

Тема 6. Регресійний аналіз і планування експериментів

Основи регресійного аналізу

Регресійний аналіз є найбільш поширеним статистичним методом обробки експериментальних даних з метою отримання емпіричних залежностей.

Метою регресійного аналізу є отримання за експериментальними даними математичної моделі, що описують поведінку деяких характеристик об'єкта в залежності від множини факторів:

$$Y = F(X_1, X_2, \dots, X_n) + E \quad (6.1)$$

Y – відгук об'єкта дослідження;

X_i – фактори, незалежні змінні, що характеризують об'єкт;

E – випадкова похибка.

В основі використання регресійного аналізу покладено дослідження об'єкта за допомогою методу чорної скриньки, на вхід якої діють вектор незалежних змінних X (рис. 6.1).

На виході є вектор параметрів Y , що характеризують об'єкти дослідження.

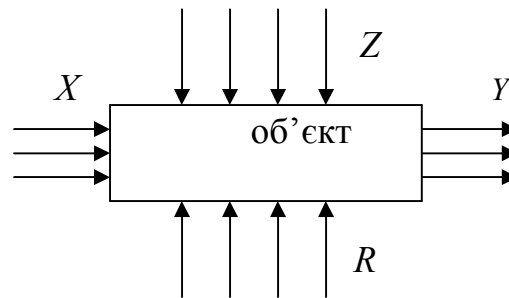


Рис. 6.1. Схема «чорної скриньки».

Крім того, на об'єкт діють некеровані фактори Z , що є незалежними змінними, які можна виміряти, але керувати не можна, а також вектор неконтрольованих факторів R , який впливає на об'єкт дослідження, але ми його не можемо виміряти.

Таким чином, в регресійному аналізі використовують вектор X та вектор Y .

Вхідні параметри (фактори), що використовуються в регресійному аналізі повинні задовольняти таким вимогам:

- 1) вхідні фактори X повинні впливати на вихідну величину Y ;
- 2) вхідні фактори X повинні бути керованими, тобто їх можна задавати по бажанню експериментатора;
- 3) фактори X повинні бути незалежними між собою;
- 4) точність вимірювання факторів X повинна бути на порядок більша точності вимірювання вихідної величини Y .

В основі обробки результатів дослідження за методом чорної скриньки покладено метод найменших квадратів.

В регресійному аналізі зазвичай використовують лінійну багатомірну залежність вихідної величини від вхідних факторів.

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_k. \quad (6.2)$$

Тут $a_0, a_1, a_2, \dots, a_k$ коефіцієнти, що визначають вплив кожного із факторів на вихідну величину.

При проведенні досліджень розрізняють *активний* і *пасивний* експеримент.

Якщо експеримент зводиться до отримання результатів поведінки об'єкта при випадкових значеннях вхідних параметрів X , експеримент є пасивним.

Якщо при проведенні експерименту всі параметри вимірюються по раніш складеному плану, то експеримент є активним.

При плановому експерименті об'єкт дослідження повинен мати такі властивості:

- 1) результати експерименту повинні бути відновлюваними;
- 2) об'єкт повинен бути керованим.

Експеримент називається *відновлюваним*, якщо при фіксованих умовах дослідження, якщо в різний час ми отримуємо одне і теж значення вихідної величини в межах відносно не великої похибки 2 – 5%.

Для виявлення відповідного експерименту цей експеримент повторюється кілька раз через нерівні проміжки часу.

Розкид значень вихідної величини характеризують відновлюваність результатів дослідження.

Якщо це значення не перевищує заданої величини, то об'єкт – задовольняє властивості відновлюваності.

Для двох факторів $k=2$ область досліджень ω обмежена значеннями X_{1min}, X_{1max} і X_{2min}, X_{2max} (рис. 6.1).

