Л6 Дисперсійний аналіз. ①
Однофакторна модель 6

Дисперсійний аналіз поєднує значну кількість задач математичної статистики, в яких аналізується вплив різних факторів на остаточний результат.
Зараз розглянемо найпростішу модель в якій перевіряється гіпотеза про вплив одного фактора.

Нехай маємо k незалежних вибірок

$$(\xi_{11}, \xi_{12}, \dots, \xi_{1n_1}) \in \Phi_{\alpha_1, \sigma^2}$$
$$(\xi_{21}, \xi_{22}, \dots, \xi_{2n_2}) \in \Phi_{\alpha_2, \sigma^2}$$
$$\dots\dots\dots$$
$$(\xi_{k1}, \xi_{k2}, \dots, \xi_{kn_k}) \in \Phi_{\alpha_k, \sigma^2}.$$

Всі параметри $\alpha_1, \dots, \alpha_k, \sigma^2$ невідомі.
Перевіряються гіпотези

$H_1: \alpha_1 = \alpha_2 = \dots = \alpha_k$,

H_2: існують індекси $i \neq j$ такі, що $\alpha_i \neq \alpha_j$.

Це задача може виникнути, наприклад, в такій ситуації. Нехай на k станках здійснюється виготовлення однакових деталей. Для кожної виготовленої деталі вимірюється деякий параметр, наприклад діаметр. Цей параметр є випадковою величиною через неминучі відхилення від стандарту.
Таким чином отримуємо k вибірок. Передбачається, що на i-му станку виготовляється n_i деталей.
Гіпотеза H_1 стверджує, що не має значення на якому станку виготовлена деталь, фактор станка не відіграє жодної ролі.

Це відповідає тому, що середнє значення усіх вибірок однакові. Гіпотеза, що конкурує з H_1 вказує на присутність систематичних відхилень для деяких станків.

Далі наші дії такі: будемо будувати із спостережень випадкову величину, яка за умови справедливості H_1 розподілена за законом Фішера з відповідним числом ступенів свободи. В результаті це визначає розв'язок.

Позначимо

$$N = \sum_{i=1}^{k} n_i, \quad \overline{\xi}_i = \frac{1}{n} \sum_{j=1}^{n_i} \xi_{ij}, \quad \overline{\xi} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \xi_{ij}.$$

<u>Теорема</u>. Якщо правильна гіпотеза H_1, то

$$\frac{(N-k) \sum_{i=1}^{k} n_i \left(\overline{\xi}_i - \overline{\xi} \right)^2}{(k-1) \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\xi_{ij} - \overline{\xi}_i \right)^2} \in F_{k-1, N-k}.$$

<u>Доведення</u>. Припустимо на деякий час, що нам відомі всі параметри $\alpha_1, \dots, \alpha_k, \sigma^2$, і застосуємо до кожного спостереження стандартизацію. Нехай

$$\eta_{ij} = \frac{\xi_{ij} - \alpha_i}{\sigma} \in \Phi_{0,1}, \; i = 1, \dots, k, \; j = 1, \dots, n_i,$$

$$\overline{\eta}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \eta_{ij}.$$

Запишемо вираз для вибіркової дисперсії побудованої за i-ю вибіркою із стандартизованих спостережень:

$$\frac{1}{n_i}\sum_{j=1}^{n_i}(\eta_{ij}-\overline{\eta}_i)^2=\frac{1}{n_i}\sum_{j=1}^{n_i}\eta_{ij}^2-(\overline{\eta}_i)^2.$$

Аналогічно доведенню теореми про властивості вибірок із нормального розподілу, за допомогою леми Фішера встановлюємо, що

$$\sum_{j=1}^{n_i}(\eta_{ij}-\overline{\eta}_i)^2=\sum_{j=1}^{n_i}\eta_{ij}^2-(\sqrt{n_i}\,\overline{\eta}_i)^2\in\chi^2_{n-1}$$

і ця величина не залежить від $\overline{\eta}_i$.

