роботи закладається в комплексному використанні принципів великомасштабної селекції, досягнень популяційної генетики, математичних методів та ПЕОМ.

Використання комп'ютерів у селекційній роботі дозволяє оперативно збирати й оброблювати великий обсяг інформації для оцінки і відбору за кращими репродуктивними показниками плідників і маточного поголів'я тварин. Велику роботу у цьому напрямі було проведено в період масового створення АСУ на базі великих ПЕОМ.

Методичні рекомендації розроблено для здобувачів ступеня вищої освіти ступеня доктора філософії зі спеціальності 204 «ТВППТ»

Тема №1

Інформаційно-обчислювальні системи в селекції

1. Проблеми племінного обліку та аналізу селекційно-генетичних параметрів

Організація великомасштабної селекції тварин молочних порід на базі використання досягнень популяційної генетики стала можливою завдяки застосуванню ЕОМ (електронних обчислювальних машин) для збирання, накопичення і обробки даних племінного обліку. Індексна оцінка племінних якостей тварин, моделювання й оптимізації програм великомасштабної селекції, оцінка бугаїв за потомством та ряд інших заходів великомасштабної селекції практично нездійсненні без використання ЕОМ.

Обробка даних зоотехнічного і племінного обліку в молочному скотарстві ґрунтується на відповідних принципах. Господарство або держплемоб'єднання, виділене для машинної обробки даних, передає в обчислювальний центр бланки обліку (щомісячного) або картки (один раз на рік) на тварин (форми 1-мол і 2-мол), пристосовані для обробки даних на ЕОМ.

В обчислювальному центрі дані переносяться на технічні носії інформації (магнітні стрічки, диски та ін.). Після перенесення даних на технічні носії інформації і контролю їх правильності бланки щомісячного обліку здають в архів ОЦ, а індивідуальні картки на тварин повертають у господарства та держплемоб'єднання для подальшого заповнення. Форми 1-мол і 2-мол пристосовані для обробки даних на ЕОМ і тому є робочим документом селекціонерів й обліковців, а також джерелом передачі інформації в ОЦ. За допомогою технічних носіїв інформації дані про тварин вводяться в пам'ять ЕОМ і проводиться їх обробка.

При щомісячному надходженні інформації на ЕОМ оформляють такі документи: звіт про результати отелень і осіменінь корів, розвиток молодняку, продуктивності корів за контрольними доїннями, кількість і якість спермопродукції бугаїв, запліднювальну здатність сперми плідників, поповнення і витрати спермобанку та ін. Крім того, на ЕОМ на кожний наступний місяць складається план осіменінь та отелень (для кожної корови), комплектування виробничих груп тварин тощо.

На основі накопиченої інформації про тварин за рік на ЕОМ складають річні звіти про продуктивність, бонітування тварин, результати їх племінного використання.

На основі накопичених даних шляхом щомісячної обробки інформації або один раз на рік за формами 1-мол і 2-мол створюють банк даних породи, на основі якого на ЕОМ вирішується ряд завдань з племінної роботи: матеріали бонітування тварин; оцінка бугаїв за розвитком, відтворною здатністю і за потомством; оцінка селекційно-генетичних параметрів, результатів селекції і т.і.

Вітчизняний та зарубіжний досвід показує, що найбільш ефективна щомісячна обробка даних зоотехнічного і племінного обліку. Така система дає змогу ретельно контролювати стан, давати спеціалістам господарств необхідну інформацію для ефективного управління виробничими процесами.

Застосування ЕОМ сприяє підвищенню ефективності племінної роботи: підвищуються точність оцінки племінних якостей тварин, ефективність добору, підбору тощо. Але особливо великий потенціал різкого підвищення ефективності племінної роботи закладається в комплексному використанні принципів великомасштабної селекції, досягнень популяційної генетики, математичних методів та ЕОМ. Продуктом такого комплексного підходу є **автоматизована інформаційна система (АІС) в селекції**.

Проблема створення АІС полягає не лише у заміні ручної праці автоматизованою обробкою даних племінного обліку на ЕОМ, а й у вирішенні принципово нових завдань, що підвищують ефективність галузі. Наприклад, підвищення точності визначення племінної цінності бугаїв на основі корекції інформації на вплив генетичних та середовищних факторів, генетико-економічної оптимізації програми великомасштабної селекції та ін. Такий підхід є основним принципом у створенні АІС.

Це здійснюють із використанням спеціалізованого обчислювального центру, укомплектованого необхідною технікою для підготовки та обробки даних на ЕОМ, бази даних за породами на технічних носіях інформації, вирішення комплексу завдань великомасштабної селекції з допомогою ЕОМ. Так, при Інституті розведення і генетики тварин УААН створено Республіканський селекційно-обчислювальний центр. Він укомплектований ЕОМ ЕС та ПЕОМ. Із племінних господарств України, а також з господарств, виділених для оцінки бугаїв за потомством, обласних племоб'єднань у обчислювальний центр регулярно надходять первинні дані про тварин, де їх переносять на технічні носії інформації. Накопичену інформацію обробляють на ЕОМ для вирішення комплексу завдань великомасштабної селекції.

АІС в селекції тварин на базі ПЕОМ будується за допомогою простого блоку до ПЕОМ типу "Модем", завдяки якому по каналах зв'язку можна передавати й отримувати інформацію незалежно від відстані між абонентами. В Республіканському інформаційно-обчислювальному центрі селекції тварин установлюється накопичувач пам'яті великої потужності, що на основі "Модем" приймає і накопичує інформацію про всіх підконтрольних тварин регіону або країни через ПЕОМ, що встановлені у племінних господарствах, племоб'єднаннях, селекційних центрах та інших підприємствах.

2. Інформаційно-обчислювальні системи в селекції та історія їхнього розвитку

Останнім часом успіх багатьох видів діяльності людей значною мірою залежить від рівня комп'ютеризації розрахунків у сфері їх діяльності. Це стосується і галузі тваринництва.

Використання комп'ютерів у селекційній роботі дозволяє оперативно збирати й оброблювати великий обсяг інформації для оцінки і відбору за кращими репродуктивними показниками плідників і маточного поголів'я тварин. Значну роботу в цьому напрямі було проведено в період масового створення АСУ на базі великих ЕОМ.

Концепція інформаційного процесу управління селекцією має такий вигляд:

1. Підхід до селекції як до динамічної системи управління процесом генетичного вдосконалювання великих і нечисельних популяцій. (Система повинна бути гнучкою і швидко реагувати на різні сторонні зміни).
2. Орієнтація на сучасні методи селекції. (Математичне забезпечення системи повинне базуватися на теорії селекції тварин, узагальнених лінійних моделях і економіко-математичних методах).
3. Орієнтація на фермерів, селекціонерів господарств, фахівців племпідприємств, селекційних центрів чи асоціацій за породами. (Система повинна надавати особам, що приймають рішення необхідну інформацію в зрозумілій формі).
4. Орієнтація на максимальну генетичну й економічну ефективність. (Система повинна забезпечити осіб, що приймають рішення, такою поточною і перспективною інформацією, яка б гарантувала максимальну ефективність розведення тварин).
5. Незалежність системи від ведення обліку даних. (Програмне забезпечення повинне працювати з будь-якою базою даних).

В інформаційній системі управління селекцією можна виділити такі блоки завдань:

- статистичний аналіз даних;
- генетична оцінка тварин;
- генетичний аналіз популяцій;
- оптимальна програма селекції;
- оптимізація відбору і підбору тварин;
- оцінка ефективності.

Найвідомішою на той період була система для молочного скотарства „СЕЛЕКС” (А.А. Цалітис, вісімдесяті роки минулого століття). Пізніше ця система під назвою „Селекс-Мікро” була адаптована для персональних ЕОМ, але через високу вартість широкого розповсюдження не набула.

Основа роботи системи "СЕЛЕКС" – автоматизований облік – ведення "електронних" картотек тварин і журналів, у яких реєструється виконання технологічних операцій (контрольні годівлі, запліднення, перевірки на тільність, запуски й ін.), ветеринарні заходи, рекомендації з коректування режиму змісту й експлуатації тварин.

Практично в Україні використовується тільки розроблена під операційну систему DOS система подібного призначення „ОРСЕК” (М.М. Майборода, С.Г. Германчук).

Для збирання інформації і визначення племінної цінності тварин використовуються також ПЕОМ на великих птахофабриках. Є досвід застосування ЕОМ у племінній роботі галузей вівчарства та свинарства („АССТР”, В.А. Коротков і К.Г. Корабельников).

Облікова інформація, що збирається в ході племінної роботи, одночасно використовується і для господарського обліку виходу продукції від окремих тварин, груп, ферм тощо. Тому підтримувати високий рівень племінної роботи на базі ПЕОМ – одна з найважливіших складових підвищення продуктивності тварин і підвищення рентабельності їх утримання.

Іншою важливою складовою в даному сенсі є раціональне використання кормової бази, яка займає найбільший відсоток у собівартості продукції тваринництва. Для комп’ютерних розрахунків кормових раціонів годівлі тварин та їх оптимізації можуть використовуватися такі системи, як „АРМ зоотехніка” (Борисенко В.Г.), „Раціон-1” (Коротков В.А.), „WinMix” (ОО „Вітамекс”) тощо.

Поліпшити планування, облік, економічний аналіз і регулювання в галузі тваринництва можуть системи типу „Прайс” (Ю.С. Коваленко, В.В. Матвеев), „Project Expert”, „1С”, „Бест” та ін.

Проте використання вказаних систем пов’язане з рядом недоліків, оскільки вони впроваджені лише в поодиноких господарствах сільськогосподарського профілю, більшість програмних продуктів написані під застарілу операційну систему DOS, у галузі явний дефіцит обчислювальної техніки, а наявна переважно застаріла.

Тому, для підвищення ролі комп’ютерної техніки й інформаційних технологій у поліпшенні економічних показників галузі потрібен комплекс заходів, серед яких можна виділити:

- необхідність суттєвих інвестиційних надходжень у галузь для придбання сучасного комп’ютерного обладнання;
- розроблення нових комп’ютерних інформаційних систем для виконання розрахунків у названих вище аспектах із застосуванням сучасних графічних та інших технологічних можливостей ПЕОМ;
- використання в інформаційних системах тваринницької галузі поряд з базами даних з інформацією про походження, розвиток і продуктивність тварин та їх предків багатовимірних сховищ даних, що є особливо цінним для племінної справи;
- поширення набутого досвіду і широке впровадження нових комп’ютерних розробок не лише для поліпшення племінної справи, а й для систематичного обґрунтування раціонів годівлі тварин, організаційного управління;
- активне впровадження в тваринництво автоматизованих систем управління технологічними процесами;
- інтенсивне впровадження INTERNET-технологій, електронної комерції, комп’ютерного цінового моніторингу продовольчого ринку.

3. Основні принципи і методи створення БД

В АІС базу даних розглядають як інформаційну модель об'єкта (наприклад, селекційний процес у породі). При цьому, ефективність системи керування об'єктом багато в чому залежить від точності і вірогідності інформації, що складає базу даних. З технічної точки зору, **база даних (БД)** – це сукупність взаємозалежних даних, що зберігаються в пам'яті ЕОМ, і можуть бути використані для одного чи декількох рішень.

Основними важливими характеристиками БД є:

- їхнє постійне розширення і зміна;
- принцип незалежності БД і програм, що їх використовують;
- швидкий довільний пошук інформації;
- принцип багаторазового використання БД для вирішення різних завдань і користувачів;
- простота доступу до інформації;
- низька вартість збереження і використання даних, мінімальні витрати на внесення змін;
- здійснення постійного контролю за вірогідністю внесених нових даних;
- таємність і конфіденційність даних, що зберігаються.

За структурою БД складається з **елементів** і **атрибутів**. Елементи БД можуть бути матеріальні (тварина, підприємство) і нематеріальні (подія, племінна цінність, продуктивність тварин). Сукупність однорідних об'єктів (запис індивідуальних номерів об'єктів) називають набором об'єктів. Атрибути останніх відбивають характеристику об'єктів – родовід тварин, місце народження, продуктивність і ін.

Зразкова БД за породою повинна мати наступний вигляд:

1. БД на бугаїв, що перевіряються (родоводи, характеристики розвитку, екстер'єр, фізіологічна оцінка сперми, результати запліднення корів і телиць, дані про продуктивність дочок і ровесниць, племінна цінність за набором ознак);
2. БД на дорослих бугаїв-плідників (родоводи, показники продуктивності, племінної цінності й їхніх предків протягом 4-х поколінь);
3. БД на корів племінних стад (родоводи, дані про продуктивність, племінну цінність корів і їхніх предків протягом 4-х поколінь);
4. Дані за групами крові племінних корів і бугаїв-плідників.
5. Параметри, що характеризують популяцію в цілому (зоотехнічні, селекційно-генетичні й економічні).

Системи управління базами даних (СУБД), уже розроблені в даний час, дозволяють виконувати багато завдань. Найбільш придатні для цих цілей СУБД FOX PRO та її аналоги. Ця СУБД дозволяє створювати файли DBF, які потім можна обробляти в програмних пакетах Microsoft Office Excel, а також програмами статистичної обробки даних SAS, HARVY.

Крім цього, програма СЕЛЕКС може використовуватися для первісного збирання і накопичення даних і подальшій їхній обробці в FOX PRO.

Питання для самоперевірки:

1. Які існують проблеми племінного обліку та аналізу селекційно-генетичних параметрів?
2. Яка історія розвитку інформаційно-обчислювальні системи в селекції?
3. Які основні принципи і методи створення БД?

Тема №2

Сучасні ППП: їхнє призначення, структура і головні особливості функціонування

Комп'ютерні системи для аналізу даних - пакети прикладних програм (ППП) - є, порівняно з іншими наукомісткими програмами, мабуть, найбільш широко застосовуваними в практичній і дослідницькій роботі у різноманітних областях людської діяльності.

Була запропонована класифікація статистичних пакетів по *чотирьох групах*:

- інтегровані пакети загального призначення;
- спеціалізовані пакети;
- предметно- (чи проблемно-) орієнтовані пакети;
- навчальні програми.

Існує до тисячі розповсюджуваних на світовому ринку пакетів, що вирішують того чи іншого виду завдання статистичного аналізу даних, у середовищі DOS чи Windows. Із західних *універсальних пакетів* найбільш відомі й добре відпрацьовані комп'ютерні системи SPSS, NCSS, Statgraphics, Statistica та ін.

До ППП висувають такі вимоги:

- принцип модульності програмного забезпечення;
- використання простої проблемно-орієнтованої мови для формулювання завдання користувача;
- автоматична організація процесу оброблення даних і зв'язків модулями пакета;
- наявність засобів ведення банку даних користувача і результатів проробленого аналізу;
- можливість діалогового режиму роботи користувача з пакетом; сумісність з іншим програмним забезпеченням.

1. ППП STATISTICA та його головні модулі

Загальна інформація. Пакет вимагає володіння статистичною термінологією. На ринку поширені версії 5.5, 6.0 і 7.0 для MS Windows.

Statistica за своєю структурою складається з декількох зв'язаних між собою «міні-пакетів»: *Базова статистика, Непараметрична статистика, Керування даними, Дисперсійний аналіз, Множинна регресія, Нелінійна регресія* і багато інших. Ці «міні-пакети» взаємодіють один з одним, маючи однаковий формат системних файлів.

У плані функціонального наповнення ППП (порівняно, наприклад, з програмою StatGraphics) відрізняється великою розмаїтістю; наприклад, він містить у собі ряд непараметричних методів аналізу, методи багатомірного аналізу: дискримінантного, факторного, кластерного логлінійного й ін.

Особливості керування пакетом. Дані легко ввести в середовище пакета, відносно легко відредагувати, створити нові змінні («ознаки»), вибрати окремі спостереження. Завдяки великій панелі інструментів, для виконання більшості завдань досить кілька разів клацнути мишкою, тому що майже для всіх функцій пакета є піктограми. Крім того, натиском правої кнопки мишки можна викликати додаткові підменю, що істотно прискорюють роботу з пакетом.

Використання OLE технології обміну між Windows-продуктами дозволяє легко інтегрувати результати, наприклад, у WinWord із Statistica.

Найбільш сильною стороною пакета є графіка і засоби редагування графічних матеріалів. Представлено сотні типів графіків: типу 2D чи 3D (є навіть графіки типу 4-D), матриці і піктограми. Можна розробити свій дизайн графіка і додати його в меню.

Крім того, пакет має тритомну документацію в 3000 сторінок і коротке керівництво. До екранного довідника входить майже вся інформація друкованої документації.

2. ППП «NCSS» та його основні модулі

ППП «NCSS-2003» (Number Cruncher Statistical System) розроблений Дж.Хінтце (США) у 2001 р. Один з найбільш поширених програмних продуктів у світі. Однією з важливих переваг даного ППП є дуже повний і детальний опис, наведений в Help'і.

ППП «NCSS-2003» являє собою багатокомпонентний інтерфейс, що у блоці «Аналіз» представлений такими модулями:

- **ANOVA** (простий дисперсійний аналіз, загальна лінійна модель (GLM), Лог-лінійні моделі, множинний дисперсійний аналіз та ін.);
- **Кластеризація** (різні методи кластеризації – дендрограми, ієрархічна кластеризація, кластеризація за K-середніми);
- **Описова статистика** (описова статистика для кількісних перемінних, кросстабуляція для якісних);
- **Аналіз часових рядів** (представлена модель ARIMA, ауто- і крос-кореляція, спектральний аналіз та ін.);
- **Багатовимірний аналіз** (Аналіз Головних Компонентів (PCA), Факторний аналіз, Канонічні кореляції й ін.);
- **Аналіз якісних перемінних** (оцінка частот в одне- і двохвибіркових аналізах, критерій Хі-квадрат, тести Мантеля-Хензеля, Мак-Немара й ін.);
- **Контроль якості** (методи Промислової статистики);
- **Регресія і Кореляція** (Лінійна і нелінійна, проста і множинна регресія, кореляційні матриці);
- **Аналіз таблиць виживаності;**
- **T-тести** (одно- і двовибірковий, парний, множинний T2 Хотеллінга та ін.).

3. ППП SPSS і його головні модулі

Загальна інформація. Пакет призначений у першу чергу для статистиків-професіоналів, тому що має досить могутній апарат статистичного аналізу. Завдяки переорієнтації розробників на платформу Windows, програма SPSS стала в даний час одним з лідерів серед універсальних статистичних пакетів.

Для роботи з ППП потрібно ПЕОМ з ОС Windows 95 чи вище і монітором VGA.

SPSS для Windows (версії 7.0-13.0) відрізняється різноманітними можливостями щодо керування даними і маніпулювання отриманими результатами, та роботи з електронними таблицями. SPSS надає досить зручну графіку (понад 50 типів діаграм), а також розвиті засоби підготовки звітів. Аналітичні параметри відображаються на екрані у виді простих і доступних меню і діалогових вікон.

Для ефективного застосування пакета потрібні методи з модулів SPSS BASE і Professional Statistics. Істотно ж підвищити ефективність класифікації і прогнозу може застосування частини (або всіх) з модулів Advanced Statistics. Перший модуль містить, зокрема, моделі логістичної регресії, а також ряд методів, суміжних із класифікацією. Модуль CHAID знаходить значимі зв'язки і вирішує завдання ієрархічного кластерного аналізу, внаслідок чого будуються дендрограми..

У SPSS також вирішено питання обміну з іншими Windows-продуктами і підтримується взаємообмін з більшістю форматів баз даних. Тож, можна, не виходячи із середовища WinWord, одночасно працювати в середовищі SPSS і, навпаки, дуже легко переносити отримані тестові чи графічні результати з SPSS у документ системи Word.

Достоїнства і недоліки. На думку розробників пакета у своїй *повній конфігурації* SPSS для Windows є пакетом з найбільш високим значенням параметра **потужність**: він має *дуже повний* набір статистичних (усього їх більш 60-ти) і графічних процедур, а також процедур створення звітів. Також, творці пакета гордяться інтерфейсом SPSS з користувачем, вважаючи його дуже простим і зручним. Крім того, традиційно пакет відрізняється високою точністю обчислень.

Однак, за підвищені комфорт і потужність потрібно заплатити чималі суми. Так, модулі «Advanced Statistics» і CHAID коштують біля \$500 і \$700. Набір з модулів для вирішення завдань класифікації в складі SPSS BASE і Professional Statistics коштує біля \$1100.

4. ППП Statgraphics і його головні модулі

Загальна інформація. Пакет призначений в основному для тих користувачів, які вже мають визначений досвід у статистиці. Особливо це

стосується модуля з багатомірними методами. Пакет був спочатку розроблений для платформи IBM PC і розрахований, у першу чергу, на графічні можливості комп'ютерної статистики. Однак, постійне його вдосконалювання в плані функціональних алгоритмів і способів керування даними істотно підсилило його базову **потужність**, що суттєво підвищило його конкурентноздатність.

Його найважливішим «плюсом» вважається вдале поєднання математичного апарата обробки даних із сучасною інтерактивною графікою. Інші, менш істотні його переваги — це широкі можливості взаємодії з електронними таблицями і СУБД (типу dBASE), а також з різноманітною периферією. Обмін з таблицями в Windows-версії виконується через стандартний буфер обміну.