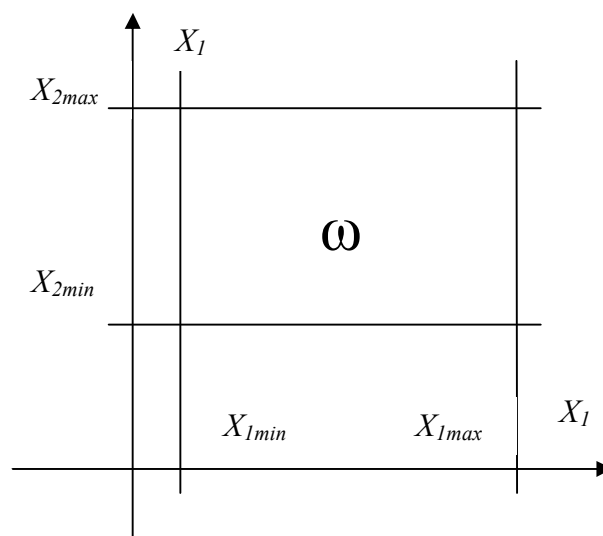


Рис. 6.1. Область досліджень для двох факторів в натуральній шкалі виміру.

Фактори можуть мати різні за величиною значення. Тому доцільно перейти від натуральних значень вхідних факторів до кодованих, що рівнозначно заміні системі координат (рис. 6.2).

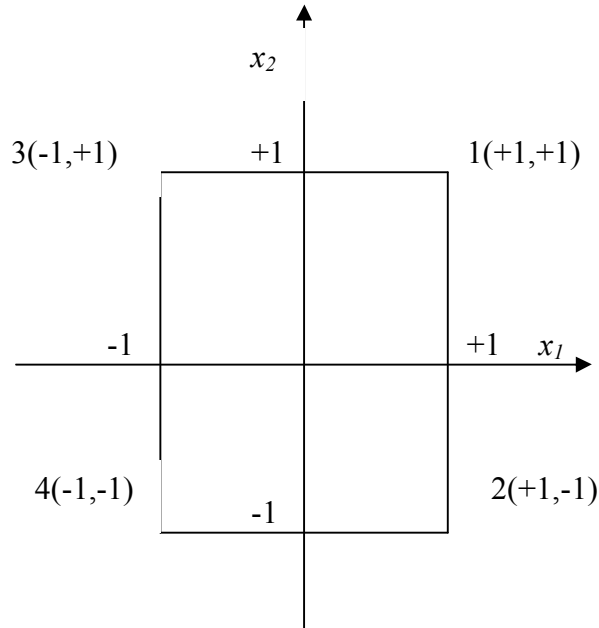


Рис. 6.2. Область досліджень функції двох змінних в кодованій шкалі виміру.

Для переходу від натуральних значень до кодованих використовують такі формули:

$$x_{i0} = \frac{X_{in} - X_{i0n}}{I_{in}}$$

$$X_{i0n} = \frac{X_{in \max} + X_{in \min.}}{2} \quad (6.3)$$

$$I_{in} = \frac{X_{in \max} - X_{in \min.}}{2}$$

X_{in} – поточне значення i -го параметру в натуральній шкалі виміру;

X_{i0n} – середина інтервалу дослідження i -го фактору натуральної шкали;

I_{in} – величина інтервалу дослідження (зміни) i -го параметра.

Модель, що отримують в результаті дослідження об'єкту за допомогою регресійного аналізу є залежністю виду:

$$y = f(x_1, x_2, \dots, x_n, a_0, a_1, \dots, a_k)$$

x_1, x_2, \dots, x_k – вхідні фактори;

a_0, a_1, \dots, a_k – коефіцієнти, що характеризують вплив кожного із факторів на вихідну величину y .

Ця модель може бути, як лінійною, так і нелінійною:

$$y = a_0 + \sum_{i=1}^k a_i x_i,$$

$$y = a_0 + \sum_{i=1}^k a_i x_i + \sum_{ij=1}^k a_{ij} x_i x_j + \sum_{i=1}^k a_i^2 x_i^2. \quad (6.4)$$

Найбільш часто метою регресійного аналізу є отримання математичної моделі, яка зазвичай має лінійний вигляд.

$$y = a_0 + \sum_{i=1}^k a_i x_i.$$

На експеримент, що виконується за принципом чорної скриньки, діють не контрольовані та випадкові фактори. Тому в моделі вихідна величина y та коефіцієнти $a_0, a_1, a_2, \dots, a_k$ теж є випадковими величинами. В результаті обробки експериментів за методом регресійного аналізу отримуємо, модель такого вигляду:

$$\tilde{y} = b_0 + \sum_{i=1}^k b_i x_i, \quad (6.5)$$

де $b_0, b_1, b_2, \dots, b_k$ – статистичні оцінки коефіцієнтів $a_0, a_1, a_2, \dots, a_k$ рівняння

6.2.

