Практична робота №5
МЕТОДИКА ВИЗНАЧЕННЯ РІВНЯНЬ І КОЕФІЦІЄНТІВ РЕГРЕСІЇ

Мета роботи – оволодіти методикою визначення рівнянь і коефіцієнтів регресії.

1 Вказівки з самопідготовки до роботи

1.2 Завдання для самостійної підготовки

Вивчити:

– апроксимація результатів експериментальних досліджень;

– регресивний аналіз результатів експериментальних досліджень.

– *Скласти звіт по роботі:*

– номер, найменування та мета роботи;

– апроксимація результатів експериментальних досліджень;

– регресивний аналіз результатів експериментальних досліджень.

1.2 Питання для самопідготовки

1.2.1 Апроксимація результатів експериментальних досліджень.

1.2.2 Регресивний аналіз результатів експериментальних досліджень.

1.3 Рекомендована література

1. Адаменко М. І., Бейлін М. В. Основи наукових досліджень. Х.: ХНУ імені В. Н. Каразіна, 2014. 188 с.

2. Бобилєв В. П., Іванов І. І., Пройдак Ю. С. Методологія та організація наукових досліджень: навчальний посібник. Дніпропетровськ: Системні технології, 2008. 264 с.

3. Кислий В. М. Організація наукових досліджень: навчальний посібник. Суми: Університетська книга, 2011. 224 с.

ВКАЗІВКИ ДО САМОСТІЙНОЇ ПІДГОТОВКИ

1 Апроксимація результатів експериментальних досліджень

Поняття апроксимації. Процес одержання на основі результатів експериментальних досліджень математичної залежності $y=\varphi(x)$, яка з достатньою точністю відтворює досліджувану закономірність

y = f(x), називається ***апроксимацією.*** Функціональні залежності, одержані способом апроксимації експериментальних даних, називаються *емпіричними.*

Емпірична залежність *y=φ(x)* по суті є *математичною моделлю* процесу дослідження, результати якої дійсні тільки в межах зміни аргументу, тобто в інтервалі варіації фактора $x_1, x_2, \ldots, x_к$.

Необхідність в емпіричних залежностях виникає тоді, коли аналітичні залежності вважаються складними і вимагають громіздких обчислень для практичного використання або ж тоді, коли аналітичні залежності взагалі відсутні. Можна вважати, що емпіричні залежності – це наближене виявлення аналітичних, а процес апроксимації – спосіб заміни складного або неможливого процесу одержання точних аналітичних виразів.

Виконання апроксимації результатів експериментальних досліджень складається з двох основних і послідовних етапів, а саме:

- *етап* – вибір загального вигляду типової функціональної залежності (апроксиманти);
- *етап* – розрахунок числових значень параметрів (коефіцієнтів) апроксиманти.

Вибір загального вигляду рівняння апроксимації. Для того, щоб з`ясувати, до якого класу функцій належить шукана апроксиманта *y=f(x),* необхідно звернутись до графічного зображення результатів експерименту. Графік будується в декартовій системі координат *x* і *y.* Значення фактора відкладаються на осі абсцис, значення параметра оцінки – на осі ординат, а власне результати позначаються точками (рис. 2).

З`єднавши точки прямою 1, одержують графік результатів експерименту. Якщо відмітити серединні точки кожного з відрізків, то через них можна провести пряму 2 (або криву), яка буде приблизним уявленням графіка шуканої апроксиманти.

Далі порівнюють отриманий графік з графіками типових функцій і обирають загальний вигляд рівняння апроксиманти, яке буде найбільш подібно описувати досліджувану залежність.