Пакет Statgraphics побудований за модульним принципом. *Базовий* модуль містить ряд загальних процедур, а також процедури лінійної регресії. У розділі *Описової статистики* можна проаналізувати одну чи декілька змінних, підібрати ті чи інші теоретичні функції розподілу, одержати статистики, що нас цікавлять. Є можливість розрахувати коефіцієнти кореляції Пірсона і рангової кореляції Спірмена, але інших засобів ранжирування змінних немає.

Традиційно пакет має великі й дуже гнучкі графічні можливості: у наявності не тільки 2D кольорові, але і 3D графіки, діаграми, графіки, частотні гістограми, діаграми розсіювання, стовпчасті і кругові діаграми.

Багатомірний аналіз винесений, як і ряд інших розділів статистики, у додаткові модулі. Усього їх чотири: *Контроль якості*, *Планування експериментів*, *Часові ряди* і *Багатомірні методи*.

Недоліки пакета. Незначними недоліками ППП Statgraphics є нечіткості в його довідковій системі і видача результатів розрахунків з точністю до 4–5 значущих цифр. Є і більш серйозні мінуси. Наприклад, пакет некоректно працює при проведенні парних порівнянь на основі t-критерію. У плані коректності обчислень, очевидно, пакет трохи уступає своїм найближчим конкурентам (Statistica, SPSS).

5. Пакет аналізу генетичних даних TFPGA і його основні модулі

Пакет аналізу генетичних даних TFPGA (Tools For Population Genetic Analysis) був розроблений Марком Міллером з університету Північної Аризони. Версія 1.3 даного пакета (2000 р.) вільно поширюється через Інтернет із сайту автора.

Пакет призначений для аналізу Диплоїдних даних (ко-домінантних чи доміантних) і Гаплоїдних (гаплотипів). При цьому, можлива одно-, дво- і триярусна ієрархічна організація даних типа: Популяції – субпопуляції – субсубпопуляції.

Уведення і редагування даних проводиться за допомогою вбудованого редактора чи програми “Блокнот” MS Windows.

Перед початком аналізу програма нагадує, що слід вказати загальні характеристики вихідних даних: число локусів, максимальне число алелів на локус, число популяцій (суб-, субсубпопуляцій), а також тип даних.

Пакет має такі модулі:

- модуль «**Описова статистика**», що дає можливість одержати оцінки частот алелей за локусами, популяціями і всьому набору даних у цілому й аналогічні оцінки гетерозиготності (фактичної й очікуваної), а також ступеня поліморфізму локусів і популяцій;
- модуль «**F-статистики**» дозволяє одержати оцінки F-статистик на основі алгоритму Вейра і Кокерхама з помилками і довірчим інтервалом;
- модуль «**Генетичні дистанції**» дозволяє оцінити генетичні дистанції між елементами вихідної бази даних на основі алгоритмів Нея (1972, 1978), Роджерса (1978) і Рейнольдса (1983);
- модуль «**ХРВ**» призначений для оцінювання стану генетичної рівноваги Гарді-Вайнберга як для окремих популяцій, так і усїєї вибірки в цілому; при цьому можуть бути використані як критерій Хі-квадрат, так і точний тест (із процедурами Марковських ланцюгів і методу Монте-Карло);
- модуль «**UPGMA**» призначений для побудови дендрограм генетичної подібності між популяціями (чи субпопуляціями) на основі матриць генетичної дистанції на основі алгоритмів Нея (1972, 1978), Роджерса (1978) і Рейнольдса (1983);
- модуль «**Точний тест**» розрахований на оцінювання ступеня подібності популяцій на основі точних тестів, проведених для частот генотипів і алелів, за алгоритмом Россета і Рейнольдса (1985).

6. Пакет аналізу генетичних даних GenAIEx і його основні модулі

Пакет аналізу генетичних даних **GenAIEx** (Genetic Analysis in Excel) був розроблений Р.Пеакаллом з Австралійського національного університету (Австралія) і Р.Смоусом (Університет Рутжерса, США). Версія 6.0 даного пакета (2006 р.) вільно поширюється через Інтернет із сайту автора.

Пакет призначений для аналізу Диплоїдних даних (ко-домінантних чи домінантних) і Галоїдних (гаплотипів). Він вбудовується в MS Excel і введення та редагування даних проводиться за допомогою цього табличного редактора.

Пакет має такі модулі:

- модуль «**Описова статистика**», що дає можливість одержати оцінки частот алелей за локусами, популяціями і всьому набору даних у цілому й аналогічні оцінки гетерозиготності (фактичної й очікуваної), а також ступеня поліморфізму локусів і популяцій, оцінки ефективного числа алелей, частоти приватних алелей, оцінки F-статистик на основі частот генотипів;

- модуль «ХРВ» призначений для оцінювання стану генетичної рівноваги Гарді-Вайнберга як для окремих популяцій, так і усїєї вибірки в цілому; при цьому використовується тільки критерій Хі-квадрат Пірсона;
- модуль «Генетичні дистанції» дозволяє оцінити генетичні дистанції між елементами вихідної бази даних на основі алгоритмів Нея (1972, 1978);
- модуль «АМОВА» використовується для проведення аналізу молекулярних даних за алгоритмом Екскоффієра та ін. (1992) з оцінкою як загального показника генетичної диференціації (Ф-статистики), так і попарних оцінок між кожною парою вибірок із визначенням рівня значущості цих оцінок методом перетасування (пермутацій);
- модуль «Тест Мантеля» призначений для оцінки коефіцієнта кореляції між елементами двох матриць (критерію Мантеля) з визначенням рівня значимості цих оцінок методом перетасування (пермутацій);
- модуль «Аналіз головних координат (РСоА)» дає можливість провести ординацію об'єктів (популяцій) у просторі перших двох головних координат, використовуючи матрицю дистанцій Нея чи матрицю оцінок попарних генетичних диференціацій (Ф- чи F-статистики).

7. Модуль «Пакет аналізу» MS Excel і його структура

Модуль «Пакет аналізу» входить до числа представлених у MS Excel надбудов (Add-In) (стандартний набір). Активізується через опцію «Надбудови». Працює із даними, набраними у електронній таблиці MS Excel. Результати видає у окремих вікнах.

Модуль містить такі модулі:

- **Однофакторний дисперсійний аналіз** (простий фішерівський дисперсійний аналіз);
- **Двофакторний дисперсійний аналіз без повторностей** (для аналізу матриць з одним спостереженням у кожній клітинці);
- **Двофакторний дисперсійний аналіз із повторностями;**
- **Кореляція** (видає трикутні кореляційні матриці, але без оцінок рівня значимості);
- **Описова статистика** (представлені головні вибіркові статистики – середнє, мода, медіана, варіанса, дисперсія, асиметрія й ексцес та ін.);
- **Двовибірковий F-тест для дисперсії** (варіанси);
- **Гістограма** (оцінка абсолютних і відносних частот для побудови варіаційного ряду і гістограм розподілу);
- **Регресія** (лінійна, проста і множинна, з оцінками значимості коефіцієнтів регресії, коефіцієнта детермінації і помилки регресії);
- **Парний двовибірковий t-тест для середніх;**
- **Двовибірковий t-тест з однаковими дисперсіями;**

- **Двовибірковий t-тест із різними дисперсіями.**

Питання для самоперевірки:

1. Які основні модулі та функції має ППП Statistica?
2. Які основні модулі та функції має ППП NCSS?
3. Які основні модулі та функції має ППП SPSS?
4. Які основні модулі та функції має ППП StatGraphics?
5. Які основні модулі та функції має ППП TFPGA?
6. Які основні модулі та функції має ППП GenAIEx?
7. Які основні модулі та функції має модуль MS Excel «Пакет аналізу»?

Тема №3

Основні автоматизовані інформаційні системи (АІС) в тваринництві

1. Коло завдань, що вирішується АІС у селекції

Досвід останніх років показує, що використання ПЕОМ сприяє підвищенню ефективності племінної роботи: підвищуються точність оцінки племінних якостей тварин, ефективність відбору і підбора і т.і. Особливо помітне підвищення ефективності за комплексного використання принципом великомасштабної селекції, досягнень популяційної генетики, математичних методів і ЕОМ. Продуктом такого комплексного підходу є **автоматизовані інформаційні системи (АІС) у селекції**.

Під час розробки АІС необхідно використовувати системний підхід, тобто чітко визначити мету і критерії племінної роботи, виявити комплекс питань, що необхідно вирішувати, визначити обсяг, структуру і терміни надходження інформації, а також її обробки на ЕОМ, та систему контролю точності даних племінного і зоотехнічного обліку, передбачити принцип безупинного розвитку АІС.

До технічних засобів АІС відносяться власне ЕОМ, програмне забезпечення, засоби передачі даних в ОЦ, служба забезпечення (інженери, техніки, програмісти, оператори, зоотехніки-селекціонери і науковці).

Основними завданнями з племінної роботи, які можна і потрібно вирішувати в АІС є:

- оцінка селекційно-генетичних параметрів (мінливості, успадкування, кореляції, повторюваності, селекційного диференціала) за вибірками;
- факторний аналіз мінливості господарсько-корисних ознак у популяціях;
- оцінка племінної цінності тварин на підставі великої кількості інформації;
- оцінка і добір племінних тварин за генетичною резистентністю до хвороб і спадкоємних аномалій;
- оцінка прогнозу і результатів селекції, взаємодії «генотип х середовище», міжстадних генетичних розходжень та інших селекційно-генетичних параметрів;
- добір тварин для запису в ДПК;
- складання планів замовленого спарювання з урахуванням заданих параметрів і селекційно-генетичних показників;
- складання плану лінійно-ротаційного підбору для товарних черід;
- атестація походження племінних тварин за імуногенетичними групами крові;
- оптимізація програми селекції за породою;

- підготовка до друку форм 1-мол і 2-мол;
- складання зведених матеріалів з бонітування худоби.

2. Автоматизована інформаційна система «ІНСЕЛ»

АІС «ІНСЕЛ» була розроблена фахівцями колишнього Всесоюзного НДІ розведення і генетики тварин (Н.З. Басовський, 1985; 1987).

Як полігон для розробки, перевірки і впровадження в практику АІС «ІНСЕЛ» була використана Ленінградська область.

Створення БД по чорно-рябій породі на технічних носіях інформації, що охоплює 20-літній період використання племінних тварин, а також розробка пакета прикладних програм (ППП) для ЕОМ у рамках АІС «ІНСЕЛ» дозволили вирішувати багато завдань із селекції тварин:

- оптимізацію програм селекції;
- оцінку бугаїв за якістю нащадків з використанням сучасних досягнень популяційної генетики;
- оцінку результатів відбору, підбору і т.і.

Функціонування АІС «ІНСЕЛ» у поєднанні з комплексом заходів щодо племінної роботи сприяло постійному росту генетичного потенціалу худоби.

3. Інтегрована автоматизована система «СЕЛЕКС»

Інтегрована автоматизована система «СЕЛЕКС» (селекція, економіка, система) була розроблена на початку 70-х років минулого сторіччя Латвійською Аналітичною станцією з племінної роботи (А.А.Цалітис, Л.К.Ернст, 1982) за участю Всесоюзного інституту тваринництва.

Пізніше ця система під назвою „Селекс-Мікро” була адаптована для персональних ЕОМ, але через високу вартість широкого розповсюдження не набула.

Основа роботи системи "СЕЛЕКС" - автоматизований облік - ведення "електронних" картотек тварин і журналів, у яких реєструється виконання технологічних операцій (контрольні годівлі, запліднення, перевірки на тільність, запуски й ін.), ветеринарні заходи, рекомендації з коректування режиму змісту й експлуатації тварин.

Відповідно до цієї системи, селекціонери господарств щомісяця передають в ОЦ бланки, у яких зафіксована інформація про отелення і запліднення тварин, контролі молочної продуктивності та т.і.

На основі цієї інформації на ЕОМ складаються щомісячні звіти про продуктивність і відтворну здатність худоби в межах республіки, районів, господарств і ферм. По закінченні господарського року складаються підсумкові показники як по кожній тварині, так і по регіонах. Підсумкові показники про продуктивність корів за 305 днів лактації використовуються для складання звітів щодо бонітування худоби, оцінки бугаїв за якістю нащадків і вирішення інших завдань з племінної роботи.

Крім того, пізніше, у систему «СЕЛЭКС» були додані програми по автоматизації обліку і звітності з ветеринарії і племінного обліку у свинарстві.

У сучасному стані інформаційно-аналітична система обробки даних у молочному скотарстві «СЕЛЕКС» виконує такі завдання:

- оперативне керування виробництвом;
- інформаційне обслуговування селекційно-племінної роботи;
- визначення генетичного потенціалу тварин;
- визначення племінної цінності тварин;
- формування оперативної звітності підприємства;
- формування річної звітності (форма 7-мол) по підприємству;
- аналіз результатів бонітування по підприємству;
- прогноз продуктивності;
- формування племсвідоцтв та племкарток (1-мол, 2-мол);
- формування генотипу молодняка;
- видача на запит фахівців таблиць довільної форми;
- обмін інформацією з базами даних регіонального рівня.

4. Інформаційно-обчислювальна система «Еліта»

ІОС «Еліта» була розроблена Литовським інститутом економіки сільського господарства (В.І. Плунге, В.Ю. Раманаускене, 1980; А.П. Баніс, 1986).

Ця система має такий комплекс завдань:

- облік і оперативний аналіз даних контролю молочної продуктивності худоби, придатності корів до машинного доїння, стану здоров'я, відтворної здатності, росту і розвитку, м'ясних якостей молодняка;
- складання звітів про результати племінної роботи в розрізі всіх рівнів керування;
- генеалогічний аналіз стад і порід;
- оцінка бугаїв за якістю потомства;
- оцінка селекційно-генетичних параметрів та ін.

Збирання інформації здійснюється на підставі надходження з господарств у ОЦ чотирьох бланків: відомості контрольних надоїв, відомості дослідження вимені корів, опису стада корів та відомості змісту жиру і білка в молоці.

Дана система дає можливість одержання інформації з племінної роботи в розрізі господарств, районів і трестів з виділенням племзаводів і племгоспів.

5. Система управління в молочному скотарстві ОРСЕК (СУМС “Орсек-ПП”)

Пакет програм СУМС “Орсек” призначений для оперативного управління виробничим і селекційним процесами в молочному стаді корів.

“Орсек” - це автоматизована селекція з використанням новітніх методів оцінки, відбору і підбору тварин, ефективного відтворення стада і підвищення його продуктивності.

Система забезпечує:

- Оригінальне, зручне і просте введення даних з автоматизованим, максимально можливим контролем як в індивідуальному, так і в груповому режимах.
- Оперативний вихід на будь-яку інформацію про тварину за допомогою системи меню або функціональних клавіш.
- Поєднання режимів перегляду, введення, коректування, видалення і друку даних під час роботи з будь-якою екранною формою.
- Автоматизований розподіл тварин за групами використання і призначення.
- Загальну оцінку тварин за продуктивністю та відтворювальною здатністю.
- Автоматизоване формування родоводів для тварин з визначенням породи, породності, лінії, родини, кровності та надання кличок молодняку.
- Формування звітів про продуктивність та відтворювальну здатність тварин на будь-яку дату року за доярками, фермами, господарствами.
- Автоматизоване проведення бонітування тварин.
- Одержання інформації про тварин за допомогою гнучкої системи запитів користувача.
- Прогноз очікуваної продуктивності корів за поточну лактацію.
- Прогноз надоїв на фуражну корову за групами корів.
- Автоматизовану видачу форм контролю за осіменінням, отеленням, запуском, продуктивністю, оцінкою екстер'єру і властивостями вимені корів.
- Автоматизоване формування і друкування карточки племінної корови і племсвідоцтва.
- Оцінку бугаїв за якістю потомства.
- Групування корів за генеалогією.
- Формування планів підбору тварин.
- Широки сервісні функції.

6. АІС у м'ясному скотарстві

АІС з племінної роботи в м'ясному скотарстві розроблена співробітниками колишніх УкрНДІплем (В.І. Власов та ін.) і Української сільськогосподарської академії (М.Г. Садко та ін.).

Функціональна структура АІС підрозділюється на ряд підсистем, що розраховані на вирішення таких завдань:

- підсистема «Плідник» - для оцінювання і відбору бугайців за походженням; оцінювання і відбору бичків за власною продуктивністю (інтенсивністю росту, живою масою, типом конституції й екстер'єру, відтворними здібностями та ін.);
- підсистема «Корова» – для оцінки і відбору телиць за походженням на основі ІПЦ предків, за власною продуктивністю, інтенсивності росту, живій масі, молочності, відтворним здібностям і ін.;
- підсистема «Бонітер» – для оцінювання і визначення комплексного класу окремої тварини, визначення їхнього призначення, складання генеалогічної структури;
- підсистема «Підбір» – для оцінювання результатів індивідуального підбору з урахуванням ступеня інбридингу, одержання продовжувачів ліній із заданим генотипом, оцінювання ефективності сполучуваності родинних груп і ліній та ін.;
- підсистема «Маркер» – для аналізу генотипів тварин за імуногенетичними маркерами (ІГМ), розрахунків коефіцієнта гомозиготності, ступеня консолідації, диференціації стад і популяцій, у т.ч. для оцінювання ефекту зчеплення ІГМ із племінними і продуктивними якостями тварин;
- підсистема «Ветеринар» – для організації збирання даних про захворюваність тварин; оцінювання плідників за частотою захворювань потомства; оцінювання репродуктивних якостей тварин, для контролю спадкоємних аномалій і т.п.;
- підсистема «Прогноз» – для розроблення оптимальної програми селекції, постійного контролю за генетичною ситуацією в череді;
- підсистема «Керування стадом» – для одержання оперативної інформації при керуванні стадом; використання тварин не основі первинних зоотехнічних даних;
- підсистема «Економіст» – для економічної оцінки прийомів і методів відбору тварин і всієї програми селекції в цілому.

7. АІС у свинарстві

Розроблення АІС у свинарстві займався Полтавський НДІ свинарства і НДІ тваринництва Лісостепу і Полісся України. Ними був розроблений пакет прикладних програм (ППП) зведення на рівні республіки і колишнього СРСР результатів оцінювання кнурів методом контрольної відгодівлі. Цей ППП здатний видавати таку інформацію:

- кількість господарств, що оцінювали своїх кнурів-плідників методом контрольної відгодівлі за роками і регіонами;
- поголів'я кнурів, що оцінювалися, і їхніх нащадків за роками і регіонами;
- середні показники контрольної відгодівлі свиней за роками і регіонами із визначенням поголів'я, віку досягнення 100кг (днів),

середньодобового приросту і витрат кормів (к.од.) на 1кг приросту за період досягнення маси від 30 до 100кг, довжини туші (см), товщини шпику (мм) і маси задньої третини напівтуші (кг) при забої в 100 кг;

- результати контрольної відгодівлі різних порід, груп і типів свиней за роками із вказівкою віку досягнення 100кг (днів), середньодобового приросту і витрат кормів (к.од.) на 1кг приросту за період досягнення ваги від 30 до 100кг, довжини туші (см), товщини шпику (мм) і маси задньої третини напівтуші (кг) при забої в 100 кг та ін.;
- дані про кількість підсвинків на відгодівлі, число і % від загального поголів'я перевірених кнурів по 12 і більше нащадках;
- результати індивідуальної оцінки кнурів за якістю нащадків з видачею на друк даних по кожному кнуру, господарству, чи пункту станції контрольної відгодівлі області й країни в цілому.

8. АІС у птахівництві

У птахівництві, де як ні в якій іншій галузі тваринництва, великий потік інформації, яку необхідно обробити і проаналізувати за короткий термін, АІС досягли широкого застосування.

Так, для селекційно-генетичного аналізу первинних даних зоотехнічного і племінного обліку використовують АІС, що були розроблені фахівцями колишніх ВНІТП, УкрНДІП, Прибалтійської і Казахської зональних дослідними станціями по птахівництву.

АІС «Сигма» була розроблена в середині 80-х років минулого сторіччя в УкрНДІП (І.А. Степаненко).

Дана АІС складається з двох підсистем:

- підсистема «Інформаційне забезпечення» включає форми первинного обліку селекційної інформації яєчних і м'ясних видів птиці, довідники і таблиці з даними про племзаводи, види і породи птиці, описи показників племінного обліку, а також машинні програми введення, контролю, перетворення і накопичення даних на всіх етапах селекційно-племінної роботи з птицею;
- підсистема «Оцінка і генетико-математичний аналіз продуктивності птиці» забезпечує оброблення інформації й оцінює племінну цінність птиці за продуктивністю, відтворними якостями і збереженістю та ін.

Для збирання первинної інформації розроблено п'ять форм обліку для яєчних курей і шість для м'ясних порід птиці.

У цілому, ця АІС забезпечила підвищення оперативності надходження селекційної інформації, значно розширила її інформативність за оцінювання племінної цінності птиці і дозволила вирішувати такі завдання:

- оцінювання птиці за масою яєць (з видачею середньої, її помилки, сигми по кожній несучці);
- оцінювання відтворної здатності за даними інкубації (з видачею кількості виведених курчат, закладених яєць, заплідненості, виведення молодняку,

виводимості яєць за несучкою, гніздом, лінією), а також оцінку плідників за якістю нащадків;

- комплексну оцінку продуктивності і збереженості птиці за рядом господарсько-корисних ознак за будь-які періоди часу;
- підсумкову і помісячну яйцєносність, живу масу, масу яєць, інкубаційні показники;
- оцінки успадкування (за Плохинським і Хейзелем-Лашем), генетичні і фенотипові кореляції і регресії;
- оцінку плідників методом «дочки-ровесниці»;
- оцінку продуктивності збереженості молодняка м'ясних порід птиці за всіма показниками продуктивності й ін.