Отримана таким чином модель (6.5) для визначення вихідної величини y називається *рівнянням регресії*.

В більшості випадках при експериментальних дослідженнях об'єкти є багатофакторними.

При цьому використовують класичний (факторний) експеримент, а також проведення експериментів за наперед встановленим планом.

Класичний експеримент є послідовністю однофакторних експериментів, в яких всі змінні крім однієї вважаються постійними, але такий експеримент не дозволяє визначити вплив взаємодії факторів на вихідний параметр, тому використовується багатофакторний експеримент за факторним планом.

Основи обробки результатів планового факторного експерименту розроблені Фішером.

Багатофакторне планування базується на розвинутому математичному апараті і дозволяє вирішувати такі задачі:

- 1) отримати математичну модель поведінки об'єкта дослідження;
- 2) виявити об'єктивні закономірності і отримати додаткову інформацію про об'єкт дослідження;
- 3) перевірити адекватність отриманої математичної моделі, що визначає інтерполяційну залежність.

Крім того, на основі отриманої математичної моделі, що представлена рівнянням регресії, можна розв'язати задачу оптимізації вхідних параметрів досліджуваного об'єкта за заданим критерієм.

Планування експериментів. Основні поняття та визначення

Плановий експеримент дозволяє розкрити механізм явищ, що закладений в області дослідження, а також визначення найбільш доцільну модель для його опису.

Змінні x_1, x_2, \dots, x_k прийнято називати *факторами*, що змінюються за бажанням експериментатора.

Вихідну величину Y називають *відгуком* або ж *функцією мети* або ж параметром оптимізації.

Дослідження здійснюється в деякій області при цьому область повинна бути такою, щоб інтервал зміни вхідних параметрів $x_i, i=1, 2, \dots, k$ значно перевищував похибку встановлену цим фактором.

Вихідна величина при переході одного значення параметру $x_i, i=1, 2, \dots, k$ до іншого змінювались значимо.

Зазвичай для отримання лінійного рівняння регресії використовується повний факторний експеримент, згідно з яким кожен фактор варіюється на двох рівнях ПФЕ 2^k .

Для отримання коефіцієнта $b_i, i=1, 2, \dots, k$ використовується метод найменших квадратів.

В задачах регресійного аналізу приймаються такі припущення:

- 1) керовані змінні $x_i, i=1, \dots, k$ є незалежними;
- 2) значення керованих змінних задаються абсолютно точно;
- 3) керовані змінні $x_i, i=1, 2, \dots, k$ не є випадковими величинами;
- 4) вихідна величина $y_U, U=1, \dots, W$ є випадковою і розподілена по нормальному закону;
- 5) дисперсії вихідної величини y_U в різних точках факторного простору є однорідними.

Побудова планів повного факторного експерименту

Повний факторний експеримент – це керований експеримент в якому всі фактори комбінуються на всіх можливих рівнях.

Тут використовуються два рівні: нижчий, який має кодоване значення -1 , та верхній $+1$.

Тоді загальна кількість дослідів дорівнює $N = 2^k$, k – кількість факторів, що використовуються в експерименті.

План проведення експерименту зручно представити в вигляді матриці, розмірності $(N \times k)$, N – кількість дослідів, k – кількість факторів.

Для підвищення точності математичної моделі доцільно повторювати експеримент кілька раз на кожному із рівнів.

З метою зменшення впливу на результати моделювання, експерименти необхідно виконувати в випадковій послідовності.

Процес організації випадкової послідовності виконання експериментів називається *рандомізацією*.

Приведемо план ПФЕ 2^3 для $k = 3$.

Таблиця 6.1. План проведення експериментів ПФЕ 2^3 .

Номер досліджу	x_1	x_2	x_3	y_{U1}	...	y_{Un}
1	-1	-1	-1	y_{11}	...	y_{1n}
2	+1	-1	-1	y_{21}	...	y_{2n}
3	-1	+1	-1	y_{31}	...	y_{3n}
4	+1	+1	-1	y_{41}	...	y_{4n}
5	-1	-1	+1	y_{51}	...	y_{5n}
6	+1	-1	+1	y_{61}	...	y_{6n}
7	-1	+1	+1	y_{71}	...	y_{7n}
8	+1	+1	+1	y_{81}	...	y_{8n}

В таблиці 6.1 n – кількість повторень дослідів на кожному із рівнів; U – номер досліджу в експерименті.