Визначаючи суму лівої частини за i отримаємо

$$Q_1=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\eta_{ij}-\overline{\eta}_i)^2\in\chi^2_{N-k}.$$

Зазначимо, що Q_1 не залежить від $\overline{\eta}_1,\dots,\overline{\eta}_k$.

Далі введемо

$$\overline{\eta}=\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\eta_{ij}=\frac{1}{N}\sum_{i=1}^{k}n_i\overline{\eta}_i.$$

Тоді маємо

$$Q_2=\sum_{i=1}^{k}n_i(\overline{\eta}_i-\overline{\eta})^2=\sum_{i=1}^{k}n_i\overline{\eta}_i^2-2\overline{\eta}\sum_{i=1}^{k}n_i\overline{\eta}_i+$$

$$+(\overline{\eta})^2\sum_{i=1}^{k}n_i=\sum_{i=1}^{k}(\sqrt{n_i}\,\overline{\eta}_i)^2-2\overline{\eta}N\overline{\eta}+N\overline{\eta}^2=$$

$$=\sum_{i=1}^{k}(\sqrt{n_i}\,\overline{\eta}_i)^2-(\sqrt{N}\,\overline{\eta})^2.$$

Нам відомо, що $\overline{\eta}_i\in\Phi_{0,\frac{1}{n_i}}$, тому $\sqrt{n_i}\,\overline{\eta}_i\in\Phi_{0,1}$.

Далі маємо

$$\sqrt{N}\,\overline{\eta} = \frac{\sqrt{N}}{N}\sum_{i=1}^{k} n_i \overline{\eta_i} = \sum_{i=1}^{k} \frac{\sqrt{n_i}}{\sqrt{N}}\sqrt{n_i}\,\overline{\eta_i} =$$

$$= \left(\frac{\sqrt{n_1}}{\sqrt{N}}, \dots, \frac{\sqrt{n_k}}{\sqrt{N}}\right)\left(\sqrt{n_1}\,\overline{\eta_1}, \dots, \sqrt{n_k}\,\overline{\eta_k}\right)^T.$$

Вектор $\left(\frac{\sqrt{n_1}}{\sqrt{N}}, \dots, \frac{\sqrt{n_k}}{\sqrt{N}}\right)$ має довжину рівну 1, тому його завжди можливо добудувати до ортогональної матриці, в якій він буде першим рядком. Застосуємо лему Фішера і отримаємо, що

$$Q_2 = \sum_{i=1}^{k}\left(\sqrt{n_i}\,\overline{\eta_i}\right)^2 - \left(\sqrt{N}\,\overline{\eta}\right)^2 \in \chi^2_{k-1}.$$

Через те, що Q_1 і Q_2 незалежні, випадкова величина

$$\frac{\frac{Q_2}{k-1}}{\frac{Q_1}{(N-k)}}$$

— розподілена за законом Фішера $F_{k-1,N-k}$.

Наразі повернемося до початкових спостережень.

$$\eta_{ij} - \overline{\eta_i} = \frac{\xi_{ij} - a_i}{\sigma} - \frac{\overline{\xi_i} - a_i}{\sigma} = \frac{\xi_{ij} - \overline{\xi_i}}{\sigma},$$

тому

$$Q_1 = \frac{1}{\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\xi_{ij} - \overline{\xi_i}\right)^2.$$

Далі припустимо, що H_1 правильна, тобто $\alpha_1=\alpha_2=\dots=\alpha_k=\alpha$. Тоді маємо

$$\bar{\eta}=\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{\xi_{ij}-\alpha_i}{\sigma}\right)=\frac{\bar{\xi}-\alpha}{\sigma}.$$