Питання для самоперевірки:

1. Яке коло задач вирішують АІС у тваринництві?
2. Які основні модулі та функції має АІС «ІНСЕЛ»?
3. Які основні модулі та функції має АІС «СЕЛЕКС»?
4. Які основні модулі та функції має АІС «ЕЛІТА»?
5. Які основні модулі та функції має АІС «ОРСЕК»?
6. Які основні модулі та функції має АІС у м'ясному скотарстві?
7. Які основні модулі та функції має АІС у свинарстві?
8. Які основні модулі та функції має АІС у птахівництві?

Тема №4

Основні завдання інформаційно-обчислювальних систем в селекції

Оцінювання міжстадної генетичної мінливості

Методи оцінювання міжстадної генетичної мінливості засновані на припущенні, що в популяції є стада з різними генотипами, а ідентичні генотипи можуть бути в різних стадах.

Основні причини генетичних відмінностей між стадами – це використання плідників із різною племінною цінністю в окремих стадах, а також внутрішньостадна селекція матерів корів.

Один із методів, що заснований на цьому принципі, запропонований МакДжиллиардом у 1952 році. Його принцип полягає в порівнянні варіанси продуктивності між стадами, оціненої за неспорідненими тваринами, з міжстадною варіансою по споріднених:

$$h_B^2 = \frac{\sigma_{HG}^2}{\sigma_G^2}.$$

Варіансу між стадами (σ_n^2) на матеріалі неспоріднених тварин розраховують по моделі однофакторного дисперсійного аналізу:

продуктивність = середнє для популяції + вплив стада +
+ випадкове відхилення.

Для оцінки варіанси між стадами на матеріалі споріднених тварин (полусибсів за батьком) використовується двофакторний дисперсійний аналіз:

продуктивність = середнє для популяції + вплив плідника + вплив стада в
межах нащадків плідника + випадкове відхилення.

Інший, більш простий метод розрахунку міжстадних генетичних змін, був запропонований і обґрунтований Робертсоном і Ренделем у 1954 р.

Він пов'язаний з оцінкою регресії продуктивності дочок плідника на рівень продуктивності відповідних стад:

$$h_B^2 = 2(1 - b).$$

Оцінювання бугаїв за якістю нащадків модифікованим методом “дочки-ровесниці”

Основою будь-якої системи селекції є оцінка племінної цінності потенційних батьків. Мета цієї оцінки - одержати якомога точніший прогноз генетичної цінності особи. Чим достовірніша генетична оцінка, тим суворіший відбір на підставі цієї оцінки, і чим інтенсивніше використання генетично кращої тварини, тим більше буде впевненості у селекціонерів у поліпшенні тварин від покоління до покоління.

Племінну цінність розраховують тільки для бугаїв і розрахунок її можна провести методом **порівняння із ровесницями**, запропонованим Альтшулером і Сухановим у 1936 році.

При використанні цього методу продуктивність дочок бугая порівнюється із продуктивністю ровесниць, які мали лактації в аналогічних умовах.

Згідно з цим методом, дані по кожній корові активної частини популяції можна записати так:

$$y_{ijkl} = \mu + 1/2BVg + 1/2BVgi + 1/4BVPgi + 1/2BVDgijk + \\ + 1/4 BVDgijk + Hk + e_{ijkl}$$

де y_{ijkl} – скорегована на систематичні фактори зовнішнього середовища (окрім впливу стада) продуктивність l-ої первістки у k-тому стаді, дочки j-тої матері та i-того батька, що відноситься до групи бугаїв g-того року народження й генотипу (походження); μ - середня продуктивність популяції; BVg – вплив групи; $BVgi$ – вплив племінної цінності батька; $BVPgi$ – вплив племінної цінності родичів батька; $BVDgijk$ – вплив племінної цінності матері, яка була запліднена спермою бугая відповідного генотипу; $BVDgijk$ – вплив племінної цінності діда материнської лінії; Hk – вплив стада; e_{ijkl} – залишковий вплив.

Елімінація впливу стада за оцінювання племінної цінності плідника досягається шляхом порівняння продуктивності його дочок із продуктивністю їх ровесниць.

Цей метод легко застосовується на практиці, але має ряд недоліків.

Зокрема, він ефективний і забезпечує безпомилкові оцінки племінної цінності бугаїв тільки у разі:

- якщо середня генетична цінність батьків перевесниць однакова для дочок усіх оцінюваних бугаїв;
- усі оцінювані бугаї відбираються з однієї популяції (закрита система розведення);
- у популяції немає генетичного тренда, тобто ефект селекції дорівнює нулю.

Оцінювання бугаїв-плідників методом BLUP

Системи управління базами даних (СУБД) дозволяють маніпулювати зібраними даними і переробити їх у необхідну для оцінки і прогнозу інформацію. Для аналізу таких даних використовуються статистичні моделі змішаного типу (ММЕ). Ці моделі містять у собі фіксовані і випадкові (рандомізовані) фактори. Результатом вирішення моделі із змішаними факторами (ММЕ) є **найкращий незміщений лінійний прогноз (BLUP; Best Linear Unbiased Prediction)** для рандомізованих, і **найкраща незміщена лінійна оцінка (BLUE; Best Linear Unbiased Estimation)** – для фіксованих факторів.

Перший показує імовірність прояву того чи іншого випадкового фактора, а друга - важливість обліку і силу дії того чи іншого фіксованого фактора. Ці методи розроблені американським вченим Ч. Хендерсоном у 1974 р.

Для досягнення найбільш точного прогнозу слід враховувати такі вимоги:

- залучити всі наявні дані;

- визначити, оцінити й усунути з оцінки значимі чи найбільш значимі фактори навколишнього зовнішнього середовища;
- використовувати статистичні методи, що забезпечують незміщений прогноз генотипу.

Дослідженнями було встановлено, що генетична оцінка молочної худоби за методом BLUP сприяє підвищенню ідентифікації генотипу на 12-40% і більше.

Статистичні моделі, що використовуються для методу BLUP, складаються для кожної популяції індивідуально, однак загалом цю модель можна описати.

Наприклад, надій дочки бугая це:

$$\text{надій} = \text{середнє для стаду} + \text{ефект року} + \text{сезону отелення} + \text{ефект бугая} + \text{інші невраховані фактори.}$$

Модель - це рівняння, що показує, як незалежні перемінні (стадо, рік, сезон отелення) впливають на залежну перемінну - ознаку (надій, % жиру). Вона необхідна для того, щоб описати фактичну ситуацію в популяції, тобто, якомога повніше і точніше показати фактори, що впливають на продуктивність тварини.

Вирішення статистичної моделі забезпечує найкращий незміщений лінійний прогноз 1/2 адитивної генетичної цінності бика, що не залежить від впливу змісту і годівлі дочок на фермі, року і сезону отелення, а також усіх взаємодій між цими факторами.

У даний час реальна генетична цінність особин не може бути оцінена з 100% точністю, тому безпосередньо виміряти ефективність того чи іншого методу неможливо. Можна лише порівняти різні методи з тим, що споконвіку вважався більш точним. З погляду математики, статистики і селекції метод BLUP вважається найбільш обґрунтованим і має під собою кращу теоретичну базу. Оцінки племінної цінності, розраховані за методом BLUP, мають мінімальну дисперсію помилки. Тому можна вважати, що, при інших рівних умовах, оцінки племінної цінності за BLUP будуть з найбільшою імовірністю відбивати реальну генетичну цінність особини.

Оцінювання генетичних змін за показниками молочної продуктивності

Існує кілька методів оцінки генетичних змін у популяціях с/г тварин:

- створення подібних умов середовища для експериментальної і контрольної популяцій;
- порівняння експериментальної популяції з тією, що вільно розмножується (контрольною) популяцією;
- селекція експериментальної популяції в двох напрямках (плюс- і мінус-варіанти);
- метод однойцевих близнюків та ін.

Однак майже всі вони придатні для лабораторних тварин.

1. Найбільш придатний метод оцінки генетичних змін у популяціях с/г тварин запропонував С. Сміт (1962; 1965). Він заснований на тому, що племінна цінність тварини не змінюється в часі, тобто сперма бугая, яка зберігається протягом багатьох років, є носієм постійного генотипу бугая-плідника. Тому зміни в продуктивності його дочок, що народилися в різні періоди часу, і будуть характеризувати генетичні зміни в популяції.

На підставі даного принципу Сміт запропонував таку формулу:

$$\Delta q = -2b_{(S-P) \cdot T},$$

де Δq - генетичне зрушення продуктивності за одиницю часу, $b_{(S-P) \cdot T}$ – зважена за плідниках регресія відхилень продуктивності дочок (S) від перевесниць (P) на час (T).

У нашій країні генетичні зміни в молочному скотарстві оцінювали за допомогою модифікованого методу Сміта.

2. Генетичні зміни можуть бути також оцінені на підставі порівняння нащадків-однолітків двох різних поколінь бугаїв-плідників.

3. В останні роки як у нашій країні, так і за кордоном для оцінки генетичних змін широко використовується метод BLUP, при якому після корекції даних продуктивності дочок на вплив середовищних факторів, оцінюють регресію різниці між показниками дочок і ровесниць на рік народження бугаїв.

Оцінка тварин за екстер'єром і типам конституції

У зв'язку з впровадженням у практику промислової технології підвищилися вимоги до екстер'єру і конституції тварин, особливо до якості вим'я і кінцівок тварин.

В останні роки в США, Канаді, Німеччині, Швеції й ін. країнах у національні програми селекції введена лінійна система оцінювання типу екстер'єру. Було розроблено методика оцінювання типу за статями. Як еталон породи селекціонери вибрали той тип тварин, що забезпечує високу продуктивність, збереження здоров'я, відтворної функції й інтенсивне використання.

Кожна ознака (наприклад, у США прийнято 13, а в Німеччині 15 статей екстер'єру) оцінюється за біологічними крайностями; наприклад, у шкалі прикріплення вимені поряд із середніми значеннями даються дві крайні оцінки – високе і низьке. Така біологічна оцінка окремих статей екстер'єру дозволяє найбільш точно оцінити ступінь їх успадкування, а також на основі порівняння з ровесницями визначити племінну цінність корів за власними показниками.

Дослідження американських і канадських учених показали, що генетичний прогрес за типом тварин у два рази вищий, ніж за молочною продуктивністю.

В Україні також була розроблена методика оцінки молочних порід за типами екстер'єру, відповідно до якої комплексну оцінку статури і морфо-функціональних особливостей худоби проведено за 100-бальною шкалою.

При цьому проводиться оцінка 10 основних статей, що характеризують екстер'єр, конституцію і морфо-функціональні властивості організму.

Підвищення ефективності селекції за рахунок збільшення точності оцінки племінних якостей худоби

Ефективність селекції залежить від трьох основних факторів:

- генетичної мінливості в популяції;
- точності оцінки генотипу тварин;
- інтенсивності добору і використання племінних тварин (насамперед, бугаїв-плідників).

У будь-якій популяції (породі) ступінь генетичної мінливості перебуває приблизно в однакових межах, тому першим фактором можна зневажити.

Другий фактор – точність оцінки генотипу тварин – залежить від точності племінного і зоотехнічного обліку, умов середовища, числа ознак, що піддаються селекції, обсягу вибірки й ін. Якщо названі фактори мають оптимальні значення, то точність оцінки генотипу залежатиме лише від методів оцінки племінних якостей тварин.

При цьому слід враховувати таке:

- оцінку бугаїв за якістю нащадків необхідно планувати так, щоб у короткий термін (1-2 року) на великому поголів'ї дочок (не менше 100) і в досить великій кількості стад одержати достовірні дані про їхню племінну цінність;
- після оцінки за якістю нащадків сперму бугаїв варто використовувати для запліднення корів протягом одного-двох років, тому що під впливом генетичного тренда з кожним роком його генетична перевага в породі знижується;
- за відбору батьків бугаїв слід використовувати велику інформацію про їхню племінну цінність, отриману в різних стадах активної частини популяції.

Корекція племінної цінності тварин на вплив генетичного тренда особливо велике значення має при складанні плану замовленого спарювання для одержання ремонтних бугайців. Така корекція дозволяє точніше прогнозувати результати відбору, підбору і тим самим підвищувати його ефективність.

Точність оцінки племінних якостей бугаїв залежить також від кількості стад, у яких вони оцінюються, а також від рівня їхньої продуктивності і племінної цінності. Установлено, що бугаї з високим генетичним потенціалом з поліпшенням умов годівлі і утримання їхніх дочок підвищують свою перевагу, а бугаї з низьким генетичним потенціалом – навпаки, в кращих стадах показують гірші результати.

Оцінка генетичних розходжень між племінною цінністю биків різних ліній і їхніх кросів

Як відомо, у молочному скотарстві генеалогічна структура породи будується на основі ліній, що розводять у племінних заводах. Бугаї, виведені в заводських лініях, надходять на станції штучного запліднення, за допомогою яких і формується генеалогічна структура породи. Від того, яким методом і якою якістю отримані бугаї в племінних заводах, буде залежати результат їхнього використання в цілому по породі.

Аналіз генеалогічної структури популяцій ряду порід молочної худоби (чорно-рябої, симентальської й ін.) показав, що породи складаються з численних ліній і родинних груп (20-25), що дуже відрізняються по величині поголів'я. У кожному племзаводі також розводять велику кількість ліній (4-6), що приводить, відповідно, до використання безлічі бугаїв-плідників (20-30 голів на 500-1000 корів).

Ряд учених і практиків зазначають, що наявність великої кількості ліній у племзаводах обумовлене пошуком удалих сполучень (кросів).

Інші, навпаки, вважають, що в молочному скотарстві при чистопорідному розведенні сполучуваність ліній не має практичного значення.

Оцінювання генетичної мінливості в стадах і популяціях

Оцінювання результатів племінної роботи, прогноз її ефективності, визначення племінної цінності тварин і моделювання селекційних програм можна зробити лише на основі адекватних оцінок рівня генетичної мінливості в стадах і популяціях.

Серед популяційно-генетичних параметрів для теорії і практики племінної роботи найбільше значення має **коефіцієнт успадкування**. Він указує на ступінь генетичної мінливості в загальному фенотиповому розмаїтті популяції за будь-якою господарсько-корисною ознакою.

Найбільш поширеними методами оцінки коефіцієнта успадкування є:

$$h^2 = 2r_{PO},$$

$$h^2 = 2R_{PO}.$$

Відповідно до теорії шляхових коефіцієнтів С. Райта, вони засновані на подвоєнні кореляції (регресії) «нащадки-батьки».

Використання цих формул без урахування впливу середовищних факторів часто приводять до перевернутих оцінок, що далеко виходять за рамки біологічної норми.

Для оцінювання успадкування господарсько-корисних ознак у популяціях молочної худоби найбільш ефективним є метод:

$$h^2 = 4r,$$

де r – коефіцієнт внутрішньокласової кореляції:

$$r = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_C^2 + \sigma_U^2},$$

де перший, другий і третій доданок у знаменнику – це варіанса між групами батьківських напівсибсів, варіанса впливу факторів середовища і випадкова варіанса, відповідно.

Тоді, оцінка коефіцієнта успадкування (після елімінації впливу середовища) може бути отримана за формулою:

$$h^2 = 4 \left[\frac{\sigma_A^2}{\sigma_A^2 + \sigma_U^2} \right].$$

Вірогідність оцінки коефіцієнта успадкування і племінної цінності тварин підвищується, якщо контрольована ознака має кілька вимірів. Наприклад, якщо продуктивність корів враховується за m лактаціями, те успадкованість при цьому визначається за формулою:

$$h_m^2 = \frac{mh^2}{1 + (m-1)t},$$

де t – коефіцієнт повторюваності ознаки.

Оцінка взаємодії між генотипом і середовищем у молочному скотарстві

Під взаємодією «генотип – середовище» розуміють зміну рангів бугаїв-плідників, тобто одні тварини показують кращі результати в будь-яких умовах середовища, тоді як інші – навпаки, в одних умовах вони гірші, а в інших – кращі. Коли така взаємодія існує, то методи оцінки племінної цінності плідників і програма їхнього використання повинна бути змінена.

Основними методами вивчення взаємодії “генотип – середовище” є методи, засновані на дисперсійному і регресійному аналізах.

Коли між генотипом і середовищем має місце залежність, фенотипова оцінка особини стає не просто:

$$P = G + E,$$

а вже буде включати компоненту взаємодії:

$$P = G + E + I_{GE}.$$

А. Харві (1964) зазначав, що варіансу взаємодії можна оцінити в перехресно-класифікованих моделях гніздового (ієрархічного) аналізу методом найменших квадратів.

Оцінку компонентів варіанси взаємодії між плідником і стадом, що використовують як міру взаємодії генотипу і середовища, можна отримати на підставі варіанси плідників у окремому стаді і генотипової варіанси за плідниками.

Для цього використовується модель:

$$Y = \mu + S + H + e,$$

де μ – середнє значення для популяції, S – випадковий ефект плідника; H – випадковий ефект стада; e – інші випадкові ефекти.

Питання для самоперевірки:

1. Які методи використовують для оцінювання міжстадної мінливості?
2. Як відбувається оцінювання бугаїв-плідників за якістю нащадків модифікованим способом «дочки-ровесниці»?
3. Як відбувається оцінювання бугаїв-плідників на підставі методу BLUP?
4. Як можна підвищити ефективність селекції за рахунок збільшення точності оцінювання племінних якостей худоби?
5. Як провести оцінювання генетичної мінливості в стадах і популяціях?
6. Як оцінити взаємодії між генотипом і середовищем у молочному скотарстві?

Тема №5

Методика математико-статистичного аналізу імуногенетичних даних для великої рогатої худоби

Імуногенетичні дані, отримані після проведення серологічних досліджень для великої рогатої худоби, мають деякі особливості, що інколи утрудняють їх аналіз із популяційно-генетичної точки зору. Останній підхід потребує такої інформації:

- оцінок частот фенотипів, генотипів та алелей за певними локусами (тобто, для відповідних еритроцитарних антигенів груп крові);
- оцінок рівня генного різноманіття у найширшому розумінні (наприклад, за М.Неєм);
- визначення ступеня генетичного інбридингу;
- оцінки рівня генетичної диференціації між групою популяцій або інших зоотехнічних одиниць;
- установа рівня генетичної подібності між групами популяцій або інших зоотехнічних одиниць (наприклад, за М.Неєм, Л.Животовським та ін.);
- та ін.

На відміну від результатів аналізу білкового поліморфізму крові чи молока, або молекулярно-генетичних маркерів із кодомінантним типом успадкування (наприклад, мікросателітів, тощо), при проведенні імуногенетичного аналізу великої рогатої худоби дуже важко отримати повну генотипову формулу у вигляді A/B , де A та B – різні алельні форми для певного локуса. Для цього необхідно мати результати не лише досліджуваної тварини (пробанда), а його матері та батька. Такий повний генетичний аналіз часто-густо не проводиться й тому результатом імуногенетичної експертизи тварини здебільшого є лише її **фенотип**, тобто формульний запис, який складається із переліку антигенів крові, що були відмічені при серологічному аналізі певної тварини.

Другою важливою особливістю імуногенетичного аналізу великої рогатої худоби є дуже висока кількість алелей, зареєстрованих для деяких антигенних систем; наприклад, за 12 генетичними системами антигенів великої рогатої худоби на теперішній час виділено понад 100 антигенів, у яких враховано більше, ніж 500 алелей (більшість яких входять до систем B та C). Наявність великого алельного різноманіття, можливість прояву у особини декількох алелей за однією системою призводить до ситуації, коли в популяції неспоріднених тварин кожна зареєстрована алель (насамперед, за системою B) є унікальною, що унеможливує проведення популяційно-генетичного аналізу за тим планом, що викладеним вище.

Тому нами було запропоновано новий, більш адекватний і технічно спрощений, підхід до аналізу імуногенетичної мінливості великої рогатої худоби.

У цілому нами під час аналізу було виявлено наявність 53-х антигенів (табл. 5.1). У такому разі всю імуногенетичну формулу повністю (тобто, коли зустрічаються всі 53 антигени) можна записати так:

$$A_1 A_2 Z' B_2 G_2 G_3 K I_1 I_2 O_1 O_2 P_2 Q T_1 T_2 Y_2 A'_1 A'_2 B' D' E'_2 \dots U U' H'' U''.$$

Ми пропонуємо перейти від такої загальноприйнятої форми запису імуногенетичного фенотипу будь-якої тварини до розгляду його у вигляді складного бінарного запису, де для кожної алелі, що було відмічено у групі тварин, яка аналізується, необхідно ставити символ “1”, якщо ця алель серологічно фіксується, або символ “0”, якщо аглютинація не мала місця. В цьому разі, наприклад, для особини, що має за системою *A* лише одну алель – *A*₂, формула буде такої: 0 1 0 ... Аналогічно кодується наявність/відсутність кожної із 53 антигенів (алелей), що зареєстровані у аналізованих тварини.

