

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧЕРНІВЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ЮРІЯ ФЕДЬКОВИЧА**

**Факультет математики та інформатики
Кафедра математичного моделювання**

**ЗАСТОСУВАННЯ НЕЛІНІЙНОЇ РЕГРЕСІЇ ДЛЯ ДОСЛІДЖЕННЯ
МОДЕЛЬНИХ ПРИКЛАДІВ У СЕРЕДОВИЩІ R**

Кваліфікаційна робота

Рівень вищої освіти – перший (бакалаврський)

Виконав:

студент 4 курсу, 407 групи

Скороход Володимир Анатолійович

Керівник:

кандидат фізико-математичних наук,
доцент Дорошенко І.В.

*До захисту допущено
на засіданні кафедри математичного моделювання
протокол № 16 від 30 травня 2023 р.
Зав. кафедрою _____ проф. Черевко І.М.*

Чернівці – 2023

Анотація

У кваліфікаційній роботі розглянуто основні концепції нелінійної регресії, ілюстровані різними прикладами. Самостійною частиною є реалізація модельних прикладів у середовищі R.

Ключові слова: регресійний аналіз, експоненціальна модель, метод нелінійних найменших квадратів, гіпотеза, довірчі інтервали.

Annotation

In the qualifying work, the main concepts of nonlinear regression are considered, illustrated by various examples. An independent part is the implementation of model examples in the R environment.

Keywords: regression analysis, exponential model, nonlinear least squares method, hypothesis, confidence intervals.

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів наукових досліджень інших авторів мають посилання на відповідне джерело.

_____ В.А. Скороход
(підпис)

ЗМІСТ

ВСТУП.....	4
1. Основні означення та мотивуючі приклади	5
2. Метод нелінійних найменших квадратів	15
3. Алгоритм Гаусса-Ньютона	19
4. Статистичні властивості оцінки МНК.....	23
4.1 Властивості для великої вибірки	23
4.2 Властивості малої вибірки	28
4.3. Асимптотичні довірчі інтервали та перевірка гіпотез	30
4.4. Три методи статистичного висновку в великій вибірці	36
5. Модельні приклади.....	38
ВИСНОВКИ.....	50
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	51.
ДОДАТКИ.....	52

Вступ

Нелінійна регресія є потужним статистичним інструментом, рідко описуваним в традиційних підручниках зі статистики. Нелінійна регресія є не лише практично важливою технікою, але й важливим прикладом статистичної моделі з реальними даними, де класична теорія безперебійної оцінки і достатніх статистик не працюють. Несумнівно, лінійна модель є чемпіоном серед статистичних методів при моделюванні зв'язку між змінними. Однак іноді зв'язок не є лінійним, наприклад, коли відповідь має сигмоїдну форму, тоді застосовується нелінійна регресія. У відмінність від лінійної регресії, нелінійна регресія є складною статистичною моделлю, де дослідження властивостей малооб'ємних вибірок є складним завданням, тому ми спираємося на асимптотичні властивості. Основним методом оцінки є нелінійний метод найменших квадратів, який, на відміну від лінійного методу найменших квадратів, вимагає ітераційного процесу.

В нелінійній регресійній моделі виникає ряд числових проблем:

- (1) пошук задовільного початкового значення,
- (2) існування розв'язку нелінійного методу найменших квадратів,
- (3) наявність кількох локальних мінімумів для суми квадратів залишків.

1. Основні означення та мотивуючі приклади

Нелінійна регресія є очевидним узагальненням лінійної регресії (лінійної моделі). Основна відмінність полягає в тому, що в нелінійній регресії очікуване значення залежної змінної є нелінійною функцією параметрів. Наприклад, квадратична регресія

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$

є лінійною щодо параметрів і, отже, може бути оцінена за допомогою методу найменших квадратів, застосованого до лінійної моделі після зміни змінної, $z_i = x_i^2$. Регресійні моделі, які можуть бути зведені до лінійної моделі шляхом введення нових змінних, називаються криволінійними моделями. Модель

$$Y_i = g(\beta^0 + \beta_1 x_i + \beta_2 z_i) + \varepsilon_i, \quad (1.1)$$

де g - монотонна функція, називається квазілінійною і не може бути зведена до лінійної моделі.

Внутрішньолінійні моделі є нелінійними регресійними моделями, які можуть бути зведені до лінійної моделі після репараметризації:

$$Y_i = g_1(\beta_1, \beta_2)x_i + g_2(\beta_1, \beta_2)z_i + \varepsilon_i, \quad (1.2)$$

де g_1 і g_2 - взаємно-однозначні функції. Насправді, при покладанні $\gamma_1 = g_1(\beta_1, \beta_2)$ і $\gamma_2 = g_2(\beta_1, \beta_2)$ ми зводимо модель до лінійної. Внутрішньонелінійні моделі є тими, які не можуть бути зведені до лінійної моделі шляхом репараметризації.

Якщо x_i позначає вектор незалежних/пояснювальних змінних/предикторів, нелінійна регресія визначається як

$$Y_i = f(x_i; \beta) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1.3)$$

де $\{\varepsilon_i\}$ - (ненаблюдавані) некорельовані випадкові члени з нульовим середнім значенням і постійною дисперсією σ^2 . Функція f називається *функцією регресії*, а

іноді - функцією відгуку. Лінійна регресія є спеціальним випадком (1.3) з $f(x_i; \beta) = \beta'x_i$, але в нелінійній регресії навіть не потрібно вимагати співпадіння розмірностей β і x_i . Крім того, зазвичай використовується позначення

$$f_i(\beta) = f(x_i; \beta) \quad (1.4)$$

для скорочення формул; тоді ми пишемо

$$Y_i = f_i(\beta) + \varepsilon_i.$$

Це позначення означає, що нелінійні регресійні функції f_i можуть бути будь-якими і не обов'язково пов'язані з передбачувальними змінними. Зазвичай, $f_i(\beta)$ - це неперервні та диференційовані функції, але не завжди, як ми побачимо пізніше. Припускаємо, що m -вимірний вектор параметрів β належить до всього R^m або до опуклого m -вимірного параметричного набору $\Theta \subset R^m$. Для однозначної оцінки параметрів нам потрібно припустити, що нелінійна регресія є ідентифікованою: $f_i(\beta_1) = f_i(\beta_2)$ для всіх $i = 1, 2, \dots, n$, впливає $\beta_1 = \beta_2$.

Іноді функція регресії є нелінійною, як у випадку квазілінійної регресії (1.1), але може бути зведена до лінійної функції шляхом застосування оберненої функції g^{-1} до обох сторін регресійного рівняння. Наведений нижче приклад підкреслює, як змінюється статистична модель при такій трансформації. В результаті, метод найменших квадратів, застосований до $g^{-1}(Y_i)$, не дає ті ж самі оцінки, які були отримані в оригінальній нелінійній регресійній моделі.

Приклад 1.1. Оцінка експоненціальної моделі [2]. Прибуток компанії зростає відповідно до експоненційної тенденції $y = e^{a+\beta t}$ протягом років з певними варіаціями. Порівняємо дві регресійні моделі: (a) термін помилки є множниковим, $\ln Y_t = \alpha + \beta t + \delta_t$, та (b) термін помилки є адитивним, що призводить до квазілінійної регресії, $Y_t = e^{a+\beta t} + \varepsilon_t$, де обидва терміни помилки мають нормальний розподіл з нульовим середнім значенням.

Розв'язання. Модель (а) підтверджує, що Y_t має логнормальний розподіл зі сподіванням $E(Y_t) = e^{\alpha + \beta t + \sigma^2/2}$ та дисперсією $\text{var}(Y_t) = (e^{\sigma^2} - 1)e^{2(\alpha + \beta t) + \sigma^2}$, що означає, що співвідношення стандартного відхилення до середнього є постійним. Ця модель зводиться до лінійної моделі після застосування логарифмічного перетворення. З іншого боку, модель (b) передбачає постійне стандартне відхилення через адитивну помилку (рисунок 1.1). Ця модель є квазілінійною з функцією $g = \exp$. Мультиплікативна модель помилки передбачає, що варіація пропорційна значенню тенденції, тоді як адитивна модель помилки має постійну варіацію (стандартне відхилення). Вибір не є очевидним, але добре підтверджено, що реальні дані часто підкоряються логнормальному розподілу, тому модель з мультиплікативною помилкою є більш обґрунтованою.

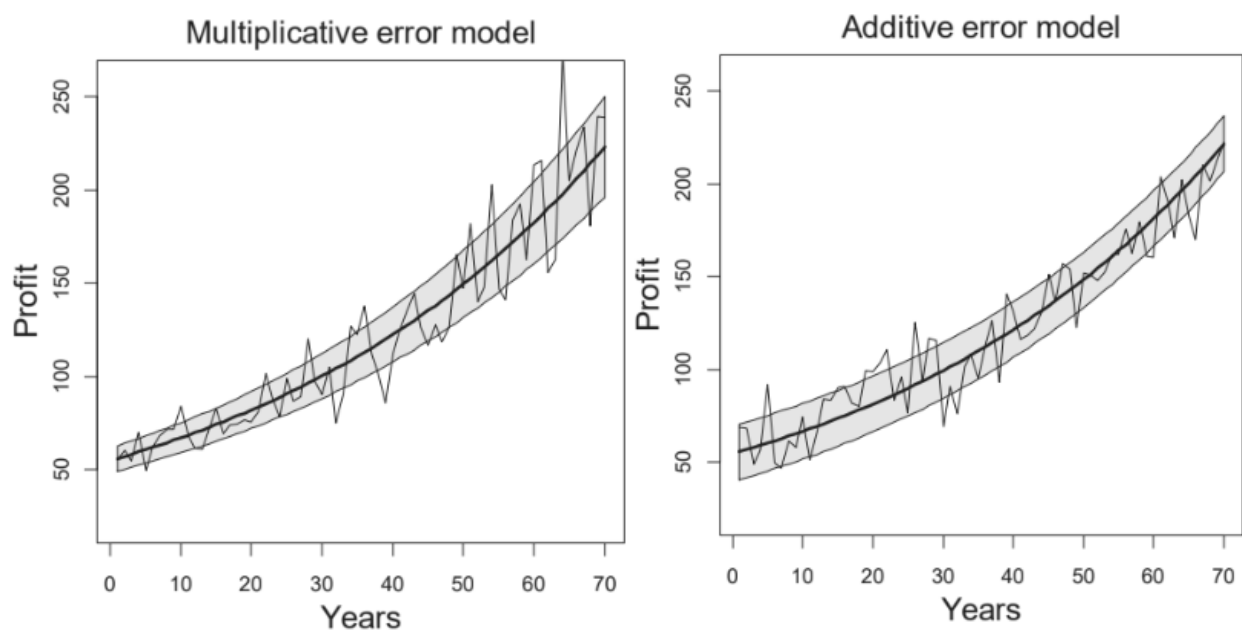


Рисунок 1.1: Дві моделі похибок для експоненційної тенденції: мультиплікативна модель передбачає, що варіація прибутку пропорційна тенденції, тоді як адитивна модель має постійну варіацію прибутку навколо тенденції

При роботі з нелінійною регресією вибір функції f завжди є важливим питанням. Зазвичай нелінійна регресія виникає у науці та інженерії, де фізичні закони визначають відношення. У таких випадках функція регресії є розв'язком

підлежного диференціального рівняння. Нижче наведено два мотивуючі приклади.

Приклад 1.2. Температура в чашці з чаєм [2]. Потрібно отримати модель падіння температури в чашці з чаєм після наливання окропленої води. Надати інтерпретацію параметрів та знайдіть час комфортного пиття, коли температура чаю становить 50 °C. Встановіть нелінійну регресію для оцінки параметрів.

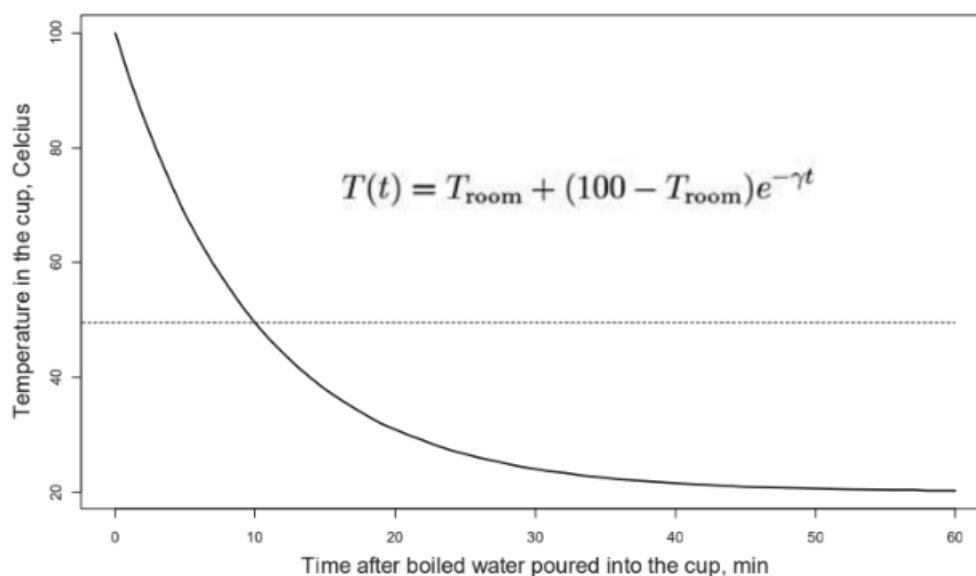


Рисунок 1.2: Нелінійна модель регресії для температури чашки з чаєм

Розв'язання. Відповідно до закону охолодження Ньютона, температура зменшується пропорційно до різниці між температурою води, $T = T(t)$, і температурою повітря в кімнаті, $T_{\text{кімн}}$. Оскільки швидкість охолодження моделюється першою похідною, цей закон записується у вигляді звичайного диференціального рівняння

$$\frac{dT}{dt} = -\gamma(T - T_{\text{кімн}}), \quad t \geq 0, \quad (1.5)$$

де t - це час, що минув після наливання кип'яченої води в чашку, а γ - параметр охолодження ($\gamma > 0$). При $t = 0$ маємо початкову умову $T(0) = 100^\circ\text{C}$. Рівняння (1.5) є одновимірним диференціальним рівнянням першого порядку. Щоб розв'язати це рівняння, ми відокремлюємо T і t і інтегруємо:

$$\int \frac{dT}{T - T_{\text{кімн}}} = -\gamma \int dt,$$

Отримуємо $T(t) = T_{\text{кімн}} + e^{-\gamma t + C}$, де C - постійна. Враховуючи граничну умову, ми знаходимо, що $e^C = 100 - T_{\text{кімн}}$, тобто кінцева модель для температури води у чашці має вигляд

$$T(t) = T_{\text{кімн}} + (100 - T_{\text{кімн}})e^{-\gamma t}, \quad (1.6)$$

де параметр γ підлягає оцінці (рисунки 1.2).