Визначення коефіцієнтів математичної моделі (рівняння регресії)

З врахуванням раніш наведених припущень покажемо приклад отримання коефіцієнтів рівняння регресії (математичної моделі) для випадку двох факторів, тобто $k = 2$. Тоді це рівняння буде мати такий вигляд:

$$y = b_0 + b_1x_1 + b_2x_2$$

План проведення експериментальних досліджень ПФЕ 2^2 для отримання коефіцієнтів моделі буде мати такий вигляд (таблиця 6.2):

Таблиця 6.2. План проведення експериментів плану ПФЕ 2^2 .

Номер	x_0	x_1	x_2	y_U
1	+1	-1	-1	y_1
2	+1	+1	-1	y_2
3	+1	-1	+1	y_3
4	+1	+1	+1	y_4

В цей план введено додатковий стовпчик з фактором x_0 , який у всіх експериментах приймає значення +1.

Тоді рівняння регресії буде мати такий вигляд:

$$y = b_0 x_0 + b_1 x_1 + b_2 x_2$$

Згідно з методом найменших квадратів визначають :

$$E = \sum_{U=1}^N (y_U - \tilde{y}_u)^2 = \sum_{U=1}^N (y_U - (b_0 x_{0U} + b_1 x_{1U} + b_2 x_{2U}))^2$$

Для визначення невідомих коефіцієнтів ми розв'язуємо:

$$\begin{cases} \frac{\partial \xi}{\partial b_0} = 0 \\ \frac{\partial \xi}{\partial b_1} = 0 \\ \frac{\partial \xi}{\partial b_2} = 0 \end{cases}$$

Якщо взяти часткові похідні, ми отримали таку систему:

$$\begin{cases} b_0 \sum_{U=1}^N x_{0U}^2 + b_1 \sum_{U=1}^N x_{1U} x_{0U} + b_2 \sum_{U=1}^N x_{2U} x_{0U} = \sum_{U=1}^N y_U x_{0U} \\ b_0 \sum_{U=1}^N x_{0U} x_{1U} + b_1 \sum_{U=1}^N x_{1U}^2 + b_2 \sum_{U=1}^N x_{1U} x_{2U} = \sum_{U=1}^N y_U x_{1U} \\ b_0 \sum_{U=1}^N x_{0U} x_{2U} + b_1 \sum_{U=1}^N x_{1U} x_{2U} + b_2 \sum_{U=1}^N x_{2U}^2 = \sum_{U=1}^N y_U x_{2U} \end{cases}$$

Перевагою використання планів повного фактичного експерименту є те, що вони мають такі властивості:

1) *ортogonalність* – це сума рядкових добутків елементів будь – яких стовпчиків матриці дорівнює 0:

$$\sum_{U=1}^N x_{Ui} x_{Uj}, i, j = \overline{0, k};$$

2) *нормування* – сума квадратів елементів будь – якого стовпчика дорівнює кількості дослідів:

$$\sum_{U=1}^N x_{xi}^2 = N, i = \overline{0, k}$$

3) *симетрія* – алгебраїчна сума елементів реального стовпчика дорівнює 0:

$$\sum_{n=1}^N x_{Ui} = 0$$

4) *ротатабельність* – дисперсії значень вихідної величини на різних відстанях від центру плану є постійні і мінімальні:

$$\begin{cases} b_0 N + b_1 \cdot O + b_2 \cdot O = \sum_{U=1}^N y_U x_{0U} \\ b_0 \cdot O + b_1 \cdot N + b_2 \cdot O = \sum_{U=1}^N y_U x_{1U} \\ b_0 \cdot O + b_1 \cdot O + b_2 \cdot N = \sum_{U=1}^N y_U x_{2U} \end{cases}$$

тоді:

$$b_0 = \frac{\sum_{U=1}^N y_U}{N}, \quad b_1 = \frac{\sum_{U=1}^N y_U \cdot x_{1U}}{N}, \quad b_2 = \frac{\sum_{U=1}^N y_U \cdot x_{2U}}{N}$$

$$b_i = \frac{\sum_{U=1}^N y_U x_{iU}}{N}$$

Врахування нелінійності добутку факторів

Використовуючи план ПФЕ можна отримати рівняння регресії:

$$\tilde{y} = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i,j=1}^k b_{ij} x_i x_j$$

Для отримання коефіцієнта такої моделі план повного факторного експерименту (ПФЕ) доповнюється додатковими стовпчиками.

