Міністерство освіти і науки України

Національний лісотехнічний університет України

Кафедра інформаційних технологій

Методичні вказівки

до лабораторної роботи №5

“ Ряди динаміки ”

з дисципліни

“Інтелектуальний аналіз даних”

для студентів з напрямку підготовки 6.050101 **“Комп’ютерні науки”**

Львів-2017

Лабораторна робота №5

Ряди динаміки

Теоретичні відомості

Алгоритм аналізу ряду динаміки

I. Середній рівень ряду динаміки

1. *Середній рівень моментного часового ряду* з рівними проміжками часу між рівнями визначають за формулою:

$$\bar{y} = \frac{1}{n-1}\left(\frac{y_1 + y_n}{2} + \sum_{i=1}^{n} y_i\right). \qquad (1)$$

Середній рівень інтервального ряду з рівними проміжками часу між рівнями визначається середнім арифметичним рівнів ряду динаміки:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\,. \qquad (2)$$

2.Мінливість рівнів ряду відносно середнього рівня характеризується середньоквадратичним відхиленням:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(y_1 - \bar{y})^2}{n-1}} \qquad (3)$$

3. Безрозмірний показник мінливості, який називається ***коефіцієнтом варіації***:

$$v = \frac{s}{\bar{y}} \qquad (4)$$

II. Показники порівняння рівнів часового ряду

1. *Абсолютний приріст ланцюгового показника:*

$$\Delta i = y_i - y_{i-1} \qquad (5)$$

2. *Абсолютний приріст базисного показника:*

$$\Delta i = y_i - y_0 \qquad (6)$$

де y_i — поточний рівень ряду; y_{i-1} — рівень, що передує y_i ; y_0 — базисний рівень;

3. *Темп росту ланцюгового показника:*

$$k_i = y_i / y_{i-1} \tag{7}$$

4. *Темп росту базисного показника:*

$$k_{i0} = y_i / y_0 \tag{8}$$

5. *Темп приросту ланцюгового показника:*

$$T_i = \Delta i / y_{i-1} \tag{9}$$

6. *Темп приросту базисного показника:*

$$T_{i0} = \Delta i_0 / y_0 \tag{10}$$

7. *Середній ланцюговий приріст* розраховують за формулою:

$$\Delta\bar{y} = \frac{1}{n-1}\sum_{i=2}^{n}\Delta i = \frac{y_n - y_1}{n-1} \tag{11}$$

8. Середній ланцюговий темп *росту є середньогеометричною величиною коефіцієнтів росту:*

$$\bar{k} = \sqrt[n-1]{k_2 \cdot k_3 \cdot \ldots \cdot k_n} = \sqrt[n-1]{y_n / y_1} \tag{12}$$

9. Знаючи середній ланцюговий темп росту, можна розрахувати (прогнозувати) *j*-й рівень ряду динаміки:

$$y_i = y_0 \bar{k}^{\,j-1} \tag{13}$$

10. *Середній темп приросту* визначається таким чином:

$$\bar{T} = \bar{k} - 1 \tag{14}$$

III. Автокореляція

1. Взаємозв'язок рівнів ряду динаміки з відповідними рівнями того ж ряду називають автокореляцією.

Вивчаємо взаємозалежність рівнів у парах $(y_i; y_{i+L})$, $i = 1, 2, \ldots, n - L$, тобто взаємозв'язок рівнів ряду динаміки з рінями того ж ряду динаміки, які зміщені на L одиниць часу. Величина L називається ***лагом***, вона може набувати значень $1, 2, \ldots, k$. Звичайно $k < \frac{n}{4}$.

Мірою автокореляції служить коефіцієнт автокореляції.

Обчислюють коефіцієнт автокореляції за формулою:

$$r(L) = \frac{\overline{y_t \cdot y_{t+L}} - \bar{y}_t \cdot \bar{y}_{t+L}}{\sqrt{(\overline{y_t^2} - \bar{y}_t^2)(\overline{y_{t+L}^2} - \bar{y}_{t+L}^2)}} \tag{15}$$

де

$$\overline{y_t \cdot y_{t+L}} = \frac{1}{n-L}\sum_{i=1}^{n-L} y_i y_{i+L}\,; \quad \overline{y_t}^{\,2} = \frac{1}{n-L}\sum_{i=1}^{n-L} y_i^{\,2}\,; \quad \overline{y_{t+L}^2} = \frac{1}{n-L}\sum_{i=1}^{n-L} y_{i+L}^2 \qquad (16)$$

2. Критичне значення коефіцієнта автокореляції виписуємо з табл. 1 додатка:

$$r^* = r(\alpha = 0{,}05;\ n)$$

Якщо $r(L) < r^*$, то автокореляція рівнів часового ряду з лагом є незначною.