Таблиця 5.1

Перелік антигенів великої рогатої худоби, виявлених під час дослідження

Системи								
A	B	C	F	J	L	M	SU	Z
<i>A</i> ₁	<i>B</i> ₂	<i>C</i> ₁	<i>F</i>	<i>J</i>	<i>L</i>	<i>M</i>	<i>S</i> ₁	<i>Z</i>
<i>A</i> ₂	<i>G</i> ₂	<i>C</i> ₂	<i>V</i>				<i>H'</i>	
<i>Z'</i>	<i>G</i> ₃	<i>E</i>					<i>U</i>	
	<i>K</i>	<i>R</i> ₁					<i>U'</i>	
	<i>I</i> ₁	<i>R</i> ₂					<i>H''</i>	
	<i>I</i> ₂	<i>W</i>					<i>U''</i>	
	<i>O</i> ₁	<i>X</i> ₁						
	<i>O</i> ₂	<i>X</i> ₂						
	<i>P</i> ₂	<i>C'</i>						
	<i>Q</i>	<i>L'</i>						
	<i>T</i> ₁							
	<i>T</i> ₂							
	<i>Y</i> ₂							
	<i>A'</i> ₁							
	<i>A'</i> ₂							
	<i>B'</i>							
	<i>D'</i>							
	<i>E'</i> ₂							
	<i>G'</i>							
	<i>I'</i>							
	<i>K'</i>							
	<i>J'</i> ₂							
	<i>O'</i>							
	<i>P'</i>							
	<i>Q'</i>							
	<i>Y'</i>							
	<i>B''</i>							
	<i>G''</i>							

У цьому разі загальна матриця із вихідними даними при запису їх у табличному редакторі MS Excel буде мати такий вигляд, як показано на рис. 5.1.

З одного боку така форма запису уніфікує дані, отримані для різних тварин, а з іншого – вже дозволяє провести розрахунки популяційно-генетичних параметрів, що були висвітлені вище.

У такому вигляді (набір одиниць чи нулів) дана форма запису імуногенетичних досліджень певної тварини може бути визначена як її **гаплотип**. Майже аналогічний підхід до визначення гаплотипного різноманіття використовується й у разі аналізу структури послідовності ДНК за участю певного набору ендонуклеаз рестрикції. В більш широкому розумінні гаплотипом називають комбінацію алелей тісно зчеплених локусів.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG						
1		Система А																Система В																					
2		A1	A2	Z'	B2	G2	G3	K	I	I2	O1	O2	P2	Q	T1	T2	Y2	A'1	A'2	B'	D'	E'2	G'	I'	K'	J'2	O'	P'	Q'	Y'	B''	G''	C1						
3																																							
4	кличка и номер																																						
5	актриса	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
6	русалка	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
7	добрая	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
8	роботка	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	палатка	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	десятка	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	лаванда	1	1	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	касатка	1	1	0	1	1	1	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	разлука	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	балерина	0	0	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	кукла	0	0	0	1	1	1	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	веха	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	нимфа	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	рассада	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	красуня	1	1	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	рыбка	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	василек	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	иволга	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	река	0	0	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	арка	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	лирика	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	зухра	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	лиска	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис. 5.1. Зовнішній вигляд електронного аркуша MS Excel із результатами імуногенетичного аналізу дослідної популяції великої рогатої худоби

Така складна бінарна форма запису вихідних даних має ряд переваг.

1. Вона відображає увесь набір антигенів, що були встановлені для певної тварини, одночасно, тобто є комплексною, що має важливу перевагу при аналізі генетичної структури популяції чи групи популяцій. При цьому враховується незалежне успадкування кожного антигенного фактора, а також незалежне успадкування самих систем антигенів.

2. Вона відмічає наявність того чи іншого антигена, хоча при цьому важко визначити – чи є певна тварина гомозиготною або гетерозиготною за

даним геном (антигеном), оскільки немає інформації про генотип її батьків. Але ми можемо використати навіть таку неповну інформацію, якщо врахуватимемо, що система успадкування кожного алелі є домінантна. У цьому разі, тварини із відсутніми тими чи іншими антигенами можна розглядати як рецесивні гомозиготи і на підставі їх частоти визначити частоту певної алелі.

3. Вона дозволяє проаналізувати й характер зчеплення груп антигенів, формуючих алелі. При цьому для матриці, подібної до тієї, що наведена на рис. 5.1, можна проводити два різних типи аналізів. R-аналіз дозволяє проаналізувати розподіл окремих тварин (чи їх груп) у багатовимірному просторі виявлених антигенів (чи їх груп) на підставі матриці їх подібності, розрахованої за допомогою міри Хеммінга. З іншого боку, Q-аналіз дозволяє на підставі матриці подібності між окремими антигенами у різних тварин проаналізувати ступінь зчепленого успадкування між антигенами та встановити вклад яких антигенів (чи їх груп) більш вагомий при диференціації тварин різного походження.

4. За появи нового антигенного фактора, його дописують у кінці ряду антигенів відповідної системи в повній бінарній формулі, а для попередніх тварин та їх груп у цьому стовпчику матриці проставляють нулі.

5. Нарешті, такий підхід дозволяє значно автоматизувати як процес накопичення та збереження імуногенетичної інформації (наприклад, за допомогою редактора електронних таблиць MS Excel), а також використовувати сучасне програмне забезпечення для проведення популяційно-генетичного аналізу імуногенетичних даних.

Розрахунок популяційно-генетичних параметрів для імуногенетичних даних з використанням програми GenAIEx

Усі розрахунки ми проводили з використанням програми “Генетичний аналіз і Excel’е” (GenAIEx v. 6.0), що була розроблена австралійськими вченими-генетиками Р.Піколлом та П.Смаусе у 2006 році (Peakall, Smouse, 2006). Ця програма має декілька переваг. По-перше, вона являє собою програму типу Add-In, тобто, вбудовується в табличний редактор MS Excel й використовує дані, що набрані у цьому редакторі. А по-друге, ця програма поширюється безкоштовно і її можна вільно брати в Інтернеті на сайті авторів (www.anu.edu.au/BoZo/GenAIEx). Нарешті, по-третє, вона здатна розрахувати популяційно-генетичні параметри, що нас цікавлять, на підставі бази даних щодо імуногенетичних даних, яка записана у бінарній формі як, наприклад, наведено на рис. 5.1.

Але для проведення необхідних розрахунків, по-перше, необхідно встановити програму. Для цього слід відкрити нову сторінку MS Excel. У головному Меню знайти опцію “Сервіс” й вибрати процедуру “Надстройки” (рис. 5.2).

Далі необхідно натиснути кнопку “Обзор”. Знайти папку “GenAIEx_6”, відкрити її, вибрати файл **GenAIEx6.xla** та клацнути кнопку **ОК** (рис. 5.3).

Програма GenAIEx 6 з'явиться у переліку Надбудов на вашому комп'ютері. У віконці напроти цієї програми повинна стояти позначка. Для того, щоб її запустити, слід натиснути на кнопку **OK** (рис. 5.4). Тоді на у переліку головного Меню з'явиться нова опція “**GenAIEx**”, а програма почне працювати, внаслідок чого на фоні аркуша MS Excel з'явиться її сторінка (рис. 5.5), яка зникне через декілька секунд.

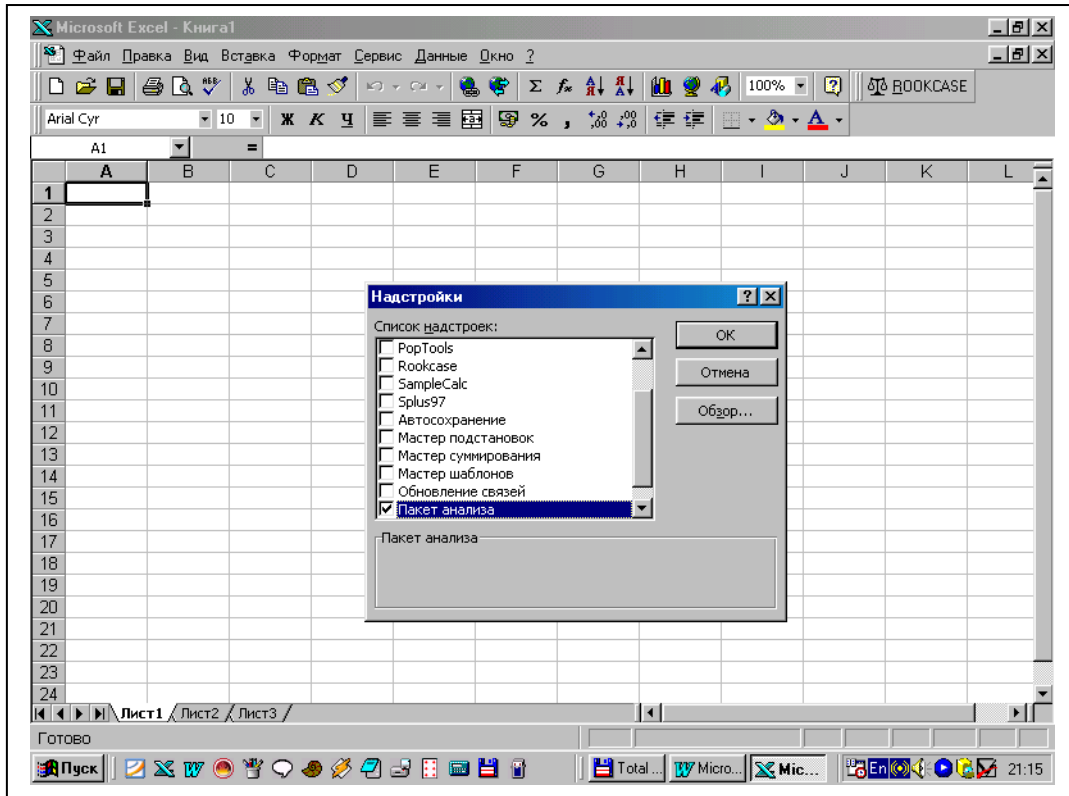


Рис. 5.2.

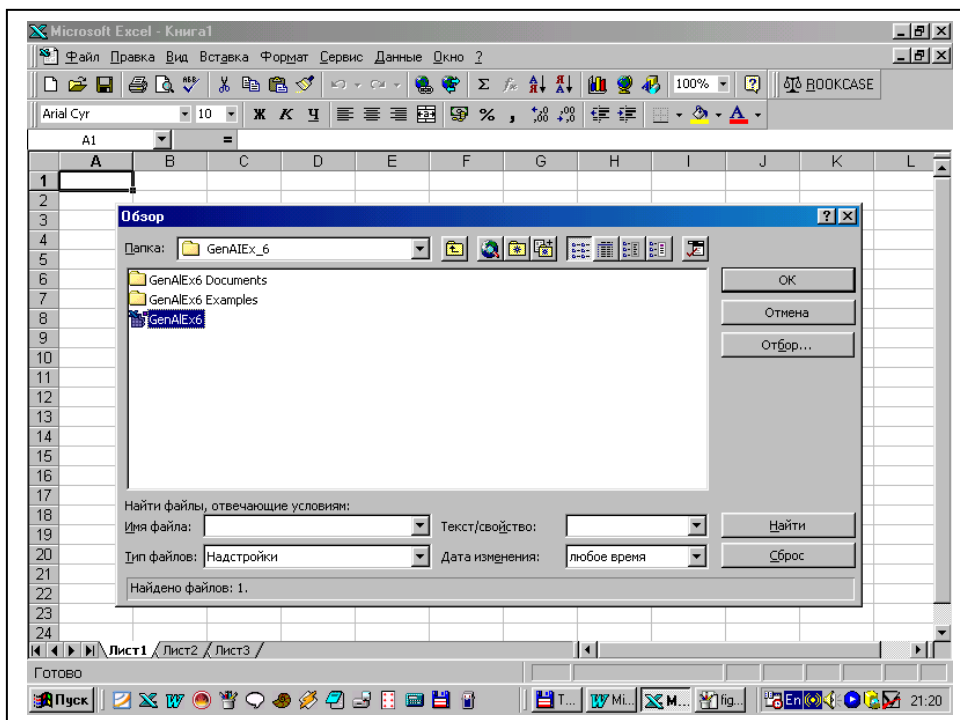


Рис. 5.3.

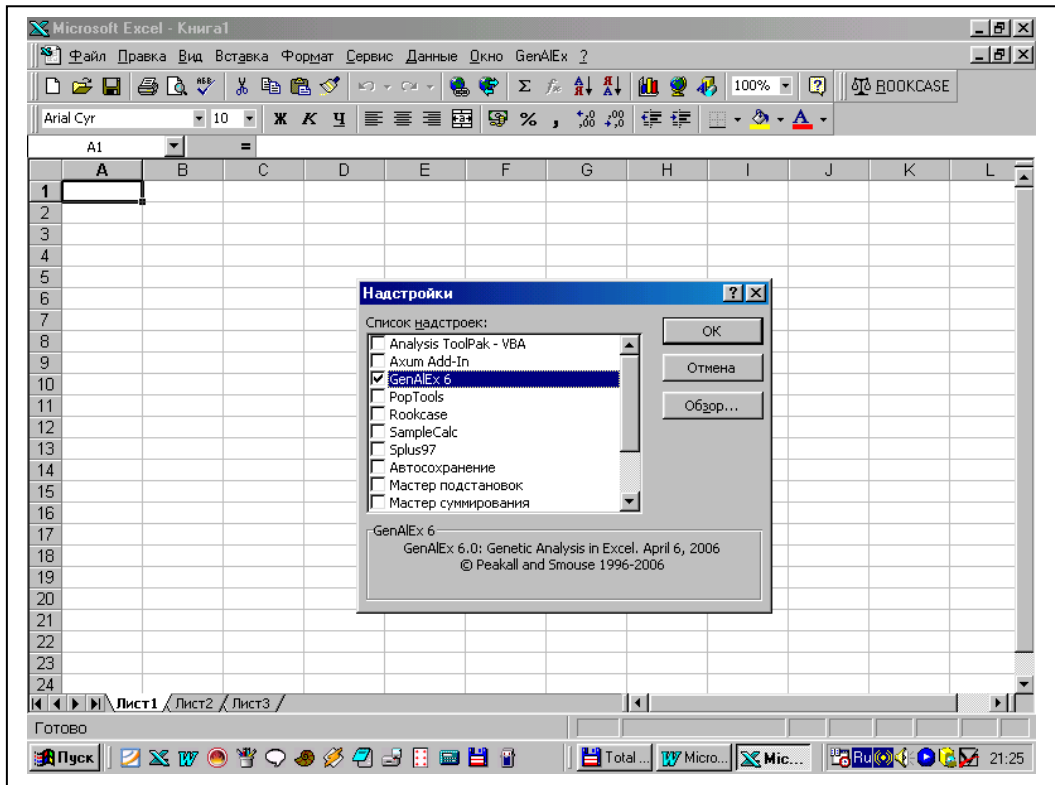


Рис. 5.4.

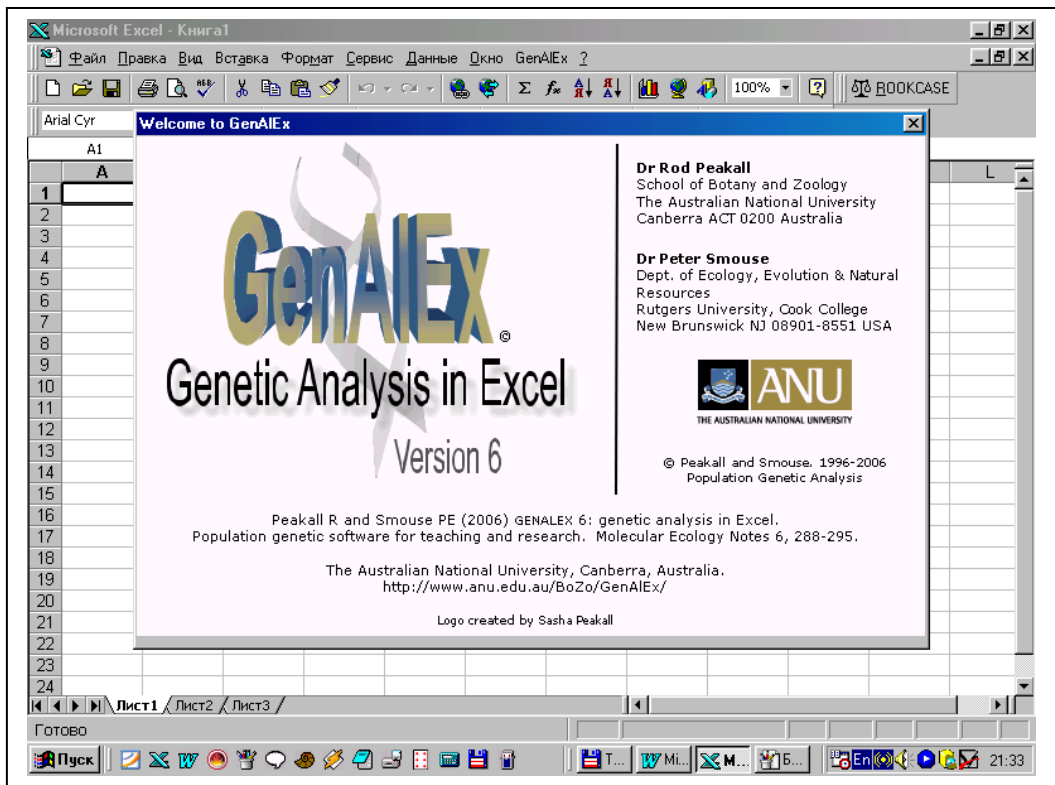


Рис. 5.5.

Далі необхідно сформувати аркуш з вихідними даними так, щоб їх могла прочитати програма. У загальному випадку на такому аркуші повинні бути такі елементи (рис. 5.6):

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	53	120	5	10	7	28	28	47	1	120					
2	Gill_Data			Pop1	Pop2	Pop3	Pop4	Pop5		Region1					
3	Sample	Pop	Locus1	Locus2	Locus3	Locus4	Locus5	Locus6	Locus7	Locus8	Locus9	Locus10	Locus11	Locus12	Locus13
4	1	Pop1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	2	Pop1	2	2	1	2	1	1	1	1	1	1	1	1	2
6	3	Pop1	1	1	1	2	1	1	1	1	1	1	1	1	2
7	4	Pop1	2	2	1	1	1	1	1	1	1	2	1	1	1
8	5	Pop1	2	2	1	1	1	1	1	1	1	1	1	1	1
9	6	Pop1	2	2	1	2	1	1	1	1	1	1	1	1	2
10	7	Pop1	2	2	1	2	2	2	2	2	1	1	1	1	1
11	8	Pop1	2	2	1	1	1	2	1	1	1	2	2	2	1
12	9	Pop1	1	1	1	1	1	2	1	1	1	2	2	2	1
13	10	Pop1	1	1	1	1	1	1	1	1	1	1	1	2	1
14	11	Pop2	2	2	1	1	1	1	1	1	1	2	2	2	1
15	12	Pop2	1	1	1	2	2	2	2	1	1	1	1	1	1
16	13	Pop2	2	2	1	2	1	1	1	1	1	2	2	2	2
17	14	Pop2	2	2	1	2	1	1	1	1	1	1	1	1	1
18	15	Pop2	1	1	1	2	1	1	1	1	1	1	1	1	2
19	16	Pop2	1	1	1	2	2	2	2	2	1	2	2	2	1
20	17	Pop2	1	1	1	2	1	1	1	1	1	2	2	2	1
21	18	Pop3	1	1	1	2	1	1	1	1	1	1	1	1	2
22	19	Pop3	2	2	1	2	1	1	1	1	1	1	1	1	1
23	20	Pop3	2	2	1	2	1	1	1	1	1	1	1	1	2
24	21	Pop3	2	2	1	1	1	1	1	1	1	1	1	1	1
25	22	Pop3	2	2	1	2	1	1	1	1	1	1	1	1	2
26	23	Pop3	2	2	1	2	2	2	1	1	1	1	1	1	1
27	24	Pop3	2	2	1	2	1	1	1	1	1	2	2	2	1
28	25	Pop3	2	2	1	2	2	2	1	1	1	2	2	2	2
29	26	Pop3	2	2	1	2	1	1	1	1	1	1	1	1	1

Рис. 5.6.