Нехай $k = 2$, тоді матриця планового експерименту буде мати такий вигляд (таблиця 6.3) :

Таблиця 6.3. План планового експерименту *ПФЕ* 2^2 з врахуванням нелінійності добутку факторів.

Номер експер.	x_0	x_{1U}	x_{2U}	$x_{1U}x_{2U}$	y_U
1	+1	-1	-1	+1	y_1
2	+1	+1	-1	-1	y_2
3	+1	-1	+1	-1	y_3
4	+1	+1	+1	+1	y_4

Отриманий план відповідає раніш вказаним властивостям плану повного факторного експерименту (ПФЕ):

$$b_{ij} = \frac{\sum_{U=1}^N y_U x_{iU} x_{jU}}{N}, \quad i, j = \overline{1, k}$$

тоді,

$$b_{ije} = \frac{\sum_{U=1}^N y_U x_{iU} x_{jU} x_{eU}}{N}, \quad \begin{matrix} i \neq j \neq e \\ i, j, e = \overline{1, k} \end{matrix}$$

План проведення експериментальних досліджень *ПФЕ* 2^3 з врахуванням добутку факторів наведено в таблиці 6.4.

Таблиця 6.4. План проведення експериментів ПФЕ 2^3 з врахуванням добутку факторів.

Номер досліджу	x_0	x_{1U}	x_{2U}	x_{3U}	$x_{1U} x_{2U}$	$x_{1U} x_{3U}$	$x_{2U} x_{3U}$	y_U
1	+1	-1	-1	-1	+1	+1	+1	y_1
2	+1	+1	-1	-1	-1	-1	+1	y_2
3	+1	-1	+1	-1	-1	+1	-1	y_3
4	+1	+1	+1	-1	+1	-1	-1	y_4
5	+1	-1	-1	+1	+1	-1	-1	y_5
6	+1	+1	-1	+1	-1	+1	-1	y_6
7	+1	-1	+1	+1	-1	-1	+1	y_7
8	+1	+1	+1	+1	+1	+1	+1	y_8

Дробовий факторний експеримент (ДФЕ)

Відрізняється від ПФЕ тим, що дозволяє суттєво зменшити кількість дослідів.

В теорії планування експериментів є методи побудови частин планів ПФЕ, які мають властивості планів ПФЕ.

Такі плани називаються *дробовими*. Якщо використовується половина плану $\frac{1}{2}$ ПФЕ 2^k , то такий план називається ДФЕ 2^{k-1} , називається *напівреплікою* ПФЕ.

Якщо $\frac{1}{4}$ ДФЕ 2^{k-2} називається *четвертною реплікою* ПФЕ.

Будь – які дробові репліки повинні мати властивості ортогональності, нормування, симетрія та ротатабельність.

При створенні планів ДФЕ спочатку будується план ПФЕ, але для меншої кількості факторів. А потім несуттєве: взаємодія факторів представляє відсутній фактор.

На прикладі побудови ДФЕ 2^{4-1} будемо план для меншої кількості ПФЕ 2^3 (таблиця 6.5)

В результаті обробки експерименту отримаємо рівняння регресії:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3$$

Нехай взаємодія 1 і 3 фактора буде не суттєвою, тобто $x_2 x_3$ приблизно дорівнює нулю. Тоді для реалізації плану ДФЕ 2^{4-1} замінюємо відсутній фактор x_4 цією несуттєвою взаємодією – $x_4 = x_1x_3$. В результаті цього

отримаємо такий план проведення експерименту, який матиме назву дробовий факторний експеримент ДФЕ 2^{4-1} , а вираз $x_4=x_1x_3$ називають його генеруючим співвідношенням.

План його проведення ДФЕ 2^{4-1} наведено в таблиці 6.6.

Таблиця 6.5. План експерименту з меншою кількістю факторів – ПФЕ 2^3 .

Номер досліджу	x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	$x_1 x_2 x_3$
1	-1	-1	-1	+1	+1	+1	-1
2	+1	-1	-1	-1	-1	+1	+1
3	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	-1	+1	-1	-1	-1
5	-1	-1	+1	+1	-1	-1	+1
6	+1	-1	+1	-1	+1	-1	-1
7	-1	+1	+1	-1	-1	+1	-1
8	+1	+1	+1	+1	+1	+1	+1

Таблиця 6.6. План проведення експериментів ДФЕ 2^{4-1} .

