

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
МИКОЛАЇВСЬКИЙ НАЦІОНАЛЬНИЙ АГРАРНИЙ УНІВЕРСИТЕТ

**Факультет технології виробництва і переробки
продукції тваринництва, стандартизації та біотехнології**

**ЗАПРОВАДЖЕННЯ МЕТОДІВ ГЕНОМНОЇ СЕЛЕКЦІЇ ДЛЯ
ПІДВИЩЕННЯ ЯКІСНИХ ОЗНАК М'ЯСА СВИНЕЙ ТА ОВЕЦЬ
(виробничо-практичні рекомендації)**



**Миколаїв
2020**

УДК 636.4.082
З-33

Рекомендовано до друку рішенням вченої ради Миколаївського національного аграрного університету від «22» грудня 2020 р., протокол № 4.

Укладачі:

- С. І. Луговий - д-р с.-г. наук, доцент, в.о. завідувача кафедри генетики, годівлі тварин та біотехнології, Миколаївський національний аграрний університет;
- С. С. Крамаренко - д-р біол. наук, доцент, професор кафедри генетики, годівлі тварин та біотехнології, Миколаївський національний аграрний університет;
- О. С. Крамаренко - канд. с.-г. наук, ст. викладач кафедри технології переробки, стандартизації і сертифікації продукції тваринництва, Миколаївський національний аграрний університет;
- В. Я. Лихач - д-р с.-г. наук, доцент, провідний науковий співробітник, Миколаївський національний аграрний університет;
- І. П. Атаманюк - д-р техн. наук, професор, завідувач кафедри вищої та прикладної математики, Миколаївський національний аграрний університет.

Рецензенти:

- Р. Л. Сусол - д-р с.-г. наук, доцент, завідувач кафедри технології виробництва та переробки продукції тваринництва, Одеський державний аграрний університет;
- О. О. Стародубець - канд. с.-г. наук, доцент, доцент кафедри птахівництва, якості та безпечності продукції, Миколаївський національний аграрний університет.

З М І С Т

ВСТУП	4
1. ІСТОРИЧНИЙ АСПЕКТ РОЗВИТКУ МЕТОДІВ АНАЛІЗУ КІЛЬКІСНИХ ОЗНАК	5
2. ОЦІНКА ПАРАМЕТРІВ	9
3. ВИПАДКОВІ І ФІКСОВАНІ ФАКТОРИ. ЗМІШАНА МОДЕЛЬ	20
ЗАКЛЮЧЕННЯ	34
СЛОВНИК ТЕРМІНІВ	35
СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ	39

ВСТУП

Виявлення локусів кількісних ознак (QTL) і оцінки параметрів, потребувало розробок в статистиці, «класичній генетиці» або розведенні, і біохімії.

Найбільш повно оцінити кількісні ознаки та оцінити вплив генетичних та паратипових факторів можливо за рахунок процедури найкращого лінійного незміщеного прогнозу (Best Linear Unbiased Prediction – BLUP), заснованої на лінійних статистичних моделях змішаного типу.

Оцінки племінної цінності, що розраховані за методом BLUP, мають найменшу дисперсію помилки, та при інших рівних умовах будуть з найбільшою ймовірністю відображати справжню генетичну цінність особини.

Змінні, як правило, поділять на дві групи: фіксовані і випадкові. «Параметри» визначаються як фіксовані змінні, що описують статистичний розподіл популяції. Наприклад, популяція з нормальним розподілом описується двома параметрами: середнім і дисперсією. Як правило, параметри оцінюються на основі вибірки, отриманої з популяції. Оцінки параметрів, отримані з даних вибірки, позначаються «статистиками».

У розділі 2 ми розглянемо принципи оцінки параметрів як фіксованих, так і випадкових величин.

Основа теорії селекційного індексу в тому, що полігенні племінні цінності для кількісних ознак слід вважати випадковими, так як ці фактори «взяті» з нормального розподілу факторів з відомою дисперсією. У генетичній оцінці на основі польових записів, також буде необхідно включити фіксовані фактори, такі як стадо або блок, в модель. Таким чином, аналізовані моделі будуть включати як фіксовані, так і випадкові ефекти на додаток до залишкових.

У розділі 3 ми коротко опишемо аналіз змішаних моделей, моделей, до яких входять як фіксовані, так і випадкові величини.

1. ІСТОРИЧНИЙ АСПЕКТ РОЗВИТКУ МЕТОДІВ АНАЛІЗУ КІЛЬКІСНИХ ОЗНАК

Сучасна генетика, як звичайно вважається, почалася з повторного відкриття робіт Менделя в 1900 році. Проте, існували значні зрушення і в статистиці, і в цитогенетиці до цієї межової дати, важливість яких стала очевидною лише пізніше. В області статистики, Пірсон в 1890 році дав визначення коефіцієнту кореляції, і показав, що він може бути використаний для опису взаємозв'язку між двома змінними. Протягом останніх десятиліть 19-го століття, були зроблені також важливі успіхи в цитології: були виявлені хромосоми, і були виявлені і описані етапи, як мейозу, так і мітозу.

Повторне відкриття законів Менделя призвело до швидкого першого синтезу генетики, статистики та цитології. Бовери (1902) і Саттон (1903) вперше запропонували "хромосомну теорію спадкування", припускаючи, що фактори Менделя були пов'язані з хромосомами. Використовуючи дрозофілу, Морган (1910) показав, що Менделевські гени були пов'язані, і можуть бути відображені лінійною групою зчеплення, числом рівному гаплоїдному набору хромосом. Гарді і Вайнберг в 1908 році представили своє знамените рівняння для опису розподілу генотипів в ізольованих популяціях в рівноважному стані. У 1919 році Холден вивів формулу для перетворення частоти рекомбінації в адитивні «одиниці карт», що позначені «морганідами» або «сантіморганідами», припускаючи, випадковий розподіл подій рекомбінацій вздовж хромосоми.

Більшість ознак, що цікавлять, показують безперервну зміну, на відміну від дискретного розподілу пов'язаного з менделевськими генами. Незважаючи на ранній синтез між менделевською генетикою і цитогенетикою, здавалося, немає очевидного зв'язку між генетикою Менделя, з одного боку, і кількісними змінами та природним відбором з іншого. Експерименти Johanssen (1903) з квасолею показали, що фактори навколишнього середовища є основним джерелом зміни у кількісних ознак, що призвело до висновку, що фенотип для цих ознак не є надійним показником для генотипу. Юл в 1906 першим запропонував, що безперервну зміну можна пояснити сукупною дією багатьох менделевських генів, кожен з яких має невеликий вплив на показники ознаки. Фішер в 1918 році показав, що поділ кількісних генів в ауткросних популяціях буде генерувати кореляції між родичами. Рауне (1918) показав, що X-хромосома обраних ліній дрозофіли містить безліч факторів, які впливають на кількість щетинок. Таким чином, до 1920 року, було створено основні теорії необхідні для виявлення окремих генів, що впливають на кількісні ознаки.

В експерименті Сакса 1923 з бобами він показав, що вплив індивідуального локусу на кількісну ознаку може бути відокремлено за допомогою серії схрещувань, в результаті чого буде рандомізовано генетичний фонд по відношенню до всіх генів не пов'язаних з генетичними маркерами, що спостерігаються. Навіть притому, що всі його маркери були морфологічними маркерами насіння з повним домінуванням, він зміг показати істотний вплив на масу насіння, пов'язану з деякими його маркерами.

З початку історії, селекція рослин і тварин була заснована на відборі особин з бажаним фенотипом як батьків для наступного покоління. Порівняння між одомашненими популяціями і їх дикими предками свідчить, що штучний відбір був досить успішним у зміні фенотипу без будь-якого формального знання генетики. Райт, Холдейн і Фішер, завершивши синтез між дарвінізмом і менделізмом в серії робіт з 1924 по 1931 років, показали, як природний відбір міг працювати на менделевських факторах, контролюючи кількісні ознаки, що піддавалися селекції. Фішер також показав, що фактори Менделя можуть прояснити фенотипічну схожість між родичами. Ці принципи стали основою для наукової селекції тварин і рослин з 1930-х років.

Використовуючи генетичні та статистичні знання, накопичені аж до 1940 року, Лаш і Хейзел розробили принципи селекційного індексу для оптимізації штучного відбору на основі відомих відносин між особинами і фенотипічною інформацією ознаки. Селекційний індекс виявився дуже ефективною і гнучкою методологією для практичної селекції рослин і тварин. Не тільки селекція могла бути економічно оптимізована, а й очікувані доходи від селекції також могли бути передбачені.

Теорія селекційного індексу була дуже мало пов'язана з генетикою Менделя. "Нескінченна модель" розроблена Фішером (1918) передбачала, що кожна кількісна ознака контролювалася багатьма незалежно сегрегуючими менделевськими генами, діючими всі разом адитивним чином, і кожен окремий локус вносив нескінченно малий внесок у загальну генетичну дисперсію. Проте, майже ідентичні результати могли бути отримані, якщо ознака контролювалася тільки деякими локусами. Тільки «адитивна» генетична варіація була розглянута в базовій моделі. Домінування (взаємодії між алелями в межах гену), і взаємодія між генами (епістаз) були за рамками селекційного індексу.

Ця біометрична методологія була розроблена протягом 1950-х, 1960-х і 1970-х років головним чином Ч.Р. Хендерсоном і його колегами. Використовуючи матричний запис, Хендерсон розробив рівняння «змішаної моделі», що поєднують оцінку найменших квадратів з селекційним індексом для того, щоб отримати незсунені точні оцінки генетичних значень особин відібраних в різних середовищах, таких як стада або блоки. Він також розробив методи для отримання незміщених оцінок генетичних і середовищних компонентів варіанс, необхідних для вирішення цих рівнянь. Нарешті, він розробив простий алгоритм знаходження оберненої матриці «визначника відносин». Це зробило можливим включення інформації про всіх відомих родичів для отримання генетичних оцінок. Жодна з цих методологій, однак, не вимагає будь-якої інформації про конкретну генетичну структуру ознаки, що селекціонують.

Протягом 50-річного періоду (1930-1980), було порівняно мало успішних експериментів, які знайшли маркер - ЛКП зв'язок в популяціях рослин і тварин, і з них ще менше були незалежно повторюваними. Одна з основних проблем, яка триває по сьогоднішній день - це відносно невеликий розмір більшості

експериментів. У більшості випадків, в яких QTL ефекти не були знайдені, потужність була занадто низькою, щоб знайти сегрегацію QTL прийнятної величини (Сольєр та ін., 1976). У період 1960-1980 були зроблені важливі методичні досягнення у виявленні QTL і оцінювання параметрів, хоча відсутність сегрегуючих маркерів була, поза всяким сумнівом, основним обмежуючим фактором для цієї технології.

У 1961 році Neimann - Soressen і Робертсон запропонували конструкцію полусибсів для виявлення QTL в товарних популяціях молочної великої рогатої худоби. Хоча фактичні результати були такими, що розчаровують, це була перша спроба виявити QTL в існуючій сегрегаційній популяції. Всі попередні дослідження базувалися на експериментальних популяціях, розроблених спеціально для виявлення QTL. Це дослідження було також новаторським і в інших аспектах. Це було перше дослідження, де використовувалися групи крові, а не морфологічні маркери, і запропоновані статистичні аналізи, χ^2 тест (хі-квадрат), заснований на квадраті суми нормальних розподілів, і аналіз варіанс (ANOVA) були також унікальні. Це було перше дослідження, яке намагалось оцінити силу для виявлення QTL, і розглянути проблему множинних порівнянь. Law (1965) завершив перший успішний експеримент по відображенню QTL (на відміну від простого виявлення) у сільськогосподарських видів. Він локалізував QTL в пшениці за допомогою заміни ліній.

Jaayakar (1970) припустив, що метод максимальної правдоподібності може бути використаний для відображення QTL. Два роки по тому, Газемана і Елстон (1972) запропонували сібс-парний метод аналізу для виявлення QTL в людських популяціях. Вони також представили функцію правдоподібності для оцінки частоти рекомбінації і параметрів QTL. Сольєр та ін. (1976) і Сольєр і Genizi (1978) розробили формули для оцінки статистичної потужності QTL для кросів між інбредні лініями і сегреційними популяціями. Для сегреції популяцій вони розглядали великі сім'ї напів- сибсів і повних - сибсів. Їх дослідження ясно показали, що дуже великі вибірки, як правило, більше 1000 фізичних особин, були зобов'язані отримати достатню потужність виявлення QTL, пояснюючи 1% від фенотипічної дисперсії.

Дослідження механізму зчеплення Маркер - QTL зажадало поліморфні гени з класичним менделевським успадкуванням. У дрозофіли лінії, що несуть кілька мутацій служили цієї мети дуже ефективно. Однак це був не той випадок для людини або сільськогосподарських видів. У рослин, спочатку єдиними доступними маркерами були гени, які призводили до морфологічних відмінностей. Очевидно, вони були недостатні для покриття всього генома. Крім того, більшість морфологічних маркерів показували повне домінування. Нарешті, прямий ефект на фенотип більшості з цих маркерів було досить значним. Таким чином, навіть коли було знайдено вплив на ознаку, що цікавить, пов'язаний з маркером, було дуже ймовірно, що це був плейотропний ефект маркера. У сільськогосподарських тварин, дослідження зв'язку маркер-ЛКП, як правило, здійснюються в межах популяцій, і вимагають в якості маркерів локуси, поліморфні популяції, що цікавлять. До 1980 року єдиними

підходящими менделевськими локусами були групи крові, які, природно превалювали у всіх популяціях, часто були багатоалельні, і не мали ніякого видимого ефекту на фенотип для будь-яких цікавих ознак. Проте, зрештою стало ясно, що загальне число поліморфних локусів крові було цілком обмежене. Таким чином, групи крові не були рішенням для виявлення QTL в популяціях тварин.