Тому

$$Q_2=\sum_{i=1}^{k}n_i\left(\frac{\bar{\xi}_i-\alpha}{\sigma}-\frac{\bar{\xi}-\alpha}{\sigma}\right)^2=$$

$$=\frac{1}{\sigma^2}\sum_{i=1}^{k}n_i\left(\bar{\xi}_i-\bar{\xi}\right)^2.$$

Отже, якщо справедлива гіпотеза H_1, то

$$(*)\quad \underset{\text{дзета}}{\zeta}=\frac{Q_2/(k-1)}{Q_1/(N-k)}=\frac{(N-k)\sum_{i=1}^{k}n_i\left(\bar{\xi}_i-\bar{\xi}\right)^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\xi_{ij}-\bar{\xi}_i\right)^2}\in$$

$$\in F_{k-1,N-k}.$$

Таким чином теорема доведена.

Далі перейдемо до побудови критерію. У виразі $(*)$ для ζ саме чисельник чутливий до систематичних відхилень між вибірками, тому ми реагуватимемо на великі значення ζ. За таблицями розподілу $F_{k-1,N-k}$ знаходимо число $q>0$ таке, що $F_{k-1,N-k}(q)=1-\varepsilon$. Інакше, якщо правильна гіпотеза H_1, то подія $\{\zeta\geq q\}$ малоймовірна. Тому відкидаємо H_1, якщо $\zeta\geq q$, і приймаємо цю гіпотезу в протилежному випадку.

При цьому $\beta_1=P_1(\zeta\geq q)=\varepsilon$.

Задачі лінійної регресії.

Постановка задачі.

Припустимо, що в результаті повторення n разів експерименту отримаємо вибірку $(\eta_1, \dots \eta_n)$. Також відомо, що значення величини η, що спостерігається, лінійно залежить від деяких відомих невипадкових числових факторів $x_1, \dots, x_k$, та ще від деякого випадкового фактора, присутність якого пояснюється випадковими похибками в роботі вимірювальних приладів, або його присутність закладена в основі експерименту. Іншими вираз
$$\eta = \theta_1 x_1 + \dots + \theta_k x_k + \varepsilon,$$
назвемо основним співвідношенням. Величини $x_1, \dots, x_k$ можуть набувати відомих нам значень у кожному експерименті. Невідомі лише коефіцієнти залежності $\theta_1, \dots, \theta_k$. Визначення цих коефіцієнтів становить основну задачу. Ця задача була б проста, аби не заважали випадкові відхилення. Здійснюючи експерименти за умови різних значень $x_1, \dots, x_k$ отримуємо спостереження
$$\eta_1 = \theta_1 x_{11} + \dots + \theta_k x_{1k} + \varepsilon_1$$
$$\eta_2 = \theta_1 x_{21} + \dots + \theta_k x_{2k} + \varepsilon_2$$
$$\dots \qquad \dots$$
$$\eta_n = \theta_1 x_{n1} + \dots + \theta_k x_{nk} + \varepsilon_n.$$
Необхідно здійснювати $n > k$ спостережень, інакше не можливо добре оцінити всі коефіцієнти.

Припускається, що
Випадкові величини $\varepsilon_1, \ldots, \varepsilon_n$ незалежні і однаково розподілені. Також $M\varepsilon_i = 0$, дисперсія $D\varepsilon_i = \sigma^2$ часто припускається невідомою.
Запишемо отримане вище співвідношення у векторному вигляді. Нехай

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{pmatrix},$$

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_k \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

тоді маємо співвідношення

$$\eta = X\theta + \varepsilon. \quad (*)$$

Матриця X називається регресором, вона складається із відомих нам чисел, які ми задаємо в процесі здійснення експерименту. Регресор має n рядків і k стовпців; його елементи вибираються так, щоб стовпці були лінійно незалежними.
Випадковий вектор ε неминуче присутній в цих співвідношеннях, але його значення не відомі.
Вектор η називається відгуком, він складається із випадкових величин, що спостерігаються нами.
Нарешті θ — вектор невідомих параметрів, які мають бути оцінені.
Зазначимо, що на відміну від попередньо викладеного, тут ми маємо справу з вибіркою, яка складається із різно розподілених спостережень,

через те, що

$$M\eta_i = \theta_1 x_{i1} + \dots + \theta_k x_{ik}$$

залежить від i.