Якщо $r(L) > r^*$, то автокореляція рівнів часового ряду з лагом є істотною.

3. Графічне зображення сукупності коефіцієнтів автокореляції точками $(L;\ r_L)$ на площині, з’єднаних відрізками прямих у вигляді многокутника, називають *корелограмою*.

IV. Вирівнювання часового ряду шляхом усереднення

1. Формула вирівнювання часового ряду способом усереднення за трьома рівнями *(m* = 3) така:

$$\widetilde{y}_i = \frac{1}{3}(y_{i-1} + y_i + y_{i+1});\ i = 2, 3, \ldots, n-1 \qquad (17)$$

Усереднені за (10.23) рівні доповнюються початковим і кінцевим вирівняними рівнями, які визначаються за такими формулами:

$$y_1 = \frac{1}{6}(5y_1 + y_2 - y_3);$$

$$y_n = \frac{1}{6}(-y_{n-2} + 2y_{n-1} + 5y_n). \qquad (18)$$

2. Формула вирівнювання часового ряду способом усереднення за п’ятьма рівнями *(m* = 5) має такий вигляд:

$$\widetilde{y}_i = \frac{1}{5}(y_{i-2} + y_{i-1} + y_i + y_{i+1} - y_{i+2}),\ i = 3, 4, \ldots, n-2. \qquad (19)$$

Усереднені за (10.25) рівні доповнюються першим, другим, передостаннім і останнім вирівняними рівнями, які визначають за такими формулами:

$$\widetilde{y}_1 = \frac{1}{5}(3y_1 + 2y_2 + y_3 - y_4);$$

$$\widetilde{y}_2 = \frac{1}{10}(4y_1 + 3y_2 + 2y_3 + y_4);$$

$$\tilde{y}_{n-1} = \frac{1}{10}(y_{n-3} + 2y_{n-2} + 3y_{n-1} + 4y_n);$$
$$\tilde{y}_n = \frac{1}{5}(-y_{n-3} + y_{n-2} + 2y_{n-1} + 3y_n). \qquad (20)$$

Лінія, що з'єднує усереднені точки, є ***емпіричною лінією тренда***.

V. Експонентне вирівнювання

1. Формула для розрахунку експонентно вирівняних значень має такий ви-
гляд:

$$\tilde{y}_i = \alpha y_i + (1-\alpha)\tilde{y}_{i-1}, \; i = 1, 2, 3,, \dots, n. \qquad (21)$$

Величину α у формулі (21) називають *параметром згладжування.* Значення а вибирають у діапазоні від 0,1 до 0,3.

Початковий вирівняний рівень або визначають за формулами (18) чи (20), або просто прирівнюють до першого рівня ряду динаміки.

2. Згладжування рівнів ряду динаміки за експонентним вирівнюванням дає можливість короткострокового прогнозування. Для прогнозованого згладженого рівня $\tilde{y}_{n+1}$ записуємо таке рівняння:

$$\tilde{y}_{n+1} = \tilde{y}_n + \Delta\tilde{y}\;, \qquad (22)$$

де $\Delta\tilde{y}$ — похибка прогнозу.

Похибка прогнозу в (22) оцінюється за такою формулою:

$$\Delta\tilde{y} = \pm t^* s\sqrt{\frac{\alpha}{2-\alpha}}, \qquad (23)$$

де t^* — коефіцієнт Стьюдента; s — середньоквадратичне відхилення (3); α — параметр згладжування.