Звичайно, ця модель є спрощенням. Об'єм, форма чашки і матеріал, з якого вона виготовлена, можуть впливати на результат, але ми не беремо до уваги ці деталі. Тепер ми можемо дати більш точну інтерпретацію параметра γ як швидкості (відносного зменшення), з якою падає температура в чашці:

$$\gamma = -\frac{dT/dt}{T - T_{\text{кімн}}}.$$

Модель (1.6) може відповісти на питання, коли чай прийнятний для пиття. Припустимо, що комфортна температура (не занадто гаряча і не занадто холодна) дорівнює $T_{\text{комф}} = 50^\circ\text{C}$. Тоді оптимальний час буде:

$$t_{\text{опт}} = \frac{1}{\gamma} \ln \frac{100 - T_{\text{кімн}}}{T_{\text{комф}} - T_{\text{кімн}}}.$$

Тепер ми встановлюємо нелінійну регресію. Припустимо, що є доступний датчик температури (важливо, щоб цей датчик не впливав на температуру води в чашці), і було зроблено n вимірів температури води в моменти часу t_1, t_2, \dots, t_n . Завдяки неминучій похибці вимірювання, модель набуває форми нелінійної регресійної моделі:

$$T_i = T_{\text{кімн}} + (100 - T_{\text{кімн}})e^{-\gamma t_i} + \varepsilon_i, \quad (1.7)$$

де $E(\varepsilon_i) = 0$ і $\text{var}(\varepsilon_i) = \sigma^2$. Якщо температура в кімнаті не відома, ми повинні оцінити два параметри, $T_{\text{кімн}}$ та γ , щоб регресійна функція була

$$f_i(T_{\text{кімн}}, \gamma) = T_{\text{кімн}} + (100 - T_{\text{кімн}})e^{-\gamma t_i}.$$

Приклад 1.3 Регресія з точкою зміни: Поріг відновлення ДНК. Вплив на клітинну ДНК іонізуючого випромінювання завдає різних типів пошкоджень, що призводить до зниження виживаності клітин. Клітини можуть відновлювати ДНК при низькій дозі радіації, але не можуть зробити це після перевищення певного порогу. Для оцінки порогу були проведені експерименти на клітинах з послідовним збільшенням дози радіації та вимірюванням виживаності *in vitro* (рисунок 1.3). Потрібно отримати модель нелінійної регресії для виживаності клітин як функції введеної радіації.

Розв'язок. Час виживання (час до смерті) має скошене вправо розподілення і, відповідно до експоненційної функції виживання, може бути моделюваний на логарифмічній шкалі. Найзручніше використовувати логарифм з основою 10 для зручності читання: значення 2 інтерпретується як 100, -1 як 0,1 і так далі. Позначимо через L логарифмічну шкалу виживання на \log_{10} типової клітини, яка піддається одній дозі випромінювання r в момент часу 0. Якщо радіація низька, очікується, що $L = \alpha - \beta r$, де α - це очікувана тривалість життя (у годинах на \log_{10} шкалі) піддаємих випромінюванню клітин, а $\beta > 0$ - швидкість виживання. Відомо, що після досягнення певного порогу r_T швидкість виживання знижується на $\delta > 0$. Таким чином, залежність між часом виживання і випромінюванням має наступний вигляд:

$$L = L(r; \alpha, \beta, \delta, r_T) = \begin{cases} \alpha - \beta r & \text{якщо } r < r_T \\ \alpha - \beta r - \delta(r - r_T) & \text{якщо } r \geq r_T \end{cases}$$

де ні α , β , δ , або r_T не відомі.

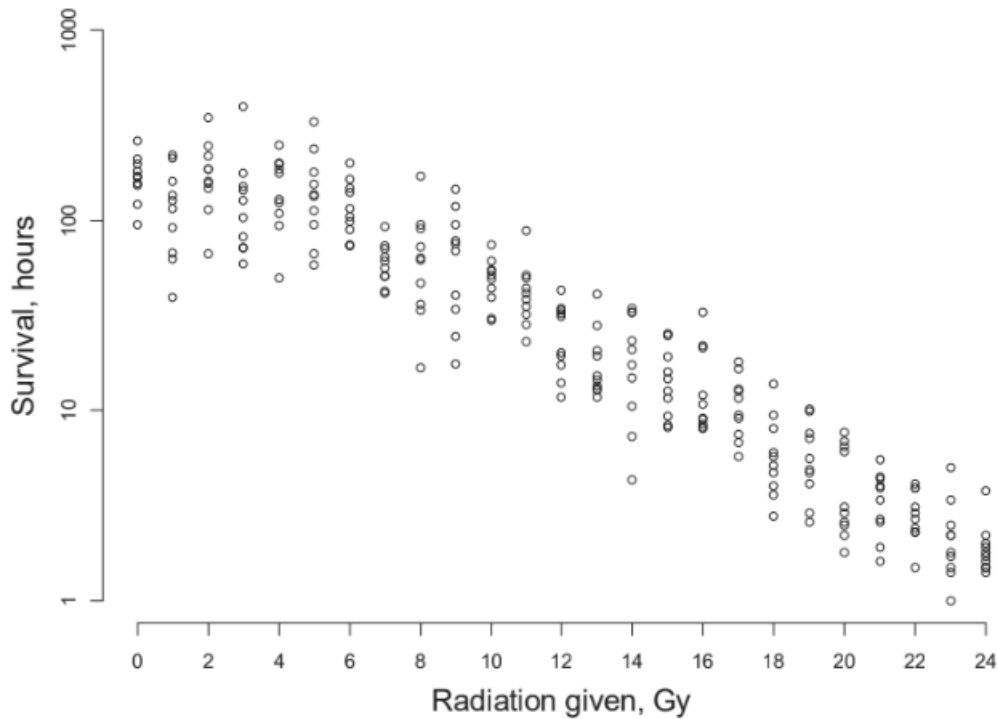


Рисунок 1.3: Регресія з точкою зміни: час виживання клітини як функція одного опромінення

Очевидно, функцію L можна подати більш компактно у такому вигляді:

$$L = \alpha - \beta r - \delta \max(r - r_T, 0).$$

Питання: при заданій серії *in vitro* експериментів $\{(r_i, Y_i), i = 1, 2, \dots, n\}$, де Y_i - вимірний час виживання, як ми оцінюємо задані параметри α, β, δ та зокрема r_T як поріг (точку зміни)? Оскільки час виживання підлягає похибці вимірювання та, що є найважливіше, реакції клітин на опромінення, неодмінно виникають відхилення вимірювань від моделі, що приводить до нелінійної регресії.

$$Y_i = \alpha - \beta r_i - \delta \max(r_i - r_T, 0) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1.8)$$

де $\{\varepsilon_i, i = 1, 2, \dots, n\}$ - некорельовані випадкові помилки з нульовим середнім значенням та постійною дисперсією. Параметр зміни r_T центральний для проблеми оцінювання і має властивість нелінійності. Фактично, якщо r_T було

відомо, то регресія могла б бути зведена до лінійної регресії з трьома коефіцієнтами:

$$Y_i = \alpha, -\beta r_i - \delta M_i + \varepsilon_i$$

де $M_i = \max(r_i - r_T, 0)$. Модель (9.8) є певною нелінійною регресійною моделлю, для якої стандартна функція nls не працює.

Давньогрецький філософ Арістотель (384—322 до н.е.) твердив, що швидкість вільно падаючого предмета пропорційна його масі [3]. Це твердження Арістотеля було перевірено майже через дві тисячі років. Великий італійський вчений Галілео Галілей (1564—1642) продемонстрував, відмінно від аристотелівської фізики, що два важкі об'єкти різної маси, випущені зі Схилу Пізанської вежі, падають на землю майже одночасно. Він продовжив експерименти з кульками, що котилися по нахилах, і докладно виміряв "повних сто разів", досягнувши точності, коли відхилення між двома спостереженнями не перевищувало однієї десятої пульсації. Це було неймовірним досягненням, оскільки в той час не існувало годинника. Ці експерименти довели його до висновку, що "вільне падіння є рівномірно прискореним рухом". Для багатьох Галілео Галілей був першим вченим, а для мене він був першим статистиком, оскільки він зібрав дані, проаналізував їх і зробив висновок. В той самий рік, коли помер Галілео, народився Ісаак Ньютон (1642—1726), який запропонував формулу для відстані, пройденої вільно падаючим тілом,

$$S(t) = \frac{1}{2}gt^2, \quad t \geq 0, \quad (1.9)$$

де g - константа, близька до $9,8 \text{ м/с}^2$, яка входить до складу його теорії механіки з трьома законами. Хоча його теорія спочатку була підтверджена обертанням планет, формула (1.9) суперечить здоровому глузду: згідно з цією формулою пір'я і металева куля падають з однаковою швидкістю. Тому аристотелівське вчення панувало в людському мисленні понад дві тисячі років! Тільки пізніше експерименти в вакуумній трубці і, нарешті, відеофіксований експеримент на

Місяці командиром астронавтом Девідом Скоттом у 1971 році, коли він випустив молоток і перо з однакової висоти і вони вдарилися об землю одночасно, підтвердили формулу Ньютона.

Причиною того, що формула Ньютона не працює на Землі, є опір повітря. За роки було запропоновано кілька моделей вільного падіння для корекції опору повітря. Одна з таких моделей, яка використовується в наступному прикладі, призводить до нелінійної регресії.

Приклад 1.4. Рівняння вільного падіння: Арістотель чи Галілей? Аллмараз та ін.[4] запропонували модель, яка надалі буде називатися С-моделлю ,

$$S(t; C) = \frac{1}{C} \ln \frac{e^{t\sqrt{gC}} + e^{-t\sqrt{gC}}}{2}, \quad t \geq 0 \quad (1.10)$$

щоб описати відстань, пройдену вільно падаючим об'єктом, де $C \geq 0$ - константа, пов'язана зі співвідношенням об'єму до маси об'єкта, а $g = 9.7935 \text{ м/с}^2$ - гравітаційна (прискорення вільного падіння) константа Землі. Доведіть, що ця модель поєднує закони Арістотеля та Ньютона і побудуйте нелінійну регресію для оцінки параметра С.

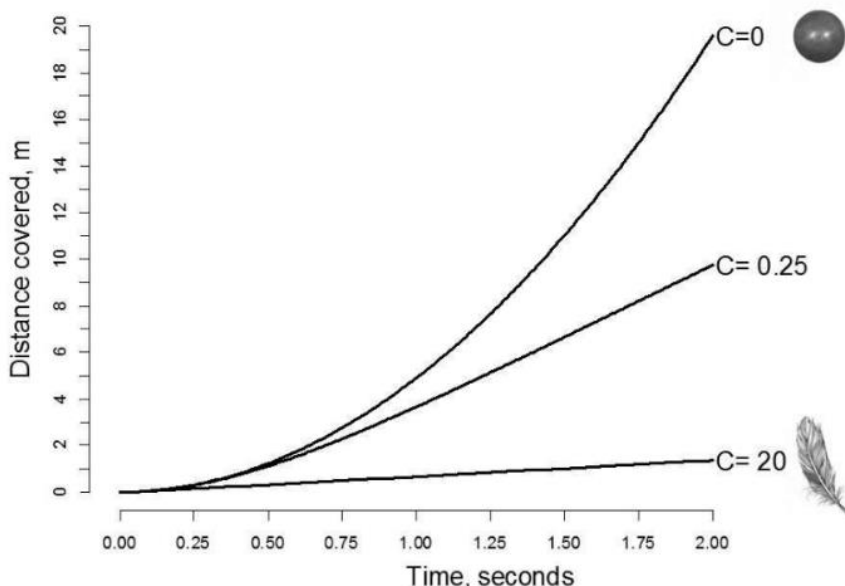


Рисунок 1.4: Відстань, пройдена вільно падаючими об'єктами

Свинцева кулька має $C = 0$, і відстань, пройдена нею, відповідає Ньютоному закону тяжіння $0.5gt^2$. Пір'я має велике $C = 20$ і пройшло лише близько метра після 2 секунд вільного падіння, тоді як звичайний об'єкт з $C = 0.25$ пройшов близько 9 метрів.

Розв'язання. В (1.10) навмисно вказано, що S залежить від параметра C , оскільки він буде предметом оцінювання. Спочатку ми покажемо, що коли $C \rightarrow 0$ (відсутній опір повітря або маса об'єкта нескінченно велика), (1.10) збігається до закону Ньютона (1.9). Застосуємо правило Лопітала,

$$\lim_{C \rightarrow 0} \frac{1}{C} \ln \frac{e^{t\sqrt{gC}} + e^{-t\sqrt{gC}}}{2} = t\sqrt{g} \lim_{C \rightarrow 0} \frac{1}{2\sqrt{C}} \frac{e^{t\sqrt{gC}} - e^{-t\sqrt{gC}}}{e^{t\sqrt{gC}} + e^{-t\sqrt{gC}}}$$

Тепер розкладемо $e^{t\sqrt{gC}}$ та $e^{-t\sqrt{gC}}$ у ряди Тейлора,

$$e^{t\sqrt{gC}} = 1 + t\sqrt{g}\sqrt{C} + \frac{1}{2}t^2gC + o(C), \quad e^{-t\sqrt{gC}} = 1 - t\sqrt{g}\sqrt{C} + \frac{1}{2}t^2gC + o(C),$$

і отримуємо

$$\begin{aligned} \lim_{C \rightarrow 0} \frac{1}{C} \ln \frac{e^{t\sqrt{gC}} + e^{-t\sqrt{gC}}}{2} &= t\sqrt{g} \lim_{C \rightarrow 0} \frac{1}{2\sqrt{C}} \frac{e^{t\sqrt{gC}} - e^{-t\sqrt{gC}}}{e^{t\sqrt{gC}} + e^{-t\sqrt{gC}}} \\ &= t\sqrt{g} \lim_{C \rightarrow 0} \frac{2t\sqrt{g}\sqrt{C}}{4\sqrt{C}} = \frac{1}{2}gt^2. \end{aligned}$$

Друге, ми досліджуємо, як летить об'єкт з $C > 0$ після певного періоду часу, або, більш конкретно, знаходимо швидкість об'єкта, коли $t \rightarrow \infty$,

$$\lim_{t \rightarrow \infty} \frac{S(t)}{t} = \lim_{t \rightarrow \infty} \frac{dS(t)}{dt} = \frac{\sqrt{gC}}{C} \lim_{t \rightarrow \infty} \frac{e^{t\sqrt{gC}} - e^{-t\sqrt{gC}}}{e^{t\sqrt{gC}} + e^{-t\sqrt{gC}}} = \sqrt{\frac{g}{C}}.$$

Це означає, що на Землі вільно падаючий об'єкт досягає своєї максимальної швидкості, яка дорівнює $\sqrt{g/C}$, і більше не прискорюється - саме це вчив Арістотель. На рис. 1.4 показано формулу (1.10) і два граничних випадки. Перо має дуже велике C і, отже, після падіння швидко досягає своєї лінійної траєкторії

і сповільнено спускається під парашутом. З іншого боку, свинцева куля має C близьке до нуля і падає відповідно до закону Ньютона.