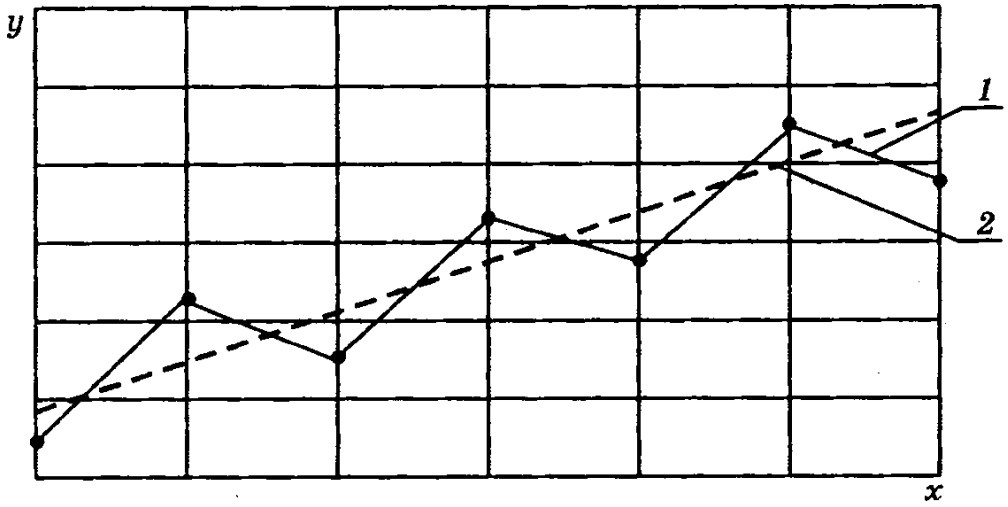

Рис. 2 – Графічне зображення результатів експерименту:
1 – експериментальна лінія; 2 – уявна апроксиманти

Методи визначення коефіцієнтів апроксиманти. Після вибору загального вигляду апроксимуючої функції переходять до розрахунку числових значень її коефіцієнтів. Залежно від типу обраної функції та вимог щодо точності результатів розрахунку застосовують такі методи:

- графічний метод;
- метод середніх;
- метод найменших квадратів.

Графічний метод застосовується для лінійних функцій та функцій, що зводяться до лінійних методом вирівнювання. Для цього вводять нові змінні:

$$x`=f_1(x,y);\ \ y`=f_2(x,y),$$

за допомогою яких рівняння набуває вигляду лінійної залежності: $y`=b_0 + b_1 x`$.

Вирівнюванню підлягають такі залежності, як гіперболічна, показникова, степенева, логарифмічна та ін.

Метод середніх завдяки своїй простоті дозволяє у більшості випадків замінити громіздкий метод найменших квадратів і одержати достатньо задовільні за точністю результати. Цей метод полягає у тому, що після того як визначено тип функції й виконано, в разі необхідності, вирівнювання, одержують лінійну залежність типу: $Y=b_0 + b_1X$. На основі одержання залежності, підставляючи значення X

та Y з кожного досліду, складають умовні рівняння, число яких дорівнює кількості дослідів N.

Усі рівняння з невідомими b_0 та b_1 розбивають на дві рівні групи і кожну з них почленно сумують. У результаті одержують рівняння, з яких складають систему:

$$\begin{cases} \sum_{i}^{\frac{N}{2}} Y_i = \sum_{i}^{\frac{N}{2}} b_0 + \sum_{i}^{\frac{N}{2}} b_1 X_i \\ \sum_{\frac{N}{2}}^{N} Y_i = \sum_{\frac{N}{2}}^{N} b_0 + \sum_{\frac{N}{2}}^{N} b_1 X_i \end{cases} \quad (18)$$

Із системи цих рівнянь знаходять значення невідомих коефіцієнтів b_0 та b_1 і записують остаточне рівняння апроксиманти.

Сутність *методу найменших квадратів* полягає у тому, що для двох функціонально зв'язаних величин x та y відомо N пар відповідних значень $(x_1;y_1), (x_2;y_2) \dots (x_n;y_n)$. Вимагається в наперед заданій формулі $y=f(x, b_0, b_1 \dots b_m)$ визначити кількість $(m+1)$ параметрів $b_0, b_1 \dots b_m$ $(m<n)$ так, щоб ця формула дозволила з найбільш точною відповідністю поновлювати значення вихідних параметрів y для заданих значень x.