Перший біохімічний поліморфізм був виявлений для серповидно-клітинної анемії Полінга в 1949 році. Левонтін і Хуббі показали в 1966 році, що електрофорез може бути використаний для розкриття більшої кількості природного поліморфізму ферментів у дрогофілі. Майже всі проаналізовані ферменти показали деякий поліморфізм, що міг бути виявлений швидкістю пересування в електричному полі. Це велика кількість природних поліморфізмів створила справжній шок в науковому співтоваристві. Так як, здавалося, не було ніякого адекватного пояснення, чому ця зміна була підтримана. Пізніше дослідження з одомашненими видами рослин і тварин виявили, що електрофоретичні поліморфізми були набагато менш поширені в сільськогосподарських популяціях. У 1980-х був ряд досліджень по виявленню QTL у сільськогосподарських рослин на основі ізоферментів за допомогою схрещування різних ліній або навіть видів в цілях отримання достатньої кількості електрофоретичних поліморфізмів (Tanksley та ін, 1982; Келеровим і Wherhahn, 1986; Edwards і ін., 1987; Weller та ін., 1988). Було ясно, однак, що, природні біохімічні поліморфізми були недостатні для повного аналізу генома в популяціях, що цікавили.

Перші виявлені поліморфізми ДНК - рівня були поліморфізми довжин рестрикційних фрагментів (ПДРФ). Grodzicker ін. (1974) вперше показали, що довжини рестрикційних фрагментів можуть бути використані для виявлення генетичних відмінностей у вірусах. Кан і Dozy (1978), використовували методи, розроблені Саус (1975) для виявлення поліморфізмів поблизу гена гемоглобіну людини. У подальшому році Соломон і Бодмер (1979) і Ботстайн ін. (1980) запропонували RFLP в якості головного джерела поліморфізму, який може бути використаний для генетичного картування. Хоча ПДРФ є діалельним, початкові теоретичні дослідження показали, що вони можуть бути присутніми по всьому геному. Бекман і Сольєр (1983) запропонували використовувати ПДРФ для виявлення і відображення QTL. Перше геномне сканування для QTL за допомогою ПДРФ було проведено на помідорі Патерсоном і співавт. (1988). З тих пір багато досліджень з аддитивного картування QTL на основі ПДРФ було успішно проведено на видах рослин. У видів тварин, проте, ПДРФ маркери, через їх діалельну природу, були гомозиготними в більшості особин, і тому виявилися не настільки корисні для відображення QTL, як спочатку очікувалося.

Основний прорив стався наприкінці десятиліття з відкриттям ДНК мікросателітів. Мюлліса ін. (1986) запропонував «полімеразну ланцюгову реакцію» (ПЛР) щоб спеціально розмножити будь-яку конкретну коротку послідовність ДНК. Використовуючи ПЛР, досить велика кількість ДНК може

бути згенерована так, щоб стандартні аналітичні методи могли бути застосовані для виявлення поліморфізмів, що складаються лише з одного нуклеотиду. З 1960-х років було відомо, що ДНК вищих організмів містить часті повторювані послідовності. У 1989 році три лабораторії незалежно виявили, що короткі послідовності повторюваної ДНК були вельми поліморфними по відношенню до кількості повторів повторюваної одиниці (ЛПТ і Luty, 1989; Тауц, 1989; Weber и Мая 1989). Найбільш поширеними з цих повторюваних послідовностей були полі-(TG), і було виявлено, що вони дуже поширені у всіх вищих видах. Ці послідовності були позначені як «прості послідовні повтори» (SSR) або «ДНК мікросателіти».

Нарешті, майже всі головні генетичні маркери були на руках. Мікросателіти були поширені у всіх геномах, що цікавлять. Майже всі полі-(TG) сайти були поліморфними, навіть у товарних популяціях тварин. Ці маркери були за визначенням «кодомінантними». Тобто, гетерозиготний генотип можна було відрізнити від будь-якого гомозиготного. Крім того, мікросателіти були майже завжди поліалельні. Таким чином, більшість особин було гетерозиготні. Коротше кажучи, "Те, що лікар прописав!". Генетичні карти щільності на основі мікросателітів були отримані для більшості сільськогосподарських видів і їх можна знайти в Інтернеті (наприклад, <http://www.marc.usda.gov/genome/>).

З 1995 року нові класи маркерів також стали вживатись. Головними з них є «поодинокі нуклеотидні поліморфізми» (SNP) (розглянуті Брукс, 1999). SNP, як правило, визначається як місце пари основ (п.о.), в якому частота найбільш загальної пари основ нижче, ніж 99%. На відміну від мікросателітів, які зазвичай мають кілька алелей, SNP, як правило, діалельні, але набагато більш поширені в усьому геномі, за оцінками частота 1 SNP на 300-500 п.о.

2. ОЦІНКА ПАРАМЕТРІВ

2.1. Принципи оцінки параметрів

Основні поняття оцінки параметрів, як правило, розкриті тільки дуже біглим способом у вступних курсах статистики. Так як ці принципи є центральними в методах, які розглядатимуться у наступних розділах, представлений огляд принципів оцінок параметрів. У розділі 2.2 ми пояснимо бажані властивості оцінок параметрів.

У цьому розділі ми будемо розглядати тільки основні методи, які були використані для оцінки параметрів QTL. Деякий з матеріалу вимагає знання матричної алгебри. Протягом решти цього тексту ми постараємося дотримуватися умовностей, що параметри позначаються грецькими символами, тоді як статистики позначені латинськими символами.

У розділах, які використовують матричну алгебру, вектори будуть позначатися малими літерами жирним шрифтом, а матриці будуть позначатися у верхньому регістрі жирним шрифтом. Транспонування матриці будемо позначати через апостроф. Обернену матрицю будемо позначати в ⁻¹ індексі.

Оцінка найменших квадратів і максимальної правдоподібності (ML) буде описано детально в розділах 2.4-2.13. Метод за оцінками «моментів», Байєсовські оцінки і мінімальні відмінності (MD) оцінки будуть описано в більш поверхневій формі.

2.2 Бажані властивості оцінки параметрів QTL

При загальному розгляді питання оцінки параметрів, існує чотири основних бажаних властивості оцінок: незміщеність, мінімальна дисперсія помилки оцінювання, оцінка в просторі параметрів і стійкість. Для простих ситуацій можливе отримати оцінювання з усіма цими властивостями, але при більш складних випадках, що не буде можливим отримання оцінок з усіма бажаними властивостями, і це буде питанням про компроміси. Опишемо ці властивості докладніше.

Незміщеність: припустимо, що $\hat{\theta}$ є оцінкою параметра θ . $\hat{\theta}$ є незміщеними, якщо $E(\hat{\theta})=\theta$, тобто, очікування оцінки дорівнює значенню параметра. Наприклад, при оцінці варіанси заснованої на середній вибірки, ми ділимо суму квадратів на $n-1$, де n розмір вибірки. Якщо замість цього ми поділимо на n , то оцінка буде зміщеною оцінкою варіанси.

Мінімальна варіанса помилки оцінювання: визначається як значення θ , для яких $E[(\hat{\theta}-\theta)^2]$ є мінімальним. Ця властивість лежить в основі оцінки найменших квадратів. Оцінювання з мінімальною варіансою помилки оцінювання також називають «найкращою» оцінкою.

Оцінки в рамках простору параметрів: прості приклади оцінок за межами простору параметрів це негативні оцінки компонент варіанси, кореляційні оцінки > 1 або < -1 або оцінки для частот рекомбінації < 0 або $> 0,5$. Хоча вимога оцінок в рамках простору параметрів може здатися тривіальним, це часто не так. У багатьох випадках не можливо, щоб отримати оцінку, яка є і незміщеною і в просторі параметрів. Проблеми параметра оцінки поза простором параметрів будуть розглянуті більш докладно в розділі 3, в контексті оцінки компонент варіанс і коваріанс. Оцінки ML (MLEs) завжди в просторі параметрів, тому що оцінка параметра поза простором параметрів має ймовірність рівну нулю за визначенням.

Стійкість: оцінка, $\hat{\theta}$, яка вважається «стійкою», якщо $\hat{\theta}$ прагне до θ , коли розмір вибірки прагне до нескінченності. Оцінка може бути стійкою, навіть якщо вона є зміщеною. Розглянемо наведений вище приклад з оцінки варіанси вибірки. Якщо ми розділимо на n замість $n-1$, оцінка є зміщеною, але стійкою, тому як n прямує до нескінченності, n прагне $n-1$. Хоча ця властивість також здається тривіальним, це особливо важливо для виявлення QTL, через неповний зв'язок між QTL і генетичними маркерами. У більшості випадків, ефект від кількісної ознаки, асоційованої з генетичним маркером є нестійкою оцінкою QTL ефекту.

Додатковою бажаною властивістю оцінок є надійність. Ця властивість міри як оцінювання залежить від неточностей в припущеннях, що використовуються для виведення оцінки. Наприклад, більшість методологій оцінки, які будуть використані, припускають що в основі лежить нормальний

розподіл залишків. Звичайно немає змінних з повністю нормальним розподілом. Одна з потенційних проблем це «викиди», спостереження, які відхиляються далі, ніж очікувалося від середнього, у зв'язку з наслідками не включаються в модель аналізу. Ці спостереження можуть потенційно мати дуже значний вплив на оцінки параметрів, особливо якщо оцінка заснована на мінімізації квадратів різниць. Як правило, надійність зменшуватиметься, коли більш конкретні припущення зроблені відносно передбачуваного розподілу.

2.3 Метод моментів оцінювання

Цей метод на даний час не має загального користування, і його інтерес можна розглядати чисто історично. Метод моментів оцінки був використаний Жученко та ін. (1979а, б) для оцінки параметрів QTL при зворотному схрещуванні. М-тий центральний моменти вибірки, T^m , обчислюється таким чином:

$$T^m = (1/N) \sum (y - y_{cp})^m \quad (2.1)$$

де N є розмір вибірки і y_{cp} є середнім вибірки. Перший центральний момент дорівнює нулю, а другий центральний момент вибірки є оцінкою варіанси розподілу. Статистики g_1 і g_2 , які використовуються для оцінки асиметрії та ексцесу розподілу, є похідними від третього і четвертого центральних моментів, відповідно.

Переваги методу моментів є те, що він легкий в застосуванні, оцінки незсунені і немає ніяких припущень, які робляться щодо властивостей, що лежать в основі розподілу QTL. До недоліків можна віднести, що оцінки параметрів можуть бути отримані поза простором параметра, наприклад, негативні оцінки дисперсії, або рекомбінації частот поза діапазону 0-0,5, і що не вся інформація в даних використана. Багато з оцінок параметрів, отриманих за Жученко та ін. (1979а, б) були поза простору параметрів.

2.4 Оцінювання параметрів методом найменших квадратів

Ми будемо використовувати матричні позначення, щоб коротко описати оцінку найменших квадратів. Припустимо, що існує ряд спостережень для деякої змінної y , які ми хочемо моделювати з точки зору інших змінних, для яких дані також доступні. Ми будемо позначати y як «залежну змінну», а інші змінні як «незалежні змінні». Мета полягає в тому, щоб «пояснити» залежну змінну з точки зору серії параметрів оцінок, що пов'язують залежні змінні з незалежною змінною. Тобто, щоб вивести функцію незалежних змінних, яка буде апроксимувати спостереження y . Як правило, не можливо повністю пояснити y через залежні змінні. Різниця між оцінками y , заснованих на незалежних змінних і оцінками параметрів, позначається як «помилка» або «залишок» моделі.

Оцінка найменших квадратів заснована на знаходженні оцінок параметрів, які мінімізують сподівання суми квадратів помилок. Таким чином, за визначенням цей метод має мінімальний розкид варіанс помилок. У матричній формі повністю загальну модель можна записати наступним чином :

$$y = f(\theta') + e \quad (2.2)$$

де y - вектор спостережень, θ - вектор параметрів, $f(\theta')$ - деяка функція θ , і t_i - вектор залишків. Рішення найменших квадратів, θ , це вектор, який зводить

до мінімуму $[y - f(\theta)]^2 = e^2$. Для лінійної моделі, рівняння (2.2) може бути записано таким чином:

$$y = X\theta + e \quad (2.3)$$

де X являє собою матрицю коефіцієнтів θ . Ефекти в лінійних моделях можуть приймати одну з двох форм, класову або безперервну. Дискретні ефекти, такі як конкретні стада, блоки або стать позначаються «ефектами класу». Хоча рівні цих ефектів можуть бути пронумеровані, немає ніякого зв'язку між номером певного стада і ефектом, пов'язаним з цим номером. Для безперервних ефектів передбачається лінійна залежність між значенням для незалежної змінної і залежною змінною. Кожен рядок X відповідає коефіцієнту θ для певного запису в y . Для ефектів класу елементів в X були б або нуль або одиниця. Для безперервних ефектів, кожен елемент в X відповідає спостережуваному значенню для незалежної змінної.