№ досліджу	x_0	x_1	x_2	x_3	$x_4=x_1x_3$	x_1x_2	x_1x_3	x_1x_4	x_2x_3	x_2x_4	x_3x_4	$x_1x_2x_3x_4$
1	+1	-1	-1	-1	-1	+1	+1	+1	+1	+1	+1	+1
2	+1	+1	-1	-1	+1	-1	-1	+1	+1	-1	-1	+1
3	+1	-1	+1	-1	+1	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	+1	-1	-1	+1	-1	-1	-1	-1	+1	+1
5	+1	-1	-1	+1	+1	+1	-1	-1	-1	-1	+1	+1
6	+1	+1	-1	+1	-1	-1	+1	-1	-1	+1	-1	+1
7	+1	-1	+1	+1	-1	-1	-1	+1	+1	-1	-1	+1
8	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1

Тоді математична модель буде представлятись рівнянням регресії такого виду:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4$$

Недоліком цього плану є те, що деякі із стовпчиків співпадають, тому їх треба викреслити, а це означає, наприклад, що b_1 є оцінкою коефіцієнтів $a_1 + a_{34}$:

$$b_3 \longrightarrow a_1 + a_{14}$$

$$b_1 \longrightarrow a_4 + a_{13}$$

Для виявлення цього ефекту змішування оцінок введено поняття визначального контрасту, який отримано шляхом множення лівої та правої частин генеруючого співвідношення на введenu змінну (в нашому випадку це x_4).

Тоді визначальний контраст матиме такий вигляд:

$$I = x_1 x_3 x_4$$

Визначальний контраст дозволяє визначити всі змішування оцінок коефіцієнтів математичної моделі шляхом множення лівої та правої частин його виразу на відповідні змінні або їх комбінації, наприклад:

$$x_3 = x_1 x_4$$

$$x_2 = x_1 x_2 x_3 x_4$$

$$x_1 = x_3 x_4$$

Тобто можна визначити змішування оцінок, що для наведених виразів має такий вигляд:

$$b_1 \longrightarrow a_1 + a_{34}$$

$$b_2 \longrightarrow a_2 + a_{1234}$$

$$b_3 \longrightarrow a_3 + a_{14}$$

$$x_1 x_2 = x_2 x_3 x_4$$

$$b_{12} \longrightarrow a_{12} + a_{234}$$

Дробові репліки дозволяють значно зменшити кількість дослідів.

Можна зменшити кількість дослідів до такого рівня, коли вона буде дорівнювати кількості факторів.

План в якому кількість дослідів дорівнює кількості факторів називається *насиченим*.

Обробка і оцінка експериментальних даних

Нехай при проведенні повного факторного експерименту отримано значення вихідної величини y_{U_i} , U' – номер дослідів; $U = \overline{1, N}$, $N = 2^3$;

i – номер повторення дослідів по кожному із рівнів, $i = \overline{1, n}$;

n – кількість повторень.

Через наявність похибок вимірювання і дії неконтрольованих факторів всі значення y_{U_i} є незалежними випадковими величинами.

Обробка цих результатів дослідження з метою отримання математичної моделі здійснюється на основі теорії математичної статистики та регресійного аналізу.

Ця обробка експериментальних даних в теорії планування експериментів повністю формалізована і легко реалізується на ЕОМ. Вона складається із окремих етапів, які часто містять перехід до наступного етапу обробки при виконанні даних умов.

Якщо вказана умова не виконується, то подальші обчислення припиняються і здійснюються необхідні зміни в плановий експеримент.

Характер змін і напрямок подальших дій визначає експериментатор.

Розглянемо окремі етапи і можливі варіанти неформальних рішень.

Етап 1.

Виключення грубих похибок

Нехай в результаті проведення експериментальних досліджень згідно з планом ПФЕ 2^3 отримано значення вихідної величини y_{ui} , де кількість повторень експериментів на кожному рівні однакова і дорівнює n .

План проведення експериментів та значення вихідної величини наведено в таблиці 6.7.

Таблиця 6.7. План ПФЕ 2^3 та результати його проведення.