Тепер дослідимо ще один крайній випадок, коли $t \approx 0$. Знайдемо прискорення вільно падаючих об'єктів на Землі як другу похідну відстані (1.10), обчислену при $t = 0$,

$$\left. \frac{d^2 S}{dt^2} \right|_{t=0} = \left. \frac{4ge^{-2t\sqrt{gC}}}{(1 + e^{-2t\sqrt{gC}})^2} \right|_{t=0} = g.$$

Це означає, що коли починається падіння, кінетика відповідає закону Ньютона (1.9), незалежно від геометрії об'єкта. Загалом, що стосується вільно падаючих об'єктів на Землі, Арістотель, був правий при нескінченності, а Галілей був правий при нулі. Для будь-яких даних про вільне падіння вимірювання часу t_i та відстані, пройденої за цей час, S_i ми повинні оцінити об'єкт-специфічний параметр C у нелінійній регресії

$$S_i = \frac{1}{C} \ln \frac{e^{t_i\sqrt{gC}} + e^{-t_i\sqrt{gC}}}{2} + \varepsilon_i.$$

2. Метод нелінійних найменших квадратів

Нелінійні найменші квадрати (ННК) є очевидним узагальненням методу лінійних найменших квадратів (МНК) і мінімізує суму квадратів залишків

$$S(\beta) = \sum_{i=1}^n (Y_i - f_i(\beta))^2. \quad (2.1)$$

Мінімізатор $\hat{\beta} \in \Theta$ називається нелінійною оцінкою найменших квадратів для β . Ми можемо виразити суму квадратів у векторній формі таким чином:

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad f(\beta) = \begin{bmatrix} f_1(\beta) \\ f_2(\beta) \\ \vdots \\ f_n(\beta) \end{bmatrix}$$

так що функцію регресії можна розглядати як відображення з Θ в R^n , $f: \Theta \rightarrow R^n$. Оскільки ми припускаємо, що нелінійна регресія є ідентифікованою, відображення f є однозначним: різні точки з $\Theta \subset R^m$ відображаються в різних точки R^n .

У векторній нотації нелінійна регресія може бути компактно записана як

$$y = f(\beta) + \varepsilon$$

де RSS задається формулою

$$S(\beta) = \|y - f(\beta)\|^2, \quad (2.2)$$

квадрата Євклідової норми. У геометричній мові, $S(\beta)$ є квадратом відстані між вектором даних y та поверхнею регресії $f(\beta)$. На рисунку 2.1 графічно ілюструються нелінійні найменші квадрати для випадку, коли кількість спостережень $n = 3$, а кількість невідомих параметрів (розмір вектора β) $m = 2$. Поверхня регресії визначається як $f: R^2 \rightarrow R^3$ та параметризується як $f = f(\beta)$. Розв'язок нелінійних найменших квадратів - це точка на поверхні f , яка найближча до вектора даних y . Нагадаємо, що для лінійної моделі поверхня $f(\beta)$ буде площиною, яку утворюють передиктори x_1, \dots, x_m . Коли $m = 1$, поверхня регресії перетворюється в криву, параметризовану як $\beta_1 = \beta$.

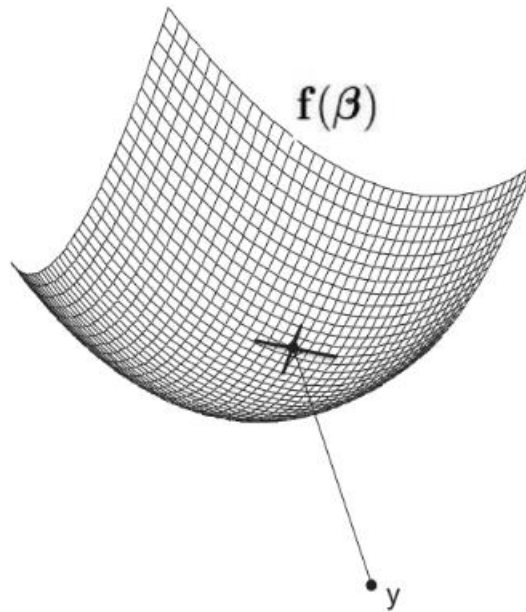


Рисунок 2.1: Геометрія методу нелінійного найменших квадратів

RSS досягає мінімуму там, де вектор перших похідних $\partial(S\beta)/\partial\beta$ дорівнює нулю. Таким чином, ми приходимо до визначення нормального рівняння, яке може бути записане у матричній формі як

$$J'(\beta)(y - f(\beta)) = 0, \tag{2.3}$$

де J - це *Якобіан*, матриця розміром $n \times m$, яка містить похідні функції регресії,

$$J = \frac{\partial f}{\partial \beta} = \begin{bmatrix} \frac{\partial f_1}{\partial \beta_1} & \frac{\partial f_1}{\partial \beta_2} & \dots & \frac{\partial f_1}{\partial \beta_m} \\ \frac{\partial f_2}{\partial \beta_1} & \frac{\partial f_2}{\partial \beta_2} & \dots & \frac{\partial f_2}{\partial \beta_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial \beta_1} & \frac{\partial f_n}{\partial \beta_2} & \dots & \frac{\partial f_n}{\partial \beta_m} \end{bmatrix} = \left[\frac{\partial f}{\partial \beta_1}, \frac{\partial f}{\partial \beta_2}, \dots, \frac{\partial f}{\partial \beta_m} \right],$$

де $\partial f/\partial\beta_j$ - це j -тий стовпець Якобіану ($j = 1, \dots, m$). Ми припускаємо, що J - матриця повного рангу (стовпці-вектори лінійно незалежні). Для лінійної моделі, $J = X$ і не залежить від β . Нормальне рівняння (2.3) є необхідною умовою першого порядку для мінімуму СКЗ. Це означає, що якщо $\hat{\beta}$ - оцінка ННК, то

вектор залишкових значень $y - f(\hat{\beta})$ ортогональний до дотичної площини поверхні $f(\beta)$ при $\beta = \hat{\beta}$. Для лінійної моделі розв'язок нормального рівняння повертає мінімум суми квадратів. Для нелінійної моделі це може бути не так: $S(\beta)$ може мати декілька локальних мінімумів, і оцінка ННК може навіть не існувати.

Коли похибки розподілені нормально, оцінки ННК та максимальної правдоподібності (МП) збігаються.

Дійсно, у векторній нотації маємо $y \sim N(f(\beta), \sigma^2 I)$, і функція лог-правдоподібності, за винятком постійного члена, має наступний вигляд

$$l(\beta) = -\frac{1}{2} \left(n \ln \sigma^2 + \frac{1}{\sigma^2} \|y - f(\beta)\|^2 \right). \quad (2.4)$$

Очевидно, що $l = \max$, коли $S(\beta) = \min$. Іншими словами, ННК = МП, якщо помилки розподілені нормально.

МП-оцінка для σ^2 має вигляд

$$\hat{\sigma}^2 = \frac{1}{n} \|y - f(\hat{\beta})\|^2.$$

Можна використовувати $n - m$ в знаменнику, але, на відміну від лінійної моделі, це не робить $\hat{\sigma}^2$ неупередженим.

Навіть якщо Y_i є незалежним та нормально розподіленим, оцінка ННК не є ефективною для скінченного n , на відміну від лінійної моделі. Вивчення властивостей нелінійної регресії для малих вибірок є складною задачею.

Мінімізація суми квадратів (2.1) не має розв'язку в явному вигляді. Замість цього слід використовувати ітераційні алгоритми. Іноді, коли моделі складні, алгоритми не збігаються — початкові значення вектора параметрів є важливими.

3. Алгоритм Гаусса-Ньютона

Алгоритм Гаусса-Ньютона (ГН) є популярним і ефективним способом знаходження мінімуму суми квадратів у нелінійній регресії. Ідея алгоритму базується на повторному наближенні нелінійної функції регресії лінійною функцією (лінійною моделлю) та використанні формули для лінійного методу найменших квадратів для отримання наступного наближення вектора параметрів. Дійсно, якщо \mathbf{b}_k є вектором параметрів на ітерації k , то ми наближаємо функцію регресії наступним чином:

$$f(\beta) \simeq f(\mathbf{b}_k) + J_k(\beta - \mathbf{b}_k), \quad (3.1)$$

де J_k є Якобіаном або матрицею розмірності $n \times t$, що містить похідні функції регресії, обчислені в точці \mathbf{b}_k ,

$$J_k = \left. \frac{\partial f}{\partial \beta} \right|_{\beta=\mathbf{b}_k}$$

Лінійне наближення (3.1) використовується як для обчислювальних цілей, так і для наближеного статистичного висновку в наступних розділах. Використовуючи наближення (3.1), ми замінюємо початкову нелінійну суму квадратів (2.1) на лінійну суму квадратів

$$\|y - f(\mathbf{b}_k) - J_k(\beta - \mathbf{b}_k)\|^2$$

що призводить до розв'язку

$$\beta - \mathbf{b}_k = (J_k' J_k)^{-1} J_k'(y - f(\mathbf{b}_k)).$$

Ітерації ГН мають наступний вигляд

$$\mathbf{b}_{k+1} = \mathbf{b}_k + (J_k' J_k)^{-1} J_k'(y - f(\mathbf{b}_k)), \quad k = 0, 1, 2, \dots \quad (3.2)$$

Для початку ітерацій необхідно мати початкове значення \mathbf{b}_0 .

Теорема 3.1 Алгоритм Гаусса-Ньютона еквівалентний алгоритму оцінки Фішера, коли похибки мають нормальний розподіл.

Доведення. Логарифмічна функція правдоподібності для нелінійної регресії з помилками, які мають нормальний розподіл, задається формулою (2.4). Похідні, виражені через Якобіан J та $\varepsilon = y - f(\beta)$, можна записати як

$$\frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2} J' \varepsilon, \quad \frac{\partial l}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{\|\varepsilon\|^2}{2\sigma^4},$$

де $\varepsilon \sim N(0, \sigma^2 I)$. Тому,

$$\text{cov}\left(\frac{\partial l}{\partial \beta}\right) = \frac{1}{\sigma^4} J' \text{cov}(\varepsilon) J = \frac{1}{\sigma^2} (J'J), \quad \text{cov}\left(\frac{\partial l}{\partial \beta}, \frac{\partial l}{\partial \sigma^2}\right) = 0,$$

$$\text{var}\left(\frac{\partial l}{\partial \sigma^2}\right) = \frac{1}{4\sigma^8} \text{var}(\|\varepsilon\|^2) = \frac{2n\sigma^4}{4\sigma^8} = \frac{n}{2\sigma^4}$$

оскільки $E(\varepsilon_i \varepsilon_j \varepsilon_k) = 0$ та $\text{var}(\varepsilon_i) = 2\sigma^4$ для будь-якого i, j та k , як у лінійній моделі регресії. Нарешті, $(m + 1) \times (m + 1)$ матриця інформації Фішера для y є блок-діагональною і виглядає так само, як у лінійній моделі:

$$T = \begin{bmatrix} \frac{1}{\sigma^2} (J'J) & 0 \\ 0' & n/(2\sigma^4) \end{bmatrix}, \quad (3.3)$$

з відмінністю, що для лінійної регресії матриця $J = X$ є стала, а для нелінійної регресії J є функцією від β . Інверсійна матриця,

$$T^{-1} = \begin{bmatrix} \sigma^2 (J'J)^{-1} & 0 \\ 0' & (2\sigma^4)/n \end{bmatrix},$$

дає нижню границю Крамера—Рао для неупереджених оцінок. Ми інтерпретуємо $\sigma^2 (J'J)^{-1}$ як обернення очікуваної Гессіана, тому алгоритм оцінки Фішера має вигляд

$$\begin{aligned} \mathbf{b}_{k+1} &= \mathbf{b}_k + \sigma^2 (J'_k J_k)^{-1} \left(\frac{\partial l}{\partial \beta} \Big|_{\beta=\mathbf{b}_k} \right) = \mathbf{b}_k + \sigma^2 (J'_k J_k)^{-1} \frac{1}{\sigma^2} J'_k (y - f(\mathbf{b}_k)) \\ &= \mathbf{b}_k + (J'_k J_k)^{-1} J'_k (y - f(\mathbf{b}_k)), \end{aligned}$$

такий самий, що і ітерації Гауса-Ньютона.

Теорема 3.2 Для лінійної моделі ітерації Гауса-Ньютона збігаються після першої ітерації незалежно від точки початку.

Доведення. Для лінійної моделі маємо $f(b) = Xb$, і Якоб'ян є фіксованою матрицею, $J = X$. Якщо b_0 є точкою початку, на першій ітерації ми отримуємо

$$\begin{aligned} b_k &= b_0 + (X'X)^{-1}X'(y - Xb_0) = b_0 + (X'X)^{-1}X'y - (X'X)^{-1}(X'X)b_0 \\ &= (X'X)^{-1}X'y = \hat{b}, \end{aligned}$$

оцінку методом найменших квадратів.

Щоб виконати ітерації (3.2), матриця Якобіана повинна мати повний ранг, тобто матриця $J'_k J_k$ повинна бути невинродженою на кожній ітерації. Таким чином, ми розширюємо умову ідентифіковності, припускаючи, що Якоб'ян є неперервною функцією на Θ і має повний ранг, або еквівалентно,

$$|J'(\beta)J(\beta)| \neq 0 \quad \forall \beta \in \Theta. \quad (3.4)$$

Немає гарантії, що ітерації Гауса-Ньютона збігаються для кожної нелінійної регресії, починаючи з будь-якої точки. Однак, ми можемо довести, що якщо ітерації збігаються, то градієнт суми квадратів на межовій точці дорівнює нулю.

Теорема 3.3 Якщо ітерації (3.2) збігаються, $\lim_{k \rightarrow \infty} b_k = b_*$, тоді градієнт S в точці b_* дорівнює нулю,

$$\left. \frac{\partial S}{\partial \beta} \right|_{\beta=b_*} = 0.$$

Доведення. Поділивши обидві сторони (3.2) на $k \rightarrow \infty$, отримуємо

$$b_* = b_* + (J'_* J_*)^{-1} J'_* (y - f(b_*))$$

завдяки неперервності матриці J . Це означає, що

$$\left. \frac{\partial S}{\partial \beta} \right|_{\beta=b_*} = -2J'_* (y - f(b_*)) = 0,$$

теорема доведена.