Клітка аркуша	Інформація
A1	- кількість локусів, що включені до аналізу (у нашому випадку – загальна кількість еритроцитарних антигенів: 53), тобто, загальна кількість стовпчиків матриці вихідних даних;
B1	- загальна кількість особин, що використовується при аналізі, тобто, загальна кількість рядків матриці вихідних даних: 120;
C1	- кількість груп (популяцій), що використовується в аналізі;
D1 – H1	- чисельність кожної групи (в нашому випадку їх п'ять, тому використано п'ять клітинок; зрозуміло, якщо груп більше, то використовуються додатково клітки праворуч);
I1	- кількість сукупностей груп (в програмній мові – регіонів); в загальному випадку – це наступна клітина після останньої, що містить об'єм останньої групи;
J1	- загальна кількість особин в кожній із сукупностей груп; в

	нашому випадку таких сукупностей одна, тому її чисельність дорівнює загальній чисельності досліджених тварин; але, якщо б таких сукупностей було більше – їх чисельності необхідно було б занести послідовно в клітини праворуч;
A2	- пояснення для набору вихідних даних (у нашому випадку – це дані, що використані в дослідженнях М.І.Гиль (2008), тому ми дали назву – Gill_Data);
D2 – H2	- імена для груп тварин (згідно вимог програми їм надаються такі імена – Pop1, Pop2, Pop3 і т.д.); якщо груп тварин (популяцій) більше, то використовуються наступні клітки праворуч;
J2	- ім'я для сукупності груп (згідно вимог програми для таких сукупностей прийнято ім'я Region); якщо таких сукупностей груп більше, то використовуються наступні клітки праворуч;
A3	- заноситься слово “Sample”, тобто “проба” англійською;
A4 – A123	- послідовно вносяться номери тварин (для нашого випадку від 1 до 120);
B3	- вноситься слово “Pop”, тобто скорочення від “популяція” англійською;
B4 – B123	- послідовно вносяться кодові позначення за належністю кожної тварини до її групи (для нашого випадку використовуються позначення Pop1, Pop2, ..., Pop5); таким чином зрозуміло, що тварина із порядковим номером, наприклад, №22 відноситься до третьої групи (має код Pop3) і т.д.
C3 – C3	- ім'я для кожного локусу (антигену); згідно вимог програми використовуються позначення Locus1, Locus2, ..., Locus53 (у нашому випадку використовується 53 антигени);
C4 – C3123	- код для результатів серологічної реакції для кожної тварини по кожному антигену: у разі, коли реакція негативна, тобто, антиген відсутній – заносимо цифру “1”, а коли реакція позитивна, тобто, антиген присутній – заносимо цифру “2” (згідно вимог програми цифра “0” означає відсутні або пропущені дані); таку матрицю з даними легко можна отримати із матриці, що наведена на рис. 5.1, додавши до кожного значення одиницю.

Після цього аркуш, що містить готову для аналізу вибірку даних, слід перейменувати й дати йому ім'я “D”, від англійського “Data” – данні.

Тепер наші імунологічні дані повністю готові до популяційно-генетичного аналізу.

Питання для самоперевірки:

1. Які особливості аналізу імуногенетичних даних?
2. Які основні етапи інсталяції ППП GenAIEx?
3. Які особливості формату має файл із даними для ППП GenAIEx?

Тема №6

Автоматизація аналізу варіаційного ряду

Автоматизований аналіз варіаційного ряду має такі етапи.

1. Відкрити файл “**Basa-data1.xls**”. Запам’ятати його під новим ім’ям. Він містить дані первинного зоотехнічного обліку для 390 голів корів, а також показники продуктивності предків (рис. 6.1).

	А	В	С	Е	Ф	Н	Ж	К	М	О	Р	Р	
1	Номер	Кличка Б	МАТЕРІВ			МАТЕРІВ		МАТЕРІВ	МАТЕРІВ			БАТЬКІВ	КІЛ, ЛАКТ
2			ЖИР		НАДІЙ	ЖИР		НАДІЙ	ЖИР		КІЛ, ЛАКТ		
3			НАДІЙ	%		КГ	%		КГ	%		КГ	
4	12	левкой	7436	3,77	280	5034	3,55	179	6203	4,28	265	4	
5	14	мінус	5426	4,04	219	5542	3,7	205	4725	4,8	227	7	
6	15	луч	7620	3,72	283	8366	3,47	290	6303	4,03	254	3	
7	17	левак	7039	4,05	285	7110	3,6	256	6311	4,17	263	5	
8	23	закон	6555	4,21	213	6155	4,02	247	6664	3,95	263	6	
9	24	мінус	6205	3,9	242	6063	3,92	238	4725	4,8	227	10	
10	32	бостон	5406	4	216	6050	3,49	211	6640	4,08	271	7	
11	34	козирь	8657	3,88	336	6975	3,93	274	5937	4,14	246	6	
12	36	ян	5730	4,03	231	5640	4,03	227	6798	4,07	277	6	
13	42	мінус	5106	3,99	204	5677	3,58	203	4725	4,8	227	8	
14	62	роджер	5819	3,88	226	5656	3,94	223	5865	4,97	291	5	
15	67	мінус	7284	3,72	271	7210	3,58	258	4725	4,8	227	8	
16	74	мастер	6560	3,44	226	4513	3,74	169	6459	4,2	271	7	
17	75	стевен	5447	4,01	230	4825	3,74	180	5464	4,65	254	5	
18	84	мастер	7502	3,71	278	4900	3,79	186	6459	4,2	271	12	
19	85	мінус	7923	3,78	299	6330	3,47	220	4725	4,8	227	6	

Рис. 6.1. Зовнішній вигляд файлу **Basa-data1.xls** із вихідними даними

Провести аналіз змінної “**Надій матерів**”; вся вихідна інформація міститься на аркуші “**Лист 1**” в інтервалі **С4:С393**.

Насамперед необхідно надати власне ім’я цього аркуша та створити новий аркуш, на якому буде розташовуватися таблиця варіаційного ряду. Для цього, клацніть правою кнопкою миші (ПКМ) по назві “**Лист1**”; у меню, що з’явиться, обрати опцію “**Переименовать**”. Назву “**Лист1**” буде виділено білим кольором на чорному фоні. З клавіатури набрати “**База даних**” та клацнути кнопку **Enter**.

Для того, щоб створити новий аркуш необхідно: клацнути по опції головного меню “**Вставка**” та обрати опцію “**Лист**”. Новий аркуш, що з’явиться, буде мати назву “**Лист1**”. Змінити його ім’я на “**Гистограмма**”.

2. Для того, щоб побудувати варіаційний ряд попередньо необхідно визначити кількість інтервалів та ширину інтервалу. При цьому розрахунок цих вибірових ознак повинен залежати як від вихідних даних, так і від обсягу вибірки. При зміні кількості значень у вибірки чи зміні будь-якого значення, всі параметри варіаційного ряду повинні змінюватися автоматично, без втручання людини.

Для цього на аркуші “**База даних**” у його клітинах створіть таблицю, що матиме такий вигляд (рис. 6.2).

	A	B	C	E	F									
389	1869	атхо	5994	4,24	233									
390	1880	стевен	7705	4,09	315									
391	1885	дар	6671	3,86	257									
392	1891	сибер	6121	3,94	242									
393	2326	листик	7715	3,79	292									
394	<table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>N=</td></tr> <tr><td>Min=</td></tr> <tr><td>Max=</td></tr> <tr><td>R=</td></tr> <tr><td>interval=</td></tr> <tr><td>h=</td></tr> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td>X0=</td></tr> </table>					N=	Min=	Max=	R=	interval=	h=			X0=
N=														
Min=														
Max=														
R=														
interval=														
h=														
X0=														
395														
396														
397														
398														
399														
400														
401														
402														
403														
404														
405														

Рис. 6.2

Розрахуємо необхідні для побудови варіаційного ряду допоміжні величини. Для цього у клітину **C396** введемо: $\text{=СЧЁТ}(C4:C393)$. Це можна зробити або з клавіатури (для цього перейти на російський алфавіт), або через опцію головного меню “**Вставка функции**” f_x .

У останньому випадку у клітину **C396** введіть знак дорівнює =, потім клацніть по кнопці f_x , відкриється “**Мастер функций**”. Далі у лівому віконці оберіть категорію “**Статистические**”, а після цього у правому - функцію “**СЧЁТ**”; клацніть по кнопці **ОК**. У вікні, що відкриється, наберіть **C4:C393** (тільки з англійського алфавіту!) та клацніть **ОК**. У клітці **C396** повинна з’явитися цифра 390.

3. У клітці **C397** набрати $\text{=МИН}(C4:C393)$, а у клітці **C398** – $\text{=МАКС}(C4:C393)$. Або знову ж використати “**Мастер функций**”; у категорії “**Статистические**” обрати функції “**МИН**” та “**МАКС**”. В обох випадках введіть область розрахунку **C4:C393**. Якщо після цього у базі даних з’явиться значення що більше чи ще менше, ніж у поточній базі даних, значення клітинок **C397** (або **C398**) зміниться. Знаходимо розмах варіювання; у клітці **C399** наберемо =C398-C397 .

4. Далі розраховуємо кількість інтервалів варіаційного ряду. Згідно з формулою Старджеса це значення залежить від обсягу вибірки:

$$i = 1 + 3,32 \lg n.$$

Але при цьому воно повинно бути цілим числом та непарним. Тому в клітці **C400** наберемо $\text{=НЕЧЁТ}(ЦЕЛОЕ(1+3,32*LOG10(C396)))$.

5. Розрахуємо ширину інтервалу. Для цього у клітину **C401** вводимо $\text{=ОКРВВЕРХ}(C399/C400;10)$. Таким чином, ми задаємо наступний режим обчислення: показник розмаху варіювання (клітина **C399**) буде віднесено до кількості інтервалів (клітка **C400**), а отримане значення округлене до найближчого більшого числа, кратного 10. Оскільки $9473/9 = 1052,56$, в даному разі це значення буде дорівнювати 1060.

Остання дія підготовчого етапу – це розрахунок нижньої межі першого інтервалу варіаційного ряду. Для цього мінімальне значення (клітина **C397**) необхідно округлити до найменшого кратного краще за все в даному разі 50. Тому у клітину **C403** вводимо =OKРВНИЗ(C397;50) .

6. Для зручності таблицю варіаційного ряду та гістограму фактичного та теоретичного нормального розподілів розташуємо на окремому аркуші – на аркуші “**Гистограмма**”, який ми вже створили раніше. Для того, щоб перейти з аркуша “**База данных**” на аркуш “**Гистограмма**”, клацніть лівою кнопкою миші (ЛКМ) по назві аркуша – “**Гистограмма**”. На цьому чистому аркуші необхідно побудувати таку таблицю (рис. 6.3).

	A	B	C	D	E	F	G	H	I	J	K
1	интервал	граница		середина	накопленная	абсолютная	нормирован.	f(t)	теоритича	теоритич.	D
2		нижняя	верхняя	интервала	частота	частота	отклонен.		абс. частота	накопл.	
3		Xmin	Xmax	Xi	Si	Pi	t		частота	частота	
4	1										
5	2										
6	3										
7	4										
8	5										
9	6										
10	7										
11	8										
12	9										
13											
14											
15											

Рис. 6.3.

7. Розраховуємо нижні межі кожного інтервалу. Оскільки ці значення відрізняються між собою на величину інтервалу, можна використати таку формулу для розрахунку значень нижніх інтервалів ($X_{min, i}$):

$$X_{min, i} = (i - 1) \cdot h + X_{min, 1},$$

де i – номер інтервалу, а h – ширина інтервалу, $X_{min, 1}$ – нижня межа першого інтервалу. Номери інтервалів наведено у клітинах **A4:A12** аркуша “**Гистограмма**”, ширина інтервалу – у клітці **C401** аркуша “**База данных**”, а нижня межа першого інтервалу – у клітці **C403** аркуша “**База данных**”.

Тому в клітину **B4** аркушу “**Гистограмма**” вводимо:

$$\text{=((A4-1)*'База данных'!C401+'База данных'!C403)+1}$$

та клацаємо **Enter**.

Для того, щоб заповнити клітини **B5:B12**, зафіксуйте курсором мишки нижнього правого кута клітини **B4** та розтягніть його донизу таблиці. Цей стовпчик таблиці заповниться автоматично розрахованими значеннями нижньої межі відповідних інтервалів.

Для розрахунку верхньої межі першого інтервалу в клітину **C4** аркушу “**Гистограмма**” введіть:

$$\text{=B4-1+'База данных'!C401}$$

Для того, щоб заповнити інші клітки цього стовпчика, знову використайте операцію розтягування формули, як у попередньому випадку.

У даному разі при розрахунку нижньої межі ми додавали до розрахункового значення ще 1, яка тут відігравала роль найменшої одиниці вимірності. Тому нижня межа кожного наступного інтервалу на 1 більша, ніж верхня межа кожного попереднього інтервалу. При аналізі вмісту жиру найменшою одиницею вимірності буде 0,01 %.

8. Розрахунок середин інтервалів. Для розрахунку середини першого інтервалу необхідно у клітині **D4** ввести вираз:
$$=(B4-1+C4)/2$$
. Для розрахунку середин наступних інтервалів знову розтягнути формулу з клітини **D4** вниз по всьому стовпчику.

9. Для розрахунку частот варіаційного ряду в клітині **E4** введіть формулу:
$$=ЧАСТОТА('База данных'!\$C\$4:\$C\$393;C4)$$
.

Згідно за цією формулою програма буде підраховувати кількість значень у інтервалі **C4:C393** на аркуші “База данных”, які будуть меншими, ніж верхня межа першого інтервалу (в даному разі – меншими, ніж 1860). Якщо цю формулу розтягнути на у клітині **E5:E12**, то ми отримаємо накопичені частоти інтервалів варіаційного ряду. У найнижчій клітині накопичена частота дорівнюватиме 390, тобто всі варіанти вибірки повинні бути меншими, ніж верхня межа останнього інтервалу ряду.

Абсолютні частоти варіаційного ряду можна розрахувати на підставі накопичених частот. Для першого інтервалу завжди абсолютна частота дорівнює накопиченій; таким чином у клітині **F4** необхідно ввести
$$=E4$$
. Абсолютна частота будь-якого іншого інтервалу дорівнює різниці між накопиченою частотою цього інтервалу та накопиченою частотою попереднього інтервалу. Тому у клітині **F5** вводимо
$$=E5-E4$$
. Для заповнення інших клітин цього стовпчика розтягуємо цю формулу донизу.

10. Далі, для перевірки відповідності розподілу нашого варіаційного нормального розподілу нам необхідно розрахувати теоретичні частоти (як абсолютні, так і накопичені). Тому спершу розраховуємо для кожного інтервалу нормоване відхилення.

Для цього у клітині **G4** вводимо у один рядок формулу:

$=(D4-CPЗНАЧ('База данных'!\$C\$4:\$C\$393))/СТАНДОТКЛОН('База данных'!\$C\$4:\$C\$393)$
--

Клацаємо **Enter**. Далі розтягуємо цю формулу донизу стовпчика.

У клітин **H4** вводимо формулу:
$$=НОРМРАСП(G4;0;1;ЛОЖЬ)$$
 і розтягуємо цю формулу вниз стовпчика. Програма розрахує нам ординати нормальної стандартизованої кривої (з середнім арифметичним рівним 0, та сигмою рівною 1) для значення нормованого відхилення кожного інтервалу.

Для того, щоб від стандартизованої нормальної кривої перейти до фактичної, слід розрахувати теоретичні абсолютні частоти. Для цього ординати нормальної кривої слід помножити на обсяг вибірки та на ширину інтервалу й віднести до вибіркової сигми. Тому в клітині **I4** вводимо в один рядок формулу:

$=H4*'База данных'!\$C\$396*'База данных'!\$C\$401/СТАНДОТКЛОН('База данных'!\$C\$4:\$C\$393)$
--

та клацаємо **Enter**. Розтягуємо клітину **I4** вниз стовпчика.

11. Розраховуємо теоретичні накопичені частоти. Накопичена частота першого інтервалу дорівнює його абсолютній частоті. Тому в клітину **J4** вводимо вираз: **=I4**. Накопичена частота будь-якого іншого інтервалу дорівнює сумі накопиченої частоти попереднього інтервалу та абсолютній частоті цього. Тому в клітину **J5** вводимо: **=J4+I5**. Розтягнемо цю формулу донизу стовпчика.

12. Перевірка відповідності емпіричного розподілу нормальному проводиться за допомогою критерію Колмогорова-Смирнова. Для цього слід розрахувати абсолютні різниці накопичених частот варіаційного ряду та теоретичних накопичених частот по всіх інтервалах. Для цього в клітину **K4** вводимо формулу: **=ABS(J4-E4)** та клацаємо **Enter**. Розтягуємо клітину **K4** вниз стовпчика.

Далі потрібно визначити найбільше з цих значень. Для цього в клітину **K16** вводимо формулу: **=МАКС(K4:K12)**. Для того, що розрахувати значення критерію Колмогорова-Смирнова слід в клітину **K19** ввести формулу: **=K16/('База данных'!C396)^0,5**. Якщо розраховане значення буде менше, ніж 1,36 – вибіркового розподіл не відрізняється від нормального. Якщо розраховане значення критерію буде перевищувати 1,36 – вибірка має ненормальний розподіл. У цьому разі її не можна аналізувати класичними методами.

13. Побудова гістограми. Для того, щоб зобразити у графічній формі отримані результати можна побудувати гістограму фактичного розподілу та теоретичного (нормального).

Для цього необхідно клацнути по піктограмі Головного меню “**Мастер диаграмм**”. У віконці, що відкриється обрати тип “**Гистограмма**”. Клацнути по кнопці “**Далее**”. У віконці “**Диапазон**” ввести: **=Гистограмма!\$F\$4:\$F\$12; Гистограмма!\$I\$4:\$I\$12** Клацнути по кнопці “**Готово**”. На аркуші з’явиться гістограма фактичного розподілу та теоретичного.

В підсумку аркуш “**Гистограмма**” повинен мати наступний вигляд (Додаток А).

Питання для самоперевірки:

1. Які модулі містить Головне меню MS Excel?
2. Як створити та редагувати аркуш в MS Excel?
3. Як побудувати таблицю в MS Excel?
4. Для чого використовуються функції СЧЕТ (масив), МИН (масив) та МАКС (масив)?
5. Для чого використовуються функції НЕЧЕТ (число), ОКРВВЕРХ (число) та ОКРВНИЗ (число)?
6. У чому полягають особливості використання функції ЧАСТОТА (масив; адрес) та як її використовують для розрахунку абсолютних та накопичених частот?

Тема №7

Автоматизація біометричного оброблення даних зоотехнічного обліку

Автоматизація біометричного оброблення даних має такі етапи.

1. Відкрити свій файл. Створити новий аркуш з назвою “**Биометрия**”. Для цього клацніть ЛКМ на кнопку Головного меню “**Вставка**” та оберіть функцію “**Лист**”. У книзі з’явиться новий аркуш з назвою “**Лист 1**”. Змінити його ім’я на “**Биометрия**”.

2. На аркуші “**Биометрия**” створити таку таблицю (рис. 7.1).

	A	B	C	D	E
1					
2	Общие характеристики выборки			Статистические ошибки	
3		Объем выборки (n)=		среднего арифметического (SE _X)=	
4		X _{мин} =		медианы (SE _{Me})=	
5		X _{маж} =		среднего квадр. отклонения (SE _S)=	
6		Размах (R)=		вариансы (SE _S ²)=	
7	Показатели центра распределения			коэффициента вариации (SE _{Cv})=	
8		Среднее арифметическое (x̄)=		коэффициента асимметрии (SE _{As})=	
9		Медиана (Me)=		коэффициента эксцесса (SE _{Ex})=	
10		Мода (Mo)=		Проверка формы распределения	
11		5% усеченное среднее ариф.=		t для коэффициента асимметрии =	
12	Показатели изменчивости			t для коэффициента эксцесса=	
13		Среднее отклонение (d')=		t табличное =	
14		Среднее квадрат. откл.(s)=		число степеней свободы (df)=	
15		Варианса (s ²)=		Доверительный интервал	
16		Дисперсия (C)=		для среднего арифметического:	
17		Кoeffициент вариации (CV)=		нижний	
18	Показатели формы распределения			верхний	
19		Кoeffициент асимметрии (As)=		для медианы:	
20		Кoeffициент эксцесса (Ex)=		нижний	
21	Структурные показатели			верхний	
22		Первая квартиль (Q1)=		для среднего квадр. отклонения:	
23		Третья квартиль (Q3)=		нижний	
24		h=		верхний	
25					

Рис.7.1.

3. Розрахунок блоку “Загальні характеристики вибірки”. Ці показники вже розраховані нами на аркуші “**База данных**”, тому ми просто їх переносимо. Для цього введемо:

- в клітину **B3**: `='База данных'!C396` .
- в клітину **B4**: `='База данных'!C397` .
- в клітину **B5**: `='База данных'!C398` .
- в клітину **B6**: `='База данных'!C399` .

4. Розрахунок блоку “Показники центра розподілу”. В цьому блоці наведено найбільш поширені та важливі середні (як ступеневі, так і структурні).

Для розрахунку середнього арифметичного в клітину **B8** введемо формулу: `=CP3HACH('База данных'!C4:C393)`. Це можна зробити або з клавіатури, або (що краще), використавши можливості “**Мастера функций**”. В такому виконайте такі дії:

- а) з клавіатури наберіть в клітці **B8** знак “**=**”;
- б) клацніть по піктограмі Головного меню **fx**;
- в) у вікні меню “**Мастера функций**”, що з’явиться, оберіть категорію “**Статистические**” (у лівому віконці) та функцію “**CP3HACH**” (у правому віконці); клацніть **OK**;

г) кнопкою “Delete” очистіть віконце “**Число 1**”;

д) клацніть по назві аркуша “**База даних**”;

е) у віконці “Число 1” після адреси аркуша “**База даних**!” уведіть з клавіатури адресу інтервалу вихідних даних – **C4:C393**;

ж) клацніть по кнопці ОК; у клітині **B8** з’явиться цифра – 6372,6461538...

Для розрахунку медіани у клітину **B9** введіть таку формулу: **=МЕДИАНА('База даних'!C4:C393)**. Або, краще, скористайтеся можливостями “**Мастера функций**” та виконайте всі кроки, як у попередньому випадку, але на кроці в) в категорії “**Статистические**” оберіть функцію “**МЕДИАНА**”. Всі наступні кроки також без змін. У клітині **B9** повинна з’явитися цифра – 6377,5.