2.5 Рішення найменших квадратів для одного параметра

Рішення найменших квадратів вирішуються шляхом знаходження оцінок параметрів, які мінімізують суму квадратів залишків. Це буде спочатку показано на наступній простій моделі лінійної регресії :

$$y = \mu + x_i b + e_i \quad (2.4)$$

де Y_i є залежною змінною для спостереження i , μ є константою, x_i незалежна змінна для спостереження i , b - коефіцієнт регресії і e - випадковий залишковий ефект. Залишкова сума квадратів обчислюється таким чином:

$$\sum (y_i - \mu - x_i b)^2 = \sum e_i^2 \quad (2.5)$$

$$\sum y_i^2 + \sum \mu^2 + \sum (x_i b)^2 - \sum (2y_i \mu) - \sum (y_i x_i b) + \sum 2(\mu x_i b) = \sum e_i^2 \quad (2.6)$$

де \sum означає підсумовування по вибірці. Рівняння (2.6) може бути ще більш спрощено, відзначивши, що константи можуть бути переміщені за межі знака підсумовування, і що сума констант дорівнює постійному розміру вибірки, N .

$$\sum y_i^2 + N\mu^2 + b^2 \sum x_i^2 - 2\mu \sum y_i - 2b \sum (y_i x_i) + 2\mu b \sum x_i = \sum e_i^2 \quad (2.7)$$

Оцінки найменших квадратів для μ і b виводяться шляхом обчислення приватних похідних рівняння (2.7) по відношенню до цих двох параметрах, і встановивши ці похідні рівними нулю. Диференціюючи за μ , і встановивши похідну рівну нулю, отримуємо:

$$2N\mu - 2\sum y_i + 2b \sum x_i = 0 \quad (2.8)$$

та рішенням найменших квадратів для μ є $(\sum y_i - b \sum x_i) / N$. Диференціюючи за b , і встановивши цю приватну похідну, рівну нулю отримуємо:

$$2b \sum x_i^2 - 2\sum (y_i x_i) + 2\mu \sum x_i = 0 \quad (2.9)$$

і рішення найменших квадратів для b - $[\sum (y_i x_i) - \mu \sum x_i] / \sum x_i^2$. Таким чином виходить система з двох рівнянь з двома невідомими. Підставляючи рішення найменших квадратів для μ в рівнянні (2.9) і переставляючи, отримуємо таке рішення для b :

$$b = (\sum (y_i x_i) - (\sum y_i \sum x_i) / N) / (\sum x_i^2 - (\sum x_i)^2 / N) \quad (2.10)$$

Рівняння (2.10) це коваріанса вибірки поділена на дисперсію x , яка є формулою для коефіцієнта регресії.

2.6 Рішення методом найменших квадратів

Для загальної лінійної моделі рішення найменших квадратів для загальної лінійної моделі, наведеної в рівнянні (2.3) отримано аналогічним чином. Залишкова сума квадратів в матричному поданні обчислюється таким чином:

$$(y - X\theta)'(y - X\theta) = e'e \quad (2.11)$$

$$y'y - 2(X\theta)'y + (X\theta)'X\theta = e'e \quad (2.12)$$

Встановивши диференціал по θ рівним нулю і вирішуючи, отримуємо :

$$\theta = (X'X)^{-1}X'y \quad (2.13)$$

Рівняння (2.13) називається «нормальним рівнянням», і широко використовується в сучасній статистиці. Якщо спостереження корелюють або не мають рівних варіанс або разом, то нормальні рівняння можуть бути модифіковані таким чином:

$$X'V^{-1}X\theta = X'V^{-1}y \quad (2.14)$$

де V є варіаційною матрицею між спостереженнями. V це діагональна матриця з рядками і стовпцями рівними кількістю спостережень. Діагональні елементи V це варіанси кожного спостереження, а недіагональні елементи є коваріаціями між відповідною парою спостережень. Рішення рівняння (2.14) називаються рішенням «узагальнених найменших квадратів», і зводить до мінімуму $e'e$, піддаючи обмеженню відому варіаційну матрицю. Ці рівняння важко застосувати, як написано, бо вони вимагають зворотню V , яку важко обчислити для великого набору даних. Для лінійної моделі оцінки параметрів також будуть незміщеними, стійкими і в просторі параметрів. Якщо y не є лінійною функцією θ , то рішення найменших квадратів можуть взагалі не бути отримані аналітично, хоча були розроблені різні ітераційні методи. Тільки ефект середнього y включається в модель, вплив на варіансу y або моменти більш високого порядку не можуть бути оцінені по найменшим квадратам.

2.7 Оцінка максимальної правдоподібності

Для одного параметра ML є набагато більш гнучкою, ніж оцінки найменших квадратів, але вимагає достатнього комплексу програмування, крім моделей, які можуть бути проаналізовані доступним програмним забезпеченням, таким як програма LE з BMDP (Елкінд співавт., 1994). Є три кроки в ML оцінці параметрів:

1. Визначення припущень, що лежать в основі статистичної моделі.
2. Побудова функції правдоподібності, яка є спільною щільністю спостережень умовних від параметрів.
3. Максимізація функції правдоподібності по відношенню до параметрів.

Основна методика ML оцінки одного параметра буде проілюстрована використовуючи приклад з біноміального розподілу. Припустимо, що зі зразка десяти спостережень, три «успішних» і сім є «помилковими». Ми хочемо вивести MLE для p , ймовірності успіху. Біноміальна ймовірність цього результату як функція від p є:

$$L = (10! (p)^3 (1-p)^7) / (3!7!) \quad (2.15)$$

де L є ймовірністю отримання цього результату, залежно від p . L позначає «функцію правдоподібності». MLE для p це величина p , яка максимізує L . MLE обчислюється шляхом диференціювання L щодо p і вирішуючи для p , з тією безліччю похідних рівних нулю. На практиці, як правило, легше обчислити і

диференціювати логарифм L . Що стосується ML , це рівносильне диференціації L , тому що функція змінної i логарифм функції будуть максимальними при тому ж значенні змінної. MLE p потім виводиться таким чином:

$$\text{Log } L = \log(10!) - \log(3!7!) + 3(\log p) + 7[\log(1 - p)] \quad (2.16)$$

$$d(\text{Log } L)/dp = 3/p - 7/(1 - p) = 0 \quad (2.17)$$

$$p = 3/10 \quad (2.18)$$

Це, звичайно, частка успіхів, отриманих у зразку. Таким чином, для цього простого випадку, MLE є інтуїтивно зрозумілим значенням оцінки. Виходячи з вищевикладеного, має бути ясно, чому MLE має перебувати в просторі параметрів. Оцінка параметра поза простором параметрів, за визначенням, мають ймовірність нуля, а, отже, не будуть MLE .

Для безперервного розподілу, ймовірність обчислюється як статистична щільність розподілу, залежна від вибірки. Статистична щільність, $f(y)$, для безперервної змінної, y , визначається як ординати функції розподілу при заданому значенні y . Наприклад, припустимо, що зразок був узятий з нормального розподілу. Для отримання MLE для середнього, необхідно обчислити статистичну спільну щільність вибірки. Для одиничного спостереження ймовірність буде :

$$L = (e^{-(y-\mu)^2/2\sigma^2})/(\sqrt{2\pi\sigma^2}) \quad (2.19)$$

де σ стандартне відхилення, e є основою для натуральних логарифмів і становить приблизно 2,72, μ є середнім, π являє собою відношення довжини кола до діаметру кола, і приблизно дорівнює 3,141 і y є змінною величиною. Для вибірки з N спостережень, ймовірність буде добутком правдоподібності для кожного окремого спостереження. Як і в попередньому випадку, MLE для μ може бути отримано шляхом обчислення похідної логарифма ймовірності по відношенню до середнього і якщо ця функція дорівнює нулю. Похідна логарифма L для вибірки з нормальним розподілом обчислюється таким чином:

$$L = \prod(e^{-(y_i-\mu)^2/2\sigma^2}/\sqrt{2\pi^2}) \quad (2.20)$$

$$\text{Log } L = \sum e^{-(y_i-\mu)^2/2\sigma^2}/\sqrt{2\pi^2} \quad (2.21)$$

$$d(\text{Log } L)/d\mu = \sum (y_i - \mu) \quad (2.22)$$

де \prod означає мультиплікативну серію, паралельну Σ , і y_i є елементом i в y . Встановивши $\sum (y_i - \mu)$ рівним нулю, ми знаходимо, що MLE μ є $(\sum y_i)/n$, вибірковим середнім, яке знову інтуїтивно правильний результат.

MLE для варіанси може бути отримано таким же чином, і знову дає інтуїтивно результат вибіркової варіанси. Цей результат буде знову розглянуто в розділі 3 по відношенню до оцінки компонентів варіанси. Хоча у двох прикладах, наведених досі, ML був використаний для отримання оцінок, які можна було б отримати за допомогою інших методів, буде показано у наступних розділах, що для більш складних завдань, ML і байєсовські оцінки є єдиними методами оцінювання, які можуть використовувати всі наявні дані.

2,8 Максимальна правдоподібність багатопараметричної оцінки

ML також може бути використано для оцінки декількох параметрів одночасно, наприклад, оцінити разом середнє і варіансу в нормальному розподілі. У цьому випадку необхідно максимізувати ймовірність щодо обох параметрів. Це може бути зроблено шляхом обчислення приватних похідних

логарифмічного правдоподібності стосовно кожного параметру, і встановивши кожную приватну похідну, що дорівнює нулю. У цьому випадку необхідно вирішити систему рівнянь, рівних числу оцінюваних параметрів. В цілому, функцію правдоподібності для оцінки m параметрів $(\theta_1, \theta_2, \dots, \theta_m)$, з вибірки N спостережень (y_1, y_2, \dots, y_n) можна записати таким чином:

$$\begin{aligned} L &= p(y_1, y_2, \dots, y_n | \theta_1, \theta_2, \dots, \theta_m) = \\ & p(y_1 | \theta_1, \theta_2, \dots, \theta_m) p(y_2 | \theta_1, \theta_2, \dots, \theta_m) \dots p(y_n | \theta_1, \theta_2, \dots, \theta_m) = \\ & \prod p(y_i | \theta_1, \theta_2, \dots, \theta_m) = \\ & \prod p(y_i | \theta) \end{aligned} \quad (2.23)$$

де $p(y_i | \theta)$ являє собою ймовірність отримання y_i , залежно від вектора параметрів. Якщо розподіл безперервний, то $p(y_i | \theta)$ буде замінено на $f(y_i | \theta)$, тобто, щільність y_i за умови θ . Таким чином, ML може застосовуватися для вирішення будь-якої проблеми, яка може бути сформульована в термінах рівняння (2.23).

Хоча, як правило, можна написати функцію правдоподібності і диференціювати логарифм L по відношенню до різних параметрів, для моделей виявлення QTL неможливо вирішити аналітично отриману систему рівнянь. Ітераційні методи для виведення рішення будуть описані в розділах 2.10-2.13.

2.9 Довірчі інтервали і перевірка гіпотез для MLE

На додаток до отримання оцінок параметрів, важливо також, визначити точність оцінок. Як правило, для цієї мети використовуються стандартні помилки оцінок. Квадрат стандартної помилки позначається як «варіанса помилки прогнозу». Наступне рівняння в загальному випадку може бути використане для отримання варіанси помилки прогнозу для MLE одного параметра :

$$\text{Var}(\hat{\theta}) = -1 / (E[d^2(\log L) / d\theta^2]) \quad (2.24)$$

де $\hat{\theta}$ є MLE для θ , і $E[d^2(\log L) / d\theta^2]$ це очікування другої похідної L по відношенню до θ . Рівняння (2.24) буде правильним, якщо перша похідна від θ є кратною різниці між істинним значенням параметра і його оцінкою. В іншому випадку варіанса помилки передбачення буде трохи більше, ніж права частина рівняння (2.24). У широкому діапазоні умов, рівняння (2.24) буде «асимптотично правильним», тобто в міру збільшення розміру вибірки, різниця між лівою і правою частинами рівняння буде прагнути до нуля. Квадратний корінь з варіанси помилки передбачення, стандартна помилка оцінки, може бути використана для визначення довірчого інтервалу оцінки.

Варіанси помилки прогнозу для оцінки декількох параметрів може бути отримана в манері аналогічної тій, що описана в рівнянні (2.24). Оцінки кожного параметра і перші похідні складатимуться з вектора з числом елементів, рівному m , числу параметрів. Другі похідні та варіанси помилок прогнозування обидва будуть квадратними $m \times m$ матрицями. Використовуючи визначники до денотатів матриць і векторів, матриця варіанс помилок прогнозу може бути обчислена за такою формулою:

$$\text{Var}[\hat{\theta}] = -(\partial^2 \text{Log } L / \partial[\theta]^2)^{-1} \quad (2.25)$$

де права частина є зворотною матрицею других приватних похідних по відношенню до $[\theta]$. Діагональними елементами будуть варіанси помилок

прогнозу оцінок і внедіагональні елементи це коваріації помилок прогнозу між елементами. Вони необхідні для перевірки гіпотез, заснованих на лінійних функціях параметрів.