Один з ознакових ознак збіжності полягає в тому, що значення суми квадратів за кожну ітерацію зменшується, але алгоритм Гауса-Ньютона не гарантує такого зменшення. Щоб забезпечити зменшення, слід використовувати частину кроку в напрямку Гауса-Ньютона,

$$\mathbf{b}_{k+1} = \mathbf{b}_k + \rho_k (\mathbf{J}'_k \mathbf{J}_k)^{-1} \mathbf{J}'_k (\mathbf{y} - \mathbf{f}(\mathbf{b}_k)), \quad (3.5)$$

де $0 < \rho_k \leq 1$ таке, що

$$S(\mathbf{b}_{k+1}) < S(\mathbf{b}_k) \text{ якщо } \left. \frac{\partial S}{\partial \beta} \right|_{\beta=\mathbf{b}_*} \neq 0.$$

Тоді можна довести, за м'яких умов, що послідовність $\{\mathbf{b}_k, k = 0, 1, \dots\}$ збігається до точки локального мінімуму функції $S(\beta)$. Загальний критерій теорії оптимізації для забезпечення збіжності є таким $\sum_{k=1}^{\infty} \rho_k = \infty$. Часто вибір $\rho_k = 1$ призводить до збіжності, якщо початкова точка досить близька до розв'язку. Для забезпечення зменшення S від ітерації до ітерації можна використовувати більш складні стратегії. Наприклад, ρ_k може бути обрано шляхом поділу на два до тих пір, поки значення суми квадратів не стане меншим, або може бути використано метод золотого перерізу. [3]

Можна використовувати кілька (нерівноцінних) критеріїв збіжності: (а) зупиняти ітерації, коли

$$\frac{|b_{k+1,j} - b_{k,j}|}{1 + |b_{k,j}|} < \varepsilon, \quad j = 1, \dots, m, \quad (3.6)$$

де ε - невелике число, наприклад, $\varepsilon = 10^{-5}$, або (б) коли виконується загальний критерій [28] кутового коефіцієнта,

$$\frac{(\mathbf{y} - \hat{\mathbf{f}}_k)' \mathbf{J}_k (\mathbf{J}'_k \mathbf{J}_k)^{-1} \mathbf{J}'_k (\mathbf{y} - \hat{\mathbf{f}}_k)}{\sqrt{m} \|\mathbf{y} - \hat{\mathbf{f}}_k\|^2} < \varepsilon. \quad (3.7)$$

За роки було розроблено кілька інших модифікацій алгоритму Гаусса-Ньютона, включаючи алгоритм Левенберга-Марквардта. Матриця $J'_k J_k$ може наближатися до сингулярності, коли параметри регресії складно визначити, навіть якщо виконується умова (3.4). Левенберг запропонував використовувати вираз $J'_k J_k + \lambda_k I$, де $\lambda_k > 0$, а при $k \rightarrow \infty$ задавати $\lambda_k \rightarrow 0$ (λ_k називається параметром регуляризації). Алгоритм Левенберга-Марквардта є популярним у інженерних застосуваннях при вирішенні невизначених обернених задач. Однак жоден з існуючих ітеративних алгоритмів не гарантує збіжності до глобального мінімуму, оскільки функція S може мати кілька локальних мінімумів. Доброю практикою є початок ітерацій з кількох початкових значень, щоб підтвердити, що вони збігаються до однієї й тієї ж точки.

4. Статистичні властивості оцінки МНК

Статистичні властивості будь-якого оцінювача, включаючи оцінювач ННК $\hat{\beta}$, можна вивчати з двох точок зору залежно від обсягу вибірки: (а) властивості для великої вибірки (або асимптотичні властивості) [5], коли розмір вибірки прямує до нескінченності, $n \rightarrow \infty$, і (б) властивості для малих вибірок, коли n є скінченним і фіксованим. Велике n призводить до нормального розподілу, тоді як розподіл $\hat{\beta}$ для скінченного n є складним. Крім того, оцінювач ННК може мати нескінченне середнє значення та дисперсію для скінченного n . Зазвичай для вивчення властивостей ННК на малих вибірках використовуються симуляції. Однак, симуляції мають свої обмеження, оскільки отримані знання стосуються конкретного вибору значень параметрів.

4.1 Властивості для великої вибірки

Великі властивості оцінювача ННК можуть бути вивчені за двома схемами. У детерміністичній схемі $\{x_i\}$ розглядаються як фіксовані, а умови, за яких

виконуються асимптотичні властивості, виражаються через $f_i = f(x_i; \beta)$. В рамках цієї схеми ми припускаємо, що ε_i є незалежними та однаково розподілені, але не обов'язково мають нормальний розподіл. Зрозуміло, що в цій схемі Y_i не мають тотожного розподілу, оскільки $E(Y_i) = f_i \neq \text{const}$, тому асимптотична теорія МП не застосовна. Нагадаємо, що ми мали схожу ситуацію в лінійній моделі. У стохастичній схемі $\{x_i\}$ вважаються випадковими iid, а ε_i є незалежними та нормально розподіленими, що дозволяє розглядати нелінійну регресію як умовне середнє. Ці припущення позначають, що (Y_i, x_i) є незалежними та однаково розподілені, тому застосовується класична теорія МП (ННК = МП). Загалом, за умови незалежними та однаково розподілені, оцінювач ННК є стійким і асимптотично нормальним. Крім того, якщо ε_i має нормальний розподіл, то ННК є асимптотично ефективним. Ця умова застосовується як до лінійних моделей, так і до особливого випадку нелінійної регресії.

Детермінована схема

Було запропоновано кілька умов для стійкості та асимптотичної нормальності методу ННК, коли x_i є фіксованими в (1.4) [2]. Було доведено, що якщо межа

$$Q(\beta, \beta_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (f_i(\beta) - f_i(\beta_0))^2 \quad \forall \beta \neq \beta_0 \quad (4.1)$$

що $Q(\beta, \beta_0) = 0$ тоді й тільки тоді, коли $\beta = \beta_0$, то метод ННК $\hat{\beta}$ є стійким. Узагальнив вищезазначений критерій: метод ННК є стійким тоді й тільки тоді, коли

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n (f_i(\beta) - f_i(\beta_0))^2 = \infty \quad \forall \beta \neq \beta_0 \quad (4.2)$$

Очевидно, що умова (4.1) впливає з (4.2).

Тепер звернемо увагу на асимптотичну нормальність. На відміну від лінійного методу найменших квадратів, нелінійний метод найменших квадратів може мати кілька розв'язків нормального рівняння (2.3). Дуже складно надати умови унікальності оцінки ННК для скінченного вибірки [2]. Принаймні, нам потрібно бути впевненими, що нормальне рівняння має єдиний розв'язок в асимптотичному сенсі, тобто існує наступний границі:

$$G(\beta; \beta_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (f_i(\beta) - f_i(\beta_0)) \left. \frac{\partial f_i}{\partial \beta} \right|_{\beta=\beta_0} \quad (4.3)$$

і $G(\beta; \beta_0) = 0$ тоді й тільки тоді, коли $\beta = \beta_0$. Далі, ми називаємо цю умову асимптотичною унікальністю. Припустимо, що, крім існування границь (4.1) і (4.3), віддалено схожих на умову Ліндеберга-Феллера, похідні рівномірно обмежені: для кожного $\delta > 0$ і кожного β_0 , існує таке K , що рівномірно,

$$\left\| \frac{\partial f_i}{\partial \beta} \right\| \leq K, \quad i = 1, 2, \dots \quad (4.4)$$

для всіх β таких, що $\|\beta - \beta_0\| \leq \delta$. Крім того, ми припускаємо, що границя

$$A(\beta) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial f_i}{\partial \beta} \right) \left(\frac{\partial f_i}{\partial \beta} \right)' \quad \forall \beta,$$

і, крім того, матриця $A(\beta)$ є невідродженою для кожного β . Тоді оцінка ННК розподілена нормально у великих вибірках (зауважте, що ε_i не обов'язково розподілені нормально):

$$\sqrt{n}(\hat{\beta} - \beta) \simeq N(0, \sigma^2 A^{-1}(\beta)), \quad n \rightarrow \infty, \quad (4.5)$$

де β - це вектор справжніх параметрів. Цей результат означає, що матрицю коваріації для $\hat{\beta}$ можна надійно оцінити як

$$\text{cov}(\hat{\beta}) = \hat{\sigma}^2 (\hat{J}'\hat{J})^{-1}, \quad (4.6)$$

де

$$\hat{\sigma}^2 = \frac{1}{n-m} \sum_{i=1}^n (Y_i - f_i(\hat{\beta}))^2 \quad (4.7)$$

це оцінка дисперсії помилки, а

$$\hat{J} = \left. \frac{\partial f}{\partial \beta} \right|_{\beta=\hat{\beta}}$$

є Якобіаном, обчисленим на оцінці ННК (вирахованому на останньому кроці алгоритму Гаусса-Ньютона). Практично, розподіл оцінки ННК наближено нормальний для великих n , що можна записати, з деякою неоднозначністю, як

$$\hat{\beta} \simeq N(\beta, \hat{\sigma}^2 (\hat{J}'\hat{J})^{-1}) \quad (4.8)$$

На жаль, на відміну від лінійної моделі, ми не можемо стверджувати безвідхиленість оцінки ННК або довести, що співвідношення між оцінкою ННК мінус справжнє значення та її стандартною помилкою має t -розподіл. Також, ми не можемо стверджувати, що оцінка ННК досягає нижньої межі Крамера-Рао і, отже, є ефективною, як у лінійній моделі, просто тому, що $\hat{\beta}$ є упередженою в більшості нелінійних регресій.

Стохастична схема

Під час цієї схеми ми розглядаємо нелінійну регресію як умовне середнє, тому

$$E(Y_i | X = x_i) = f(x_i; \beta), \quad X_i \stackrel{iid}{\sim} p(x; \theta),$$

де p - маргінальна щільність X з деяким невідомим, зайвим параметром θ . Наприклад, коли Y_i умовно розподілені нормально при заданому X_i , ми можемо подумати про загальну популяцію (Y, X) з умовною щільністю

$$Y|X = x \sim N(f(x; \beta), \sigma^2)$$

і маргінальна щільність $X \sim p(x; \theta)$. Тоді класична теорія МП застосовується з функцією лог-правдоподібності, за винятком постійного доданку, який задається так:

$$\begin{aligned} l(\beta, \theta) &= \sum_{i=1}^n \left(-\frac{1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (Y_i - f(x_i; \beta))^2 + \ln p(x_i; \theta) \right) \\ &= -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - f(x_i; \beta))^2 + \sum_{i=1}^n \ln p(x_i; \theta). \end{aligned}$$

Критична умова тут полягає в тому, що вектори β та θ не мають спільних компонент, тому максимізація l щодо β не залежить від θ . Це означає, що Y_i є однаково розподіленими та ННК = МП, а інформація Фішера є блок-діагональною, тобто оцінювання θ не впливає на оцінювання β .

За умовами теорії МП, за припущенням

$$G(\beta; \beta_0) = 0 \text{ якщо } \beta = \beta_0,$$

де

$$G(\beta; \beta_0) = E_X \left((f(X; \beta) - f(X; \beta_0)) \frac{\partial f(X; \beta)}{\partial \beta} \Big|_{\beta=\beta_0} \right), \quad (4.9)$$

то можемо вивести, що оцінка ННК асимптотично ефективна і має асимптотичний нормальний розподіл

$$\sqrt{n}(\hat{\beta} - \beta) \simeq N(0, \sigma^2 B^{-1}(\beta)), \quad n \rightarrow \infty, \quad (4.10)$$

де

$$B(\beta) = E_X \left[\left(\frac{\partial f(X; \beta)}{\partial \beta} \right) \left(\frac{\partial f(X; \beta)}{\partial \beta} \right)' \right]$$

є додатно визначеною матрицею.

На практиці, зазначені умови асимптотичної нормальності ННК в детермінованому підході (4.5) та стохастичному підході (4.10) виглядають дуже схоже лише з відмінністю у використанні емпіричного середнього та математичного сподівання відповідно. За рахунок закону великих чисел та Розширеного теорема Слутського ми можемо зробити висновок, що на практиці обидва підходи призводять до однакової апроксимації матриці коваріації оцінки ННК, яка задана формулою (4.6), і апроксимація розподілу (4.8) є придатною для обох схем.

Асимптотичний розподіл оцінки ННК в обох схемах є однаковим, з єдиною відмінністю, що за детермінованою схемою використовується значення/оцінки на вибірці для справжніх параметрів μ_x та σ_x^2 . Більше того, якщо μ_x та σ_x^2 є невідомими, обидві схеми дають однаковий результат, коли μ_x та σ_x^2 замінюються їх вибірковими оцінками.

4.2 Властивості малої вибірки

Властивості оцінювача ННК при малій вибірці (фіксоване значення n) важко дослідити аналітично через числові ускладнення, такі як недоіснування та недоунікальність розв'язку нормального рівняння. Теорія нижньої межі Крамера-Рао тут не працює просто тому, що безузбічні оцінювачі не існують для багатьох нелінійних статистичних моделей, включаючи нелінійну регресію. Згадані ускладнення ілюструються на прикладі простої нелінійної регресії, де оцінювач ННК не має скінченного математичного сподівання та дисперсії.

Приклад 4.1 Регресія Філлера. Розглянемо нелінійну модель регресії $Y_i = \nu(1 + \tau x_i) + \varepsilon_i$, де x_i - фіксовані значення, а ε - незалежні та однорідно розподілені випадкові величини з нульовим математичним сподіванням ($i = 1, 2, \dots, n$). Покажемо, що оцінювач ННК для параметра τ має розподіл Коші i , отже, має нескінченне математичне сподівання та дисперсію.

Розв'язання. Оцінювачі ННК для параметрів ν та τ мінімізують функцію СКЗ $\sum_{i=1}^n (Y_i - \nu(1 + \tau x_i))^2$. Застосувавши очевидну репараметризацію, ми зводимо цю задачу до лінійних найменших квадратів: $\beta_1 = \nu$ та $\beta_2 = \nu\tau$. Тому, якщо $\widehat{\beta}_1$ та $\widehat{\beta}_2$ є оцінками ННК перехвату та нахилу в простій лінійній регресії Y на x , оцінка ННК для τ знаходиться як відношення, $\hat{\tau} = \widehat{\beta}_2 / \widehat{\beta}_1$. Оскільки $\widehat{\beta}_1$ та $\widehat{\beta}_2$ мають нормальний розподіл, відношення має розподіл Коші, і таким чином має нескінченне математичне сподівання та дисперсію, як показав Філлер майже сто років тому.

Точна, досить складна щільність оцінювача ННК для однопараметричної нелінійної регресії з незалежними та нормально розподіленими похибками, $y = f(\beta) + \varepsilon$, за умови, що нормальне рівняння має єдиний розв'язок. Позначимо цю щільність

$$A = A(b) = \|\dot{f}(b)\|^2, \quad B = B(b; \beta) = \ddot{f}(b)f(b) - f(\beta),$$

$$C = C(b) = \dot{f}(b)\ddot{f}(b), \quad D = D(b; \beta) = \dot{f}(b)f(b) - f(\beta),$$

$$E = E(b) = \|\ddot{f}(b)\|^2, \quad Q = Q(b; \beta, \sigma^2) = \frac{A^2 + AB - CD}{\sqrt{\sigma^2 A(AE - C^2)}}$$

де одна крапка над f позначає вектор першої похідної розмірності $n \times 1$, а дві крапки позначають вектор другої похідної розмірності $n \times 1$. Щільність оцінювача ННК задана рівнянням

$$p_{EX}(b; \beta, \sigma^2) = \frac{\|\dot{f}(b)\|}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} \frac{\left(\dot{f}(b)f(b) - f(\beta) \right)^2}{\|\dot{f}(b)\|^2} \right] \times \alpha_{EX}(b; \beta, \sigma)$$

де

$$\alpha_{EX}(b; \beta, \sigma) = \left(1 + \frac{AB - CD}{A^2}\right) (2\Phi(Q) - 1) + \frac{2\sigma}{A} \sqrt{\frac{AE - C^2}{A}} \phi(Q),$$

де Φ і $\phi = \Phi'$ - стандартна нормальна функція розподілу та її щільність.

Цікавим є той факт, що точний розподіл оцінювача ННК залежить від перших двох похідних функції нелінійної регресії, практичні застосування цього результату обмежені через його складність.