Для розрахунку модального значення у клітину **B10** (з клавіатури чи за допомогою “**Мастера функций**”) введіть формулу:

=МОДА('База даних'!C4:C393).

Нарешті, для розрахунку 5% урізаного середнього арифметичного у клітці **B11** наберіть: **=УРЕЗСРЕДНЕЕ('База даних'!C4:C393;0,05)**. Для цього виконайте всі кроки, як у попередніх випадках, але після введення у віконце “**Число 1**” адреси аркушу та адреси вихідного інтервалу, після цифри 393 наберіть знак “;” та число “0,05”. Це означає, що буде розраховане середнє арифметичне на вибірці, у якій 2,5% найменших та 2,5 % найбільших значень не увійдуть у розрахунок.

5. Розрахунок блоку “Показники мінливості”. Для розрахунку середнього відхилення в клітину **B13** введіть формулу:

=СРОТКЛ('База даних'!C4:C393).

Для розрахунку середнього квадратичного відхилення (сигми) в клітину **B14** введіть формулу: **=СТАНДОТКЛОН('База даних'!C4:C393)**.

Для розрахунку варіанси у клітину **B15** введіть формулу:

=ДИСП('База даних'!C4:C393).

Для розрахунку дисперсії вибірки (суми квадратів відхилень кожної варіанти вибірки від її середнього арифметичного) у клітину **B16** введіть:

=КВАДРОТКЛ('База даних'!C4:C393).

Нарешті, для того, щоб розрахувати коефіцієнт варіації у клітці **B17**, наберіть з клавіатури таку формулу: **=(B14/B8)*100**.

6. Розрахунок блоку “Показники форми розподілу”. Розрахуємо спочатку коефіцієнт асиметрії. Для цього у клітину **B19** введіть формулу:

=СКОС('База даних'!C4:C393).

В клітину **B20** введіть формулу: **=ЭКЦЕСС('База даних'!C4:C393)**.

7. Розрахунок блоку “Структурні показники”. Для розрахунку вибіркової оцінки першої квартилі введіть в клітину **B22** формулу: **=КВАРТИЛЬ('База даних'!C4:C393;1)**. Аналогічно, для розрахунку третьої квартилі в клітину **B23** введіть формулу:

=КВАРТИЛЬ('База даних'!C4:C393;3).

Останній – третій – показник цього модуля нам буде потрібен при розрахунку довірчого інтервалу для медіани. Для розрахунку h в клітину B24 введіть з клавіатури формулу:

$$=OKPBBEPX((B3-1-1,96*(B3)^{0,5})/2;1)$$

8. Розрахунок блоку “Статистичні помилки”. Для розрахунку статистичної помилки середнього арифметичного необхідно в клітину E3 ввести формулу: $=B14/KOPENb(B3)$. Це можна зробити або з клавіатури, або використавши можливості “Мастера функций”. В останньому разі виконайте такі кроки:

- а) в клітині E3 наберіть знак “=”;
- б) клацніть по клітці B14;
- в) наберіть знак “/”;
- г) на Головному Меню клацніть по кнопці “Мастера функций”;
- д) оберіть категорію “Математические”;
- е) оберіть функцію “КОРЕНЬ” та клацніть ОК;

ж) у віконці “Число” наберіть B3 та клацніть ОК; в клітці E3 з’явиться значення помилки середнього арифметичного (58,05477).

Для розрахунку помилки медіани в клітці E4 наберіть формулу: $=(B23-B22)/KOPENb(B3)$. Для цього введіть $=(B23-B22)$, а далі повторіть кроки з г) по ж), як у попередньому випадку.

Для розрахунку помилки середнього квадратичного відхилення (сигми) в клітці E5 наберіть формулу (з клавіатури або використавши “Мастер функций”): $=B14/KOPENb(2*B3)$.

Для розрахунку помилки варіанси в клітці E6 наберіть формулу: $=B15*KOPENb(2/B3)$.

Для розрахунку помилки коефіцієнта варіації введіть в клітці E7 формулу: $=B17/KOPENb(2*B3)$.

Для розрахунку помилки коефіцієнта асиметрії в клітці E8 введіть формулу: $=KOPENb(6/(B3+3))$.

Для розрахунку помилки коефіцієнта ексцесу в клітці E9 введіть формулу: $=KOPENb(24/(B3+5))$.

9. Розрахунок блоку “Перевірка форми розподілу”. В цьому блоці ми розраховуємо одновибірковий t-критерій Ст’юдента для перевірки вірогідності відхилення оцінок коефіцієнтів асиметрії та ексцесу нулю.

Для розрахунку критерію Ст’юдента у разі оцінки коефіцієнтів асиметрії та ексцесу введіть, відповідно, в клітину E11 формулу: $=ABS(B19/E8)$, а в клітину E12: $=ABS(B20/E9)$.

У тих випадках, коли ми перевіряємо будь-яку статистичну гіпотезу, бажано було б мати табличне (критичне) критерію або одразу ж позначати, якщо отримане значення є вірогідним. Тому введемо табличне значення критерію Ст’юдента для рівня значущості $p = 0,05$, який у даному разі має число ступенів свободи $df = n - 1$; для цього в клітину E13 вводимо формулу: $=CTbЮДPACПOБP(0,05;B3-1)$. Функція, зворотна розподілу Ст’юдента розташована також у категорії “Статистические” “Мастера функций”. (Це

табличне значення критерію Ст'юдента нам ще знадобиться при розрахунку довірчих меж середнього арифметичного.)

В клітині **E14** розташуємо число ступенів свободи: $=B3-1$.

10. Розрахунок блоку “Довірчий інтервал”. Для розрахунку нижньої та верхньої межі 95 % довірчого інтервалу середнього арифметичного значення необхідно, відповідно, ввести в клітину **E17** формулу: $=B8-E13*E3$, а в **E18**: $=B8+E13*E3$.

Для розрахунку довірчого інтервалу медіани нам знадобиться величина, яку ми розраховували раніше у клітці **B24** ($h = 176$), тому що в даному випадку, нижньою межею довірчого інтервалу медіани буде 176-те найменше значення у нашій вибірці, а верхньою межею – 176-те найбільше значення у вибірці.

Тому, для розрахунку нижньої та верхньої межі 95 % довірчого інтервалу медіани, відповідно, введемо в клітину **E20** формулу:

$=\text{НАИМЕНЬШИЙ}('База данных'!C4:C393;B24)$,

а в клітину **E21**: $=\text{НАИБОЛЬШИЙ}('База данных'!C4:C393;B24)$.

Для розрахунку довірчого інтервалу для середнього квадратичного відхилення (сигми) ми використаємо розподіл Хі-квадрат. Для того, щоб розрахувати нижню та верхню межу 95 % довірчого інтервалу сигми необхідно ввести, відповідно, в клітину **E23** формулу:

$=B14*\text{КОРЕНЬ}(E14/\text{ХИ2ОБР}(0,025;E14))$,

а у клітину **E24**: $=B14*\text{КОРЕНЬ}(E14/\text{ХИ2ОБР}(0,975;E14))$.

11. Таким чином, у підсумку аркуш “**Биометрия**” повинен мати такий вигляд (Додаток Б).

Питання для самоперевірки:

1. Які особливості використання функцій СРЗНАЧ (масив), МЕДИАНА (масив), МОДА (масив) та УРЕЗСРЕДНЕЕ (масив; частка)?
2. Для чого використовуються функції СРОТКЛОН(масив), СТАНДАТКЛОН (масив), ДИСП (масив) та КВАДРАТКЛОН (масив)?
3. Для чого використовуються функції СКОС(масив) та ЭКСЦЕСС (масив)?
4. Які особливості має використання функцій КВАРТИЛЬ (масив; число) при розрахунку першого та третього кватилів?
5. Для чого використовуються функції НОРМАСП() та СТЬЮДРАСПОБР ()?
6. Для чого використовуються функції НАИМЕНЬШИЙ (масив; число) та НАИБОЛЬШИЙ (масив; число)?

Тема №8

Автоматизація оцінки бугаїв-плідників за нащадками

Для того щоб провести автоматизовану оцінку бугаїв-плідників за якістю нащадків, необхідно виконати такі операції.

1. Відкрити свій файл на аркуші “База даних”. Клацнути ЛКМ по назві заголовка таблиці первинних даних “Кличка Б”; цей стовпчик містить клички бугаїв-плідників – батьків корів, які розглядаються в аналізі.

Для того, щоб проаналізувати скільки бугаїв-плідників було використано та отримати повний список їх кличок, клацніть по опції Головного меню “Данные”, у меню, що з’явиться, оберіть опцію “Фильтр” й увімкніть режим “Автофильтр”. У верхньому правому куті клітини з назвою “Кличка Б” з’явиться сірий квадратик з чорним трикутником. Клацніть по цьому трикутнику й у віконці з’явиться перелік кличок бугаїв-плідників: “аман, астроном, атхо,” (рис. 8.1).

	А	В	С
1			
2	Нош	(Все)	
3		(Первые 10.. (Условие...))	НАДІЙ
4		аман	7436
5		астроном	5426
6		атхо	7620
7		бей	7039
8		боб	6555
9		бостон	6205
10		булат	5406
11		великан	8657
12		гектор	5730
13		грозний	5106
14		гулант	5819
15		дар	7284
16		діамант	6560
17		дінар	5447
18		75 стевен	7500

Рис. 8.1.

Якщо необхідно обрати із загальної бази даних лише дочок одного будь-якого бугая, то в цьому списку оберіть його кличку та клацніть по неї ЛКМ. Вихідна база даних трансформується у базу даних, що буде містити лише тих корів, які мають одного обраного батька.

Оберемо для порівняльного аналізу бугаїв-плідників за продуктивними якостями нащадків (за надоем за кращою лактацією) лише 16 бугаїв, які мають чисельних нащадків.

2. Створити новий аркуш, якому надати ім’я “Быки”. На цьому аркуші створіть таблицю (рис. 8.2).

3. Для того, щоб підрахувати кількість дочок для кожного бугая, наберіть в клітині **В4** формулу:

=СЧЁТЕСЛИ('База данных'!\$B\$4:\$B\$393;A4).

Далі, клацніть по клітці **B4** ЛКМ, візьміться курсором за малий чорний квадратик, що розташований в правому нижньому куті цієї клітини та розтягніть формулу донизу – до клітини **B19**. Відпустить ЛКМ й у всіх клітинах з'являться відповідні чисельності.

4. Для того, щоб розрахувати середній надій дочок для кожного бугая-плідника, по-перше, розрахуємо сумарну кількість надоеного від усіх його дочок молока, а потім віднесемо цю кількість до чисельності дочок. Для цього в клітину **C4** введіть формулу:

=СУММЕСЛИ('База данных'!\$B\$4:\$B\$393;A4;'База данных'!\$Z\$4:\$Z\$393)

(В цій формулі в інтервалі **B4:B393** розташовані індивідуальні клички бугаїв, а в інтервалі **Z4:Z393** – показники надою дочок за кращою лактацією.)

	А	В	С	Д	Е	Ф	Г	Н	І	Ж	К
1	кличка	число	Надой(лучшая лактация),кг				Жир(лучшая лактация),%				
2	бугая-	дочерей	сумма	средний	абсолютное	относительное	сумма	средний	абсолютное	относительное	
3	производител	(п)		надой	отклонение,кг	отклонение,%		жир,%	отклонение,%	отклонение,%	
4	ятхо										
5	гулант										
6	закон										
7	запас										
8	звонок										
9	золотой										
10	кулік										
11	листик										
12	лукавий										
13	луч										
14	мінус										
15	роджер										
16	сибер										
17	стевен										
18	хомяк										
19	ян										
20											
21	по стаду										
22											

Рис. 8.2.

Розтягніть цю формулу донизу – й всі клітини цього стовпчика заповняться відповідними цифрами.

Для того, щоб розрахувати середній надій, в клітину **D4** введіть просту формулу: **=C4/B4** та розтягніть цю формулу донизу.

В клітину **D21** введіть середню продуктивність по стаду в цілому; для цього наберіть формулу: **=СРЗНАЧ('База данных'!Z4:Z393)**.

5. Далі проаналізуємо рівень продуктивності нащадків окремих бугаїв-плідників порівняно із середнім показником по стаду.

Для цього в клітину **E4** введіть формулу: **=D4-\$D\$21**. Розтягніть цю формулу донизу до кінця стовпчика. З'являться цифри - позитивні чи негативні, залежно від того переважають дочки цього бугая середній показник по стаду чи поступаються йому, відповідно.

6. Оцінимо відносний рівень продуктивності дочок – нащадків обраних бугаїв-плідників. Для цього визначимо на скільки відсотків вони переважають

(або поступаються) середнім показникам по стаду. Для цього в клітину **F4** введіть формулу: $=E4*100/SD\$21$. Розтягніть цю формулу донизу.

7. Далі, для того, щоб заповнити клітини таблиці щодо вмісту жиру в молоці, знову повторіть кроки 4-6, але для стовпчика **AB4 : AB393** вихідної бази даних, де занесені дані про вміст жиру в молоці дочок за кращою лактацією.

8. Зобразимо результати нашого аналізу у вигляді графіка з двома осями ОУ. На однієї з них буде наведено відносне відхилення за надосем, а на іншій – за вмістом жиру в молоці.

Для цього виконайте такі дії:

- клацніть по гістограмі Головного меню “**Мастер диаграмм**” та оберіть діаграми “**Нестандартные**”;
- серед нестандартних оберіть діаграму типу “**Графики 2 оси**” та клацніть по кнопці “**Далее**”;
- у віконці “**Диапазон:**” уведіть:
 $=\text{Быки!}\$A\$4:\$A\$19;\text{Быки!}\$F\$4:\$F\$19;\text{Быки!}\$J\$4:\$J\19 ;
- клацніть по кнопці “**Ряд**”; у віконці “**Имя**” проти “**Ряд2**” введіть “**Жир**”, а проти “**Ряд1**” – “**Надой**”; клацніть по кнопці “**Далее**”;
- у віконці “**Размещение**” клацніть по слову “**вверху**” та по кнопці “**Готово**”.

9. Розташувати діаграму таким чином, щоб у підсумку дана сторінка мала наступний вигляд (Додаток В).

Питання для самоперевірки:

1. У чому особливості використання режиму “**АВТОФИЛЬТР**” при формуванні підвибірок?
2. Які особливості мають функції **СЧЕТЕСЛИ** (масив; число) та **СУММЕСЛИ** (масив; число; масив)?
3. Які особливості побудови графіка з двома осями?

Тема №9

Автоматизація порівняльного аналізу в ряду “предки-нащадки”

Для того, щоб провести автоматизоване порівняння в ряду «предки-нащадки» необхідно виконати такі операції.

1. Відкрити свій файл, створити новий аркуш, який назвати “Парн.сравн.” та побудувати на ньому дві таблиці (рис. 9.1).

На цьому аркуші будемо проводити порівняльний аналіз за рівнем продуктивності (за надоем) між різними генераціями тварин нашої дослідної групи. При цьому парні порівняння будемо проводити:

- у відношенні центральної тенденції (вибіркових середніх арифметичних) з використанням парного критерію Ст’юдента;
- у відношенні рівня мінливості (вибіркових варіанс) з використанням критерію Фішера.

	A	B	C	D	E	F	G
1	Генерация		Надой, кг				
2		n	Хср.	SE x	s	CV, %	
3	Матери матерей						
4	Матери отцов						
5	Матери						
6	1-ая лактация						
7	3-ья лактация						
8	Лучшая лактация						
9							
10	Парные	В отношении центра		В отношении ур-ня			
11	сравнения	распределения		изменчивости			
12		разность	ур-нь значимости	ур-нь значимости			
13	ММ-Л						
14	МО-Л						
15	ММ-МО						
16	ММ-М						
17	М-Л						
18							

Рис. 9.1.

2. Спочатку розрахуємо вибіркові показники (обсяги вибірок, середні арифметичні, середні квадратичні відхилення та інші).

3. В клітину **В3** заносимо формулу:

=СЧЁТ('База данных'!Н4:Н393)

В клітини **В4-В8** заносимо аналогічні формули, змінюємо лише адреси інтервалів вихідних даних. Наприклад, для Матерів Батьків це: **М4:М393**, й т.і.

Замість набору цієї формули з клавіатури можна використатися можливостями “Мастера функций”. Для цього виконайте такі кроки:

- а) з клавіатури наберіть в клітці **В3** знак “=”;
- б) клацніть по піктограмі Головного меню ***fx***;
- в) у вікні меню “Мастера функций”, що з’явиться, оберіть категорію “Статистические” (в лівому віконці) та функцію “СЧЕТ” (у правому віконці; клацніть ОК;

- г) у віконці, що відкриється стерти, всі записи;
- д) клацнути по імені аркуша “База данных”;
- е) після тексту 'База данных!', що з'явиться у віконці, ввести з клавіатури H4:H393 та клацнути ОК.

4. В клітину **C3** вводимо формулу:

$\boxed{=CP3HACH('База данных'!H4:H393)}$.

В клітини **C4-C8** вводимо аналогічні формули, також змінюємо лише адреси інтервалів вихідних даних.

5. В клітину **E3** заносимо наступну формулу:

$\boxed{=CTAHNDOTKJIOH('База данных'!H4:H393)}$.

В клітки **E4-E8** заносимо аналогічні формули, також змінюємо лише адреси інтервалів вихідних даних.

6. Розрахуємо тепер помилки вибіркового середнього арифметичних. Для цього в клітину **D3** слід ввести формулу: $\boxed{=E3/KOPENb(B3)}$. Замість введення з клавіатури можна використати можливість “Мастера функций”.

Для того, щоб розрахувати оцінки помилок для всіх рядів даних, розтягніть формулу із клітини **D3** в клітинах **D4-D8**.

7. Розраховуємо вибіркові коефіцієнти варіацій; для цього введіть в клітину **F3** формулу: $\boxed{=E3*100/C3}$, та розтягніть її донизу стовпчика.

8. Далі проведемо безпосередньо попарні порівняння й заповнимо другу таблицю цього аркуша.

По-перше, розрахуємо абсолютні відмінності між окремими парами тварин. Для цього в клітці **B13** розрахуємо різницю між середнім на доєм Матерів матерів та Кращою лактацією дочок; це можна зробити за формулою: $\boxed{=C3-C8}$.

Далі, в клітці **B14** розрахуємо різницю між середнім на доєм Матерів Батьків та Кращою лактацією дочок: $\boxed{=C4-C8}$.

Заповнюємо подібним чином усі клітини цього стовпчика.

Замість того, щоб вводити формули безпосередньо, можна, наприклад, зробити такі кроки:

- а) ввести з клавіатури у клітину **B13** знак “=”;
- б) клацнути лівою кнопкою миші (ЛКМ) по клітці **C3**;
- в) ввести з клавіатури знак “-”;
- г) клацнути ЛКМ по клітці **C8**;
- д) клацнути по кнопці “Enter”.

9. Процедура попарного порівняння між двома рядами повністю автоматизована. При цьому програма розраховує оцінку рівня вірогідності парного критерію Ст'юдента.

Для порівняння між на доєм Матерів матерів та Кращою лактацією дочок у клітину **C13** введіть формулу:


$\boxed{=TTECT('База данных'!H4:H393;'База данных'!Z4:Z393;2;1)}$.

Тут у дужках наведено таку інформацію:

- адреса ряду “Надій матерів матерів”;
- адреса ряду “Надій дочок за кращою лактацією”;

- далі, цифра “2” означає, що обрано двосторонній критерій Ст’юдента;
- цифра “1” наприкінці означає, що обрано модифікація тесту Ст’юдента для парного порівняння зв’язаних даних.

Замість того, щоб вводити цю формулу з клавіатури використайте можливості “**Мастера функций**”. Для цього виконайте такі дії:

- а) в клітину **C13** введіть “=”;
- б) клацніть по піктограмі Головного меню ;
- в) у вікні меню “**Мастера функций**”, що з’явиться, оберіть категорію “**Статистические**” (в лівому віконці) та функцію “**ТТЕСТ**” (у правому віконці); клацніть ОК;
- г) у віконці “**Массив1**” клацніть ЛКМ по назві аркуша “**База данных**”, а далі з клавіатури наберіть інтервал Н4:Н393;
- д) у віконці “**Массив2**” аналогічно уведіть адресу другого інтервалу значень - 'База данных'!Z4:Z393;
- е) у віконці “**Хвосты**” клацніть 2;
- ж) у віконці “**Тип**” клацніть 1; далі клацніть по кнопці ОК.


10. Подібним чином, змінюючи адреси рядів, що порівнюються, розрахуйте всі клітини цього стовпчика.

11. Процедуру парного порівняння за характером мінливості двох рядів даних також повністю автоматизовано. При цьому програма розраховує рівень значущості критерію Фішера (одностороннього).

Для цього в клітину **E13** ввести формулу:

$$=ФТЕСТ('База данных'!Н4:Н393;'База данных'!Z4:Z393)$$

Крім того, також можна використати можливості “**Мастера функций**”. Для цього виконати такі дії:

- а) в клітину **E13** введіть “=”;
- б) клацніть по піктограмі Головного меню ;
- в) у вікні меню “**Мастера функций**”, що з’явиться, оберіть категорію “**Статистические**” (в лівому віконці) та функцію “**ФТЕСТ**” (у правому віконці); клацніть ОК;
- г) у віконці “**Массив1**” клацніть ЛКМ по назві аркуша “**База данных**”, а далі з клавіатури наберіть інтервал Н4:Н393;
- д) у віконці “**Массив2**” аналогічно уведіть адресу другого інтервалу значень - 'База данных'!Z4:Z393; далі клацніть по кнопці ОК.

