

Н.Л. Дорош, Л.М. Петречук

ДОСЛІДЖЕННЯ ЧАСОВОГО РЯДУ СПЕЦІАЛІЗОВАНОЇ ІНФОРМАЦІЇ

Анотація. Розроблена надбудова Excel (XLA), яка надає можливість виконувати спектральний аналіз даних в електронних таблицях. Модулі надбудови надають можливість перегляду даних, виконання експоненціального згладжування, нелінійного перетворення Бокса-Кокса та згладжування даних з використанням алгоритмів цифрової фільтрації.

Ключові слова: надбудова MS Excel, часовий ряд, спектральний аналіз, імпульсна характеристика, циклічні зміни.

Постановка проблеми

Розповсюдженість та зручність табличного процесору Ms Excel давно зробили його одним з найпопулярніших програмних продуктів для виконання різноманітних обчислень. Можливість підключення до MS Excel різноманітних програмних модулів у вигляді надбудов – ще одна важлива його перевага [1]. Надбудова Excel (xla), яка дозволить виконувати графічне представлення даних і спектральної щільності сигналу, а також виведення результатів обробки даних методом цифрової фільтрації на робочий лист – компенсує відсутність функцій для проведення спектрального аналізу у табличному процесорі.

Формулювання цілей статті

Метою даної роботи було створення надбудови, модулі якої дозволяють виконувати згладжування часового ряду, його перетворення за методом Бокса-Кокса та згладжування цифровим фільтром з кінцевою імпульсною характеристикою (КІХ) і аналіз часового ряду.

Об'єктом дослідження є часовий ряд, який обробляється у спроектованій надбудові.

Предметом дослідження виступають методи згладжування даних, які використані у розробленій надбудові.

Основна частина

Нова надбудова забезпечує додаткові функціональні можливості обробки часових рядів у процесорі MS Excel [2] і використовує функції, які перебувають у бібліотечному dll-файлі. У функціях реалізовані такі можливості: перегляд даних часового ряду; експоненціальне згладжування часового ряду; нелінійне перетворення Бокса-Кокса (Box-Cox Transformation); згладжування цифровим фільтром з кінцевою імпульсною характеристикою (КИХ).

Функції реалізовані мовою C++ у середовищі C++Builder. Стандартні файли dll створюються за допомогою компіляторів C++Builder [3] і використовують в проекті внутрішню мову програмування VBA (Visual Basic for Applications). Функції стандартної dll неможливо викликати безпосередньо з робочого аркуша, або з аркуша макросів MS Excel [1]. Для цього побудовані спеціальні інтерфейсні функції. Діаграма компонентів, яка доводить структуру надбудови Excel, надана на рисунку 1.

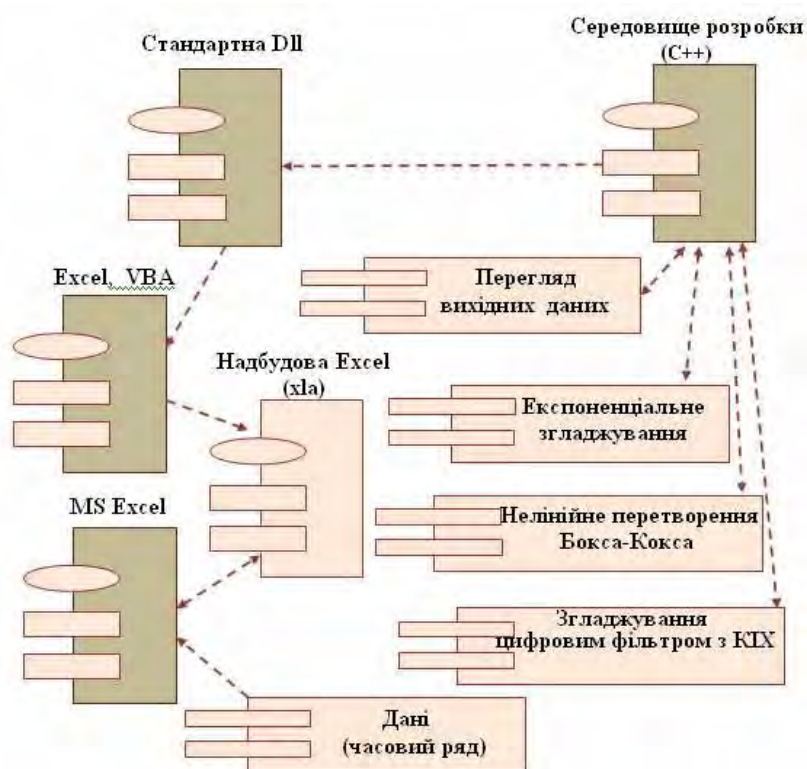


Рисунок 1 – Діаграма компонентів для надбудови Excel

Експоненціальне згладжування сигналу в модулі надбудови здійснюється за рекурентною формулою:

$$S(t) = \alpha \cdot x(t) + (1 - \alpha) \cdot S(t - 1) \quad (1)$$

де α – параметр згладжування; $S(t)$ – згладжене значення часового ряду

$$X = \{x_1, x_2, x_3 \dots x_t\}.$$

Значення параметру згладжування α знаходиться у межах $0 \leq \alpha \leq 1$, рекомендовано обирати значення α від 0,35 до 1.

Нелінійне перетворення сигналу по методу Бокса–Кокса

Згладжування часового ряду $X = \{x_1, x_2, x_3 \dots x_T\}$ здійснюється за формулою:

$$S(t) = \begin{cases} \frac{X(t)^\lambda - 1}{\lambda} & \text{якщо } \lambda \neq 0 \\ \ln(X(t)) & \text{якщо } \lambda = 0 \end{cases} \quad (2)$$

де $S(t)$ – згладжене значення, λ - параметр.