Права частина основного співвідношення (*) лінійно залежить від невідомих параметрів $\theta_1, \dots, \theta_k$, тому говоримо про задачу лінійної регресії.

Історично склалося так, що термін "регресія" не відбиває суті проблеми. Тут підійшла б назва "статистичне дослідження залежностей".

Окремим випадком є така постановка задачі. Нехай маємо набір функцій $\psi_1(t), \dots, \psi_k(t)$, а основне співвідношення має вигляд:

$$\eta = \theta_1 \psi_1(t) + \dots + \theta_k \psi_k(t) + \varepsilon.$$

Змінна t може інтерпретуватися як час або температура. Здійснюючи експерименти за умови $t = t_1, \dots, t_n$, отримуємо спостереження

$$\eta_i = \theta_1 \psi_1(t_i) + \dots + \theta_k \psi_k(t_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

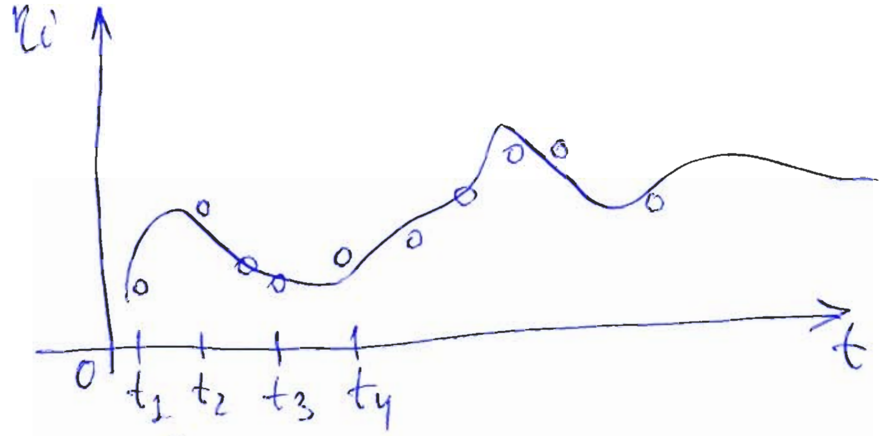
тобто $x_{ij} = \psi_j(t_i)$. Наприклад, можна взяти

$$\psi_1(t) = 1, \ \psi_2(t) = t, \dots, \psi_k(t) = t^{k-1},$$

і тоді основне співвідношення набуде вигляду:

$$\eta = \theta_1 + \theta_2 t + \dots + \theta_k t^{k-1} + \varepsilon.$$

У цьому випадку задача має просту геометричну інтерпретацію.

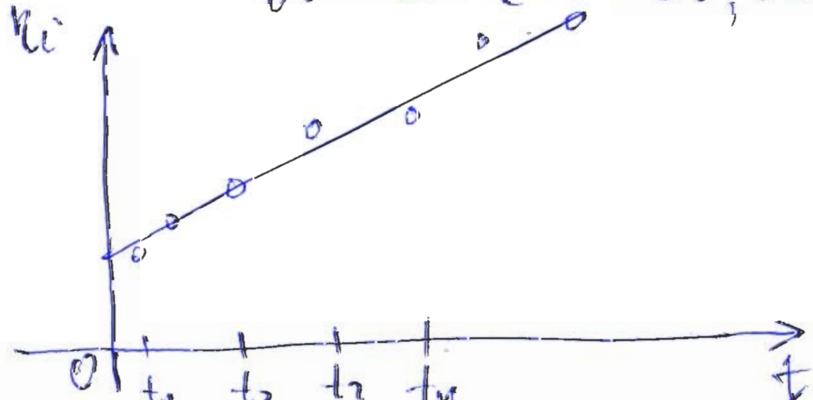


Числа $\theta_1, \ldots, \theta_k$ є коефіцієнтами полінома. Задача полягає в тому, щоб підібрати їх так, щоб графік полінома найкращим чином наближав отриману сукупність точок $(t_1, \eta_1), (t_2, \eta_2), \ldots, (t_n, \eta_n)$.