Номер досліджу	x_0	x_1	x_2	x_3	$x_1 x_2$	y_{ui}			
						y_{u1}	y_{u2}	...	y_{un}
1	+1	-1	-1	-1	+1	y_{11}	y_{12}	...	y_{1n}
2	+1	+1	-1	-1	-1	y_{21}	y_{22}	...	y_{2n}
3	+1	-1	+1	-1	-1	y_{31}	y_{32}	...	y_{3n}
4	+1	+1	+1	-1	+1	y_{41}	y_{42}	...	y_{4n}
5	+1	-1	-1	+1	+1	y_{51}	y_{52}	...	y_{5n}
6	+1	+1	-1	+1	-1	y_{61}	y_{62}	...	y_{6n}
7	+1	-1	+1	+1	-1	y_{71}	y_{72}	...	y_{7n}
8	+1	+1	+1	+1	+1	y_{81}	y_{82}	...	y_{8n}

Для включення явно поганих результатів експериментальних досліджень, які можуть в подальшому негативно позначитись на отриманому рівнянні регресії, в кожному рядку планів здійснюється перевірка однорідності рішень y_{ui} за допомогою критерію Стьюдента, значення якого визначають за формулою:

$$t_p = \frac{|y_U^* - \tilde{y}_U|}{S_U},$$

де y_U^* – найменше або найбільше значення вихідної величини y_{ui} в u -му рядку плану досліджень;

$\tilde{y}_U = \frac{\sum_{i=1}^{n-1} y_{Ui}}{n-1}$ – середнє значення в u -му рядку без врахування y_U^* ;

S_U – оцїнка середньо квадратичного вїдхилення в u -му рядку:

$$S_U = \sqrt{\frac{1}{n_U - 1} \sum_{i=1}^{n-1} (y_{Ui} - \tilde{y}_U)^2};$$

n_U – кїлькїсть повторень в u -му рядку.

Якщо розрахункове значення $t_p > t_{табл., \alpha, f}$, де α – рївень значущостї, а $f = n_u - 1$ – кїлькїсть степенїв свободи, то вважається, що y_U^* є грубою похибкою і її треба виключити з таблицї результатїв експериментальних дослїджень.

Для спрощення подальшої обробки результатїв доцїльно повторити дослїд на цьому ж рївнї, для того щоб забезпечити однакову кїлькїсть повторень на кожному з рядкїв плану.

Етап 2.

Перевїрка однорїдностї вимїрїв

Для перевїрки однорїдностї вимїрїв в кожному рядку планїв визначаємо \tilde{y}_U .

Із всїх середнїх значень визначаємо найбільше і найменше значення, а потїм обчислюють розрахункове значення критерїю Стюдента за формулою:

$$t_p = \frac{\tilde{y}_{U_{\max}} - \tilde{y}_{U_{\min}}}{S_y \sqrt{\frac{1}{n_{\max}} + \frac{1}{n_{\min}}}}$$

n_{\max}, n_{\min} – кїлькїсть повторень дослїдїв в рядках з максимальним та мїнімальним середнїми значеннями $\tilde{y}_{U_{\max}}, \tilde{y}_{U_{\min}}$ вїдповїдно;

S_y – оцїнка дисперсїї вїдтворюваностї, що визначається за формулою:

$$S_y = \sqrt{\frac{1}{N(n-1)} \sum_{U=1}^N \sum_{i=1}^n (y_{Ui} - \tilde{y}_U)^2}.$$

Якщо $t_p > t_{табл., \alpha, f}$, де α – рївень значущостї, а $f = n_{\max} + n_{\min}$ – кїлькїсть степенїв свободи, то вважають, що значення вихїдної величини в рїзних точках плану сильно вїдрїзняється. Тому цї результати експерименту можна вважати результативними і можна перейти до наступного етапу обробки результатїв експерименту.

Етап 3.

Визначення однорїдностї дисперсїй

При однаковїй кїлькостї повторень дослїдїв на кожному рївнї в кожному рядку плану, тобто:

$$n_U = n = const, \quad U = \overline{1, N}$$

перевірка однорідності дисперсій здійснюється за G - критерієм Кохрена, розрахункове значення якого визначається за формулою:

$$G_p = \frac{S_{U \max}^2}{\sum_{U=1}^N S_U^2}$$

$$S_U^2 = \frac{1}{n-1} \sum_{U=1}^N (y_{Ui} - \tilde{y}_U)^2$$

S_U^2 – дисперсія в u -му рядку плану;

$S_{U \max}^2$ – максимальне значення дисперсії із всіх визначених.