4.3. Асимптотичні довірчі інтервали та перевірка гіпотез

Оскільки властивості нелінійної регресії в скінченних вибірках складні, ми спираємося на асимптотичні властивості для перевірки гіпотез та побудови довірчих інтервалів (ДІ). При припущенні про нормальний розподіл, застосовується теорія МП, і, фактично, методи, описані нижче, можна отримати з асимптотичної нормальності оцінювача ННК = МП. Розглянемо три типи наближень: метод Вальда, відношення правдоподібності та критерій оцінки результату.

Метод Вальда

Побудова ДІ Вальда та пов'язана перевірка гіпотез з асимптотичним нормальним наближенням (при великому значенні n) є простою: якщо (4.6) - оцінена матриця коваріації, то $(1 - \alpha)100\%$ симетричний ДІ для β_j дорівнює $\hat{\beta}_j \pm q_{1-\alpha/2} SE_j$, де $q_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, а SE_j - оцінка стандартної похибки оцінки ННК, обчислена як квадратний корінь з (j, j) -го діагонального елемента матриці (4.6).

$$SE_j = \hat{\sigma} \sqrt{(\hat{J}'\hat{J})^{jj}}, \quad j = 1, 2, \dots, m,$$

де $\hat{\sigma}^2$ є оцінкою для σ^2 , обчисленою як середньоквадратичні залишки (4.7), а верхній індекс позначає діагональний елемент оберненої матриці. Відповідно, якщо нульова гіпотеза має вигляд $H_0: \beta_j = \beta_{j0}$ як впливає з (4.8), гіпотеза

відхиляється, якщо $|\hat{\beta}_j - \beta_{j0}|/SE_j > q_{1-\alpha/2}$, при цьому значення p -показника дорівнює $2\Phi(-|\hat{\beta}_j - \beta_{j0}|/SE_j)$. Вбудована функція `R nls` повідомляє про стандартну похибку SE_j , t -статистику $\hat{\beta}_j/SE_j$ та значення p -показника при нульовій гіпотезі $\beta_{j0} = 0$.

Перевагою підходу Вальда є те, що матриця коваріації оцінюється один раз після останньої ітерації методу Гаусса-Ньютона і потім використовується для обчислення ДІ та перевірки гіпотез для всіх параметрів m . Тест Вальда легко модифікувати для перевірки лінійної функції параметрів, скажімо, $a'\beta = c$, де a - це фіксований вектор розмірності $m \times 1$, а c - константа. Тоді Z -статистика набуває вигляду

$$Z = \frac{a'\hat{\beta} - c}{\hat{\sigma}\sqrt{a'(\hat{J}'\hat{J})^{-1}a}} \approx N(0,1), \quad (4.11)$$

з p -value обчислюється як $2\Phi(-|Z|)$.

Наближена довірча область $(1 - \alpha)$ для параметрів впливає з наближення (4.8) і має еліптичну форму з центром в оцінці ННК $\hat{\beta}$,

$$\{\beta \in R^m: (\beta - \hat{\beta})'(\hat{J}'\hat{J})(\beta - \hat{\beta}) \leq \sigma^2\chi^{-2}(1 - \alpha, m)\}, \quad (4.12)$$

де $\chi^{-2}\{1 - \alpha, m\}$ є $(1 - \alpha)$ квантилем розподілу хі-квадрат з m ступенями вільності. Ми очікуємо, що довірча область з ймовірністю не менше $1 - \alpha$ покриває справжній вектор параметрів β , принаймні для достатньо великого n . Деякі автори пропонують використовувати розподіл Фішера для правої нерівності в (4.12), що означає, що квантиль хі-квадрат замінюється на m -кратний квантиль Фішера, $mF^{-1}(1 - \alpha, m, n - m)$, щоб імітувати довірчу область лінійної моделі. Зверніть увагу, що для великого n обидва набори будуть близькими, оскільки при зростанні другого ступеня вільності розподіли хі-квадрат і Фішера збігаються.

Апроксимація шляхом відношення правдоподібностей

Більш складний тест відношення правдоподібностей (LR) включає оцінку додаткової нелінійної регресії для перевірки гіпотези $\beta_j = \beta_{j_0}$ для кожного фіксованого $j = 1, 2, \dots, m$. Нагадаємо, що для загального тесту ВП використовується формула:

$$-2 (\max l_{restricted} - \max l_{unrestricted}) \simeq \chi^2(k),$$

де l - це функція логарифму правдоподібності, $\max l_{unrestricted}$ - це МП-оцінка з незалежними значеннями параметрів β , $\max l_{restricted}$ - це МП-оцінка з обмеженими значеннями параметрів β , а k - кількість обмежень/обмежувальних умов. Ми застосуємо цей загальний тест до перевірки $\beta_j = \beta_{j_0}$. Позначимо \hat{S}_{j_0} та \hat{S} мінімальні значення суми квадратів залишків при обмеженні $\beta_j = \beta_{j_0}$ та без обмежень відповідно,

$$\hat{S}_{j_0} = \min_{\beta: \beta_j = \beta_{j_0}} \|y - f(\beta)\|^2, \quad \hat{S} = \min_{\beta} \|y - f(\beta)\|^2, \quad (4.13)$$

Після деяких алгебраїчних перетворень, використовуючи вираз (2.4), ми отримуємо

$$-2 (\max l_{restricted} - \max l_{unrestricted}) = n \ln(\hat{S}_{j_0}/\hat{S}),$$

де \hat{S}_{j_0} обчислюється за допомогою нелінійної регресії з фіксованим значенням j -го параметра $\beta_j = \beta_{j_0}$. Тому для великих значень n ми отримуємо, що при нульовій гіпотезі $\beta_j = \beta_{j_0}$,

$$n \ln(\hat{S}_{j_0}/\hat{S}) \simeq \chi^2(1) \quad (4.14)$$

разом з p -value обчислюється як $1-pchisq$, де $pchisq$ обчислюється при значенні $n \ln(\hat{S}_{j_0}/\hat{S})$ з одним ступенем вільності. Нумеричним недоліком тесту на основі відношення правдоподібності є те, що для тестування лінійної комбінації

параметрів $a'\beta = c$ потрібне мінімування суми квадратів з обмеженням. Оскільки $Z \sim N(0,1)$, це означає, що $Z^2 \sim \chi^2(1)$, ми можемо переписати наближення (4.14) таким чином

$$\text{sign}(\hat{\beta}_j - \beta_{j0}) \sqrt{n \ln \left(\frac{\hat{S}_{j0}}{\hat{S}} \right)} \simeq N(0,1), \quad (4.15)$$

де $\hat{\beta}_j$ є j -тою компонентою оцінки ННК $\hat{\beta}$. Наближення (4.15) буде називатися *наближенням на основі знакового відношення правдоподібності*. Зауважте, що число під коренем є невід'ємним, тому що $\hat{S}_{j0} \geq \hat{S}$, оскільки \hat{S}_{j0} - це СЗК за обмеженням.

Тест на основі відношення правдоподібності дає інтервал довіри з нижнім та верхнім обмеженнями для параметра β_j як розв'язок рівнянь

$$\ln \left(\hat{S}_{j0}(\beta_j) \right) - n \ln \hat{S} - \chi^{-2}(1 - \alpha, 1) = 0, \quad (4.16)$$

де $\hat{S}_{j0}(\beta_j)$ є сумою квадратів залишків, коли j -ий параметр фіксується на значенні β_j , а $\chi^{-2}(1 - \alpha, 1)$ є $(1 - \alpha)$ квантилем розподілу хі-квадрат з одним ступенем вільності. Цей метод буде називатися довірчим інтервалом на LR-профіль. Можна побудувати графік лівої частини рівняння як функцію від β_j , щоб знайти точку перетину з віссю x . Альтернативно, можна використовувати функцію *uniroot* в R для розв'язання цього рівняння. Метод профілювання, де сума квадратів залишків виражається як функція єдиного параметра β_j , також використовується у наближенні на основі підхідності.

Довірчий інтервал для параметрів регресії може бути отриманий з вищезазначеного наближення на основі розподілу хі-квадрат ($k = m$),

$$\left\{ \beta \in R^m : \ln \|y - f(\beta)\|^2 - n \ln \hat{S} \leq \frac{1}{n} \chi^{-2}(1 - \alpha, m) \right\}. \quad (4.17)$$

Можна отримати альтернативний довірчий інтервал, який моделює лінійну модель,

$$\{\beta \in R^m : \|y - f(\beta)\|^2 - \hat{S} \leq \hat{\sigma}^2 \chi^{-2}(1 - \alpha, m)\}. \quad (4.18)$$

Отримаємо дещо більший довірчий регіон, замінивши χ^{-2} на mF^{-1} , як у підході Вальда вище, щоб врахувати скінчені ступені вільності.

Апроксимація оцінки

Основним результатом теорії МП вважається те, що за припущенням незалежності та однорідності розподілу, розподіл похідної функції log-подібності, обчисленої на справжньому значенні параметра, збігається до мультіваріативного нормального розподілу з нульовою середньою та коваріаційною матрицею, що дорівнює матриці інформації Фішера. Цей факт веде до використання тесту на основі оцінки. Оригінально цей тест був розроблений відомим індійським статистиком [7]. В економетриці його називають тестом множників Лагранжа. Оскільки з доведення теореми 1.2 випливає, що похідна функції log-подібності (2.4) дорівнює $\sigma^{-2}J'(y - f(\beta))$ а матриця інформації Фішера дорівнює $\sigma^{-2}(J'J)$, і оскільки $\hat{\beta} \rightarrow \beta$ з ймовірністю 1, ми можемо зробити висновок, що для великого значення n , в дусі тесту на основі оцінки,

$$(J'J)^{-\frac{1}{2}}\hat{J}'(y - f) \simeq N(0, \sigma^2 I),$$

де $\hat{f} = f(\hat{\beta})$ і $\hat{J} = J(\hat{\beta})$. Це буде називатися *апроксимацією оцінки*.

Крім того, оскільки розв'язок МП задовольняє $\hat{J}'(y - \hat{f}) = 0$, вищезазначена апроксимація може бути еквівалентно записана як

$$(\hat{J}'\hat{J})^{-\frac{1}{2}}\hat{J}'(\hat{f} - f) \simeq N(0, \sigma^2 I), \quad (4.19)$$

Ця апроксимація, в рамках нелінійної регресії, була названа Пазманом [8] наближено точною апроксимацією. Апроксимація оцінки (4.19) дає можливість побудувати довірчий регіон з наближеною довірчою ймовірністю $(1 - \alpha)100\%$.

$$\{\beta \in R^m: (\hat{f} - f(\beta))' \hat{J} (\hat{J}' \hat{J})^{-1} \hat{J}' (\hat{f} - f(\beta)) \leq \hat{\sigma}^2 \chi^{-2} (1 - \alpha, m)\}. \quad (4.20)$$

Апроксимація (4.19) може бути використана для перевірки нульової гіпотези про j -ий параметр $H_0: \beta_j = \beta_{j0}$. Позначимо через $\hat{\beta}_j$ оцінку ННК, яка мінімізує \hat{S}_{j0} у (4.13), де β_j тримається фіксованим (обмеженим). Обчисліть обмежене підігнане значення $\hat{f}_j = f(\hat{\beta}_j)$, оцінку дисперсії $\hat{\sigma}_j^{-2} = \|y - \hat{f}_j\|^2 / (n - m + 1)$, і обмежений Якобіан $\hat{J}_j = J(\hat{\beta}_j)$. Обчисліть статистику перевірки, яка визначається як

$$\hat{\sigma}_j^{-2} = (\hat{f} - \hat{f}_j)' \hat{J}_j (\hat{J}_j' \hat{J}_j)^{-1} \hat{J}_j' (\hat{f} - \hat{f}_j), \quad (4.21)$$

де $\hat{f}_j = f(\hat{\beta}_j)$ є незалежним підігнаним значенням, як вище. Тоді, якщо спостережуване значення (4.21) більше, ніж квантиль розподілу хі-квадрат з 1 ступенем вільності, ми відхиляємо нульову гіпотезу $H_0: \beta_j = \beta_{j0}$.

Щоб знайти довірчий інтервал за допомогою апроксимації за критерієм збігу для β_j , ми використовуємо ідею профілю, виражаючи інші параметри через β_j , як у випадку перевірки відношення правдоподібностей ($1 \leq j \leq m$ фіксовано). У певному сенсі, вихідна регресія з m параметрами зводиться до одного параметра β_j . Алгоритм для обчислення довірчого інтервалу за допомогою апроксимації за критерієм збігу для параметра β_j виглядає наступним чином.

1. Визначаємо статистику критерію як функцію від β_j ,

$$T_j(\beta_j) = \hat{\sigma}_j^{-2} (\hat{f} - f(\tilde{\beta}_j))' \hat{J}_j (\hat{J}_j' \hat{J}_j)^{-1} \hat{J}_j' (\hat{f} - f(\tilde{\beta}_j)). \quad (4.22)$$

2. $(1 - \alpha)$ довірчий інтервал за допомогою критерію збігу для параметра β_j є розв'язком рівняння

$$T_j(\beta_j) - \chi^{-2} (1 - \alpha, 1) = 0. \quad (4.23)$$

Як і в випадку LR довірчого інтервалу, для розв'язання рівняння (4.23) використовується функція `uniroot`.

Прогнозування за допомогою нелінійної регресії

Важливим застосуванням регресії є прогнозування залежної змінної $Y = f(x; \beta) + \varepsilon$ при заданому значенні передбачуваної змінної x . Прогноз очевидний: $\hat{Y} = f(x; \hat{\beta})$. Ми використовуємо метод дельта для наближення стандартної похибки прогнозу, як

$$SE(\hat{Y}) \simeq \hat{\sigma} \sqrt{1 + j'(\hat{J}'\hat{J})^{-1}j}, \quad (4.24)$$

де $\hat{J} = J(\hat{\beta})$ - це Якобіан розмірності $n \times m$, а $j = \partial f(x; \beta) / \partial \beta$ - це похідна розмірності $m \times 1$ при прогнозуванні x , обидва оцінені за допомогою ННК - оцінки $\hat{\beta}$. В результаті отримуємо наближений $(1 - \alpha)$ симетричний довірчий інтервал (ДІ) для Y

$$f(x; \hat{\beta}) \pm \Phi^{-1}(1 - \alpha/2)SE(\hat{Y}).$$

Зауважемо, що цей ДІ охоплює майбутні спостереження Y , тобто індивідуальні прогнози. Щоб отримати ДІ для регресійного/середнього прогнозу, потрібно видалити 1 з підкореневого виразу у формулі (4.24).

4.4. Три методи статистичного висновку в великій вибірці

1. *Еквівалентність між (4.12) та (4.18)*. Розкриємо вираз $\|y - f(\beta)\|^2$ за формулою Тейлора навколо оцінки ННК до другого члена. Оскільки похідна дорівнює нулю при значенні $\hat{\beta}$, маємо

$$\begin{aligned} \|y - f(\beta)\|^2 - \|y - f(\hat{\beta})\|^2 &\simeq \frac{1}{2}(\beta - \hat{\beta})' \left. \frac{\partial^2 \|y - f(\beta)\|^2}{\partial \beta} \right|_{\beta=\hat{\beta}} (\beta - \hat{\beta}) \\ &\simeq (\beta - \hat{\beta})'(\hat{J}'\hat{J})(\beta - \hat{\beta}). \end{aligned}$$

Друге наближення зумовлене тим, що очікуваний Гессіан дорівнює $\hat{J}'\hat{J}$, що підтверджує твердження про еквівалентність алгоритмів Гаусса-Ньютона та Фішера для нелінійної регресії.