Навіть якщо варіанса помилки прогнозу обчислюється технічно, ML як і раніше можна використовувати для тестування гіпотез, за допомогою «критерію відношення правдоподібності». У тесті відношення правдоподібності ML отримано при двох альтернативних порівнюваних гіпотезах. У нульовій гіпотезі, один або більше з параметрів, які максимізуються в альтернативній гіпотезі передбачаються фіксованими. Наприклад, середнє встановлюється рівним нулю. Альтернативна гіпотеза називається «повна» модель, тому що MLE виводяться для всіх параметрів, у той час як нульова гіпотеза іменується «редукована» модель, тому що деякі з параметрів значень є фіксованими. У припущенні, що нульова гіпотеза вірна, натуральний логарифм відношення ML з повних і скорочених моделей буде асимптотично розподілений як $(1/2)\chi^2$, де χ^2 є статистикою хі-квадрат. Число ступенів свободи (df) дорівнюватиме кількості параметрів, які максимізуються в альтернативною гіпотезі, але фіксуються в нульовій гіпотезі. Це відношення буде мати розподіл χ^2 тільки якщо нульова гіпотеза «вкладена» в альтернативну гіпотезу. Гіпотези є «вкладеними», якщо деякі параметри, які фіксовані в нульовій гіпотезі встановлюють свої ML значення в альтернативній гіпотезі, але всі параметри, які фіксовані в альтернативній гіпотезі, також фіксовані в нульовій гіпотезі.

2.10 Методи для максимізації функції правдоподібності

Численні ітераційні методи були запропоновані, для максимізації багатопараметрової функції правдоподібності. Початкові рішення для всіх методів вибираються довільно. Ці методи будуть порівнюватися на основі легкості застосування, швидкості збіжності та ймовірності збіжності. З усіх методів, які будуть розглянуті нижче, тільки очікування максимізації (EM) гарантовано сходиться до максимуму, за умови, що максимум існує в просторі параметрів. Проте, навіть для EM, точка збіжності може бути тільки локальним максимумом. Цілком можливо, що можуть бути й інші максимуми з більш високими значеннями. Як правило, проблема множинних максимумів виникає через ітерації для декількох різних наборів вихідних значень. Якщо всі рани сходяться до однієї і тієї ж оцінці параметрів, то ймовірно, але не завжди, цей набір параметрів є головним максимумом.

Ітераційні методи максимізації можна розділити на три категорії: похідно - вільні методи, методи, засновані на обчисленні перших похідних і методи, засновані на обчисленні других похідних. Для всіх методів, заснованих на похідних, оцінки параметрів і - тій ітерації обчислюються шляхом вирішення системи рівнянь рівних з кількістю з числом оцінюваних параметрів.

Ці редукування рівняння самі є функціями оцінок параметрів з попередньої ітерації. Як правило, ітерація триває доти, поки зміни між раундами не опускаються нижче досить малого значення. Хоча це загальноприйнятий критерій наближення конвергенції, це не обов'язково так.

Якщо збіжність повільна, то можливо, що зміни між послідовними раундами ітерації може бути невеликими, навіть якщо оцінки не близькі до фактичного рішення. Конвергенція, як правило, найшвидша для методів другої похідної, але це не гарантується, навіть якщо є максимум в просторі параметрів. Ми будемо розглядати першими похідно-вільні методи, потім методи, засновані на обчисленні других похідних і, нарешті, методи, засновані на обчисленні перших похідні .

2.11 Похідно-вільні методи

Було розроблено кілька алгоритмів загального призначення, для знаходження максимуму функції без обчислення похідних. Ці методи доступні в багатьох програмних пакетах, і можуть бути застосовані практично до будь-якої неперервної функції. Вони засновані на прогнозуванні напрямки максимумів, заснованих на безлічі існуючих рішень. Наприклад, припустимо, що ймовірність є функцією двох параметрів. Значення ймовірності отримані для трьох наборів вихідних рішень. Ці три точки визначають площину на поверхні правдоподібності. Напрямок найвищого збільшення ймовірності може бути тоді визначено для цих трьох точок. Новий набір рішень потім обчислюється в напрямку найшвидшого підйому функції правдоподібності. Цей метод може бути розширений на будь-яку кількість параметрів. На кожному кроці число аналізованих точок на один більше ніж число параметрів.

Хоча похідно-вільні методи можуть бути порівняно легко обчислені для будь-якої функції, вони мають серйозні недоліки. Напрямок тільки передбачається, і немає методу оцінки, як далеко йти в цьому напрямку. Таким чином, досить легко піти за межі максимуму, і новий набір рішень може мати більш низьку ймовірність, ніж у попередньому рішенні. Загалом, похідно-вільні методи, як правило, неефективні при великих вибірках або багатьох параметрів. Ці методи не гарантують збіжність до рішення.

2.12 Методи, засновані на другій похідній

У методі Ньютона - Рафсона (Далквіст і Бьорк , 1974), обидві перші похідні і матриця других похідних обчислюються аналітично. Рішення для і -го туру ітерації обчислюються шляхом вирішення наступної системи рівнянь:

$$[\hat{\theta}_i] = [\hat{\theta}_{i-1}] - \partial \text{Log } L / \partial [\theta] * (\partial^2 \text{Log } L \partial [\theta]^2)^{-1} \quad (2.26)$$

де $[\hat{\theta}_i]$ є оцінкою q , для і -й ітерації, $[\hat{\theta}_{i-1}]$ є попередньою оцінкою $[\theta]$, а інші терміни, такі, як визначено вище, з похідними, обчисленими для і-1 оцінки $[\theta]$.

Основна перевага методу Ньютона-Рафсона те, що збіжність, як правило, швидка. До недоліків можна віднести те, що алгоритм може не сходитися, навіть якщо ймовірність має максимум в межах простору параметрів і те, що обчислення матриці других похідних часто нетривіальне завдання. Ця проблема трохи пом'якшується якщо чисельні методи використовуються для оцінки диференціалів (Bailey, 1961; Дженсон, 1989). Таким чином, алгебра дещо спрощується, але є деякі жертви як в ефективності, з точки зору обчислювального часу, так і в точності оцінок і дисперсії помилок прогнозу. Як показано вище, ця матриця може бути використана для отримання оцінок стандартної помилки оцінок, що саме по собі є важливою метою .

2.13 Методи на основі перших похідних (очікування максимізації)

ЕМ заснована на обчисленні перших похідних. Принцип, що лежить в основі ЕМ це розгляд двох щільностей вибірки, перша заснована на повній специфікації даних (невідомих), а друга - на основі неповної специфікації даних (відомих). Алгоритм ЕМ складається з двох етапів: етап оцінки, в якому достатні статистики оцінюються для повної функції щільності даних і крок максимізації, в якій ця функція максимізується стосовно параметрам. «Достатня статистика» це статистика, отримана з вибірки, яка містить всю інформацію у вибірці релевантну стосовно оцінюваним параметрам. Наприклад, вибіркове середнє є достатньою статистикою для оцінки популяційної середньої.

ЕМ як правило, вважається методом вибору, тому що він гарантує збіжність до локального максимуму за умови, що він існує в просторі параметрів. Проте, швидкість збіжності може бути дуже повільною, і немає ніякої гарантії, що знайдений максимум буде глобальним максимумом. Єдиний спосіб приблизно вирішити цю проблему в тому, щоб почати ітерацію з декількох різних наборів початкових значень. Якщо всі рани сходяться в одних і тих же рішеннях, то цілком імовірно, що є тільки один максимум в просторі параметрів. Додатковою перевагою ЕМ є те, що можна включати «помилкові» параметри в аналіз, наприклад, ефект блоку або стада, навіть якщо ці параметри мають дуже велике число рівнів.

2.14 Оцінка Байеса

Байєсовське оцінювання відрізняється від ML в тому, що замість максимізації функції ймовірності, «апостеріорна ймовірність» від θ , $p(\theta|y)$ досягає максимуму, якщо функція від правдоподібності множиться на «апріорний» розподіл θ . Теорема Байеса в загальних термінах для декількох параметрів і спостережень обчислюється таким чином:

$$p(\theta_1, \theta_2, \dots, \theta_m | y_1, y_2, \dots, y_n) = p(\theta_1, \theta_2, \dots, \theta_m) p(y_1, y_2, \dots, y_n | \theta_1, \theta_2, \dots, \theta_m) \quad (2.27)$$

де $p(\theta_1, \theta_2, \dots, \theta_m | y_1, y_2, \dots, y_n)$ це «апостеріорна» ймовірність параметрів, $p(\theta_1, \theta_2, \dots, \theta_m)$ це «апріорна ймовірність» параметрів, і $p(y_1, y_2, \dots, y_n | \theta_1, \theta_2, \dots, \theta_m)$ це функція правдоподібності. Подібно ML, можна максимізувати апостеріорну ймовірність або щільність функцію по відношенню до значень параметрів. Припускаючи що апріорна інформація для параметрів доступна, байєсовські оцінки, які роблять можливим використання цієї інформації, повинні бути краще ML, який ігнорує всю апріорну інформацію про параметри.

Замість максимізації апостеріорної щільності, можна визначити «функцію втрат», яка визначається економічним значенням «втрата» неправильних оцінок параметрів. Типовими прикладами є лінійні і квадратичні функції втрат. У лінійній функції втрат, величина втрат є лінійною функцією різниці між параметрами оцінки та їх істинними значеннями. У квадратичній функції втрат, втрата збільшується квадратично як функція від різниці між оцінками параметрів і їх істинними значеннями. Мінімізація лінійної функції втрат еквівалентна максимізації апостеріорної щільності. Мінімізація квадратичної функції втрат еквівалентна максимізації середнього для апостеріорного розподілу.

Аналогічним чином, байєсовський тест альтернативної гіпотези заснований на мінімізації очікування функції втрат. Якщо рішення має бути прийняте між двома альтернативними гіпотезами, економічна вартість «втрати» визначається для кожного неправильного рішення. Очікування втрат буде ймовірнісне для кожного неправильного рішення (помилки типу I і типу II), помножене на його економічну вартість. Рішення потім ґрунтується на мінімізації очікуваних втрат.

Існує два основних недоліки оцінки Байєса. По-перше, апріорна інформація про параметри часто є невизначеною, і не можливо математично представити цю інформацію в термінах статистичної функції розподілу без додаткових припущень, які не можуть бути перевірені. По-друге, якщо багато записів, включені в аналіз, то функція правдоподібності має тенденцію до «притискання» апріорного розподілу від θ . У цьому випадку, байєсовські оцінки, як правило, сходяться до MLE.

2.15 Оцінка мінімальної різниці

Хоча байєсовські оцінки вимагають великих припущень апріорі про дані, обидва і метод найменші квадрати і ML також роблять деякі припущення про те, що лежить в основі розподілу даних. У методі найменших квадратів нормальний розподіл залишків майже завжди передбачається. Як правило, обидва методи і найменших квадратів і ML є стійкими до порушень передбачуваного розподілу. Однак, це не може бути так у присутності «викидів», спостережень, які відхиляються значно від передбачуваного розподілу.

MD оцінки це значення параметрів, які мінімізують деякі функції теоретичного розподілу, враховуючи оцінки параметрів і емпіричний розподіл даних. Найбільш поширена використовувана функція це сума квадратів відхилень, звана відстанню Крамера - Мізеса, яке обчислюється таким чином:

$$\delta = \sum_{i=1}^n [(\Phi(y_i|\theta_1, \theta_2, \dots, \theta_m) - (i - 0.5)/N)]^2 \quad (2.28)$$

де δ це відстань яке має бути мінімізована, $\Phi(y_i|\theta_1, \theta_2, \dots, \theta_m)$ це сукупна функція нормального розподілу для спостережень y_i із спостереженнями відсортованими в порядку зростання порядку i з урахуванням оцінок параметрів $\theta_1, \theta_2, \dots, \theta_m$, а інші терміни, як визначено раніше. Мета полягає в тому, щоб знайти оцінки параметрів, які мінімізують δ . Тобто, теоретичний розподіл, який найближче апроксимує емпіричний розподіл даних.

Цей метод залежить тільки від рангу спостережень, а не їх абсолютних значень, і тому менш схильний до дії екстремальних значень y . У присутності викидів, цей спосіб був знайдений, щоб бути більш стійким, що ML для оцінювання параметрів QTL у зворотному пересічному дизайні (Перес- Enciso і Того, 1999).

Теоретично, δ може бути отримано шляхом прийняття приватних похідних лівій частині рівняння (2.28) по відношенню до параметрів і установки цієї системи рівнянь рівними нулю. На практиці, як і ML, ця система рівнянь не може бути вирішена аналітично, і потрібні ітераційні методи. Якщо число параметрів низьке, Рівняння (2.28) можна звести до мінімуму шляхом проб і помилок. Крім того, аналогічно ML, може бути застосована

апроксимація ітерації Ньютона-Рафсона. Перша і друга часткова похідна можна апроксимувати обчисленням δ над серією можливих значень параметрів, і ці похідні можуть бути використані в рівнянні (2.26), щоб отримати параметр оцінки для наступної ітерації.

3. ВИПАДКОВІ І ФІКСОВАНІ ФАКТОРИ. ЗМІШАНА МОДЕЛЬ

3.1. Випадкові і фіксовані фактори

Досі ми припускали, що всі фактори включення в модель, за винятком залишкових, є фіксованими. Для випадкових факторів, передбачається, що кожен фактор взятий з нескінченної популяції можливих ефектів з відомим розподілом.