Вірогідний проміжок для прогнозованого рівня y_{n+1} ширший і визначається таким виразом:

$$y_{n+1} = \tilde{y}_n \pm t^* s\sqrt{1 + \frac{\alpha}{2-\alpha}}. \qquad (24)$$

VI. Аналітичне моделювання тренда

1. Запишемо рівняння регресії

$$\widehat{y}(t) = b_0 + b_1 t. \tag{25}$$

Де
$$b_1 = \frac{\overline{yt} - \bar{y} \cdot \bar{t}}{\overline{t^2} - \bar{t}^2}; \quad b_0 = \bar{y} - b_1 \bar{t} \ . \tag{26}$$

Оскільки рівні ряду визначені для часової послідовності $\{t\} = \{1, 2, \dots, n\}$, то можна використати такі формули:

$$\sum_{i=1}^{n} t_i = \frac{n(n+1)}{2} \ ; \ \bar{t} = \frac{n+1}{2} \ ;$$

$$\sum_{i=1}^{n} t_i^2 = \frac{n(n+1)(2n+1)}{6} \ ; \ \overline{t^2} = \frac{(n+1)(2n+1)}{6}, \tag{27}$$

Та $\overline{yt} = \frac{1}{n} \sum_{i=1}^{n} y_i \sum_{j=1}^{n} t_j$;

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \ ;$$

2. Істотність лінійної тенденції ряду динаміки визначається дисперсійним аналізом. Складемо таблицю дисперсійного аналізу (табл. 1)

Таблиця 1

Джерело мінливості	Сума квадратів	Число ступенів	Середній квадрат	*f*-відношенн
Лінійна регресія	$SS_x = b_1^2 n \left(\frac{n^2 - 1}{12} \right)$	$\nu_x = 1$	$MS_x = SS_x$	—
Залишкове відхилення	$SS_e = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$	$\nu_e = n - 2$	$MS_e = \frac{SS_e}{n-2}$	$f_e = \frac{MS_x}{MS_e}$
Сума	$SS_T = n(\overline{y^2} - \bar{y}^2)$	$\nu_T = n - 1$	—	—

Критична точка визначається з таблиць F-розподілу:

$$f^{*} = f(p = 1 - \alpha;\ \nu_1 = 1;\ \ \nu_2 = n - 2)\,. \tag{28}$$

Якщо емпіричне значення критерію f_e більше від критичного, то є підстави стверджувати про наявність лінійної тенденції у ряді динаміки.

3. Дисперсія ряду динаміки відносно прямої найменших квадратів визначається формулою:

$$S^2 = MS_e = \frac{1}{\text{[illegible]}} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2\ . \tag{29}$$

4. Оцінки дисперсій параметрів лінійного тренда розраховують за такими формулами:

$$S_{b_1}^2 = \frac{12}{\text{[illegible]}} S^2; \tag{30}$$

$$S_{b_0}^2 = \frac{2(2n+1)}{\text{[illegible]}} S^2 = \overline{t^2} S_{b_1}^2\ . \tag{31}$$

5. Вірогідні проміжки для параметрів лінійного тренда визначають за формулами

$$\Pr\left\{ \beta_0 \in \left[\left(b_0 - t_2^{*} S_{b_0}^2 \right); \left(b_0 + t_2^{*} S_{b_0}^2 \right) \right] \right\},$$

$$\Pr\left\{ \beta_1 \in \left[\left(b_1 - t_2^{*} S_{b_1}^2 \right); \left(b_1 + t_2^{*} S_{b_1}^2 \right) \right] \right\},$$

$t_2^{*} = t(p = 1 - \frac{\alpha}{2}\ ;\ \nu = n - 2)$ — коефіцієнт Стьюдента(табл..2 додатка) (32)

6. Для окремого прогнозованого рівня ряду динаміки при $t = t_0$ з вірогідністю $(1 - \alpha)$ вірогідний проміжок є таким:

$$\left[\left(\hat{y}(t_0)-t_2^* s\sqrt{1+\frac{1}{n}+\frac{(t_0-\bar{t})^2}{n(\overline{t^2}-\bar{t}^2)}}\right);\left(\hat{y}(t_0)+t_2^* s\sqrt{1+\frac{1}{n}+\frac{(t_0-\bar{t})^2}{n(\overline{t^2}-\bar{t}^2)}}\right)\right]. \tag{33}$$

VII. Автокореляція відхилень від тренда

1. Розглядаємо часовий ряд, який має модель тренда. Залишкові відхилення від тренда

$$E_i = Y(t_i) - \hat{y}(t_i), i = 1, 2, \ldots, n. \tag{34}$$

утворюють часовий ряд.