12. Подібним чином, змінюючи адреси рядів, що порівнюються, розрахуйте всі клітини цього стовпчика.

13. Далі, використовуємо можливості опції “**Умовне форматування**”. Для цього виділіть прямокутник клітин **C13:E17**. Далі виконайте такі кроки:

- а) клацніть по кнопці Головного меню “**Формат**” та оберіть режим “**Условное форматирование**”;
- б) у меню, що відкриється, зробіть такі зміни (рис. 9.2);

в) клацніть по кнопці “**Формат**” та оберіть такі зміни: “**Начертание**” оберіть “*полуужирный курсив*”, а “**Цвет**” – червоний, та клацніть по кнопці ОК;

г) клацніть по кнопці ОК; всі значення у межах виділених інтервалі, які є меншими першого рівня значущості (тобто, 0,05) будуть виділені червоним кольором.

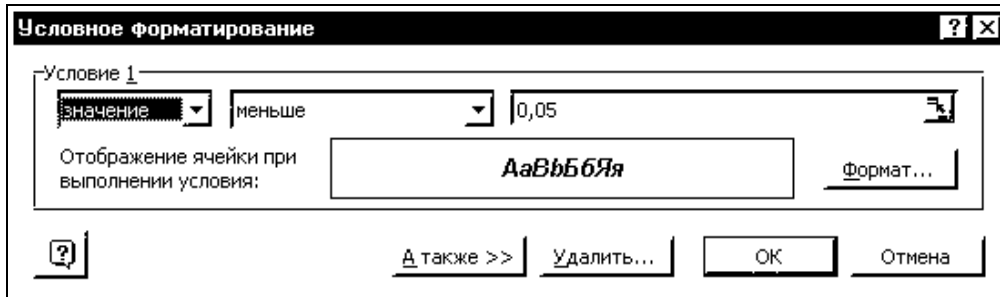


Рис. 9.2.

14. Побудуємо гістограму на підставі отриманих даних. Для цього клацнемо по піктограмі Головного меню “**Мастер диаграмм**”. Оберемо тип “**Гистограмма**” та клацнемо по кнопці “**Далее**”.

У віконці “**Диапазон**” введемо:

=Парн.сравн.!\$A\$3:\$A\$8; Парн.сравн.!\$C\$3:\$C\$8

та клацнемо по кнопці “**Далее**”.

У віконці “**Ось Y (значения)**” введіть “Надой, кг”. Клацніть по кнопці “**Готово**”.

15. Наприкінці цей аркуш повинен мати такий вигляд (Додаток Д).

Питання для самоперевірки:

1. Які особливості має функція ТТЕСТ (масив; масив)?
2. Для чого використовується функція ФТЕСТ (масив; масив)?
3. Які особливості використання режиму “Умовне формування”?
4. У чому особливості побудови та редагування точкової діаграми?

Тема №10

Автоматизація розрахунків коефіцієнтів успадкування та повторюваності

Для того, щоб провести автоматизовані розрахунки коефіцієнтів успадкування та повторюваності, необхідно виконати такі операції.

1. Відкрити свій файл та створити новий аркуш з назвою “**Корреляция**”. На цьому аркуші створити таблицю (рис. 10.1).

	А	В	С	Д	Е	F	G	Н	І	J	К	Л	М
1													
2	Генерация	Козфициент парной корреляции					Преобразование		Доверительный				
3	(лактация)	между надоем и содержанием жира					Фишера		интервал				
4		n	r	SE_r	t	p	z	SE_z	r_min	r_max			
5													
6	ММ												
7	МО												
8	М												
9	Д1												
10	Д3												
11	Дл												
12													
13		по надоем											
14											Козфициент наследуемости		
15	Пары										h2_min	h2_max	
16	ММ - М												
17	МО - М												
18	М - Дл												
19													
20											Козфициент повторяемости		
21											W_min	W_max	
22	Пары												
23	Д1 - Дл												
24	Д1 - Д3												
25	Д3 - Дл												
26													

Рис. 6.1.

2. В клітини **В6:В11** цієї таблиці перенести обсяги вибірових значень для відповідних генерацій, що були розраховані на аркуші “Парн.сравн.”. Наприклад, в клітину **В6** необхідно занести: `=Парн.сравн.!В3` й т.і.

3. Далі, розрахуємо коефіцієнти парної лінійної кореляції Пірсона-Браве між показниками надою та вмістом жиру по кожній генерації тварин дослідного стада. Для цього, наприклад, в клітину **С6** для **Матерів матерів** вносимо формулу:

$$=КОРРЕЛ('База данных'!Н4:Н393;'База данных'!J4:J393)$$

У даному разі вона містить такі дані: `'База данных'!Н4:Н393` – адреса клітинок, що містять дані у відношенні надою ММ; `'База данных'!J4:J393` – у відношенні вмісту жиру в молоці ММ.

Аналогічно, змінюючи адресацію клітинок з вихідними даними, розраховуються коефіцієнти кореляції і для інших генерацій.

4. Для розрахунку помилки коефіцієнту кореляції між надоєм та вмістом жиру в молоці **ММ** в клітину **Д6** заносимо формулу:

$$=КОРЕНЬ((1-C6^2)/(B6-2))$$

Для того, щоб розрахувати оцінки помилок коефіцієнтів кореляції й для інших генерацій цю формулу можна розтягнути донизу стовпчика.

5. Далі розраховуємо вірогідність отриманих вибірових оцінок коефіцієнтів кореляції. Для цього розраховуємо значення критерію Ст'юдента та рівень його вірогідності.

В клітину **E6** вносимо формулу: $=ABS(C6/D6)$. В клітину **F6** - формулу: $=СТЬЮДРАСП(E6;B6-2;2)$.

Для розрахунку цих показників для тварин інших генерацій обидві формули можна розтягнути.

6. Використаємо опцію “Умовне форматування” Головного меню й встановимо для клітинок **F6:F12** наступний формат: “Значення у цих клітках виділяти червоним кольором, якщо вони не перевищують 0,05” (див. вище).

7. Розраховуємо довірчий інтервал для вибірових оцінок коефіцієнтів кореляції. Для цього використаємо z-перетворення Фішера.

В клітину **G6** вносимо: $=ФИШЕР(C6)$. Далі розтягнемо цю формулу донизу. В клітину **H6** вносимо формулу для розрахунку помилки перетворення Фішера: $=1/КОРЕНЬ(B6-3)$. Цю формулу також розтягнемо донизу.

Для розрахунку нижньої межі 95% довірчого інтервалу в клітину **I6** вносимо формулу: $=ФИШЕРОБР(G6-1,96*H6)$. Для розрахунку верхньої межі 95% довірчого інтервалу в клітину **J6** вносимо формулу: $=ФИШЕРОБР(G6+1,96*H6)$.

Друга частина аналізу присвячена розрахунку коефіцієнтів успадкування та повторюваності по надою. Використовуємо три пари генерацій: **ММ-М**, **МО-М** та **М-Л**. Для розрахунку коефіцієнтів успадкування використовується формула $h^2 = 2r_{мд}$, де $r_{мд}$ – значення коефіцієнту кореляції між показниками матерів та їх нащадків.

Для оцінки вибірових значень коефіцієнтів повторюваності використовуємо дані по лактаціях Дочок: по першій, третій та кращій. Коефіцієнт повторюваності розраховується, як коефіцієнт кореляції між надоєм: по 1-й та по Кращій лактаціях, по 1-й та по 3-й лактаціях, по 3-й та по Кращій лактаціях дочок.

8. По-перше, необхідно визначити кількість парних значень, для яких буде розраховуватися значення коефіцієнтів кореляції. Наприклад, для пари **ММ-М** в клітину **B16** заносимо: $=МИН(B6;B8)$. В цій формулі перша адреса – це клітина, в якій занесено обсяг вибірки по **ММ**, а в другій – обсяг вибірки по **М**.

Аналогічно заповнюємо клітини **B17** та **B18**, а також **B23:B25**.

9. Повторюємо кроки 2-6 для розрахунку відповідних значень коефіцієнтів кореляції в клітинах **C16:J18** (для розрахунків коефіцієнтів успадкування) та **C23:J25** (для розрахунків коефіцієнтів повторюваності).

Використовуємо при цьому такі адреси вихідних даних на аркуші “База даних”:

Адресація	Ознака
C4:C393	- надій матерів;
E4:E393	- вміст жиру в молоці матерів;
H4:H393	- надій матерів матерів;
J4:J393	- вміст жиру в молоці матерів матерів;
M4:M393	- надій матерів батьків;
O4:O393	- вміст жиру в молоці матерів батьків;
S4:S393	- надій по 1-й лактації дочок;
T4:T393	- вміст жиру в молоці по 1-й лактації дочок ;
W4:W393	- надій по 3-й лактації дочок;
X4:X393	- вміст жиру в молоці по 3-й лактації дочок ;
Z4:Z393	- надій по кращій лактації дочок;
AB4:AB393	- вміст жиру в молоці по кращій лактації дочок.

10. Для розрахунку нижньої межі довірчого інтервалу для коефіцієнту успадкування використаємо режим умовного форматування (функцію “**Якщо**”). Для цього в клітину **K16** заносимо: $=ЕСЛИ(F16<0,05;2*I16;0)$. (Вона визначається наступним чином: якщо значення рівня вірогідності коефіцієнту кореляції менше, ніж 0,05, то в клітину **K16** внести результат множення 2 на значення нижньої межі довірчого інтервалу коефіцієнта кореляції (клітина **I16**), а якщо ні, то внести нуль.)

Аналогічним чином розрахуємо й верхню межу довірчого інтервалу: $=ЕСЛИ(F16<0,05;2*J16;0)$.

Обидві формули розтягнемо донизу.

11. Довірчі інтервали для коефіцієнтів повторюваності розрахуємо так, як описано в п. 10, але без множення відповідних значень коефіцієнтів кореляції на 2.

12. У підсумку цей аркуш повинен мати такий вигляд (Додаток Є).

Питання для самоперевірки:

1. Для чого використовується функція КОРРЕЛ (масив; масив)?
2. Які особливості використання функції ФИШЕР (число) та ФИШЕРОБР (число)?
3. Які особливості використання функції ЕСЛИ (умова; число; число)?

Тема №11

Автоматизація розрахунків коефіцієнтів лінійної регресії

Для того, щоб провести автоматизовані розрахунки коефіцієнтів лінійної регресії, необхідно виконати такі операції.

1. Відкрити свій файл та створити новий аркуш з назвою “Регрессия”. Створити на цьому аркуші таблицю (рис. 11.1).

	A	B	C	D	E	F
1						
2		Надой дочерей	Надой матерей			
3		(Y)	(X)			
4	Сумма					
5	Сумма квадратов					
6	Сумма произведений					
7						
8	Кол-во пар					
9						
10	Козфициенты регрессии:		Ошибки:	t	p	
11	угла наклона (b)					
12	пересечение с осью ОУ (a)					
13						
14	Остаточна сумма квадратов:					
15						
16	Козффициент корреляции:					
17						
18		нижняя гран.	верхняя гран.			
19	Козффициент наследуемости:					
20						

Рис. 11.1.

2. Розрахуємо коефіцієнти лінійної регресії між показниками надою матерів та дочок за кращою лактацією.

Для цього, по-перше, необхідно розрахувати по обох рядах вихідних даних суми та суми квадратів. А, по-друге, суму добутків між кожною парою значень обох рядів вихідних даних.

Таким чином, в клітину **B4** вносимо формулу:

$\text{=СУММ('База данных'!Z4:Z393)}$,

а в клітину **C4**: $\text{=СУММ('База данных'!C4:C393)}$.

Для розрахунку сум квадратів вводимо формули:

- у клітину **B5**: $\text{=СУММКВ('База данных'!Z4:Z393)}$,

- а у клітину **C5**: $\text{=СУММКВ('База данных'!C4:C393)}$.

Нарешті, для розрахунку суми добутків в клітину **B6** вводимо формулу:

$\text{=СУММПРОИЗВ('База данных'!Z4:Z393;'База данных'!C4:C393)}$

3. В клітину **B8** переносимо кількість пар відповідних значень з аркуша “Корреляция”: =Корреляции!B18 .

4. Розрахуємо коефіцієнти регресії . Для цього вносимо в клітину **B11** формулу:
$$=(B8*B6-B4*C4)/(B8*C5-(C4^2))$$
 ,

а у клітину **B12**:
$$=(B4*C5-B6*C4)/(B8*C5-(C4^2))$$
 .

5. Для розрахунку помилок коефіцієнтів регресії необхідно розрахувати залишкову варіансу. Для цього в клітину **B14** вносимо формулу:

$$=(B5-B12*B4-B11*B6)/(B8-2)$$
 .

Тоді, для розрахунку помилок коефіцієнтів регресії вносимо в клітину **C11** формулу:
$$=КОРЕНЬ(B14/(C5-((C4^2)/B8)))$$
 , а в клітину **C12**:
$$=C11*КОРЕНЬ(C5/B8)$$
 .

6. Розрахуємо рівень вірогідності отриманих значень вибірових коефіцієнтів регресії. Для цього розрахуємо значення критерію Ст'юдента та їх рівень значущості.

Для цього в клітину **D11** вносимо формулу:
$$=ABS(B11/C11)$$
 , а у клітину **D12**:
$$=ABS(B12/C12)$$
 .

В клітину **E11** вносимо формулу:
$$=СТЬЮДРАСП(D11;B8-3;2)$$
 , а в клітину **E12** – аналогічну, лише замінюємо **D11** на **D12** .

7. Для зручності в клітині **B16:E16** переносимо відповідні оцінки для коефіцієнта кореляції (з аркуша “**Корреляция**”).

Для клітин **E11**, **E12** та **E16** використовуємо режим “**Умовного форматування**” за стандартним алгоритмом: “Якщо значення в клітині менше 0,05, то виділити його червоним кольором та жирним курсивом”.

8. Розрахуємо довірчі інтервали для коефіцієнту успадкування. (Для розрахунку коефіцієнту успадкування використовуємо формулу:

$$h^2 = 2b_{мд}$$
 ,

де $b_{мд}$ – коефіцієнт кута нахилу ліній регресії ознаки дочки залежно від ознаки її матері.

Для розрахунку нижньої довірчої межі 95% довірчого інтервалу вносимо в клітину **B19** формулу:

$$=ЕСЛИ(E11<0,05; 2*(B11-1,96*C11);0)$$
 .

Для розрахунку верхньої довірчої межі 95 % довірчого інтервалу вносимо в клітину **C19** формулу:

$$=ЕСЛИ(E11<0,05;2*(B11+1,96*C11);0)$$
 .

9. Побудуємо точковий графік залежності надою дочок (за кращою лактацією) від рівня молочної продуктивності їх матерів та побудуємо на підставі цих даних лінію регресії.

Для цього необхідно клацнути по піктограмі Головного меню “**Мастер диаграмм**”. У вікні “**Стандартные**” обрати “**Точечная**”, у вікні “**Вид**” – “**Точечная диаграмма позволяет сравнить пары значений**”. Клацнути по кнопці “**Далее**”.

У віконці “**Диапазон**” увести:

$$='База данных'!C4:C393; 'База данных'!Z4:Z393$$

та клацнути **“Ряды в столбцах”**.

Далі, клацнути по закладці **“Ряд”**. У віконці **“Значения X:”** повинно бути записано: $\boxed{= \text{‘База данных’!SC\$4:SC\$393}}$, а у віконці **“Значения Y:”** $\boxed{= \text{‘База данных’!Z\$4:Z\$393}}$. Клацнути по кнопці **“Далее”**.

У закладці **“Заголовки”** з клавіатури ввести наступний текст: у віконці **“Ось X (категорий)”** – **“Надій матерів, кг”**, а у віконці **“Ось Y (значений)”** – **“Надій дочок, кг”**. Клацнути по кнопці **“Готово”**.

10. Очистити область побудови діаграми від ліній сітки осі ОУ та від сірого кольору.

11. Необхідно на підставі розсіювання точок вихідних даних провести лінію регресії з шуканими параметрами кута нахилу та точки перетину з віссю ОУ (див. вище). Для цього підведіть курсор на скупчення точок графіка та клацніть ПКМ. У меню, що відкриється, оберіть **“Добавить линию тренда”** та клацніть по ній ЛКМ.

У меню, що відкриється оберіть:

- у закладці **“Тип”** – **“Линейная”**;
- у закладці **“Параметры”** – клацніть ЛКМ у віконцях **“Показывать уравнение на диаграмме”** та **“Поместить на диаграмму величину достоверности аппроксимации (R^2)”**.

Клацніть ПКМ по легенді та вилучіть її з графіку.

12. В підсумку даний аркуш повинен мати такий вигляд (Додаток Ж).

Питання для самоперевірки:

1. Які особливості мають функції СУММКВ (масив) та СУММПРОИЗВ (масив)?
2. Які особливості використання функції ЕСЛИ (умова; ТО; ІНАКШЕ)?
3. У чому особливості функції СТЬЮДРАСП (число; ступені свободи; хвост)?
4. Які особливості побудови лінійного графіка із рівнянням регресії?

Тема №12

Автоматизація алгоритму автоматизованого дисперсійного аналізу

Для того, щоб провести автоматизований дисперсійний аналіз, необхідно виконати такі операції.

1. Відкрити свій файл та створити новий аркуш з назвою “Дисп.анализ”. Створити на цьому аркуші таблицю (рис. 12.1).

	A	B	C	D	E	F	G
1		Генерация					Суммы
2		ММ	МО	М	Д1	Д3	сумм
3	Объем (n)						
4	Сумма (Sx)						
5	(Sx)^2/n						
6	Сумма квадратов (Sx^2)						
7							
8	Число групп=						
9	N=						
10							
11	Источник вариации	Сумма квадр.	Число ст.св.	Средн.квadrat	дисперс.отнош	Уровень знач.	Доля влияния
12		(SS)	(df)	(MS)	(F)	(p)	(h^2)
13	Между группами						
14	Внутри групп						
15							
16	Общее варьирование						
17							
18	Поправка на неравенство						
19	объемов групп						
20	n0=						
21	Поправочный ср.квadrat						
22	Исправленная доля влиян.						
23	Ошибка доли влияния						
24							

Рис. 12.1.

2. По-перше, перенесемо з аркушу “Корреляции” обсяги кожної з груп. Для цього, наприклад, к клітинку **B3** заносимо: =Корреляции!B6 . Аналогічно заповнюємо клітинки **C3:F3**, лише змінюючи їх адресацію.

3. Розраховуємо суми. Для цього, наприклад, у клітинку **B4** заносимо формулу: $\text{=СУММ('База данных'!H4:H393)}$. Аналогічно заповнюємо клітинки **C4:F4**, лише змінюючи їх адресацію.

4. У клітинку **B5** заносимо формулу: =B4^2/B3 . Розтягнемо цю формулу вправо до клітинки **F5**.

5. Розраховуємо суми квадратів. Для цього, наприклад, в клітинку **B6** заносимо формулу: $\text{=СУММКВ('База данных'!H4:H393)}$. Аналогічно заповнюємо клітинки **C6:F6**, лише змінюючи їх адресацію.

6. Нарешті, розраховуємо суми сум по всіх рядках. Для цього у клітинку **G3** вносимо формулу: =СУММ(B3:F3) , яку розтягуємо донизу таблиці.

7. Далі, для розрахунку факторіальної, випадкової й загальної сум квадратів дисперсійного комплексу нам необхідно розрахувати допоміжну величину N. Для цього в клітинку **B9** внесемо формулу: =(G4^2)/G3 .

8. В клітинку **B8** вносимо кількість градацій дисперсійного комплексу: =СЧЁТ(B3:F3) .

9. Наступна частина роботи присвячена безпосередньому проведенню дисперсійного аналізу. Для цього заповнюємо таблицю дисперсійного аналізу.

В стовпчик “Сумма квадратів” вводимо формули:

- в клітинку **B13**: $=G5-B9$;
- в клітинку **B16**: $=G6-B9$;
- в клітинку **B14**: $=B16-B13$.

В стовпчик “Число степеней свободи”:

- в клітинку **C13**: $=B8-1$;
- в клітинку **C16**: $=G3-1$;
- в клітинку **C14**: $=C16-C13$.

В стовпчик “Середній квадрат”:

- в клітинку **D13**: $=B13/C13$;
- в клітинку **D14**: $=B14/C14$.

10. Розрахуємо значення критерію Фішера. Для цього в клітинку **E13** вносимо таку формулу для розрахунку дисперсійного відношення: $=D13/D14$.

Далі визначимо його рівень значущості; для цього у клітинку **F13** вносимо формулу: $=FРАСП(E13;C13;C14)$.

11. Нарешті, після того як дисперсійний аналіз проведено, визначимо силу впливу фактора. Для цього в клітинку **G13** вносимо формулу: $=B13/B16$, а в клітинку **G14** формулу: $=B14/B16$.