З курсу теорії ймовірностей відомо, що найкращими є ті значення параметрів функції, котрі перетворюють на мінімум суму:

$$\sum_{k=1}^{N} [f(x_n, b_0, b_1, .. b_m) - y_k]^2 = min \quad (19)$$

тобто суму квадратів відхилень значень y, які визначені за формулою, від експериментальних. Саме тому цей метод отримав назву методу найменших квадратів.

Використовуючи необхідні умови мінімуму функції багатьох змінних, вираховуємо часткові похідні функції та прирівняємо їх до нуля. Це дає систему $(m+1)$ рівнянь з $(m+1)$ невідомим, тобто

$$\sum_{k=1}^{N} [f(x_n, b_0, b_1, .. b_m) - y_k] \frac{df_{(x_n, b_0, b_1, .. b_m)}}{db_i} = 0 \quad (20)$$

А.2 Регресивний аналіз результатів експериментальних досліджень

Під *регресивним аналізом* розуміють дослідження закономірності зв`язку між двома змінними, коли одному значенню X відповідає

сукупність значень *Y*, тобто зв`язок між ними не повністю визначений.

Функцію *Y=f(X)* називають регресивною, коли значення *Y* утворюють статистичний ряд розподілу з характеристиками безперервної випадкової величини. Тому регресивний зв`язок між величинами *X* та *Y* можна визначити лише тоді, коли забезпечується можливість виконання статистичних замірів.

Статистичні залежності описують математичними моделями, тобто *рівняннями регресії*, які відтворюють зв`язок між значеннями фактору *X* і змінною характеристикою досліджуваного процесу *Y*.

Рівняння регресії, по можливості, повинні бути простими й адекватними.

Існують *однофакторні* й *багатофакторні* регресивні залежності.

Регресивний аналіз виконується у такій послідовності:

- перевірка наявності кореляційного зв`язку;
- апроксимація експериментальних даних;
- статистичний аналіз рівнянь регресії.

***Перевірка наявності кореляційного зв`язку.*** У багатьох випадках метою експериментальних досліджень є, насамперед, виявлення наявності залежності між двома змінними величинами. Якщо змінність однієї випадкової величини впливає на розподіл іншої, то вважають, що між такими випадковими величинами існує *статистичний зв`язок*. Для оцінки статистичного зв`язку між двома змінними величинами використовують *коефіцієнт кореляції.* Визначення коефіцієнта кореляції виконується на основі результатів експериментальних спостережень.

Нехай проведено *N* спостережень, і в кожному випадку відомо значення двох параметрів *x* та *y*, тобто одержано дві вибірки:

$$x_1, x_2, \ldots, x_n \qquad y_1, y_2, \ldots y_n.$$

За кожною з них знаходять середнє значення та середнє квадратичне відхилення S_x та S_y. Тоді коефіцієнт кореляції визначається за формулою:

$$R = \frac{\sum_{i=1}^{n}(x_i - \tilde{x})(y_i - \tilde{y})}{(n-1)S_x S_y} \qquad (21)$$

Величина коефіцієнта кореляції завжди перебуває у межах $-1 \leq R \leq 1$. Коефіцієнт характеризує тільки лінійну залежність між випадковими величинами. При додатному значенні коефіцієнта можна вважати, що при зростанні однієї величини інша в середньому теж зростає. При від'ємному значенні R зростання одної величини зумовлює в середньому зменшення значення іншої. При $R = +1$ або $R = -1$ між величинами x і y існує кореляційний зв`язок. Значення $R=0$ свідчить про відсутність лінійної залежності між $x$ і $y$. Щільність зв`язку вважають задовільною при $R \geq 0,5$, доброю – при $R=0,8...0,85$.

Для перевірки відповідності вибіркового значення коефіцієнта кореляції R значенню кореляції (ρ) між генеральними сукупностями x і y використовують *t – розподіл Стьюдента.* Спочатку знаходять розрахункове значення $t_{розр}$ за формулою:

$$t_{\text{розр}} = |N|\sqrt{\frac{N-2}{1-R^2}} \qquad (22)$$

і порівнюють його з табличним.