Особливо при розгляді адекватності моделі тренда є вимога про незалежність залишкових відхилень. Аналіз незалежності зводиться до перевірки гіпотез про відсутність автокореляції:

H_0: рівні часового ряду (34) незалежні;

H_1: має місце автокореляція рівнів часового ряду (34); α— рівень значущості.

Для перевірки гіпотез використовують критерій Дарбіна — Ватсона:

$$D = \frac{\sum_{i=2}^{n}(E_i - E_{i-1})^2}{\sum_{i=1}^{n} E_i^2}. \tag{35}$$

Значення критерію (35) обмежені проміжком (0; 4). На рис. 1 зображено графік функції щільності критерію та критичні точки.

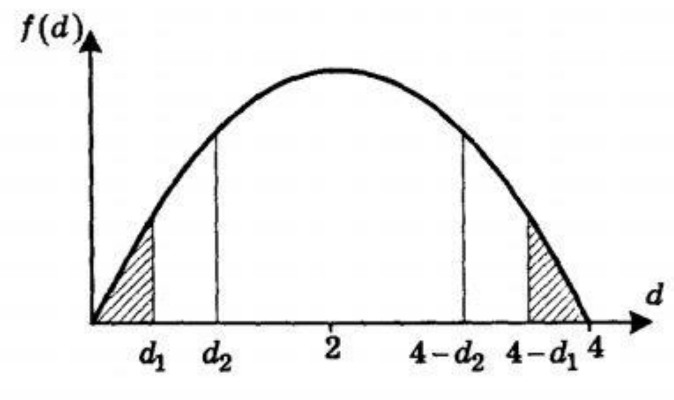

Рис. 1

Якщо у ряді динаміки залишків (33) немає автокореляціїї, то значення розрахованого за формулою (34) емпіричного значення критерію $d_в$ близьке до двох.

Можливі чотири варіанти висновку перевірки гіпотез:

1) $d_в < d_1$ ряд має додатну автокореляцію;

2) $d_в > 4 - d_1$ ряд має від'ємну автокореляцію;

3) $d_2 < d_в < 4 - d_2$ автокореляція відсутня;

4) $d_1 < d_в < d_2$ або $4 - d_2 < d_в < 4 - d_1$ однозначного висновку немає, необхідні додаткові дослідження.

Завдання до лабораторної роботи

1. Запрограмувати алгоритм аналізу рядів динаміки.
2. Вирішити завдання аналізу рядів динаміки відповідно до номеру варіанту.
3. Сформулювати висновки в термінах предметної області.
4. Оформити звіт до лабораторної роботи

За даними, наведеними в таблиці, виконати аналіз ряду динаміки згідно описаного вище алгоритму, а саме

1) знайти середній рівень, середньоквадратичне відхилення відносно середнього рівня, коефіцієнт варіації;
2) обчислити коефіцієнти автокореляції з лагами $L = 1, 2, 3, 4, 5$ і побудувати корелограму;
3) проаналізувати значущість коефіцієнтів автокореляції.
4) провести вирівнювання часового ряду способом усереднення за трьома і п'ятьма точками;
5) визначити тенденцію ряду експонентним вирівнюванням;
6) скласти за МНК рівняння лінійного тренда;
7) обчислити оцінки дисперсії параметрів моделі та їх вірогідні проміжки;
8) обчислити прогнозовані значення та їх вірогідні проміжки на один і два роки з надійністю $\gamma = 0{,}9$.
9) Оцінити якість лінійної моделі за автокореляцією відхилень від тренда.

Завдання передбачає 30 варіантів В кожному варіанті всі можливі значення Y_i величини Y змінюються на постійну величину, рівну номеру варіанта + дані з першої лабораторної.

Для контролю за деяким процесом здійснювали реєстрацію кількості сигналів через однакові інтервали часу:

T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
У	9	6	3	8	15	5	7	6	13	9	4	10	15	7	6	8

T	17	18	19	20	21	22	23	24	25	26	27	28	29	ЗО	31	32
У	12	2	4	5	11	7	7	8	12	4	6	3	14	10	5	7