Моделі, які включають в себе як фіксовані, так і випадкові чинники називаються «змішаними моделями». Розглянемо спочатку загальну стратегію вирішення змішаних моделей, на основі «моделі змішаних рівнянь» (Henderson, 1973) в розділах 3.2-3.6. У розділах 3.8-3.11 ми будемо розглядати ряд моделей, що використовуються для отримання генетичних оцінок для кількісних ознак, і звернемо увагу на переваги і недоліки кожної моделі. Ми розглянемо максимальні оцінки правдоподібності параметрів змішаних моделей в розділі 3.12.

Загалом, будемо вважати, що випадкові ефекти взяті з нормального розподілу із середнім значенням 0 і відомою дисперсією. Таким чином, оцінки для випадкових факторів можуть бути отримані, тільки якщо їх варіанси відомі. У заключних чотирьох розділах ми будемо розглядати методи оцінки компонент варіанси в змішаних моделях, на основі постійної підгонки і максимальної правдоподібності, і обмеженого максимального правдоподібності (REML).

3.2 Змішана лінійна модель

Як приклад ми спочатку розглянемо наступну просту змішану модель, використовувану для отримання племінної цінності бугаїв для виробництва молока :

$$Y_{ijk} = N_i + S_j + e_{ijk} \quad (3.1)$$

де Y_{ijk} це запис молочної продуктивності корови k в стаді i , N_i це ефект стада i , S_j це ефект батька корови j на її продуктивність і e_{ijk} є випадковим залишковим фактором. Фактор стадо будемо вважати фіксованим ефектом, а ефект батька будемо припускати випадковим. У загальних термінах змішану модель можна записати в матричній формі, таким чином:

$$y = X\beta + Zu + e \quad (3.2)$$

де β це вектор фіксованих ефектів, u - вектор випадкових ефектів, X і Z є зв'язуючими матрицями і e - вектор випадкових залишкових ефектів. Адитивні племінні цінності вважаються випадковими ефектами, з відомою варіаційної матрицею. Обидва u і e передбачаються, що мають нормальний розподіл. Таким чином, y має багатовимірний нормальний розподіл при середньому значенні $X\beta$, і варіансою V обчисленої таким чином:

$$V = ZGZ' + R \quad (3.3)$$

де G це матриця варіанса u , а R - матриця варіанс залишкових ефектів.

Якщо бики-плідники не пов'язані, то матриця варіанс ефекту батьків буде $I\sigma^2_s$, де I це одинична матриця, і σ^2_s , варіанса ефекту батька рівна одній четвертій адитивної генетичної варіанси. Варіанса ефекту батька дорівнює одній четвертій адитивної генетичної варіанси, тому що кожен батько передає половину своїх генів кожній дочці. Якщо звести в квадрат обчислену варіансу, половина адитивного генетичного ефекту стає однією четвертою адитивної генетичної варіанси.

Якщо бики - виробники пов'язані, то $G=A\sigma^2_s$, де A числова матриця відносин між виробниками. Діагональні елементи A дорівнюватимуть одиниці, і недіагональні елементи будуть відображати частину генів, що два окремих особи, відповідні певною рядку і стовпці A , які ідентичні за походженням, наприклад, 0,5 для батька і сина, і 0,25 для напів- сибсов .

Обидві A і G завжди симетричні матриці, тобто $G = G'$ і $A = A'$. Як і у фіксованій моделі, залишки зазвичай вважають некоррельованими і що мають однакову дисперсію. У цьому випадку $R = I\sigma^2_e$, де σ^2_e є залишковою варіансою. Для моделі плідників наведеної в рівнянні (3.1), якщо відносини між плідниками ігноруються розподіл конкретного запису можна записати наступним чином :

$$Y_{ijk} = [e^{-(sj)2/2\sigma^2_s} / 2\pi\sigma^2_s] [e^{-(Y_{ijk}-\mu_i-S_j)2/2\sigma^2_e} / 2\pi\sigma^2_e] \quad (3.4)$$

3.3 Моделі змішаних рівнянь

Теоретично, рішення найменших квадратів для фіксованих ефектів можуть бути отримані з рівняння (2.14). Визначивши рішення для фіксованих ефектів, як β , Хендерсон показав (1973), що рішення для випадкових ефектів потім можуть бути обчислені як: $GZ'V^{-1}(y-\hat{y})$.

Проте, дисперсійна матриця $V = ZGZ' + R$, що не діагональна, і тому не може бути обернена для великих наборів даних. Рішення для β і u для великих наборів даних може бути отримані шляхом вирішення наступної системи рівнянь, що позначається як «моделі змішаних рівнянь» (Хендерсон, 1973):

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad (3.5)$$

де R^{-1} це зворотна матриця залишкових варіанса, і G^{-1} це зворотна матриця варіанса для u . Ліва сторона цих рівнянь складається з квадратної, симетричною матриці званої «матрицею коефіцієнтів», і β и вектор рішень. Як зазначалося раніше, для аналізу однієї ознаки, як правило, передбачається, що залишкові варіанси для кожного запису рівні і не корелюють. У цьому випадку матриця залишкових варіанса дорівнює $I\sigma^2_e$, і $R^{-1} = I/\sigma^2_e$. Таким чином, модель змішаних рівнянь можна спростити, множачи обидві частини на σ^2_e наступним чином:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad (3.6)$$

Для «моделі батька», наведеної в рівнянні (3.1), $X'X$ буде діагональною матрицею з рядками і стовпцями рівними кількістю стад. Діагональним елементом кожного рядка буде число спостережень у відповідному стаді, і всі недіагональні елементи будуть рівні нулю. Точно так само, $Z'Z$ буде діагональною з кожним діагональним елементом, рівним кількості спостережень дочок кожного батька. $X'Z$ буде мати кількість рядків, рівну кількості стад, і стовпців, що дорівнює кількості батьків. Кожен елемент буде число записів у відповідній комбінації стадо \times батько. $X'u$ буде вектор довжиною рівною числу стад, і кожен елемент буде сумою значень записів у відповідному стаді. $Z'u$ буде довжиною кількістю плідників, і кожен елемент буде сума записів всіх дочок відповідного батька.

3.4 Рішення моделі змішаних рівнянь

Рішення моделі змішаних рівнянь можна отримати, помноживши вектор правої сторони на зворотну матрицю коефіцієнтів. Точне рішення вимагає інвертування матриці коефіцієнтів, але в цілому матриця коефіцієнтів буде набагато менше, ніж V . Число рядків і стовпців V дорівнює загальній кількості записів, а число рядків і стовпців у матриці коефіцієнтів дорівнює числу рівнів ефектів, включених в модель, яка, як правило, набагато менше.

Якщо багато ефектів включено в модель, наближені рішення можуть бути отримані методом ітерацій. Є кілька методів ітерації, які можуть застосовуватися для вирішення моделі змішаних рівнянь. Ітерація Гаусса-Зейделя, головний обраний метод, тому що він відносно швидкий, і гарантовано сходиться, за умови, що рівняння мають рішення (Quaas і Полла, 1980). У ітерації Гаусса-Зейделя рішення для кожного рівняння в ітерації обчислюються наступним чином :

$$V_i = (\sum Y_i - \sum_{j=1}^{i-1} V_{jci} - \sum_{j=i+1}^J V_{cjj}) / c_{ii} \quad (3.7)$$

де V_i це рішення для рівняння i на ітерації k , Y_i це права сторона для рівняння i , V_j є рішенням для рівняння j при ітерації $k - 1$, c_{ij} це елементи матриці коефіцієнтів в стовпці i та рядку j , і c_{ii} це діагональний елемент стовпця i . Чим більше діагональні елементи матриці коефіцієнтів по відношенню до недіагональних елементів, тим швидше буде зближення в ітерації Гаусса-Зейделя. Таким чином, моделі отців можуть сходиться менш ніж за десять раундів ітерації, так як діагональні елементи, як правило, дуже великі в порівнянні з недіагональними елементами. Моделі тварини, описані нижче, як правило, вимагають сотні раундів ітерації для досягнення приблизної збіжності.

Обчислення матриці коефіцієнтів і раніше вимагає звернення G . Для моделі батька $G^{-1} = A^{-1}/\sigma_s^2$, и $G^{-1}\sigma_e^2 = A^{-1}\sigma_e^2/\sigma_s^2$. σ_e^2/σ_s^2 константа, яка зазвичай передбачається має бути відома. Як зазначалося вище, σ_e^2/σ_s^2 дорівнює одній четвертій адитивної генетичної дисперсії. Таким чином, σ_e^2 включає в себе три чверті генетичної дисперсії, плюс всі розбіжності, що залишилися. Таким чином, σ_e^2 дорівнює $1 - h^2/4$ від загальної дисперсії, де h^2 успадкованість, тобто відношення адитивної генетичної дисперсії до фенотипічної дисперсії. Таким чином, в термінах успадкування, σ_e^2/σ_s^2 дорівнює $(4h^2)/h^2$. Хендерсон (1976) розробив простий алгоритм, щоб інвертувати A зі списку окремих

особин і їх батьків і матерів. Таким чином, єдина матриця, яка має бути обернена це матриця коефіцієнтів, яка буде квадратною матрицею розміром, рівному числу ефектів включених в модель.

3.5. Деякі важливі властивості моделі змішаних рівнянь

Рішення нормальних рівнянь, представлених у розділі 2, називаються «найкращі лінійні незміщені оцінки» (BLUE). Незміщеність також визначена в главі 2. Поняття «кращий» в тому, що якщо β є оцінкою β , то $E(\hat{\beta}-\beta)^2$ є мінімальним серед класу лінійних незміщених оцінок. Хендерсон (1973) називає рішення для випадкових ефектів в змішаній моделі «кращими лінійними незміщеними прогнозами» (BLUP). При передбачуваної структурі варіанси, випадкові рішення в моделі змішаних рівнянь, u , будуть «кращими» в тому сенсі, що $E(\hat{u}-u)^2$ буде зведено до мінімуму, в рамках прийнятих обмежень. Так як випадкові ефекти не є параметрами, їх рішення називаються «прогнозами», а не «оцінками». BLUP рішення мають кілька важливих властивостей, які будуть підсумовані тут.

Варіанси помилок прогнозів фіксованих і випадкових ефектів можна оцінити шляхом обернення матриці коефіцієнтів моделі змішаних рівнянь. Ця інверсія може бути розділена на чотири суб-матриці, відповідних чотирьом підматрицям в моделі змішаних рівнянь. Тобто:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \quad (3.8)$$

Діагональні елементи C_{11} будуть відповідати варіансам помилки прогнозу для вирішення фіксованих ефектів, і діагональні елементи C_{22} будуть відповідати варіансі помилки прогнозу для рішень випадкових ефектів. Рішення для фіксованих ефектів будуть мати більшу варіансу, ніж реальних ефектів, тоді як помилка прогнозування варіанси рішення випадкових ефектів, які регресували до середнього значення, буде менше, ніж дисперсія ефектів. Загалом:

$$\text{var}(u) = \text{var}(\hat{u}) + \text{pev}(\hat{u}) \quad (3.9)$$

де $\text{var}(u)$ і $\text{var}(\hat{u})$ це варіанси від u від u , і $\text{pev}(\hat{u})$ це передбачена помилка варіанси u . Хендерсон (1973) також показав, що коваріація між u і \hat{u} дорівнює $\text{var}(u)$. Таким чином, регресія від u на \hat{u} дорівнює одиниці. Тобто, якщо фактична різниця між двома випадковими ефектами дорівнює x , очікувана різниця між їхніми рішеннями також буде дорівнює x . Це не відноситься до фіксованих ефектів. Відношення $\text{var}(u)/\text{var}(\hat{u})$ називається «достовірністю» від u , і дорівнює квадрату кореляція між u і \hat{u} , тобто є коефіцієнтом детермінації. Квадратний корінь від надійності позначається «точністю» u .

3.6 Абсорбція рівнянь

Для багатьох поширених моделей в генетичному аналізі, кількість рівнянь може бути далі зменшено технікою, званої «абсорбція». Ця техніка може бути застосована як до фіксованих, так і змішаних моделей, і буде проілюстрована за допомогою наступної системи рівнянь:

$$\begin{bmatrix} X'_1 X_1 & X'_1 X_2 \\ X'_2 X_1 & X'_2 X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} X'_1 y \\ X'_2 y \end{bmatrix} \quad (3.10)$$

де індекси 1 і 2 відносяться до будь-якого поділу рівнянь на дві групи. Ці рівняння можна переписати таким чином:

$$\begin{aligned} X'_1 X_1 \beta_1 + X'_1 X_2 \beta_2 &= X'_1 y \\ X'_2 X_1 \beta_1 + X'_2 X_2 \beta_2 &= X'_2 y \end{aligned} \quad (3.11)$$

Рішення для β_1 у першому з системи рівнянні дає :

$$\beta_1 = [X'_1 y - X'_1 X_2 \beta_2] [X'_1 X_1]^{-1} \quad (3.12)$$

β_1 у другому з системи рівняння може бути замінене рішенням для β_1 в рівняння (3.11). Друга система рівнянь тепер тільки функція від β_2 і може бути вирішено окремо. Після отримання рішення для β_2 , «зворотні рішення» можуть бути отримані для β_1 , вирішуючи для β_1 в рівнянні (3.11).