В окремому випадку, коли $k=2$, маємо справу з простою регресією, в інших випадках регресія називається ~~мн~~ множинною.

Приклад. Припустимо, що ми вивчаємо залежність розчинності речовини в якійсь рідині від температури цієї рідини. Позначимо температуру літерою t і здійснюємо вимірювання розчинності за різних температур. Отримані дані (див. графік) вказують на лінійну залежність

$$\eta_i = \theta_1 + \theta_2 t_i + \varepsilon_i, \quad i = 1, \ldots, n.$$



Задача полягає в оцінюванні невідомих параметрів θ_1 і θ_2, що визначають цю залежність.

17. Метод найменших квадратів.

Оцінки невідомих параметрів $\theta_1, \dots, \theta_k$ будемо знаходити **методом найменших квадратів.**

Введемо відповідні позначення

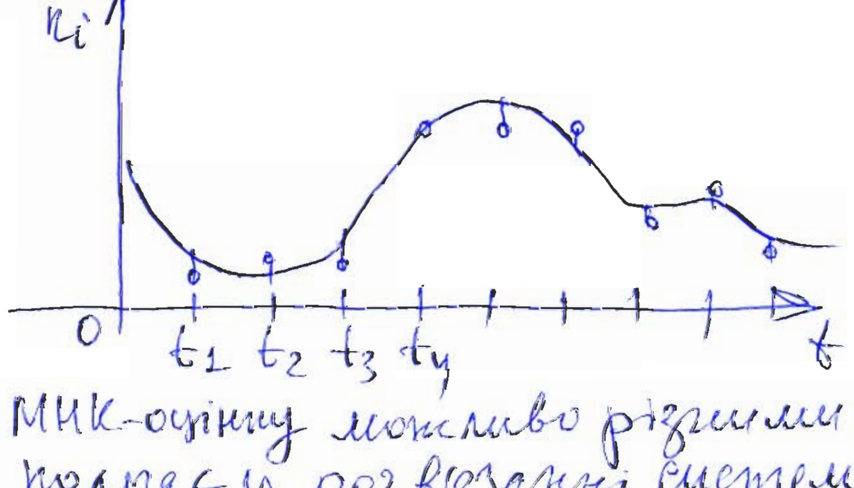
$$S(\theta) = \sum_{i=1}^{n} (\eta_i - \theta_1 x_{i1} - \dots - \theta_k x_{ik})^2 =$$

$$= |\eta - X\theta|^2,$$

Оцінкою методу найменших квадратів (МНК-оцінкою) називається те значення $\theta = \theta^*$, за якого $S(\theta)$ досягає мінімального значення:

$$S(\theta^*) = \min_{\theta} S(\theta).$$

Якщо звернутися до розглянутої вище графічної ілюстрації, ми бачимо, що числа $\theta_1, \dots, \theta_k$ підбираються так, щоб мінімальною була сума квадратів довжин вертикальних відрізків, які сполучають точки (t_i, η_i) з відповідними точками на кривій.



Знаходити МНК-оцінку можливо різними способами. Один із них полягає у розв'язанні системи так званих **нормальних** рівнянь:

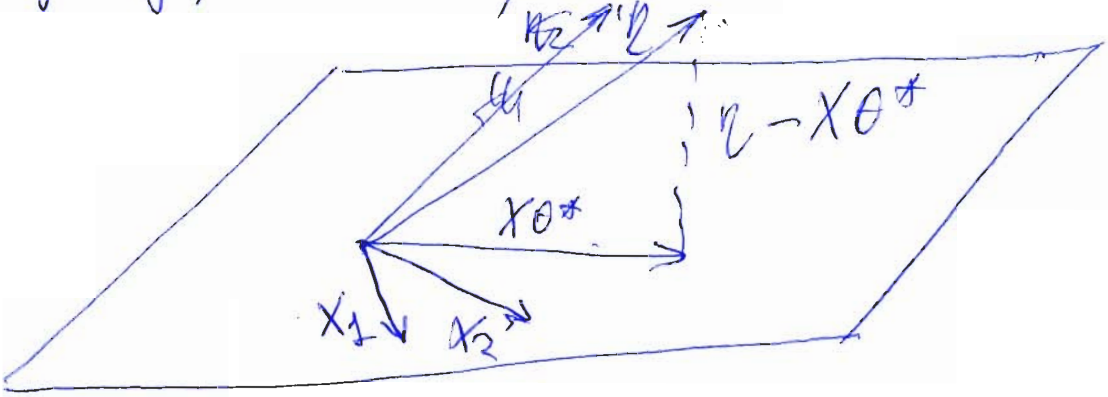
$$\frac{\partial S(\theta)}{\partial \theta_j} = 0, \quad j = 1, \dots, k,$$

що досить не просто. Спробуємо знайти оцінку із геометричних міркувань.

Позначимо через $X_1, \dots, X_k$ стовпці матриці X. Це лінійно незалежні вектори в $\mathbb{R}^n$. Через те, що $k < n$, ці вектори породжують в $\mathbb{R}^n$ підпростір $\mathbb{R}^k$. Будь-яка лінійна комбінація цих векторів знову належить тому ж $\mathbb{R}^k$, тобто для $\forall \theta$

$$X\theta = X_1\theta_1 + \dots + X_k\theta_k \in \mathbb{R}^k,$$

В тому числі $X\theta^* \in \mathbb{R}^k$. Проілюструємо все це на рисунку, коли $n=3$, а $k=2$.



У відповідності з методом найменших квадратів потрібно знайти таке значення $\theta = \theta^*$, за якого довжина вектора $\eta = X\theta^*$ буде мінімальною. Цей вектор на рисунку зображений пунктиром. Зрозуміло, що його довжина мінімальна, якщо він ортогональний площині, а таким чином і векторам, що породжують цю площину. Запишемо цей висновок для загального випадку:

$$\eta - X\theta^* \perp X_j, \quad j = 1, \dots, k.$$

Інакше це можна записати так:

$$X^T(\eta - X\theta^*) = 0,$$

де O – це нульовий вектор розмірності k. Звідси знаходимо, що

$$X^T\eta = X^T X\theta^*.$$

Квадратна матриця $X^T X$ є невиродженою. Домножимо отриману рівність зліва на $(X^T X)^{-1}$, і отримаємо МНК-оцінку

$$\theta^* = (X^T X)^{-1} X^T \cancel{X} \eta.$$

Припустимо додатково, що вектори $X_1, \ldots X_k$ ортогональні. У цьому випадку задача лінійної регресії значно спрощується. Зокрема матриця $X^T X$ стає діагональною:

$$X^T X = \begin{pmatrix} (X_1, X_1) & 0 & \ldots & 0 \\ 0 & (X_2, X_2) & \ldots & 0 \\ \ldots & \ldots & & \ldots \\ 0 & 0 & \ldots & (X_k, X_k) \end{pmatrix};$$

$(X^T X)^{-1}$ також буде діагональною;

$$(X^T X)^{-1} = \begin{pmatrix} (X_1, X_1)^{-1} & 0 & \ldots & 0 \\ 0 & (X_2, X_2)^{-1} & \ldots & 0 \\ 0 & 0 & \ldots & (X_k, X_k)^{-1} \end{pmatrix}.$$

Якщо домножимо останню матрицю на вектор

$$X^T\eta = \begin{pmatrix} (X_1, \eta) \\ (X_2, \eta) \\ \ldots \\ (X_k, \eta) \end{pmatrix}$$

то отримаємо просте вираження для компоненти МНК-оцінки θ^*:

$$\theta_i^* = \frac{(X_i, \eta)}{(X_i, X_i)}, \quad i = 1, \ldots, k.$$

Довірчі інтервали та перевірка гіпотез

Після визначення оцінок невідомих параметрів природно задатися питанням можливості побудови довірчих інтервалів та перевірки гіпотез. Для цього необхідна інформація про розподіл випадкових величин ε_i.