2. *Еквівалентність між (4.12) та (4.24).* Використовуючи елементарну апроксимацію засобами математичного аналізу, $\ln(1 + x) \simeq x$ для малих значень x , ми представляємо

$$\begin{aligned} \ln\|y - f(\beta)\|^2 - \ln \hat{S} &= \ln\left(1 + \frac{\|y - f(\beta)\|^2 - \hat{S}}{\hat{S}}\right) \\ &\simeq \frac{\|y - f(\beta)\|^2 - \hat{S}}{\hat{S}} \simeq \frac{\|y - f(\beta)\|^2 - \hat{S}}{\hat{\sigma}^2}, \end{aligned}$$

це підтверджує заявлену еквівалентність. Зауважте, що ми припускаємо, що \hat{S} та $\|y - f(\beta)\|^2$ близькі, що є правдою, коли β і $\hat{\beta}$ близькі.

3. *Еквівалентність між (4.12) та (4.20).* Апроксимуємо нелінійну функцію регресії навколо $\hat{\beta}$ лінійною функцією, як у (3.1), $f(\beta) \simeq \hat{f} + \hat{J}(\beta - \hat{\beta})$, де $\hat{f} + f(\hat{\beta})$. Підставляючи це у ліву частину виразу (4.20), отримуємо

$$(\beta - \hat{\beta})'(\hat{J}'\hat{J})(\hat{J}'\hat{J})^{-1}(\hat{J}'\hat{J}) = (\beta - \hat{\beta})'(\hat{J}'\hat{J})(\beta - \hat{\beta}),$$

як у (4.12). Ясно, що різниця може бути значною для малих значень n або великих значень σ^2 .

Важливим спостереженням є те, що всі наближення ґрунтуються на близькості між β та $\hat{\beta}$. Це означає, що функції потужності відповідних тестів повинні мати схожий вигляд в околі нульового значення, але можуть розходитися, коли відстань між нульовим та альтернативними значеннями збільшується.

5. Модельні приклади

Розглянемо застосування функції *nls* в R. Для виконання цієї функції користувач повинен вказати формулу для нелінійної регресії та початкові значення.

5.1. Модельний приклад 1. Оцінка абсолютного світового рекорду у марафоні

Використовуємо дані про світові рекорди у марафоні з 1908 по 2013 рік з веб-сайту http://en.wikipedia.org/wiki/Marathon_world_record_progression (рис 5.1). Як можна бачити, найкращий час бігу не обов'язково зменшується з року в рік (довжина марафону - 26 миль 385 ярдів). Цікаво знати, чи існує ліміт - абсолютний світовий рекорд. Потрібно: (а) використати експоненціальну функцію з асимптотою для оцінки абсолютного світового рекорду у марафоні; (б) протестувати гіпотезу, що абсолютний світовий рекорд - дві години, за допомогою Вальда тесту та LR-тесту.

Дані знаходяться у файлі `marathonWR.txt` та код R міститься у функції `marathon` з нижче наведеним вихідним результатом.

Час світового рекорду в марафоні t_i у році Y_i моделюється як експоненціальна функція з асимптотою

$$t_i = \alpha_1 + \alpha_2 e^{-\alpha_3(Y_i - 1900)} + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

Згідно цієї моделі, у році $Y = 1900$ очікуваний час рекорду складає $\alpha_1 + \alpha_2$, а передбачений абсолютний світовий рекорд становить α_1 . Параметр α_3 визначає щорічну швидкість зменшення часу бігу. Цей параметр показує, наскільки швидко річний рекорд наближається до абсолютного світового рекорду. Виклик вбудованої функції `nls` має наступний вигляд:

```
nls(ti~a1+a2*exp(-a3*(year-1900)),start=list(a1=130,a2=10,a3=.01)).
```

Аналогічно як і в лінійній моделі (функція *lm*), змінна перед \sim вказує назву залежної змінної, а вираз після \sim вказує формулу нелінійної регресії. Коли регресійна модель складніша, ніж просто формула, слід використовувати функцію *function*. Наприклад, ми можемо визначити функцію перед викликом *nls* наступним чином:

```
mwr=function(a1,a2,a3,time_year) a1+a2*exp(-a3*(time_year-1900))
```

і далі використовувати

```
out=nls(ti~mwr(a1,a2,a3,time_year=year),
```

```
start=list(a1=a1.0,a2=a2.0,a3=a3.0))
```

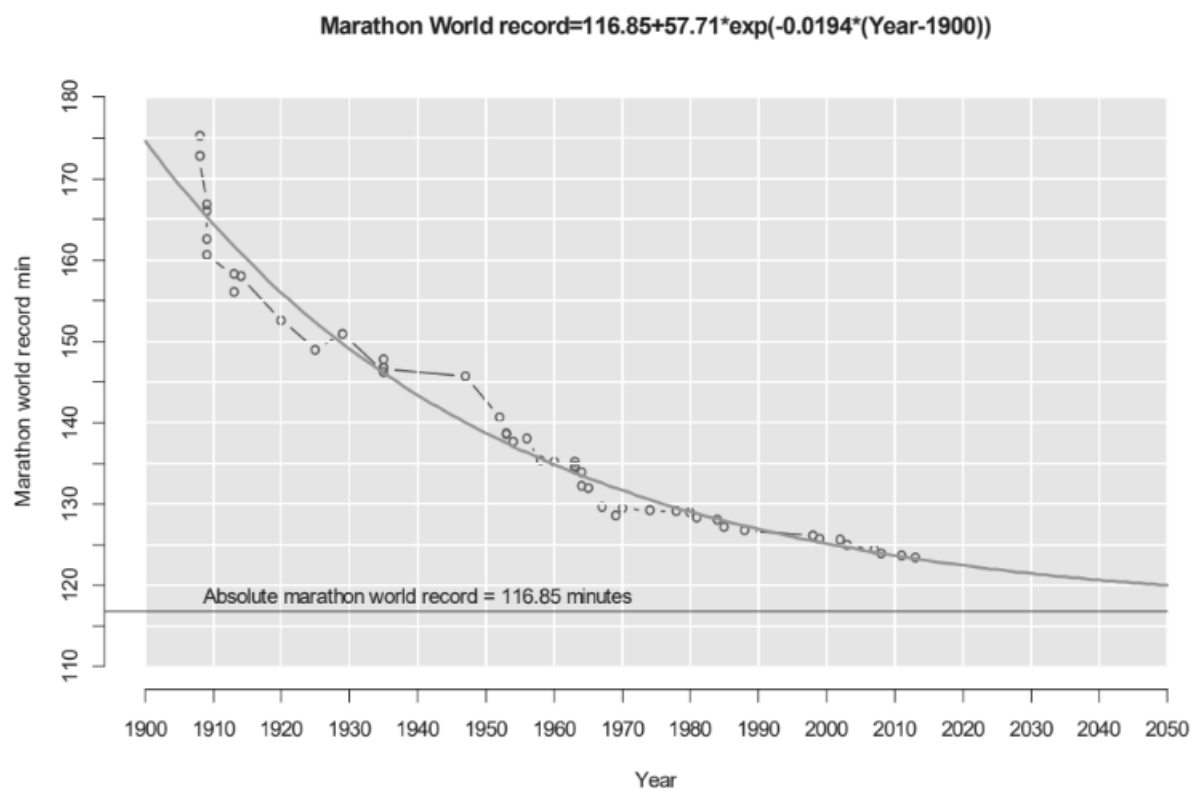


Рисунок 5.1: Підгонка світового рекорду в марафоні за допомогою функції **nls** з використанням експоненціальної функції з оціненою асимптотою як абсолютного світового рекорду

Назви параметрів можуть бути будь-якими, а їх початкові значення повинні бути вказані за допомогою параметру *start* (це відрізняється від *lm*). Коли

початкові значення дуже віддалені від оцінки ННК, алгоритм мінімізації може не збігтися. Немає загального методу для знаходження задовільних початкових значень. Однак, у випадку з одним внутрішньо нелінійним параметром можна використовувати цикл по діапазону значень та оцінювати інші параметри за допомогою функції *lm*. Фактично, у випадку з експоненціальною функцією лише параметр *a3* є внутрішньо нелінійним, і параметри *a1* та *a2* можна оцінити за допомогою функції *lm*. Наведений нижче результат генерується за допомогою команди *summary(out)*, де *out* - результат функції *nls*. *Std.Error j*-го параметра дорівнює *SE_j*, а *t-value* - це відношення *Estimate* до *Std.Error*. *Pr(> |t|)* - це *p-value*, обчислене на основі розподілу *t* з $48 - 3 = 45$ ступенями вільності за нульовою гіпотезою, що справжній параметр дорівнює нулю. Зауважемо, що, відмінно від лінійної регресії, ні стандартна помилка, ні *p-value* не є точними для нелінійної регресії. Залишкова стандартна похибка (*Residual standard error*) - це $\hat{\sigma}$, як квадратний корінь з мінімальної суми квадратів залишків, поділеної на ступені вільності. Якщо *out* - результат функції *nls*, то таблицю *Parameters* можна отримати за допомогою *summary(out)\$coefficients* або *coef(summary(out))*, а $\hat{\sigma}$ можна отримати за допомогою *summary(out)\$sigma*. Оцінки параметрів можна отримати за допомогою *coef(out)*.

Нижче наведено виведення функції


```

marathon()
      (Intercept)          x
2.63473315 117.71237352 57.26242173 0.02020202

Formula: ti ~ a1 + a2 * exp(-a3 * (year - 1900))

Parameters:
      Estimate Std. Error t value Pr(>|t|)
a1 1.168e+02  2.648e+00  44.130 < 2e-16 ***
a2 5.771e+01  2.092e+00  27.586 < 2e-16 ***
a3 1.940e-02  2.265e-03   8.566 5.26e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.66 on 45 degrees of freedom

Number of iterations to convergence: 3
Achieved convergence tolerance: 2.426e-07

```

Висновок. Згідно з цією моделлю, ми оцінюємо, що кожного року час рекорду зменшується на 2%, а абсолютний світовий рекорд складає 117 хвилин. *p*-value тестує, чи дорівнює відповідний параметр нулю. Наприклад, *p*-value для параметра *a1* не має багато сенсу, оскільки нульова гіпотеза полягає в тому, що абсолютний світовий рекорд дорівнює нулю, *p*-value для *a2* також не має багато сенсу, *p*-value для параметра *a3* тестує, чи зменшується час марафону в середньому з року в рік. Усі ці гіпотези мають незначний практичний інтерес, і не дивно, що *p* значення такі малі.

(b) Тепер ми проілюструємо, як перевіряти гіпотезу за допомогою нелінійної регресії. Перевіримо, чи точно дорівнює абсолютний світовий рекорд марафону двом годинам, $H_0: \alpha_1 = 120$. Значення *Z*-статистики для Вальда тесту обчислюється як

$$(\hat{\alpha}_1 - 120)/SE_1 = (116.8) - 120)/2.648 = -1.2085 > -1.96.$$

Тобто, нульову гіпотезу не можна відхилити з відповідним *p*-value, обчисленим як $2*(1-pnorm(1.2086)) = 0.226855$.

Для застосування тесту LR відношення правдоподібностей нам потрібно провести регресію $(t_i - 120) = a_2 e^{-a_3(Y_i - 1900)} + \varepsilon_i$. Наступні рядки коду були додані для обчислення p -value тесту LR:

```
#LRT H0:a1=120  
  
ti120=ti-120  
  
out120=nls(ti120~a2*exp(-a3*(year-1900)),  
start=list(a2=a2.0,a3=a3.0))  
  
n=length(ti)  
  
S.hat=summary(out)$sigma^2*(n-3)  
  
S.hat120=summary(out120)$sigma^2*(n-2)  
  
LHS=n*log(S.hat120/S.hat)  
  
pvLRT=1-pchisq(LHS,df=1)  
  
paste("LRT p-value =",pvLRT)
```

з вихідними даними

```
[1] "LRT p-value = 0.176937930208949"
```

Два метода дають однаковий результат: з рівнем значущості 0.05, нульова гіпотеза про те, що абсолютний світовий рекорд становить дві години, не може бути відхилена.

Зауваження щодо інтерпретації вихідних даних *nls*:

1. Причина, чому співвідношення оцінки параметра до його стандартної похибки називається t - статистикою в *nls*, а не Z -статистикою, полягає в тому, що для обчислення p -value використовується розподіл t з $n - m$ ступенями вільності, щоб імітувати лінійну модель.

2. p -value у nls , пов'язане з нульовою гіпотезою, що коефіцієнт (параметр) дорівнює нулю. У лінійній моделі нульова гіпотеза про нульовий коефіцієнт ($\beta_j = 0$) перевіряє наявність незалежної змінної (предиктора) у регресії, тоді як у нелінійній регресії параметр β_j може не мати пов'язаного з ним предиктора. Крім того, нульова гіпотеза $\beta_j = 0$ може не мати сенсу. Отже, невеликі p -value не обов'язково вказують на практично важливий результат. Наприклад, в прикладі з марафоном p -value $< 2 \times 10^{-16}$ для параметра $a1$ пов'язане з нульовою гіпотезою, що людина може подолати 26 миль миттєво.

Нелінійні моделі регресії часто використовуються в фармакокінетиці для моделювання концентрації лікарського засобу в організмі.

Декілька методів зображення, таких як магнітно-резонансна томографія або ПЕТ, вимагають введення контрастної речовини. Ставиться питання :коли слід починати зображення перед тим, як речовина починає розчинатися та виводитися? Комбінація двох процесів, дифузії речовини та її виведення, дає підстави для моделі фармакокінетики з двома параметрами.

5.2. Модельний приклад 2. Модель фармакокінетики з двома параметрами.

Введена контрастна речовина досягає життєво важливого органу відповідно до експоненційної моделі:

$$V_D(t) = V_0(1 - e^{-\alpha t}),$$

де $\alpha > 0$ - швидкість дифузії, а V_0 - максимальна концентрація лікарського засобу в органі, виміряна у мг/л. З іншого боку, лікарський засіб виводиться з організму за допомогою моделі зниження $V_D(t)e^{-\beta t}$ що призводить до трьохпараметричної моделі концентрації лікарського засобу.

$$V(t) = V_0(1 - e^{-\alpha t})e^{-\beta t}, \quad t \geq 0.$$

Потрібно знайти: (a) початкові значення для параметрів V_0 , α і β ;

(b) оцінити параметри V_0 , α і β ; (c) обчислити та відобразити 95% довірчий інтервал для встановлених значень; (d) оцінити оптимальний час зображення, його стандартну похибку та 95% довірчий інтервал.

На рисунку 5.2 показані дані та підганяюча модель. Код R можна знайти в функції *tworch*. Тут ми використовуємо локальну функцію *VtM* для обчислення двокомпаратментної моделі та її похідних/градієнту/Якобіану щодо трьох параметрів.