Звичайно, ця процедура потребує вирішення для $[X'_1 X_1]^{-1}$. У багатьох випадках, $X'_1 X_1$ діагональне, як зазначалося раніше для моделі батька, наведеною в рівнянні (3.1), а тому може бути легко оберненою. Для цієї моделі, $X'X$ матриця буде діагональною з кількістю записів в кожному стаді, як діагональним елементом, а зворотна матриця буде бути зворотне кожного діагонального елемента. Випадкові ефекти також можуть бути поглинені, за умови, що матриця коефіцієнтів має відповідну діагональну, або блок діагональну, структуру.

3.7 Аналіз багатовимірної змішаної моделі

Модель змішаних рівнянь також може бути використана для аналізу декількох взаємопов'язаних ознак, наприклад, виробництва молока і молочний жир корів. Багатоознакова модель батька може бути описана таким чином:

$$Y_{ijkl} = N_{il} + S_{jl} + e_{ijkl} \quad (3.13)$$

де Y_{ijkl} це запис продуктивності корови k в стаді i для ознаки l , N_{il} це ефект стада i на ознака l , S_{jl} це ефект батька корови j на ознаку l , а e_{ijkl} це випадковий залишковий зв'язок з ознакою l . У цьому випадку, як правило, можна припустити, що обидва адитивних генетичних ефекти і залишки мають багатовимірний нормальний розподіл. Як і у одновимірному випадку, розподіл кожного запису буде дано по розподілу з випадкових генетичних рівнів ефектів розподілу залишкових ефектів. Для двох корельованих ознак, x і y , розподіл залишків для кожної особини буде виглядати наступним чином:

$$(2\pi\sigma_x^2\sigma_y^2(1-\rho^2))^{-1/2} e^{\varphi} \quad (3.14)$$

де σ_x^2 і σ_y^2 залишкові відхилення за ознаками x і y , $\rho = \sigma_{xy}/\sigma_x\sigma_y$ це залишкова кореляція, і φ обчислюється таким чином:

$$\varphi = -1/(2(1-\rho^2))((x-\mu_x)^2/\sigma_x^2 - 2\rho(x-\mu_x)(y-\mu_y)/\sigma_x\sigma_y + (y-\mu_y)^2/\sigma_y^2) \quad (3.15)$$

де μ_x і μ_y це середні для ознак x і y , і рівні $N_{il} + S_{jl}$ для кожної ознаки. Розподіл для генетичних ефектів обчислюються аналогічним чином.

Тепер розглянемо загальну модель змішаних рівнянь, наведених у рівнянні (3.5). Матриця варіанс залишків більше не буде діагональною, але буде

«блочно-діагональною». Для двох ознак, матриця залишків матиме структуру $I \otimes R_i$, де I це одинична матриця, і R_i - матриця 2×2 з такими елементами :

$$R_i = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \quad (3.16)$$

' \otimes ' позначає «продукт Кронекера», що означає, що кожен елемент I множиться по R_i . Точно так само, варіаційна матриця ефектів батька буде $A \otimes S$, де S є такою 2×2 матрицею:

$$S = \begin{bmatrix} \sigma_{sx}^2 & \sigma_{sxy} \\ \sigma_{sxy} & \sigma_{sy}^2 \end{bmatrix} \quad (3.17)$$

де σ_{sx}^2 і σ_{sy}^2 це варіанси ефекту батька для ознак x і y , і σ_{sxy} це коваріація між ними. Хоча обидві залишкова і батьківська матриця ефектів можуть бути легко інвертовані, спрощення, отримане в рівнянні (3.6) шляхом множення на залишкову варіанса вже не можливо. Загальне число рівнянь буде дорівнює числу рівнів ефектів кількості ознак.

3.8 Модель з повторами

Модель, що приводиться в рівнянні (3.1) доцільна, якщо кожна корова має тільки один запис. Якщо у корови є кілька записів, то буде ефект, загальний для всіх записів однієї і тієї ж тварини. Один із способів впоратися з цією ситуацією, це розглянути безліч записів тварин, як корельовані ознаки. Кілька записів однієї і тієї ж тварини будуть мати ненульові залишкові і генетичні коваріанси. Число рівнянь буде потім рівним числу рівнів стад і плідників, в залежності від максимальної кількості записів на індивідуума. Ця модель також вимагає, щоб залишкові і генетичні варіанси для записів однієї і тієї ж тварини були відомі.

Рішенням дещо простіше є «повторювана модель». Ця модель передбачає, що той же самий ефект батька є загальним для всіх особин, і що залишкові коваріації серед всіх записів однієї і тієї ж особи однакові. У цьому випадку модель Рівняння (3.1) може бути змінена таким чином:

$$Y_{ijkl} = N_i + S_j + C_{jk} + e_{ijkl} \quad (3.18)$$

де N_i зараз відноситься до ефекту стадо-рік-сезону, і C_{jk} є загальним ефектом для всіх записів корови k . C_{jk} також буде випадковим ефектом з матрицею варіанси $I\sigma^2_c$, де σ^2_c це варіанса через загальний ефект корови. Ця модель передбачає, що загальні ефекти корови не корелюють, які є всього лише наближення, тому що родинні особини будуть мати позитивну коваріацію. Як зазначалося раніше ефект батька включає в себе одну четверту адитивного генетичного ефекту, в той час як ефект корови буде включати в себе інші три чверті адитивного генетичного ефекту, плюс додатковий ефект званий «ефект постійного навколишнього середовища». Цей ефект буде включати неадитивні генетичні ефекти, і для навколишнього середовища, і загальні для всіх записів особини. У матричній формі цю модель можна записати наступним чином:

$$y = X\beta + Z_{1s} + Z_{2c} + e \quad (3.19)$$

де Z_1 це матриця коефіцієнтів для генетичних ефектів батька, Z_2 це матриця коефіцієнтів для загальних індивідуальних ефектів, і s і c це вектора батьківських генетичних і загальних індивідуальних ефектів, відповідно. Рівняння змішаної моделі для цієї моделі буде :

$$\begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1 + G^{-1}\sigma_e^2 & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + I\gamma \end{bmatrix} \begin{bmatrix} \beta \\ s \\ c \end{bmatrix} = \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \end{bmatrix} \quad (3.20)$$

де $\gamma = \sigma_e^2/\sigma^2c$. Число рівнянь буде дорівнює загальній кількості рівнів. Однак, так як $Z_2'Z_2 + I\gamma$ є діагональною матрицею, ці рівняння можуть бути легко абсорбовані (Ufford співавт., 1979).

3.9 Індивідуальна модель тварини

Повторювана модель ігнорує всі відносини між самками. Хендерсон (1973) першим припустив, що моделі змішаних рівнянь можуть бути використані для оцінки полігенних значень племінної цінності для всіх тварин в популяції з урахуванням всіх відомих відносин споріднення, через «індивідуальну модель тварини» (ІАМ). Проста ІАМ наводиться нижче:

$$Y_{ijk} = N_i + a_j + p_j + e_{ijk} \quad (3.21)$$

де Y_{ijk} є запис k особини j в «стаді» або «блоці» i , N_i це фіксований ефект стада i , a_j це випадковий адитивний генетичний ефект особини j , p_j це випадковий ефект постійного навколишнього середовища для особини j і e_{ijk} це випадковий залишок пов'язаний з кожним записом. Як і в повторюваній моделі, ефект постійної навколишнього середовища потрібен, якщо особини можуть мати кілька записів, тому що там буде загальний ефект для всіх записів кожної особини взагалі. У матричному поданні ця модель може бути записана таким чином:

$$y = X\beta + Z_1a + Z_2p + e \quad (3.22)$$

де Z_1 являє собою матрицю коефіцієнтів для адитивних генетичних ефектів, Z_2 є матриця коефіцієнтів для ефектів постійного навколишнього середовища, а i і p вектори ефекти адитивні генетичні та постійні навколишнього середовища, відповідно.

У повністю фіксованій моделі, адитивні генетичні та постійні середові ефекти будуть повністю не правильні, бо кожен рівень цих двох ефектів пов'язаний з тією, окремою особиною. У ІАМ ці ефекти можуть бути оцінені окремо, тому що обидва є випадковими, а структури їх варіанс різні. Варіаційна матриця для постійних середових ефектів буде $I\sigma^2p$, де σ^2p це компонента варіанс постійного середового ефекту. Варіаційна матриця для адитивного генетичного ефекту буде $A\sigma^2_A$, де σ^2_A - адитивні генетичні варіанси. Після множення на залишкову варіансу, модель змішаних рівнянь для цієї моделі є

$$\begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1 + G^{-1}\sigma_e^2 & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + I\gamma \end{bmatrix} \begin{bmatrix} \beta \\ a \\ p \end{bmatrix} = \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \end{bmatrix} \quad (3.23)$$

де $\gamma = \sigma_e^2/\sigma^2p$.

У більшості випадків не всі особини, включені в популяцію буде мати записи. Наприклад, якщо аналізованою ознакою є молочна продуктивність, тільки самки матимуть записи продуктивності. Особини без записів, такі як батьки корів, можуть бути включені в аналіз за допомогою матриці відносин. Додаткові рівняння можуть бути додані для цих особин в змішану модель рівнянь наступним чином:

$$\begin{bmatrix} X'X & X'Z_1 & 0 & X'Z_2 \\ Z_1'X & Z_1'Z_1 + G^{11}\sigma_e^2 & G^{12}\sigma_e^2 & Z_1'Z_2 \\ 0 & G^{21}\sigma_e^2 & G^{22}\sigma_e^2 & 0 \\ Z_2'X & Z_2'Z_1 & 0 & Z_2'Z_2 + I\gamma \end{bmatrix} \begin{bmatrix} \beta \\ a_1 \\ a_2 \\ P \end{bmatrix} = \begin{bmatrix} X'y \\ Z_1'y \\ 0 \\ Z_2'y \end{bmatrix} \quad (3.24)$$

де G^{11} і G^{22} це блоки зворотних матриць генетичних варіанс, що відносяться до особин з і без записів відповідно, G^{12} і G^{21} є поза діагональні блоки, і a_1 і a_2 це рішення для особин з і без записів, відповідно. Всі інші елементи рівнянь для тварин без записів будуть рівні нулю. Загальне число рівнянь дорівнюватиме кількості стад, а також числу особин із записами, плюс загальна кількість осіб, включених в матрицю відносин.

Навіть якщо ця система рівнянь, як правило, буде дуже великою, вона також буде «досить порожньою», тобто більше 90% елементів дорівнюватиме нулю. Ці рівняння можуть бути вирішені шляхом ітерацій Гаусса- Зайделя, як описано в розділі 3.4. Як було зазначено раніше, число ітерацій, необхідних для конвергенції, є функцією розміру діагональних елементів в матриці коефіцієнтів порівняно з поза- діагональними елементами. У моделі батька діагональні елементи, як правило, досить великі, тому що кожен батько має багато дочок із записами. Це не той випадок для моделей тварин. У ІАМ, діагональний елемент складається тільки з вкладу зворотної матриці спорідненості, плюс власні записи індивіда. Таким чином, необхідно буде ще багато ітерацій в ІАМ, в порівнянні з моделлю батька, в яких діагональні елементи, як правило, набагато більше, ніж поза-діагональні елементи.

3.10 Групування осіб з невідомими предками

Хоча можна включити предків у ІАМ, у самих старих тварин є невідомі батьки. Модель змішаних рівнянь моделі тварини, наведена в рівнянні (3.24) припускала, що ці особини взяті випадкової вибіркою з «базової популяції». Однак це, як правило, не так. У «засновники» - особини також відібрані, і взагалі можуть бути не з одного і того ж покоління.

Томпсон (1979) розробив стратегію для невідомих батьків, припускаючи «фантомних» батьків, які потім групуються за загальними характеристиками особин з невідомими батьками, такими як стать і вік. Ці ефекти «генетичних груп» вважаються фіксованими ефектами, і обчислюються за допомогою матриці спорідненості. Генетична оцінка кожного індивідуума потім обчислюється як сума його генетичної групи і адитивних генетичних ефектів. Одна особина може мати внесок від ефектів декількох груп, якщо його предки мають фантомних батьків з декількох груп. Westell і співавт. (1988) розробили

алгоритм щоб безпосередньо обчислити генетичну оцінку кожної особини з моделі змішаних рівнянь.

3.11 Редукована модель тварини

Розглянемо ІАМ тільки з одним записом на одну тварину. У цьому випадку неможливо оцінити ефект постійного навколишнього середовища, так як цей ефект буде повністю плутатися з випадковим залишком. Модель аналізу тепер можна переписати таким чином:

$$Y_{ijkl} = H_i + 1/2S_j + 1/2D_k + M_{ijkl} + e_{ijkl} \quad (3.25)$$

де S_j є ефект батька j індивідуума l , D_k це ефект матері k , M_{ijkl} це інша частина адитивного генетичного ефекту для особини l не включена в ефекти батька і матері та e_{ijkl} це випадковий залишок.

Як зазначалося раніше, для конкретної особини, дисперсія ефекту батька це одна четверта адитивної генетичної дисперсії. Точно так само, дисперсія ефекту матері також одна четверта генетичної дисперсії. Таким чином, для будь-якої конкретної особини з тими ж двома батьками, ефекти батька й матері не будуть пояснювати половину генетичної дисперсії. Таким чином, цей ефект був названий ефектом «менделевської вибірки» (Quaas і Поллок, 1980), тобто, конкретна генетична компонента передається особини l , що відрізняє цю особину з повних сибсів.