12. Оскільки обсяги наших груп були не рівні, слід розрахувати оцінку сили впливу фактора з урахуванням нерівності цих груп. Для цього заповнюємо третю таблицю аркуша.

В клітинці **B20** розрахуємо середній зважений обсяг груп; для цього вводимо наступну формулу:

$$=(G3-((СУММКВ(B3:F3)/G3)))/(B8-1).$$

Далі розраховуємо поправочний середній квадрат; в клітинку **B21** вносимо формулу: $=(D13-D14)/B20$.

Нарешті, розраховуємо виправлену силу впливу фактора; в клітинку **B22** вносимо формулу: $=B21/(D14+B21)$. Його статистичну помилку розраховуємо за формулою: $=(1-B22)*C13/C14$.

13. Таким чином, у підсумку цей аркуш повинен мати наступний вигляд (Додаток 3).

Питання для самоперевірки:

1. Які особливості використання функції FРАСП (число; ступені свободи1; ступені свободи2)?
2. Як розраховується оцінка сили впливу фактора за Снедекором (при нерівності обсягів вибірок)?

Тема №13

Автоматизація аналізу терміну господарського використання тварин

Для того, щоб провести автоматизований аналіз терміну господарського використання тварин, необхідно виконати такі операції.

1. Відкрити свій файл та створити новий аркуш з назвою “**Вибраковка**”. На цьому аркуші створити таблицю (рис. 13.1).

	A	B	C	D	E	F	G	H	I
1	X _{min} =								
2	X _{max} =								
3	n=								
4									
5	Возраст	Накоплен.	Абсол.	Сохран-	Выбра-	Удельн.	Удельн.	Накоплен.	Ожидаем.
6	в лактац.	частота	сохран.	ность	ковка	выбрак.	сохран.	сохран-	продолж.
7	(x)	(Sx)	(Dx)	(Lx)	(dx)	(qx)	(px)	ность	использов.
8	1								
9	2								
10	3								
11	4								
12	5								
13	6								
14	7								
15	8								
16	9								
17	10								
18	11								
19	12								
20	13								
21									
22									
23	Средняя взвешенная выбраковка=								
24									

Рис. 13.1

2. Наступний аналіз дає можливість оцінити середній термін господарського використання тварин дослідного стада (у лактаціях) та повікову інтенсивність выбраковки тварин після кожної лактації – від першої до максимальної для стада.

По-перше, необхідно визначити мінімальну кількість лактацій для дослідних тварин, максимальну кількість лактацій та обсяг вибірки. (Нагадаємо, дані щодо кількості лактацій для кожної тварини занесено в клітинах R4:R393 аркушу “База даних”.)

Для цього в клітинку **B1** внесемо формулу:

=МИН('База даних'!R4:R393)

В клітинку **B2** - формулу: =МАКС('База даних'!R4:R393)

А в клітинку **B3** - формулу для розрахунку обсягу вибірки:

=СЧЁТ('База даних'!R4:R393)

Як бачимо, з 389 тварин дослідного стада, кількість лактацій варіювала від 1 до 13. Тому в таблиці (стовпчик “Возраст в лактациях”) вносимо цифри від 1 до 13.

3. По-друге, розрахуємо абсолютні накопичені частоти у відношенні кількості лактацій (тобто, кількість тварин, що мали не більше ніж 1, 2, 3 й т.п. лактацій). Для цього в клітинку **B8** вносимо формулу:

=ЧАСТОТА(База данных!\$R\$4:\$R\$393;A8)

й розтягнемо цю формулу донизу стовчика.

Таким чином, 7 тварин мали лише по одній лактації, 31 тварина – одну чи дві лактації, 98 тварин – одну, дві чи три й так далі.

4. Для розрахунку абсолютної збереженості використаємо ці частоти у такому вигляді. Ми визначили, що з 398 тварин дослідного стада кількість лактацій для окремих тварин становила від 1 до 13. Ті тварини, що мали по 13 лактацій мали також 12-ту, 11-ту, 10-ту, 9-ту, ..., другу й першу. Ті тварини, що мали по 12 лактацій мали також 11-ту, 10-ту, 9-ту, ..., другу й першу. І так далі.

Таким чином, всі тварини дослідного стада лактували хоча б один раз. Тоді, для віку “1-а лактація” у стовпчику абсолютна збереженість (клітка **C8**) повинно бути записано обсяг вибірки: **=B3**.

Визначена кількість тварин (ця кількість визначена в клітинці **B8**) мала лише одну лактацію й до другої були вибракувані. Тому в клітинку **C9** заносимо формулу: **=C8-B8** й розтягнемо цю формулу донизу стовпчика.

5. Далі, визначимо відносну збереженість, тобто, частку тварин, що зберігається у стаді після кожної лактації. Для цього в клітинку **D8** вносимо формулу: **=C8/\$C\$8**.

6. Розрахуємо частку вибракуваних тварин після кожної лактації (по відношенні до всього обсягу вибірки). Для цього в клітинку **E8** вносимо формулу: **=D8-D9**. Ми отримуємо цифру 0,018. Таким чином, після першої лактації зі стада вибракується 1,8 % тварин.

Розтягуємо цю формулу донизу стовпчика до **E20**. Оскільки після 13-ої лактації всі тварини вибракувані, то сума значень, що розраховані у клітинках цього стовпчика повинна дорівнювати одиниці. Перевіримо це. Для цього в клітинку **E21** вносимо формулу: **=СУММ(E8:E20)**

Якщо при розрахунку за цією формулою отримано 1,000, то всі попередні розрахунки відносної збереженості та вибракування зроблено без помилок.

7. Розрахуємо питому збереженість та питома вибракування. Перший показник визначає частку тварин, що зберігаються з кількості тварин заданого віку. А друга, навпроти, частку тварин, що вибракувані, з кількості тварин заданого віку.

Для цього в клітинку **F8** вносимо формулу: **=E8/D8** й розтягнемо її до кінця стовпчика. В клітинку **G8** – формулу: **=1-F8** й теж розтягнемо її донизу.

8. Далі розрахуємо показник середнього питомого вибракування. Для цього, спочатку, необхідно розрахувати суму коефіцієнтів збереженості. В клітинку **D22** вносимо формулу: **=СУММ(D8:D20)**. Тоді для розрахунку середнього питомого вибракування використаємо формулу, яку вносимо в

клітинку **E23**: $=1/D22$. Отримуємо значення 0,1853. Таким чином, в середньому 18,53 % тварин кожної лактації вибраковується до наступної.

9. Важливий показник при аналізі тривалості господарського використання – це очікувана тривалість використання. Але, по-перше, нам необхідно розрахувати накопичену збереженість для тварин кожного віку. Для цього в клітинку **H8** внесемо дані з клітини **D22**, тобто вводимо: $=D22$. Далі, в клітинку **H9** вводимо формулу: $=H8-D8$. Розтягнемо цю формулу до кінця стовпчика.

Нарешті, в клітинку **I8** вносимо формулу: $=H8/D8$. Її знову ж розтягнемо цю формулу донизу стовпчику.

Таким чином, ми отримали показники очікуваної тривалості використання тварин дослідного стада. Для первісток очікувана тривалість господарського використання становить трохи більше 5 лактацій (5,4). Для тварин, які вже мають одну лактацію, очікувана тривалість господарського використання становить 4,5 лактації (в середньому). Й так далі. Нарешті, для тварин 12-ї лактації, зрозуміло, очікувана тривалість господарського використання складає лише одну лактацію.

10. Зобразимо процес вибракування графічно. Для цього клацніть по піктограмі Головного меню “**Мастер диаграмм**”. У вікні, що відкриється, з набору “**Стандартные**” оберіть тип діаграми “**Точечная**”, а вид: “**Точечная диаграмма, на которой значения соединены отрезками**”. Клацніть по кнопці “**Далее**”.

У віконці “**Диапазон**” уведіть:

$= \text{Выбраковка!}\$A\$8:\$A\$20; \text{Выбраковка!}\$D\$8:\$D\20

та клацніть по кнопці “**Далее**”.

У віконці “**Ось X (категорий)**” уведіть “**Возраст (лактация)**”, а у віконці “**Ось Y (значений)**” – “**Сохранность**”. Клацніть по закладці “**Легенда**” та відключіть режим наявності легенди. Для цього клацніть по віконці “**Добавить легенду**” так, щоб у ньому зник знак “ \square ”. Клацніть по кнопці “**Готово**”.

11. Після того, як графік буде встановлено у відповідному місці, нам необхідно трохи його підправити.

Для цього підведіть курсор до осі X та клацніть правою кнопкою миші (ПКМ). З’явиться меню, де вибрати режим “**Формат оси**”. Встановити такі зміни:

- у закладці “**Шкала**”: мінімальне значення – 1; максимальне значення – 13; ціна ділення – 1;
- у закладці “**Шрифт**” обрати розмір 10. Клацнути “**ОК**”.

Аналогічно підправимо вісь OY. Тут встановимо такі значення: мінімальне – 0,0, максимальне – 1,0, ціна ділення – 0,1. Розмір шрифту також оберемо 10.

Очистимо лінії сітки та сірий фон графіка. Для цього встановимо курсор на будь-якій лінії сітки та клацнемо ПКМ. У меню, що відкриється, обрати опцію “**Очистить**”. Далі клацнемо по області побудови графіка (сірий прямокутник) та повторимо ті ж самі дії.

Питання для самоперевірки:

1. Які показники використовуються при аналізі терміну господарського використання корів?
2. Які особливості має побудова діаграм типу «Точечная»?
3. Як змінити шифр на діаграмі?

Тема №14

Моделювання селекційного процесу за допомогою ПЕОМ

Для того, щоб провести моделювання селекційного процесу за допомогою ПЕОМ, необхідно виконати такі операції.

1. Відкрити свій файл та створити новий аркуш з назвою “**Модель**”. Створити на цьому аркуші таблицю (рис. 14.1).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2	Граница	Объем		Доля		Нормиров. отклонение	$f(\xi)$	$\Phi(\xi)$	Интенс. селекции	Среднее тов.стад.	Среднее племядра	Ожидаемая продуктивн. дочерей		
3	отбора	товарн. стада	плем- ядра	товарн. стада	плем- ядра							t	i	Среднее
4	(кг)													
5														
6														
7														
8														
9														
10														
11														
12														
13														
14														
15														
16	Доля т/с	Доля п/я	Объем	t	Граница	$f(\xi)$	i	Среднее		Ожид.прод.дочерей				
17	$p(\Phi(\xi))$		п/я		отбора			тов.ст.	п/ядра	мин.	макс.			
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														

Рис. 14.1

2. В клітинки **A5:A13** перенесіть з листа “Гистограмма” значення верхньої межі кожного із дев’яти інтервалів.

3. В клітинки **B5:B13** перенесіть з листа “Гистограмма” значення накопичених частот для кожного із дев’яти інтервалів.

4. В клітинках **C5:C13** розрахуйте обсяг племінного ядра для кожного із дев’яти інтервалів, так, щоб у сумі із обсягом товарного стада, сумарний обсяг вибірки становив 390 голів.

5. В клітинках **D5:D13** та **E5:E13** розрахуйте частки особин, що входять до складу товарного стада та племінного ядра для різних інтервалів.

6. В клітинку **F5** внесіть формулу для розрахунку нормованого відхилення: $\frac{=(A5-Парн.сравн.!$C$5)/Парн.сравн.!$E$5}$ та розтягніть цю формулу до кінця стовпчика.

7. В клітинку **G5** внесіть формулу: $=НОРМРАСП(F5;0;1;0)$, а у клітинку **H5** – формулу: $=НОРМРАСП(F5;0;1;1)$ та розтягніть їх до кінця стовпчика.

8. В клітинку **I5** внесіть формулу $=G5/(1-H5)$ та розтягніть її до кінця стовпчика.

9. В клітинку **J5** внесіть формулу:

$\text{=Парн.сравн.!}\$C\$5-(G5/H5)*\text{Парн.сравн.!}\$E\$5$

а в клітинку **K5** – формулу:

$\text{=Парн.сравн.!}\$C\$5+I5*\text{Парн.сравн.!}\$E\$5$

та розтягніть їх до кінця стовпчика.

10. Нарешті, в клітинку **L5** та клітинку **M5** внесіть такі формули:

$\text{=Парн.сравн.!}\$C\$5+\text{Корреляции!}\$K\$18*I5*\text{Парн.сравн.!}\$E\$5$

$\text{=Парн.сравн.!}\$C\$5+\text{Корреляции!}\$L\$18*I5*\text{Парн.сравн.!}\$E\$5$

та розтягніть їх до кінця стовпчика.

11. Використавши “Мастер діаграм” (режим “Точечная”) побудуйте графік залежності очікуваної продуктивності дочок від границі відбору (рис. 14.2).

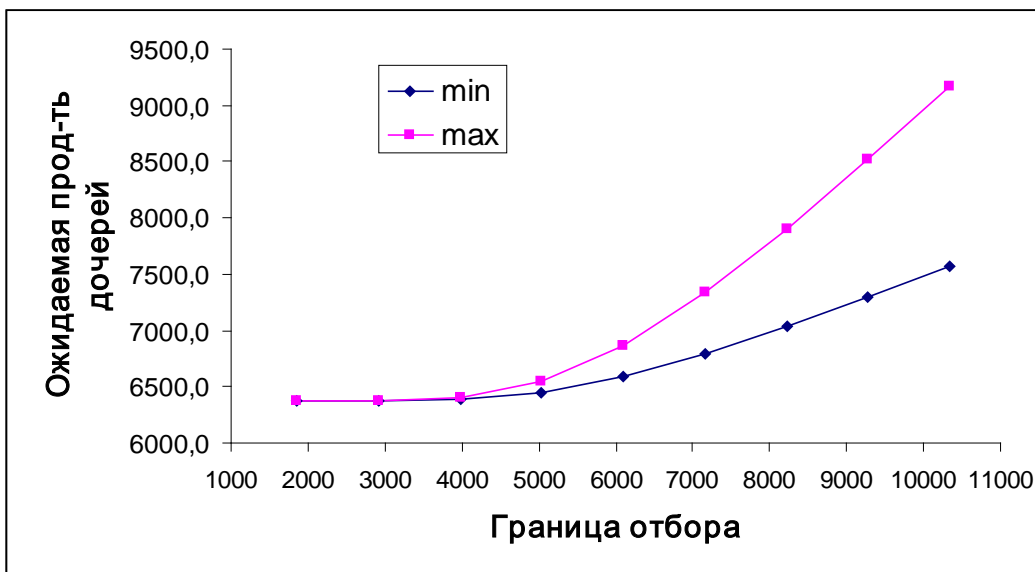


Рис. 14.2

12. В клітинки **A18:A26** внесіть значення від 0,1 до 0,9 із кроком 0,1.

13. В клітинках **B18:B26** розрахуйте значення, які у сумі з величинами із клітинок **A18:A26** даватимуть одиницю.

14. В клітинках **C18:C26** розрахуйте відповідні до частот, відповідні значення обсягу племінного ядра, з урахуванням того, що ці цифри повинні бути цілими.

15. В клітинку **D18** внесіть формулу $\text{=НОРМСТОБР}(A18)$ та розтягніть її до кінця стовпчика.

16. В клітинку **E18** внесіть формулу:

$\text{=Парн.сравн.!}\$C\$5+D18*\text{Парн.сравн.!}\$E\$5$

та розтягніть її до кінця стовпчика.

17. В клітинку **F18** внесіть формулу: $\text{=НОРМРАСП}(D18;0;1;0)$ та розтягніть її до кінця стовпчика.

18. В клітинку **G18** внесіть формулу: $\text{=F18}/(1-A18)$ та розтягніть її до кінця стовпчика.

19. В клітинки **H18** та **I18** внесіть, відповідно, такі формули:

=Парн.сравн.! $\$C\$5-(F18/A18)*$ Парн.сравн.! $\$E\5

=Парн.сравн.! $\$C\$5+G18*Парн.сравн.!\$E\5

та розтягніть їх до кінця стовпчиків.

20. В клітинки **J18** та **K18** внесіть, відповідно, такі формули:

=Парн.сравн.! $\$C\$5+Корреляции!\$K\$18*G18*Парн.сравн.!\$E\5

=Парн.сравн.! $\$C\$5+Корреляции!\$L\$18*G18*Парн.сравн.!\$E\5

та розтягніть їх до кінця стовпчиків.

21. Використавши “Мастер диаграмм” (режим “Точечная”), побудуйте графік залежності очікуваної продуктивності дочок від частки відібраних у племінне ядро тварин (рис. 14.3).

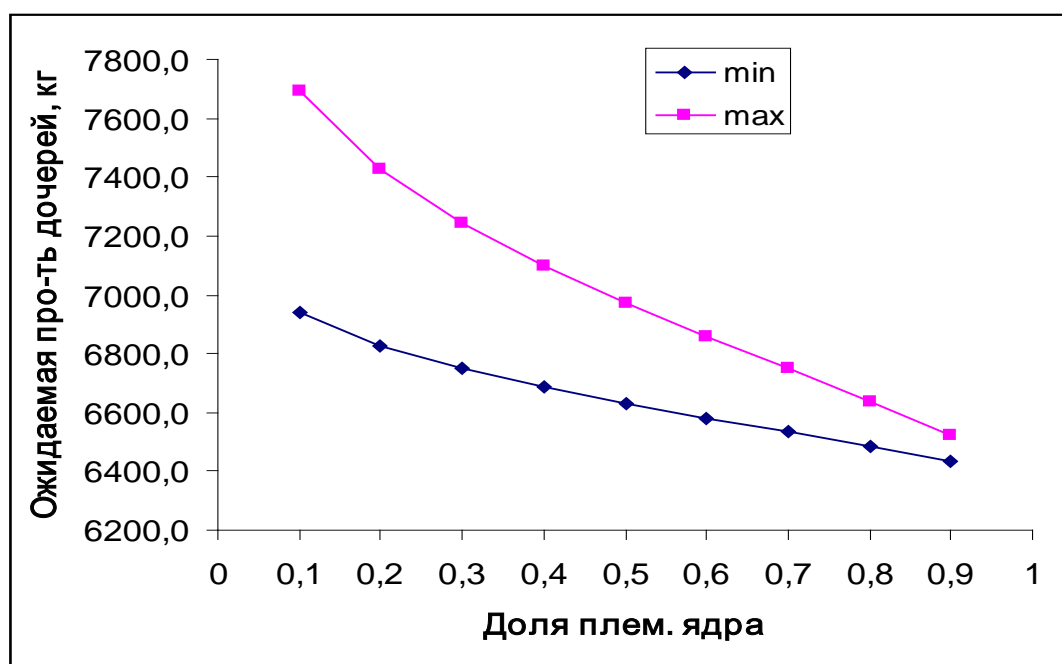


Рис. 14.3

Питання для самоперевірки:

1. Які показники використовуються при аналізі селекційного процесу?
2. Як розрахувати очікувану продуктивність дочок залежно від рівня межі відбору?
3. Як розрахувати очікувану продуктивність дочок залежно від обсягу племінного ядра?

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Боровиков В. П. STATISTICA® - Статистический анализ и обработка данных в среде Windows ® / В. П. Боровиков, И. П. Боровиков. – М. : Информационно-издательский дом «Филинь», 1997. – 608 с.
2. Бююль А. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей / А. Бююль, П. Цёфель. – СПб. : ООО «ДиаСофтЮП», 2001. – 608 с.
3. Компьютерная биометрика / под ред. В. Н. Носова. – М. : Изд-во МГУ, 1990. – 232 с.
4. Лапач С. Н. Статистика в науке и бизнесе / С. Н. Лапач, А. В. Чубенко, П. Н. Бабиц. – К. : МОРИОН, 2002. – 640 с.
5. Манько А. А. Статистический анализ в MS Excel / А. А. Манько. – М. : Издательский дом «Вильямс», 2004. – 448 с.
6. Меркурьева Е. К. Биометрия в селекции и генетике сельскохозяйственных животных / Е. К. Меркурьева. – М. : Колос, 1970. – 423 с.
7. Плохинский Н. А. Руководство по биометрии для зоотехников / Н. А. Плохинский. – М. : Колос, 1969. – 256 с.
8. Практикум по методике опытного дела в защите растений / [В. Ф. Пересыпкин, С. Н. Коваленко, В. С. Шелестова и др.]. – М. : Агропромиздат, 1989. – 175 с.
9. Комп'ютерні методи в сільському господарстві та біології: Навчальний посібник / [О. М. Царенко, Ю. А. Злобін, В. Г. Скляр и др.]. – Суми : Видавництво “Університетська книга”, 2000. – 203 с.
10. Peakall R. GenAIEx 6: Genetic analysis in Excel. Population genetic software for teaching and research / R. Peakall, P. E. Smouse / Molecular Ecology Notes. – 2006. – V. 6. – P. 288-295.

Навчальне видання

СПЕЦІАЛЬНІ ІНФОРМАЦІЙНІ СИСТЕМИ І ТЕХНОЛОГІЇ

Методичні рекомендації

Укладачі: Крамаренко Сергій Сергійович
Волосюк Юрій Вікторович

Формат 60×84.1/16. Ум. друк. арк. 0,9

Тираж ___ прим. Зам № _____

Надруковано у видавничому відділі
Миколаївського національного аграрного університету.
54020, м. Миколаїв, вул. Георгія Гонгадзе, 9

Свідоцтво суб'єкта видавничої справи ДК № 4490 від 20.02.2013 р.