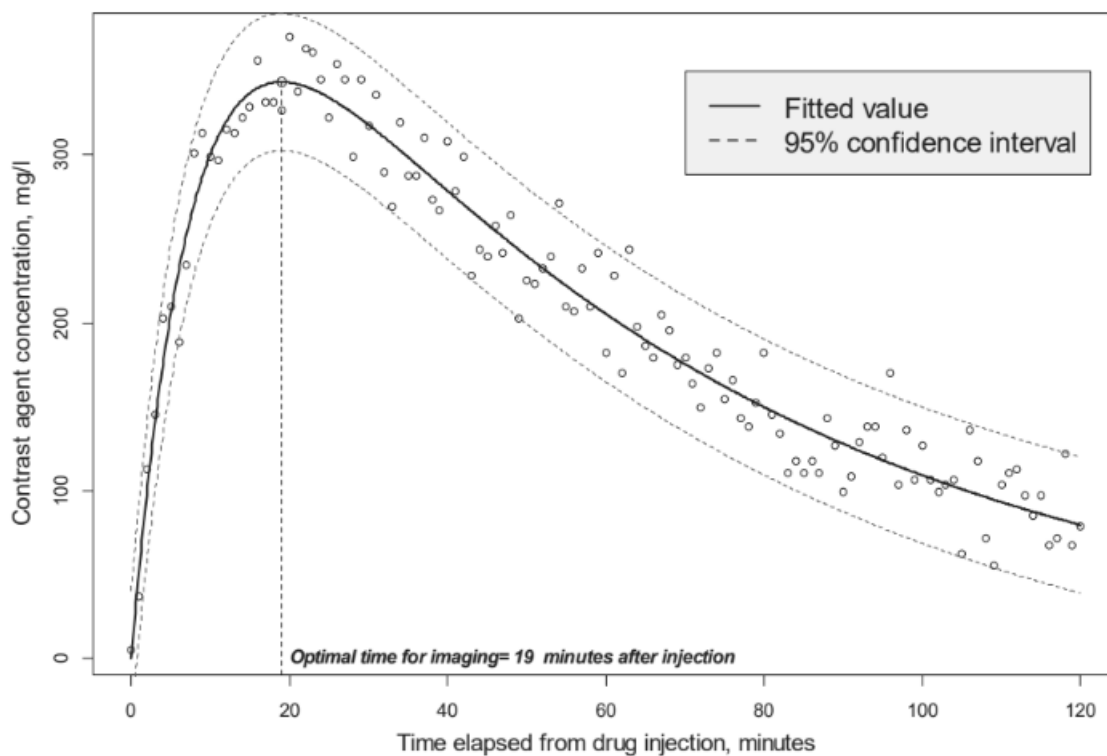


Рисунок 5.2. Модель фармакокінетики двокомпаратментної системи для концентрації препарату після введення

```
VtM=function(V0,alpha,beta,x) # Two-compartment function
```

```
{
```

```
e1=exp(-alpha*x);e2=exp(-beta*x)
```

```
Vt.ret=V0*(1-e1)*e2
```

```

dVdV0=(1-e1)*e2
dVda=V0*e1*e2*x
dVdb=-Vt.ret*x
attr(Vt.ret, &quot;gradient&quot;)=cbind(dVdV0,dVda,dVdb)
return(Vt.ret)
}

```

Зауважемо, що градієнт прикріплюється до значення, що повертається, як атрибут. Наприклад, якщо *Vall* - це значення, що повертається з *VtM*, ми можемо отримати Якобіан як *attr(Vall, "gradient")*. Нам потрібен Якобіан, щоб обчислити довірчий інтервал для припустимого значення. Виклик функції *nls()* виглядає наступним чином:

```

out = nls(Vt ~ VtM(V0, alpha, beta, x = min), start = list(V0 = V00, alpha = a0, beta = b0)).

```

- а) Прямий спосіб отримати початкові значення для трьох параметрів полягає у визначенні сітки значень для параметрів α та β за допомогою подвійного циклу, а потім знайти оптимальне значення V_0 , яке мінімізує RSS, $\sum_{t=1}^n (V_t - V_0 E_t)^2$, де $E_i = (1 - e^{-\alpha t})e^{-\beta t}$, як у моделі нахилу. Мінімум RSS еквівалентний максимуму $(\sum_{t=1}^n V_t E_i)^2 / \sum_{t=1}^n E_i^2$. За допомогою векторизованих обчислень ми створюємо подвійну сітку для двох параметрів, використовуючи функцію *expand.grid* або *rep* з параметрами *times* (за замовчуванням) та *each*, і використовуємо RSS з моделі нахилу. Після побудови подвійної (розширеної) сітки для параметрів α та β , ми обчислюємо матрицю значень $V(t)$ з кількістю рядків, рівним добутку довжини подвійної сітки та довжини *min*. Це можна зробити за допомогою матричного множення *alphs%*%(min)* та *betas%*%(min)*, де *alphs* та *betas* - це масиви подвійної сітки. Інший спосіб знаходження хороших початкових значень полягає у використанні однієї сітки для параметра α та оцінки V_0 та β за допомогою лінійної регресії $\ln(V_t/(1 - e^{-\alpha t})) = \ln V_0 - \beta t + \eta$

- b) Після визначення початкових значень використовується функція `nls` (див. функцію `twocph`).
- c) Стандартна похибка для прогнозованого значення $Y = f(x; \hat{\beta}) + \varepsilon$ обчислюється за формулою (4.24).
- d) Оптимальний час зображення - це час, коли об'єм контрастного агента досягає максимального значення. Візьмемо похідну $dV(t)/dt = V_0[\alpha e^{-\alpha t} - \beta(1 - e^{-\alpha t})]e^{-\beta t}$ та прирівняйте її до нуля, що дає $t_{opt} = \alpha^{-1} \ln(1 + \alpha/\beta)$. Матриця $(J')^{-1}$ обчислена на кожній ітерації, може бути отримана з виводу `nls` за допомогою команди `summary(out.nls)$cov.unscaled`, так що коваріаційна матриця 3×3 для NLS (C) оцінюється `summary(out.nls)$sigma^2*summary(out.nls)$cov.unscaled`. Як альтернатива, цю матрицю можна отримати за допомогою команди `vcov(out.nls)`. Оптимальний час оцінюється як $\hat{t}_{opt} = \hat{\alpha} \ln(1 + \hat{\alpha}/\hat{\beta})$ і, оскільки це нелінійна функція від $\hat{\alpha}$ та $\hat{\beta}$, можна використовувати метод дельта для оцінки дисперсії \hat{t}_{opt} .

$$\text{var}(\hat{t}_{opt}) \simeq \begin{bmatrix} \left. \frac{\partial t_{opt}}{\partial \alpha} \right|_{\alpha=\hat{\alpha}, \beta=\hat{\beta}} \\ \left. \frac{\partial t_{opt}}{\partial \beta} \right|_{\alpha=\hat{\alpha}, \beta=\hat{\beta}} \end{bmatrix}' \mathbf{C} \begin{bmatrix} \left. \frac{\partial t_{opt}}{\partial \alpha} \right|_{\alpha=\hat{\alpha}, \beta=\hat{\beta}} \\ \left. \frac{\partial t_{opt}}{\partial \beta} \right|_{\alpha=\hat{\alpha}, \beta=\hat{\beta}} \end{bmatrix}, \quad (5.1)$$

де

$$\frac{\partial t_{opt}}{\partial \alpha} = \frac{(\alpha + \beta)^{-1} \alpha - (\ln(\alpha + \beta) - \ln \beta)}{\alpha^2},$$

$$\frac{\partial t_{opt}}{\partial \beta} = \frac{(\alpha + \beta)^{-1} - \beta^{-1}}{\alpha}.$$

> `twocph()`

Starting values: 752.1721 0.06154545 0.02118182

Formula: $Vt \sim V0 * (1 - \exp(-\alpha * \text{min})) * \exp(-\beta * \text{min})$

Parameters:

Estimate Std. Error t value Pr(>|t|)
*V0 5.324e+02 1.715e+01 31.04 <2e-16 ****
*alpha 1.084e-01 7.220e-03 15.01 <2e-16 ****
*beta 1.587e-02 5.291e-04 29.99 <2e-16 ****
*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*
Residual standard error: 20.54 on 118 degrees of freedom
Number of iterations to convergence: 5
Achieved convergence tolerance: 7.238e-06
topt = 18.99124 , SE topt = 0.5476612

Висновок. Оптимальний час зображення - це 19 хвилина після введення з контрастною похибкою 0,55 і інтервалом довіри 95% $19 \pm 1,96 \times 0,55 = (18, 20)$. Орієнтовно, зображення слід виконувати протягом 18-20 хвилин.

Модельний приклад 5.3. Функція виробництва CES.

Файл містить 50 точок даних для K, L і Y. Потрібно (а) оцінити параметри функції виробництва CES, побудуйте графік поверхні разом із фактичними даними та значеннями; (б) перевірити гіпотезу про те, що функції CES та Кобба-Дугласа статистично нерозрізніми.

Розв'язання. R-код можна знайти в файлі ces.r. Нелінійна регресія для оцінки функції CES має наступний вигляд:

$$Y_i = A(\alpha K_i^\rho + (1 - \alpha)L_i^\rho)^{1/\rho} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

де $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$. (а) Підгонка за допомогою ННК для цієї функції вимагає досить точних початкових значень. Оскільки параметр A є лінійним, достатньо використовувати сітку зі значеннями α_l та ρ_j . Щоб знайти оптимальні значення, ми використовуємо подвійний цикл та обчислюємо вектор $(\alpha K_i^\rho + (1 - \alpha)L_i^\rho)^{1/\rho_j}$.

Потім ми знаходимо оптимальне значення A як коефіцієнт нахилу з лінійного методу найменших квадратів.

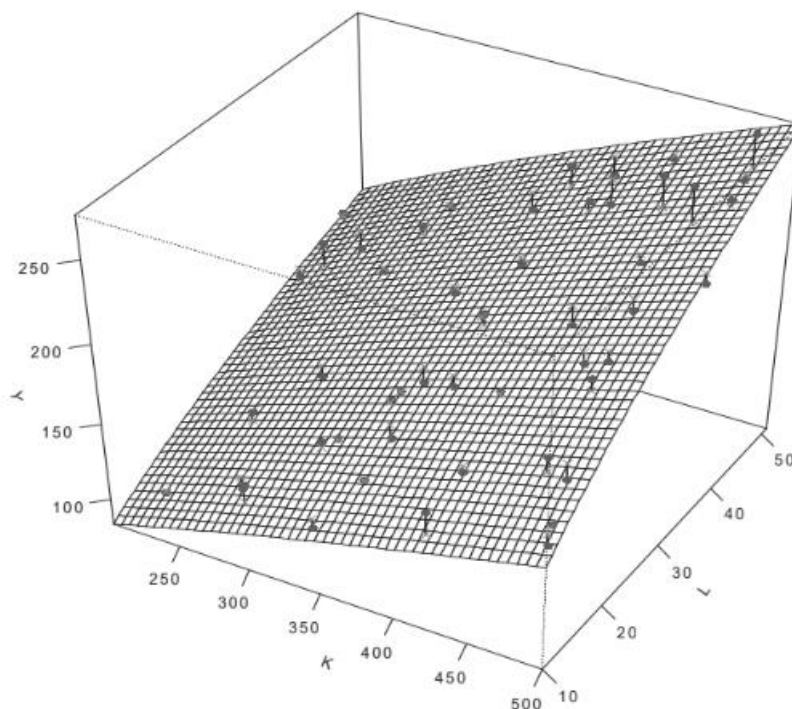


Рисунок 5.3: Поверхня для функції CES виробництва з фактичними та встановленими значеннями як проекціями

Оскільки p -value відповідає нульовій гіпотезі $H_0: \rho = 0$, p -value для тесту Вальда можна отримати безпосередньо з результатів, і воно становить 0.178355.

Для обчислення p -value для тесту LR нам потрібно оцінити функцію виробництва Кобба-Дугласа з нелінійної регресії $Y_i = AK_i^\rho + (1 - \alpha)L_i^{1-\alpha}$. Важливо зауважити, що використання лінійної регресії після логарифмічного перетворення не є тут відповідним, оскільки це мовчки передбачає множникову помилку; нам потрібно залишитися при адитивній помилці для порівняння RSS з двох моделей. Тест LR виконується по аналогії з тестуванням. На рисунку 5.3 зображено оцінену поверхню функції виробництва CES разом із фактичними трійними даними (K_i, L_i, Y_i) та оціненими значеннями як проекціями на поверхню вздовж вісі z . Як можна побачити з графіка, модель відносно добре підходить.

Модельний приклад 5.4.

У цьому прикладі продемонстровано, як обчислювати моделі нелінійної регресії та як порівнювати різні моделі, щоб вибрати ту, яка найкраще відповідає нашим даним. Метрики RMSE і R2 використовуватимуться для порівняння різних моделей. RMSE представляє похибку прогнозування моделі, тобто середню різницю між спостережуваними значеннями результатів і прогнозованими значеннями результатів. R2 являє собою квадрат кореляції між спостережуваними та прогнозованими значеннями результату. Найкращою моделлю є модель з найнижчим RMSE і найвищим R2.

Результати нижче наведеного дослідження наступні – кращими є поліноміальна та сплайн регресія для наших даних.

Модель	RMSE	R2
Лінійна	6,07	0,535
Поліноміальна	4,96	0,689
Логарифмічна	5,24	0,657
Сплайн	4,97	0,688
Узагальнена адитивна	5,02	0,684

Використовується набір даних Boston [з пакету MASS] для прогнозування середньої вартості будинку (mdev) у передмісті Бостона на основі змінної предиктора lstat (відсоток нижчого статусу населення). Код представлений в додатку.