Для особин, що не мають потомства, менделевська вибірка для варіаційної матриці з M ефектів буде $I\sigma^2_m$, де σ^2_m це варіанса менделевської вибірки. Коваріація буде нульовою для осіб без потомства. Особи з потомством віддадуть частину їх M ефектів їх потомству, і тому буде позитивна коваріація між M ефектами батьків і ефектами батька чи матері їх потомства. Для осіб без потомства, M і залишкові ефекти буде повністю прибрані, а модель може бути змінена таким чином:

$$Y_{ijkl} = 1/2H_i + 1/2S_j + D_k + \varepsilon_{ijkl} \quad (3.26)$$

$$\text{де } \varepsilon_{ijkl} = M_{ijkl} + e_{ijkl}.$$

Таким чином, тільки H , S і D ефекти включені в модель для особин без потомства. Тобто, ці особини можуть бути абсорбовані в рівняння їхніх батьків і матерів, як показано у Quaas і Поллока (1980). Вони назвали цю модель «редукована модель тварини» (RAM). Якщо число осіб без потомства включене в аналіз є відносно великим, порівняно з кількістю осіб з потомством, потім можливе істотне зниження загальної кількості рівнянь в моделі. Для особин без потомства також можна вважати «гаметичну» модель, так як ефекти, включені в цю модель, є внеском батьківської і материнської гамети кожному нащадку.

3.12 Оцінка максимальної правдоподібності змішаних моделей

У змішаній моделі, яка включає в себе як фіксовані, так і випадкові ефекти, функція правдоподібності є функцією спільної щільності інтегрованою по випадкових ефектів (Titterton співавт., 1985). Проілюструємо це на основі простої змішаної моделі, описаної в рівнянні (3.1). Статистична функція щільності для цієї моделі, що передбачає, не пов'язаних біків буде :

$$f(Y_{ijk}) = \prod \left\{ f(s_j) \prod f(y_{ijk}) \right\} \quad (3.27)$$

де $f(Y_{ijk})$ є функцією щільності для особини k , Π являє собою мультиплікативну суму, $f(s_j)$ являє функцію щільності для батька j , що з нормальною щільністю функції з середнім значенням рівним нулю, а варіансой σ^2s , $f(y_{ijk})$ являє нормальну функцію щільності для дочки k батька j , яке має середню s_j+h_i , і варіансу з σ^2e . Функція правдоподібності потім обчислюється шляхом інтегрування по відношенню до випадкових батьківських ефектів наступним чином:

$$L = \prod \int \left\{ f(s_j) \left[\prod^{IK} f(y_{ijk}) \right] \right\} ds_j \quad (3.28)$$

Рішеннями параметрів ML є ті значення фіксованих ефектів, і відхилення, що максимізують функцію правдоподібності. Немає простих алгоритмів для отримання рішення параметрів ML для великих змішаних моделей систем.

3.13 Оцінка компонентів варіанси, аналіз методів варіанса

Для того щоб вирішити змішану модель рівнянь, наведену в рівнянні (3.5) варіанси випадкових ефектів повинні бути відомі. На практиці ці компоненти варіанси повинні бути обчислені з тих же даних. Оцінка компонент варіанса розглядатиметься тут тільки в найзагальніших рисах.

Різні методи були запропоновані для оцінки компонентів варіанс. Ці методи можуть бути згруповані в методи типу «аналізу варіанс» і методи типу «максимальної правдоподібності». При аналізі методів типу варіанс, компоненти варіанс оцінюються першими обчисленими рішеннями для фіксованих і випадкових ефектів. Компоненти дисперсії потім оцінюються їх очікуваннями, які є функціями рішень.

Найбільш важливим з цих методів є метод III Хендерсона, який вперше обчислює рішення для всіх ефектів при абсолютно фіксованій моделі (Henderson, 1984). Ці рішення використовуються потім для отримання очікування для компонент дисперсії на основі скорочення сум квадратів. Оцінки компонентів варіанс, одержувані за методом III Хендерсона є об'єктивними, але не гарантується, що вони лежать з параметром у 1 просторі. Можна отримувати негативні компоненти варіанси, або генетичні або екологічні кореляції поза діапазону від -1 до +1. Це питання буде розглянуто більш докладно в розділі 3.16.

Розраховані варіанси для остаточних і випадкових ефектів в загальній змішаній моделі, σ^2e і σ^2u , обчислюються наступним чином:

$$\hat{\sigma}_e^2 = \frac{y'y - R(b, u)}{N - r - t + 1} \quad (3.29)$$

$$\hat{\sigma}_u^2 = \frac{R(u|b) - \hat{\sigma}_e^2(t - 1)}{\text{tr}[Z'Z - Z'X(X'X)^{-1}X'Z]} \quad (3.30)$$

де N - число спостережень, r - ранг (як правило, кількість рівнів - 1) фіксованих ефектів, t це ранг випадкових ефектів, 'tr' означає слід матриці, а $(X'X)^{-1}$ - це «узагальнена» інверсія $X'X$. Слід квадратної матриці це сума діагональних елементів. Для пояснення узагальнених зворотних см. Сирл

(1971). $R(b, u)$ і $R(u | b)$ скорочення сум квадратів, і обчислюються наступним чином:

$$R(b, u) = y'[XZ] \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix} \begin{bmatrix} X' \\ Z' \end{bmatrix} y \quad (3.31)$$

і:

$$R(b|u) = R(b, u) - y'Z(Z'Z)^{-1}Z'y \quad (3.32)$$

3.14 Оцінка максимальної правдоподібності з компонентів дисперсії

Сучасним методом вибору для оцінки компонентів варіанса є REML. Цей метод полягає в необхідності ітерації, так як формули, що використовуються для оцінки компонентів дисперсії є функціями рішень змішаних моделей, які обчислюються на основі попередніх оцінок значень параметрів. REML відрізняється від стандартної максимальної правдоподібності за рахунок того факту, що беруться оцінки фіксованих ефектів не рівні їх значень параметрів. Це буде пояснено нижче. Ми спочатку опишемо максимальну правдоподібність (MLE) компонентів дисперсії, а потім опишемо зміни, необхідні для обмеження MLE. Висновок даний тут уважно досліджували Лінч та Уолш (1998). Для більш докладного пояснення см. Сирл та ін (1992).

Як і в розділі 2, оцінки ML отримані шляхом побудови функції правдоподібності (функція спільної щільності спостережень), диференціюючи логарифм цієї функції по відношенню до параметрів, встановлюючи ці відмінності, рівними нулю, і вирішуючи для значення параметрів в отриманій системі рівнянь. Для змішаної моделі параметрами є компоненти варіанса і фіксовані ефекти. Статистична щільність функції для одного спостереження зі змішаної моделі було дано в рівнянні (3.4). Функція правдоподібності це спільна щільність всіх спостережень після інтегрування по випадковим ефектам.

Як зазначалося раніше, розподіл y в змішаній моделі є багатомірне нормальним. Функція правдоподібності для вибірки з одновимірним нормальним розподілом було дано в рівнянні (2.19). Враховуючи той факт, що середні і варіанси в змішаній моделі можуть бути різним для кожного спостереження, ця ймовірність стає:

$$L = \prod \left[\frac{e^{-(y_i - \mu_i)^2 / 2\sigma_i^2}}{\sqrt{2\pi\sigma_i^2}} \right] \quad (3.33)$$

де N є розміром вибірки, μ_i і σ_i^2 є середнє і дисперсія для спостереження i .

Для будь-якого ряду значень, x_1 до x_n , $\text{Pexp}(x_i) = \exp \Sigma(x_i)$, де $\exp[.]$ означає $[.]$ в ступені e . Таким чином, після видалення константи з мультиплікативної суми, рівняння (3.33) приймає вигляд:

$$L = (2\pi)^{-N/2} \left(\prod \sigma_i^2 \right)^{-1/2} \exp \left[- \sum (-(y_i - \mu_i)^2 / 2\sigma_i^2) \right] \quad (3.34)$$

Так як $P\sigma_i^2 = |V|$, де $|V|$ - визначник V (Сирл, 1982), рівняння (3.34) можна записати в матричній формі наступним чином:

$$L = (2\pi)^{-N/2} |V|^{-1/2} \exp \left[-\frac{1}{2} (y - X\beta)' V^{-1} (y - X\beta) \right] \quad (3.35)$$

Як зазначалося в рівнянні (3.3), для змішаної моделі:

$$V = ZGZ^T + R = Z(A\sigma_u^2)Z^T + I\sigma_e^2 \quad (3.36)$$

Натуральний логарифм функції правдоподібності полягає в наступному:

$$\text{Log } L = -(N/2) \ln(2\pi) - \frac{1}{2} \ln |V| - \frac{1}{2} (y - X\beta)' V^{-1} (y - X\beta) \quad (3.37)$$

У теорії, оцінки ML тепер можуть бути отримані шляхом диференціювання правої частини рівняння (3.37) по відношенню до β , σ_u^2 і σ_e^2 , і встановивши ці похідні рівними нулю. Однак, рішення ML для β самі є функціями компонент варіанси. Таким чином, ітераційне рішення буде необхідно. Диференціація рівняння (3.37) по відношенню до β дає:

$$\partial(\log L)/\partial\beta = -2X^T V^{-1} (y - X\beta) \quad (3.38)$$

Встановимо цю похідну рівній нулю, і вирішуючи для β отримуємо:

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y \quad (3.39)$$

де $\hat{\beta}$ і \hat{V} є оцінками β і V . Рівняння (3.39) є узагальненим рішенням найменших квадратів, наведених у главі 2, за винятком того, що V замінюється на \hat{V} . Як зазначалося раніше, ці рішення є функціями оцінок компонент варіанси. Диференціація $\text{Log } L$ по відношенню до дисперсії вимагає похідні $|V|$ і V^{-1} . Для будь-якої квадратної матриці, M , похідні $|M|$ і M^{-1} по відношенню до скалярної величини x обчислюються наступним чином (Searle, 1982):

$$\partial(\ln |M|)/\partial x = \text{tr}(M^{-1} \partial M / \partial x) \quad (3.40)$$

$$\partial M^{-1} / \partial x = M^{-1} (\partial M / \partial x) M^{-1} \quad (3.41)$$

Використовуючи ці рівняння, похідна $\log L$ по відношенню до σ_i^2 , вектору компонент дисперсії є:

$$\begin{aligned} \partial(\log L)/\partial\sigma_i^2 = & -1/2 \text{tr}(V^{-1} V_i) + 1/2 (y - X\hat{\beta})' V^{-1} V_i V^{-1} (y - X\hat{\beta}) + \\ & 1/2 (\hat{\beta} - \beta)' X^T V^{-1} V_i V^{-1} X (\hat{\beta} - \beta) \end{aligned} \quad (3.42)$$

де $V_i = \partial V / \partial \sigma_i^2 = I$ для $\sigma_i^2 = \sigma_e^2$ і $v_i = z_i z_i'$ для $\sigma_i^2 = \sigma_u^2$. Встановивши ці рівняння рівними нулю, $\hat{\beta} = \hat{\beta}$, і помінявши місцями, це рівняння приймає вигляд:

$$\text{tr}(V^{-1} V_i) = (y - X\hat{\beta})' V^{-1} V_i V^{-1} (y - X\hat{\beta}) \quad (3.43)$$

Для σ_e^2 рівняння (3.43) приймає вигляд:

$$\text{tr}(V^{-1}) = (y - X\hat{\beta})' V^{-1} V^{-1} (y - X\hat{\beta}) \quad (3.44)$$

і для σ_u^2 рівняння (3.43) приймає вигляд:

$$\text{tr}(V^{-1} Z A Z^T) = (y - X\hat{\beta})' V^{-1} Z A Z^T V^{-1} (y - X\hat{\beta}) \quad (3.45)$$

Рівняння (3.44) і (3.45) є функціями як $\hat{\beta}$, так і \hat{V}^{-1} , яке з'являється в обох частинах цих рівнянь. Крім того, \hat{V}^{-1} є нелінійною функцією компонент дисперсії. Таким чином, ітераційні методи рішення потрібні для вирішення цих нелінійних рівнянь. Методи, описані в главі 2, можуть бути використані для ітерації нелінійних рівнянь.

3.15 Обмеження максимальної правдоподібності з компонентами дисперсії

Проблему зі стандартним MLE можна пояснити, розглядаючи MLE варіанси для нормального розподілу, отриманого в главі 2.

Ця оцінка $(1/N)\sum(y_i - \mu)^2$. Таким чином, оцінка дисперсії є функцією μ , фактичної середньої, яка невідома. Для стандартної оцінки дисперсії вибірки ця проблема вирішена шляхом заміни μ середньої вибірки, і ділення на $N-1$, замість N .

Ділення на $N-1$ відноситься до невизначеності в оцінці істинного середнього. У змішаній моделі оцінки компонент варіанси, паралельна проблема виникає в тому, що MLE припускає, що рішення фіксованих ефектів рівні істинним значенням.