Припустимо, що випадкові величини $\varepsilon_1, \ldots, \varepsilon_n$ незалежні та розподілені за законом Φ_{0,σ^2}. В таких випадках маємо задачу нормальної регресії:

Теорема Нехай всі ε_i незалежні і розподілені за законом Φ_{0,σ^2}. Тоді оцінка методом максимальної вірогідності (ММВ-оцінка) і оцінка методом найменших квадратів (МНК-оцінка) для параметра θ співпадають.

Доведення. Запишемо функцію вірогідності вибірки. Не зважаючи на те, що спостереження тепер розподілені не однаково, принцип залишається таким же.

$$f(\theta, \sigma^2, \eta) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\eta_i - \theta_1 x_{i1} - \ldots - \theta_k x_{ik})^2\right\} =$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} S(\theta)\right\}.$$

Із даного запису випливає, що досліджувати функцію вірогідності на максимум — це те ж саме, що досліджувати $S(\theta)$ на мінімум. Теорема доведена.

Якщо маємо функцію вірогідності, знайдемо ММВ-оцінку для σ^2. Маємо

$$l(\theta,\sigma^2,\eta) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}|\eta - X\theta|^2,$$

$$\frac{\partial l(\theta,\sigma^2,\eta)}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}|\eta - X\theta|^2 = 0,$$

звідки отримуємо ММВ-оцінку для σ^2

$$(\sigma^2)^* = \frac{|\eta - X\theta^*|^2}{n}.$$

Нижче наведемо без доведення теорему, яка встановлює імовірносні властивості оцінок невідомих параметрів в задачах нормальної регресії.

<u>Теорема</u> Нехай всі ε_i незалежні і розподілені за законом Φ_{0,σ^2}. Тоді:

1) МНК-оцінка θ^* має багатовимірний нормальний розподіл, при цьому

$$M\theta^* = \theta, \quad C(\theta^*) = \sigma^2 (X^T X)^{-1};$$

2) $\dfrac{|\eta - X\theta^*|^2}{\sigma^2} \in \chi^2_{n-1};$

3) θ^* та $|\eta - X\theta^*|^2$ незалежні.

Наслідок 1. Оцінка

$$\hat{\sigma}^2 = \frac{1}{n-k}|\eta - X\theta^*|^2$$

є незміщеною для σ^2.

Доведення. У відповідності до 2) твердження сформульованої теореми, випадкова величина $\frac{|\eta - X\theta^*|^2}{\sigma^2}$ розподілена як $Z_1^2 + \dots + Z_{n-k}^2$, де всі Z_i незалежні і розподілені за законом $\Phi_{0,1}$, тому маємо

$$M\left(\frac{|\eta - X\theta^*|^2}{\sigma^2}\right) = M(Z_1^2 + \dots + Z_{n-k}^2) = n-k,$$

тобто

$$M\left(\frac{|\eta - X\theta^*|^2}{n-k}\right) = \sigma^2$$

Звідси випливає, між іншим, що знайдена вище ММВ-оцінка $(\sigma^2)^*$ є зміщеною.

Наслідок 2. Якщо стовпці регресора $X_1, \dots, X_k$ ортогональні, то матриця $C(\theta^*) = \sigma^2 (X^T X)^{-1}$ діагональна і компоненти оцінки $\theta_1^*, \dots, \theta_k^*$ незалежні. При цьому

$$\theta_i^* \in \Phi_{\theta_i, \frac{\sigma^2}{|X_i|^2}}.$$

Це твердження відразу випливає з властивостей багатовимірного нормального розподілу.