ВИСНОВКИ

В даній кваліфікаційній роботі розглянуто основні концепції нелінійної регресії, ілюстровані різними прикладами. Реалізація модельних прикладів розглянуто в середовищі R.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Літнарівч Р.М. Побудова і дослідження математичної моделі за джерелами експериментальних даних методами регресійного аналізу. Навчальний посібник, МЕРУ, Рівне, 2011.-140 с.
2. Р.Є. Майборода, О.В. Сугакова АНАЛІЗ ДАНИХ ЗА ДОПОМОГОЮ ПАКЕТА R – 2013
3. Allmaraz, M., Bangerth, W., Linhart, J.M., Polanco, J., Wang, F., Wang, K., Webster, J. and Zedler, S. (2013). Estimating parameters in physical models through Bayesian inversion: A complete example. *SIAM Review* 55:149—167.
4. Bates, D.M., Watts, D.G. (1988). *Nonlinear Regression and Its Applications*. New York: Wiley
5. Demidenko E. *Advanced Statistics with Applications in R* (2020). John Wiley & Sons Inc – 856 p.
6. Demidenko, E. (2013). *Mixed Models. Theory and Applications with R*. Hoboken, NJ: Wiley.
7. R. Xu, D.I.I. Wunsch, “Survey of clustering algorithms,” *IEEE Trans. Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

Додатки

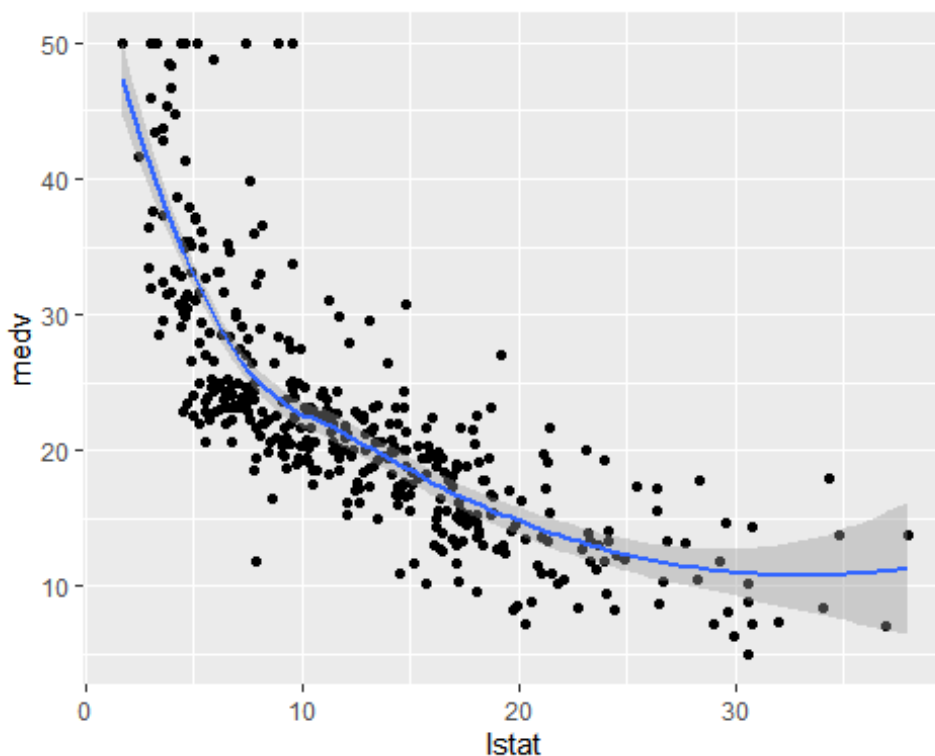
```
library(tidyverse)

library(caret)

# Load the data
data("Boston", package = "MASS")
# Split the data into training and test set
set.seed(123)
training.samples <- Boston$medv %>%
  createDataPartition(p = 0.8, list = FALSE)
train.data <- Boston[training.samples, ]
test.data <- Boston[-training.samples, ]

ggplot(train.data, aes(lstat, medv) ) +
  geom_point() +
  stat_smooth()

## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Лінійна регресія

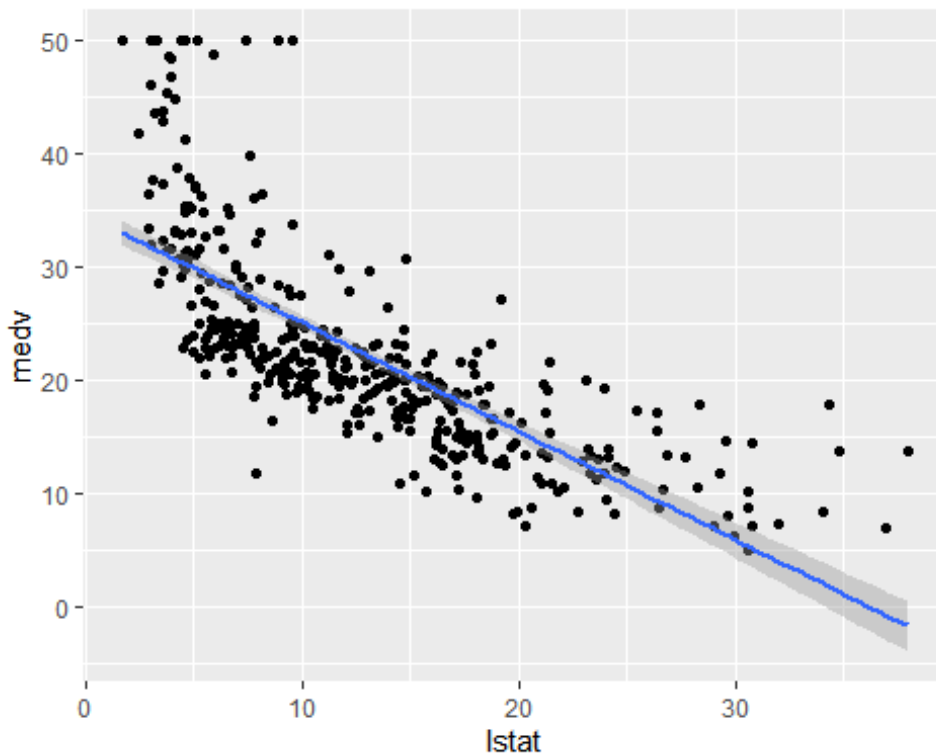
Формула лінійної регресії: $medv = b_0 + b_1 * lstat$.

```
# Build the model
model <- lm(medv ~ lstat, data = train.data)
# Make predictions
predictions <- model %>% predict(test.data)
# Model performance
```

```
data.frame(
  RMSE = RMSE(predictions, test.data$medv),
  R2 = R2(predictions, test.data$medv)
)
```

```
##          RMSE          R2
## 1 6.503817 0.513163
```

```
ggplot(train.data, aes(lstat, medv) ) +
  geom_point() +
  stat_smooth(method = lm, formula = y ~ x)
```



Поліноміальна регресія

$$medv = b_0 + b_1 * lstat + b_2 * lstat^2.$$

```
lm(medv ~ lstat + I(lstat^2), data = train.data)
```

```
##
## Call:
## lm(formula = medv ~ lstat + I(lstat^2), data = train.data)
##
## Coefficients:
## (Intercept)      lstat  I(lstat^2)
##    42.5736     -2.2673     0.0412
```

```
# Поліноміальна модель порядку 5
```

```
# Build the model
```

```
model <- lm(medv ~ poly(lstat, 5, raw = TRUE), data = train.data)
```

```
# Make predictions
```

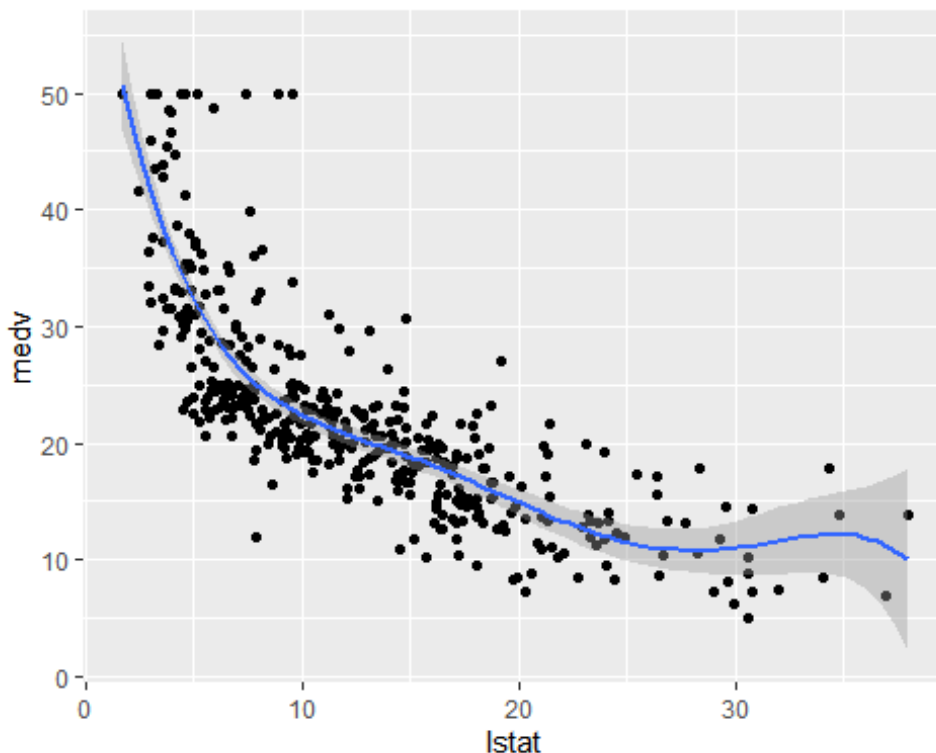
```

predictions <- model %>% predict(test.data)
# Model performance
data.frame(
  RMSE = RMSE(predictions, test.data$medv),
  R2 = R2(predictions, test.data$medv)
)

##          RMSE          R2
## 1 5.270374 0.6829474

ggplot(train.data, aes(lstat, medv) ) +
  geom_point() +
  stat_smooth(method = lm, formula = y ~ poly(x, 5, raw = TRUE))

```



Логарифмування даних. Якщо є нелінійний зв'язок, можна спробувати логарифм перетворення змінних предиктора.

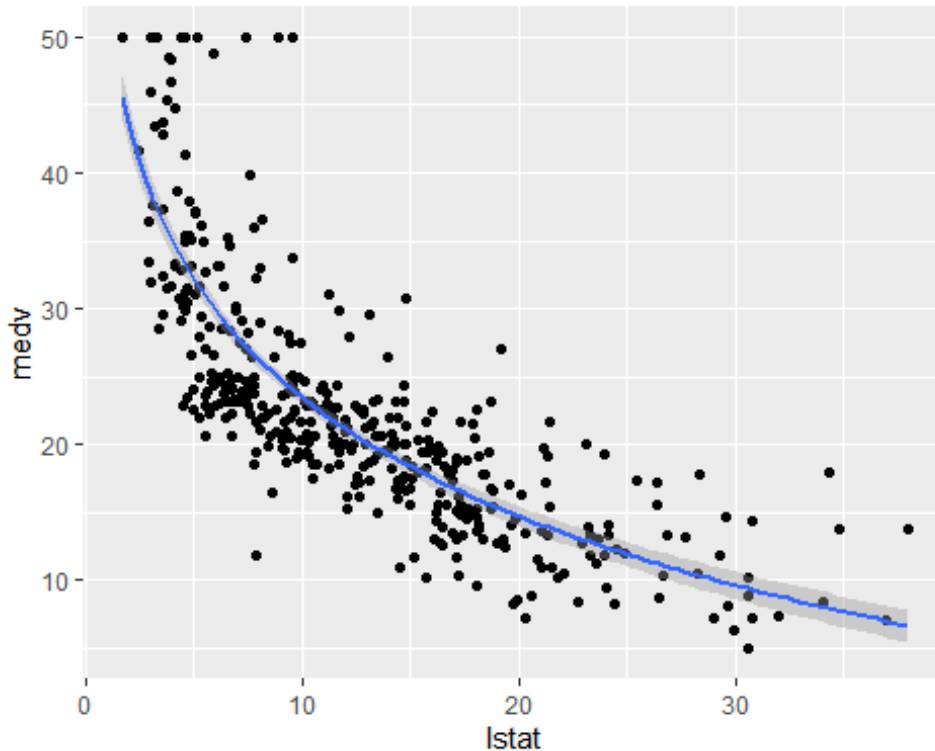
```

# Build the model
model <- lm(medv ~ log(lstat), data = train.data)
# Make predictions
predictions <- model %>% predict(test.data)
# Model performance
data.frame(
  RMSE = RMSE(predictions, test.data$medv),
  R2 = R2(predictions, test.data$medv)
)

##          RMSE          R2
## 1 5.467124 0.6570091

```

```
ggplot(train.data, aes(lstat, medv) ) +
  geom_point() +
  stat_smooth(method = lm, formula = y ~ log(x))
```



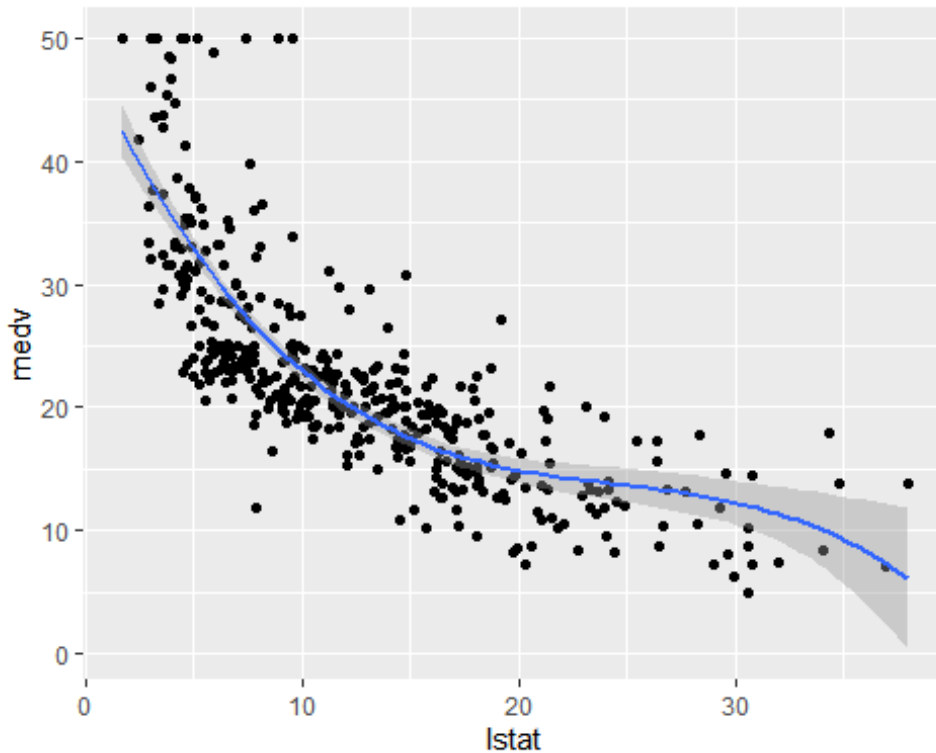
Сплайн регресія

Ми створимо модель за допомогою кубічного сплайна третього порядку.

```
knots <- quantile(train.data$lstat, p = c(0.25, 0.5, 0.75))
library(splines)
# Build the model
knots <- quantile(train.data$lstat, p = c(0.25, 0.5, 0.75))
model <- lm(medv ~ bs(lstat, knots = knots), data = train.data)
# Make predictions
predictions <- model %>% predict(test.data)
# Model performance
data.frame(
  RMSE = RMSE(predictions, test.data$medv),
  R2 = R2(predictions, test.data$medv)
)

##          RMSE          R2
## 1 5.317372 0.6786367

ggplot(train.data, aes(lstat, medv) ) +
  geom_point() +
  stat_smooth(method = lm, formula = y ~ splines::bs(x, df = 3))
```



Узагальнена адитивна модель

Узагальнені адитивні моделі, або GAM, є технікою автоматичної підгонки сплайнової регресії. Це можна зробити за допомогою пакета `mgcv` R.

```
library(mgcv)

## Загрузка требуемого пакета: nlme

##
## Присоединяю пакет: 'nlme'

## Следующий объект скрыт от 'package:dplyr':
##
## collapse

## This is mgcv 1.8-41. For overview type 'help("mgcv-package")'.

# Build the model
model <- gam(medv ~ s(lstat), data = train.data)
# Make predictions
predictions <- model %>% predict(test.data)
# Model performance
data.frame(
  RMSE = RMSE(predictions, test.data$medv),
  R2 = R2(predictions, test.data$medv)
)

##          RMSE          R2
## 1 5.318856 0.6760512
```



```
ggplot(train.data, aes(lstat, medv) ) +  
  geom_point() +  
  stat_smooth(method = gam, formula = y ~ s(x))
```