У REML, ця проблема вирішується за допомогою лінійного перетворення спостережень, яке видаляє фіксовані ефекти з моделі. Розглянемо загальну змішану модель наведену в рівнянні (3.2). Визначимо матрицю K так, що $KX = 0$. Тоді:

$$y^* = Ky = KZ_U + K_e \quad (3.46)$$

Серл та ін (1992) показують, що K задовольняє наступному співвідношенню :

$$P = K_T(KVK^T)^{-1}K \quad (3.47)$$

$$\text{де : } P = V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}XV^{-1} \quad (3.48)$$

так що :

$$y^{*T}V^{*-1}y^* = (y^TK^T)(KVK^T)^{-1}K(Ky) = y^TPy \quad (3.49)$$

Підстановка Ky для y , $KX = 0$ для X , KZ для Z і KVK^T для V в рівняння (3.43) дає наступні REML оцінки компонент дисперсії :

$$\text{tr}(\hat{P}) = y^TPy \quad (3.50)$$

для σ^2_e і:

$$\text{tr}(\hat{P}ZAZ^T) = y^T \hat{P}ZAZ^TPy \quad (3.51)$$

для σ^2_u . Як і у випадку ML, ці нелінійні рівняння можуть бути вирішені тільки шляхом ітерації. Оцінки компонентів REML дисперсії може бути отримано похідно - основними і похідними вільними методами.

3.16 Проблема компонент дисперсії поза простору параметрів

У розділі 2 ми відзначали, що ML- оцінки можуть бути зміщеними. Проте, важливою перевагою методів ML (і REML) є те, що всі оцінки параметрів повинні лежати в просторі параметрів. На перший погляд це властивість здається тривіальною. Проте, для багатоознакової моделі, особливо моделі з більш ніж двома ознаками, дуже мало ймовірно, що матриця оцінок компонентів варіанс, отриманих іншими методами, такими як Метод III Хендерсона, буде насправді дійсним матричним компонентом дисперсії.

Насамперед, зауважимо що для дійсної матриці компонентів дисперсії, всі передбачувані відхилення повинні бути позитивним, і всі кореляції, за оцінками варіанса і коваріацій, повинні бути в межах від -1 до +1. Проте ці умови не є достатніми. Дійсна матриця варіанса повинна бути позитивно певною, або, принаймні, невід'ємно визначеною. Тобто, всі власні значення повинні бути позитивними, або, принаймні, не негативними. Хендерсон (1984) наводить приклад трьох ознакових «псевдо» варіаційних матриць, у яких всі варіанси є позитивними і все кореляції в допустимому діапазоні. Тим не менш,

ця матриця не є дійсною матрицею варіанси, тому що одне з власних значень негативно. Діагональні елементи зворотної цієї матриці негативні. Тому, якщо використовується матриця для розрахунку генетичної матриці варіанси в змішаних моделях рівнянь, рішення для особин, які мають родинне ставлення, буде менш схожі, ніж рішення для не пов'язаних тварин.

Слід зазначити, що, навіть якщо компоненти варіанси обчислюються за допомогою ML методу, це не гарантує, що функції компонент варіанси будуть також лежать в межах допустимого простору параметрів. Наприклад, в батьківській моделі описаної розділі 3.2, успадкування, як правило, оцінюється як чотири компоненти варіанси батька поділені на суму всіх компонент варіанса. Хоча спадковість повинна бути менше 1, немає ніякої гарантії, що оцінка отримана з успадкування оцінкою компонент варіанс REML лежатиме в просторі параметрів.

ЗАКЛЮЧЕННЯ

Ми розглянули першими бажані властивості оцінок в цілому. Потім ми розглянули різні методи оцінювання параметрів схрещування інбредних ліній, підкреслюючи метод найменших квадратів і ML. ML не є тривіально застосовуваною, але може бути застосована до багатьох моделей, які не піддаються вирішенню за допомогою інших способів. На відміну від інших методів оцінки, MLE має бути в межах простору параметрів. Були представлені загальні методи для обчислення похибки оцінок дисперсій MLE і тестування гіпотез, пов'язаних зі значеннями параметрів.

Аналіз змішаних моделей є набагато складнішим, ніж аналіз фіксованих моделей. Ці моделі є переважними для генетичного аналізу, по-перше, тому що вони використовують інформацію про генетичні дисперсії і генетичні зв'язки серед тварин, які не можуть бути використані фіксованими моделями. По-друге, рішення випадкових ефектів може повернути до середнього. Тому що варіанса рішень зростає залежно від кількості інформації включеної в аналіз. Це дозволяє проводити точне порівняння генетичних значень особин з широко різними обсягами інформації. Зворотне вірно для рішень фіксованих моделей. Їх варіанси зменшуються, коли кількість інформації збільшується. Нарешті, для випадкових ефектів, регресії істинних значень ефектів на їх передбачення рівні одиниці. Методи були представлені для оцінки компонентів варіанси випадкових ефектів на основі ML і REML. Рівняння REML є нелінійними, і самі вони функції оцінок параметрів. Таким чином, ці рівняння можуть бути вирішені тільки шляхом ітерацій. Якщо QTL ефекти вважаються випадковими, також буде необхідно оцінити їх дисперсії через REML.

СЛОВНИК ТЕРМІНІВ

Absorption	Поглинання
Accuracy	Точність
Additive genetic effects	Адитивні генетичні ефекти
Ancestors	Предки
Animal models	Тваринні моделі
Analysis of variance (ANOVA)	Дисперсійний аналіз (ANOVA)
Assumptions	Припущення
Bayesian estimation	Байєсівського оцінювання
Best Linear Unbiased Estimates (BLUE)	Кращі Лінійні Незміщені Оцінки
Best Linear Unbiased Predictors (BLUP)	Кращі Лінійні Незміщені Прогнози
Biased estimates	Зміщені оцінки
Binomial distribution	Біноміальний розподіл
Biochemical polymorphism	Біохімічний поліморфізм
Bootstrap	Бутстреп
Breeding values	Племінні оцінки
Chromosomes	Хромосоми
Codominance	Кодомінантність
Concordance	Узгодження
Consistency	Стійкість
Continuous distribution	Безперервний розподіл
Correlated traits	Корельовані риси
Covariance matrix	Коваріаційна матриця
Cramer-von Mises distance	Крамера-Мізеса відстань
Cytogenetics	Цитогенетика
Dam additive genetic component	Аддитивна генетична компонента матері
Dam effect	Ефект матері
Data distribution	Розподіл даних
Density functions	Функції щільності
Derivative-based methods	Методи на основі похідних
Derivative-free methods	Похідно-вільні методи
Discrete distribution	Дискретний розподіл
DNA base pairs (bp)	Пар основ ДНК (п.о.)
DNA-microsatellites	ДНК-мікросателіти
Dominance	Домінування
Electrophoresis	Електрофорез
Environmental effects	Вплив навколишнього середовища
Enzyme polymorphisms	Поліморфізми ферментів
Epistasis	Епістаз

Error variances	Дисперсії помилок
Estimated breeding values (EBV)	Значення племінної цінності
First derivative based methods (EM)	Методи на основі першої похідної
Fixed effects	Фіксовані ефекти
Gamma distribution	Гамма-розподіл
Gauss-Seidel iteration	Гаусса-Зейделя ітерація
General linear model	Загальна лінійна модель
Genetic evaluations	Генетичні оцінки
Genetic map units	Генетичних карт одиниці
Genetic mapping	Генетичне картування
Genetic standard deviations	Генетичні стандартні відхилення
Genetic variance	Генетична дисперсія
Genome	Геном
Genotype distribution	Розподіл генотипів
Genotypes	Генотипи
Hardy-Weinberg distribution	Розподіл Харді-Вайнберга
Henderson's method	Метод Хендерсона
Herd effect	Ефект стада
Heritability	Спадковість
Homozygotes	Гомозиготи
Individual animal model (IAM)	Індивідуальна тваринна модель (IAM)
Joint density function	Спільна функція щільності
Kronecker product	Кронекера продукт
Kurtosis	Екссес
Least squares solutions	Рішення найменших квадратів
Likelihood function	Функція правдоподібності
Linear models	Лінійні моделі
Linear regression estimation	Лінійне оцінювання регресії
Linkage	Зчеплення
Linkage disequilibrium	Нерівновага по зчепленню
Marker alleles	Маркера аллелі
Marker-assisted selection (MAS)	Селекція за допомогою маркерів (MAS)
Markov Chain Monte Carlo (MCMC)	Ланцюги Маркова метод Монте-Карло (MCMC)
Maternal effect	Материнський ефект
Matrix algebra	Матрична алгебра
Maximum likelihood analysis	Аналіз максимальної правдоподібності
Mean squares	Середні квадрати
Meiosis	Мейоз

Mendelian genetics	Менделевська генетика
Mendelian sampling	Менделевська вибірка
Microsatellites	Мікросателіти
Minimum difference estimation	Оцінка мінімальної різниці
Mixed models	Змішані моделі
Moments method	Метод моментів
Multiple regression models	Моделей множинної регресії
	Багатовимірний нормальний розподіл
Multivariate normal distribution	Нормальний розподіл
Normal distribution	Нормальні рівняння
Normal equations	Нульова гіпотеза
Null hypothesis	Простір параметрів
Parameter space	Батьківські алелі
Paternal alleles	Батьківський ефект
Paternal effect	Постійний екологічний ефект
Permanent environmental effect	Плейотропія
Pleiotropy	Розподіл Пуассона
Poisson distribution	Полі (ТГ) сайти
Poly(TG) sites	Полімеразна ланцюгова реакція (ПЛР)
Polymerase chain reaction (PCR)	Аналіз популяційний
Population analyses	Популяції розмір
Populations, size of	Імовірність апостеріорна
Posterior probability	Помилки прогнозу дисперсія
Prediction error variance	Апріорний розподіл
Prior distribution	Потомство
Progeny	Аналіз ЛКП
QTL analysis	Випадкові ефекти
Random effects	Записи
Records	Зниження тваринна модель
Reduced animal model (ram)	Знижена оцінка максимальної ймовірності
Reduced maximum likelihood estimation (REML)	Регресійна модель
Regression model	Надійність
Reliability	Повторювана модель
Repeatability model	Залишкова сума квадратів
Residual sum of squares	Обмеження максимальної правдоподібності
Restricted maximum likelihood estimation	Довжини рестрикційних фрагментів поліморфізм (ПДРФ)
Restriction fragment length polymorphisms (RFLP)	Надійність
Robustness	Вибірка
Sample	

Sample size	Обсяг вибірки
Second derivative based methods	Методи на основі другої похідної
Sex-limited traits	Риси обмеженні статтю
Simple sequence repeats (SSR) see DNA	Прості повтори послідовності (SSR)
Single nucleotide polymorphisms (SNPS)	Одиничний нуклеотидний поліморфізм
Skewness	Асиметрія
Software	Програмне забезпечення
Standard deviation	Стандартне відхилення
Statistical density	Статистична щільність
Statistical methods	Статистичні методи
Statistical power	Статистична потужність
Unbiased estimates	Об'єктивні оцінки
Unbiasedness	Незміщеність
Variables	Змінні
Variance components	Компоненти дисперсії

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

Іовенко В. М., Скрепець К. В., Писаренко Н. Б., Харічев Д. С. Особливості поліморфізму окремих QTL-генів овець південного регіону України. *Науковий вісник «Асканія-Нова»*. 2017. Вип. 10. С. 39-48.

Харічев Д. С. Сучасні молекулярно-генетичні дослідження у вівчарстві. *Вівчарство та козівництво*. 2017. № 2. С. 215-222.

Cavanagh C. R., Jonas E., Hobbs M., Thomson P.C., Tammen I., Raadsma H. W. Mapping Quantitative Trait Loci (QTL) in sheep. III. QTL for carcass composition traits derived from CT scans and aligned with a meta-assembly for sheep and cattle carcass QTL. *Genetics Selection Evolution*. 2010. Vol. 42. P. 1-14.

Edfors-Lilja I., Wattrang E., Marklund L., Moller M., Andersson-Eklund L., Andersson L., Fossum C. (1998). Mapping quantitative trait loci for immune capacity in the pig. *The Journal of Immunology*. 1998. Vol. 161(2). P. 829-835.

Flint J. Analysis of quantitative trait loci that influence animal behavior. *Journal of neurobiology*. 2003. Vol. 54(1). P. 46-77.

Grisel J. E. *Quantitative trait locus analysis*. Alcohol Research & Health. 2000. Vol. 24(3). P. 169-174.

Johnson P. L., McEwan J. C., Dodds K. G., Purchas R. W., Blair H. T. Meat quality traits were unaffected by a quantitative trait locus affecting leg composition traits in Texel sheep. *Journal of animal science*. 2005. Vol. 83(12). P. 2729-2735.

Stinckens A., Schroyen M., Peeters L., Janssens S., Buys N. Quantitative trait mutations in cattle, sheep and pigs: a review. *Animal Science Reviews*. 2010. Vol. 59. P. 1-15.

Weller J. I. *Quantitative trait loci analysis in animals*. CABI Publishing, Oxford, 200. 387 p.

Навчально-наукове видання

**ЗАПРОВАДЖЕННЯ МЕТОДІВ ГЕНОМНОЇ СЕЛЕКЦІЇ
ДЛЯ ПІДВИЩЕННЯ ЯКІСНИХ ОЗНАК М'ЯСА
СВИНЕЙ ТА ОВЕЦЬ
(виробничо-практичні рекомендації)**

Укладачі:

Луговий Сергій Іванович
Крамаренко Сергій Сергійович
Крамаренко Олександр Сергійович
Лихач Вадим Ярославович
Атаманюк Ігор Петрович

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.