7

<u>Наслідок 3.</u> Якщо стовпці регресора $X_1, \dots, X_k$ ортогональні, то

$$\frac{(\theta_i^* - \theta_i)|X_i|}{\sigma} : \sqrt{\frac{1}{n-k}\frac{|\eta - X\theta^*|^2}{\sigma^2}} =$$

$$= \frac{(\theta_i^* - \theta_i)|X_i|\sqrt{n-k}}{|\eta - X\theta^*|} \in T_{n-k}.$$

Далі перейдемо до побудови довірчих інтервалів.

1. Довірчий інтервал для σ^2. Із таблиці розподілу χ^2_{n-k} знаходимо числа q_1 і q_2 такі, що

$$\chi^2_{n-k}(q_1) = \frac{\varepsilon}{2}, \quad \chi^2_{n-k}(q_2) = 1 - \frac{\varepsilon}{2}.$$

Тоді, у відповідності з наслідком 1,

$$P\left(q_1 < \frac{|\eta - X\theta^*|^2}{\sigma^2} < q_2\right) = \chi^2_{n-k}(q_2) - \chi^2_{n-k}(q_1) = 1 - \varepsilon,$$

звідки випливає, що

$$P\left(\frac{|\eta - X\theta^*|^2}{q_2} < \sigma^2 < \frac{|\eta - X\theta^*|^2}{q_1}\right) = 1 - \varepsilon.$$

Далі припускаємо, що стовпці регресора $X_1, \dots, X_k$ ортогональні.

2. Довірчий інтервал для θ_i за умови, що σ^2 відома.

Застосуємо факт, що $\frac{(\theta_i^* - \theta_i)|X_i|}{\sigma} \in \Phi_{0,1}$.

Нехай число q таке, що $\Phi_{0,1}(-q) = \frac{\varepsilon}{2}$, тоді маємо

$$P\left(-q < \frac{(\theta_i^* - \theta_i)|X_i|}{\sigma} < q\right) = \Phi_{0,1}(q) - \Phi_{0,1}(-q) = 1-\varepsilon,$$

тому

$$P\left(\theta_i^* - \frac{q\sigma}{|X_i|} < \theta_i < \theta_i^* + \frac{q\sigma}{|X_i|}\right) = 1-\varepsilon.$$

3. Довірчий інтервал для θ_i за умови, що σ^2 невідоме.

Застосуємо наслідок 3 і знайдемо з таблиць число q таке, що $T_{n-k}(-q) = \frac{\varepsilon}{2}$.

Тоді

$$P\left(-q < \frac{(\theta_i^* - \theta_i)|X_i|\sqrt{n-k}}{|\eta - X\theta^*|} < q\right) = T_{n-k}(q) -$$

$$- T_{n-k}(-q) = 1-\varepsilon$$

після відповідних перетворень отримаємо довірчий інтервал

$$P\left(\theta_i^* - \frac{q|\eta - X\theta^*|}{|X_i|\sqrt{n-k}} < \theta_i < \theta_i^* + \frac{q|\eta - X\theta^*|}{|X_i|\sqrt{n-k}}\right) = 1-\varepsilon,$$

Побудова довірчих інтервалів надає можливість перевірити гіпотези у відповідності з викладеною раніше концепцією. Наприклад, для перевірки гіпотези $H_1: \theta_i = C$ проти гіпотези $H_2: \theta_i \neq C$ за умови невідомої дисперсії σ^2

Беремо критичну множину

$$K = \left\{ (\eta_1, \ldots, \eta_n) : C \notin \left(\theta_i^* - \frac{q|\eta - X\theta^*|}{|X_i|\sqrt{n-k}}, \right. \right.$$

$$\left. \left. \theta_i^* + \frac{q|\eta - X\theta^*|}{|X_i|\sqrt{n-k}} \right) \right\}.$$

Тоді

$$\beta_1 = P_1((\eta_1, \ldots, \eta_n) \in K) = \varepsilon.$$