Якщо $t_{розр} > t_{табл}$ при кількості ступенів свободи $f=N-2$ і рівні значимості q=5%, то кореляційний зв`язок існує, а також підтверджується для генеральних сукупностей.

Статистичний аналіз рівняння регресії. Статистичний аналіз одержаного рівняння регресії полягає у розв'язанні двох основних завдань: оцінки значимості коефіцієнтів рівняння; перевірки адекватності рівняння регресії експериментальним даним.

Необхідною передумовою статистичного аналізу є *нормальність розподілу вихідної величини і однорідність дисперсій дослідів.*

Здійснивши перевірку забезпечення передумов, можна починати ви конання статистичного аналізу рівняння регресії. Для розв'язання задач аналізу необхідно мати *кількісну оцінку похибки експерименту* в цілому. Такою оцінкою є *дисперсія відновлення* S_y^2 , яка визначається за формулою:

$$S_y^2 = \frac{\sum_{j-1}^{N} S_j^2}{N} \qquad (23)$$

де N – кількість дослідів; S_y^2 - дисперсія j-го досліду.

Кількість ступенів свободи f_y дисперсії експерименту дорівнює сумі ступенів свободи дисперсій всіх дослідів, тобто $f_y=N(n-1)$.

Оцінка точності визначення коефіцієнтів рівняння регресії та їх значущості. Після того, як рівняння одержано і знайдено дисперсію відновленості, необхідно оцінити точність визначених коефіцієнтів регресії. Оскільки їх визначено з результатів експерименту, а результати є випадковими величинами, то значення коефіцієнтів регресії b_i теж будуть випадковими. Тому *показником точності коефіцієнта* буде його *дисперсія* S_b^2 , яка визначається за формулою:

$$S_b^2 = \frac{S_y^2}{n \cdot N} \tag{24}$$

Після цього оцінюють *значимість коефіцієнтів регресії* за допомогою *t–критерію Стьюдента.* Для кожного коефіцієнта визначають розрахункове значення *t*–критерію за формулою:

$$t_{\text{розр}} = \frac{|b_i|}{S_{b_i}} \tag{25}$$

де S_{bi} – середнє квадратичне відхилення значення коефіцієнтів:

$$S_{b_i} = \sqrt[2]{S_{b_i}^2} \tag{26}$$

За таблицями знаходять значення *t–критерію* для рівня значимості

q=0,05 та числа ступенів свободи $f_y=N(n–1)$.

Якщо виконується умова $t_{розр} > t_{табл}$, то коефіцієнти регресії значимі, а якщо ні, то такий коефіцієнт можна виключити з регресії, тобто прийняти за нуль.

Перевірка адекватності одержаного рівняння регресії необхідна для того, щоб відповісти на запитання, чи буде рівняння відтворювати значення критерію оцінки з тією ж точністю, що і результати експерименту.

Для цього використовують значення *F–критерію Фішера.* Спочатку визначають розрахункове значення *F*– критерію за формулами:

$$F_{\text{розр}} = \frac{S_{\text{ад}}^2}{S_{\text{у}}^2}, \text{ якщо } S_{\text{ад}}^2 > S_{\text{у}}^2 \tag{27}$$

$$F_{\text{розр}} = \frac{S_y^2}{S_{\text{ад}}^2}\text{, якщо } S_y^2 > S_{\text{ад}}^2 \quad (28)$$

де $S_{\text{ад}}^2$ дисперсія адекватності, яка визначається за співвідношенням:

$$S_{\text{ад}}^2 = \frac{n\sum_{j-1}^{n}(\bar{y}_j - \bar{y}_i)^2}{f_{\text{ад}}} \quad (29)$$

де $\overline{y_i}$ значення вихідної величини в *j*– досліді, визначене за рівнянням регресії; $f_{ад}$ – кількість ступенів свободи;

$f=N–P$,

де *P*– число коефіцієнтів рівняння регресії.

За таблицями визначають значення *F*– критерію для рівня значущості

q=0,05 та кількості ступенів свободи f_y і $f_{ад}$.

Рівняння вважається адекватним, якщо виконується умова: $F_{розр}<F_{табл}$.