Визначене розрахункове значення G_p порівнюється з табличним $G_{табл}$, що взяте з кількістю степенів свободи $f_1 = n-1$ і $f_2 = N$ та рівнем значущості α .

Якщо $G_p \leq G_{табл}$, то гіпотеза про однорідність дисперсій приймається, а це означає, що сильно в досліді немає дисперсій, що сильно відрізняється інших, тобто в досліді немає системних похибок.

Інакше, якщо $G_p > G_{табл}$, то вважають, що виміри не є однорідними і потрібно виконати деякі неформальні дії, а саме: перейти до іншої функції, а бо ж виключити грубі похибки серед вимірів, або ж зменшити інтервал варіювання вхідних факторів.

Етап 4.

Визначення статистичної значимості коефіцієнтів рівняння регресії

Після визначення значень коефіцієнтів рівняння регресії

$$\tilde{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{123} x_1 x_2 x_3,$$

$$\text{або} \quad \tilde{y} = b_0 + \sum_{i=1}^k b_i x_i + \sum_{\substack{n,j=1 \\ i \neq j}}^k b_{ij} x_i x_j + \sum_{\substack{i,j,e=1 \\ i \neq j \neq e}}^k b_{ije} x_i x_j x_e$$

треба перевірити їх значимість за t - критерієм Стьюдента, розрахункове значення обчислюється за формулою:

$$t_p = \frac{|b_i|}{S_b},$$

$$S_b = \frac{S_y}{\sqrt{nN}},$$

$$S_y = \sqrt{\frac{1}{N(n-1)} \sum_{U=1}^N \sum_{i=1}^n (y_{Ui} - \tilde{y}_U)^2}$$

Визначене розрахункове значення t - критерію Стьюдента порівнюється з $t_{табл}$, що взяте з рівнем значущості α та кількістю степенів свободи $f = N(n-1)$.

Якщо виконується умова, що $t_p > t_{табл, \alpha, f}$, то вважається, що відповідний коефіцієнт рівняння регресії є статистично значимим і його можна

використовувати в моделі, яка відповідає рівнянню регресії, інакше його порівнюють нулю і його не включають в отриману модель.

Значимість коефіцієнта сильно залежить від кількості дослідів N та кількості повторень дослідів n на кожному рівні плану.

При їх збільшенні точність моделі зростає.

Етап 5.

Перевірка адекватності отриманої математичної моделі, що представлена в вигляді рівняння регресії

Перевірка адекватності отриманої моделі, описуваному ним об'єкту дослідження здійснюється за F - критерієм Фішера.

Розрахункове значення цього критерію визначається за формулою:

$$F_p = \frac{S_{ad}^2}{S_y^2},$$

$$S_{ad}^2 = \frac{1}{(N - k - 1)} \sum_{U=1}^n (\bar{y}_U - \tilde{y}_U)^2$$

S_{ad}^2 – дисперсія адекватності отриманої моделі, що характеризує розкид середніх значень \bar{y}_U , що отримане в результаті проведення експериментів і значень вихідної величини \tilde{y}_U , які отримані за допомогою рівняння регресії.

k – кількість факторів в експерименті.

Розрахункове значення критерію порівнюється з $F_{табл.\alpha, f_1, f_2}$, що взяте з рівнем значущості α та кількістю степенів свободи f_1 та f_2 :

$$f_1 = N - k - 1;$$

$$f_2 = N(n - 1).$$

Якщо виконується умова, що $F_p \leq F_{табл.\alpha, f_1, f_2}$, то гіпотеза про адекватність опису досліджуваного об'єкта отриманою математичною моделлю приймається.

Якщо вказана умова не виконується, то дана модель не адекватно описує досліджуваний об'єкт.

Для отримання адекватної моделі, дослідник може виконати такі неформальні дії для отримання адекватної математичної моделі:

- збільшити кількість повторень дослідів по кожному рівні плану;
- зменшити інтервал варіювання факторів;
- змінити шкалу виміру вихідної величини y , тобто представити її в математичній моделі як $y^2, y^3, \sqrt{y}, \sqrt[3]{y}, \ln(y), e^y$ тощо;
- перейти до отримання моделі нелінійного вигляду:

$$\tilde{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3$$

$$\text{або } \tilde{y} = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i,j=1}^k b_{ij} x_i x_j + \sum_{i=1}^k b_u x_i^2.$$