Одним із способів визначення оптимальної величини параметру лямбда є максимізація логарифма функції правдоподібності (функція максимальної лог-правдоподібності):

$$f(x, \lambda) = -\frac{N}{2} \ln \left[\sum_{i=0}^{N-1} \frac{(x_i(\lambda) - \bar{x}(\lambda))^2}{N} \right] + (\lambda - 1) \sum_{i=0}^{N-1} \ln(x_i) \quad (3)$$

де N – кількість спостережень;

x – спостереження часового ряду;

\bar{x} - середнє значення спостережень часового ряду.

$$\bar{x}(\lambda) = \frac{1}{N} \sum_{i=0}^{N-1} x_i(\lambda) \quad (4)$$

Тобто необхідно вибрати таке значення параметра лямбда, при якому дана функція приймає максимальне значення.

У розробленій надбудові згладжування шляхом лінійної фільтрації вихідних даних цифровим фільтром з кінцевою імпульсною характеристикою реалізовано декількома варіантами: усереднення по n точках ($n=3-15$); метод параболи крізь декілька рівновіддалених значень сигналу, з подальшим використанням у якості згладженої величини значення параболи у середній точці ($n=5, 7, 9, 11, 13$); усереднення Спенсера по 15 та 21 точках [4].

Згладжування сигналу по трьох точках виконувалося згідно формули:

$$Y(n) = (X(n-2) + X(n-1) + X(n) + X(n+1) + X(n+2)) / 3 \quad (5)$$

Згладжування методом Спенсера по 21-й точці здійснювалося згідно формули:

$$x_i = \frac{1}{320} (-3x_{i-7} - 6x_{i-6} - 5x_{i-5} + 3x_{i-4} + 21x_{i-3} + 46x_{i-2} + 67x_{i-1} + 74x_i + 67x_{i+1} + 46x_{i+2} + 21x_{i+3} + 3x_{i+4} - 5x_{i+5} - 6x_{i+6} - 3x_{i+7}) \quad (6)$$

Спектральний аналіз сигналу призначений для оцінки частотного складу дискретного сигналу [5]. Непараметричні методи засновані на обчисленні оцінок спектральної щільності потужності (СЩП) безпосередньо за відліком вихідної послідовності, що визначає їх основні переваги – можливість застосування до широкого класу стаціонарних сигналів і шумів і високу обчислювальну ефективність за рахунок застосування алгоритмів швидкого перетворення Фур'є (ШПФ).

Для оцінки спектральної щільності потужності застосовують метод періодограм [6], який полягає в обчисленні оцінки СЩП кінцевої випадкової послідовності довжини N , яка називається періодограмою.

$$\hat{S}(\omega) = \frac{|X(e^{j\omega T})|^2}{Nf_D} \quad (7)$$

де $X(e^{j\omega T})$ – спектральна щільність кінцевої послідовності $x(n)$.

$$X(e^{j\omega T}) = \sum_{n=0}^{N-1} x(n)e^{-j\omega Tn} \quad (8)$$

Для зменшення ефекту розтікання спектру при обчисленні періодограми за допомогою дискретного перетворення Фур'є (ДПФ) і, як наслідок, згладжування періодограми, застосовують вагові функції (вікна), і видозмінена періодограма, звана модифікованою періодограмою, набирає вигляду:

$$\hat{S}_w(\omega) = \frac{\frac{1}{f_D} |X_w(e^{j\omega T})|^2}{\sum_{n=0}^{N-1} |w(n)|^2} \quad (9)$$

де $w(n)$ – вагова функція (вікно) довжини N ;

$X_w(e^{j\omega T})$ – спектральна щільність множення $x(n) \cdot w(n)$:

$$X_w(e^{j\omega T}) = \sum_{n=0}^{N-1} x(n)w(n)e^{-j\omega Tn} \quad (10)$$

Віконна функція ШПФ використовується для придушення крайових ефектів розривності реальних функцій шляхом введення вагових коефіцієнтів для вибірки даних у вікні, які забезпечують зни-

ження амплітуд крайових точок (старту і стопа) і, в результаті, поліпшення результатів ШПФ. Різні види віконних функцій (прямокутне вікно, Хеммінга, Ханнінга, Блекмана-Харріса) дають різні результати, як по точності, так і по частотному вирішенню і використовуються для різних видів сигналів [7].

В роботі застосовано вікно Хеммінга.

Нехай $X = \{x_1, x_2, x_3 \dots x_N\}$ – часовий ряд, який надано, $Y = \{y_1, y_2, y_3 \dots y_N\}$ – часовий ряд (сигнал) після згладжування.

Згортка сигналу має вигляд:

$$y(n) = \sum_{m=0}^{\infty} h(n-m)x(m) = \sum_{m=0}^{\infty} h(m)x(n-m) \quad (11)$$

де m - затримка послідовності, $h(n)$ – імпульсна характеристика (ІХ).

$H(z)$ — z -зображення ІХ $h(n)$ – обчислюється за формулою:

$$H(z) = \sum_{n=0}^{\infty} h(n)z^{-n} \quad (12)$$

Передавальна функція $H(z)$ дорівнює:

$$H(z) = \frac{Y(z)}{X(z)} \quad (13)$$

де $X(z)$, $Y(z)$, – z - зображення сигналу до- і після обробки.

Нехай $H\left(e^{j\hat{\omega}}\right)$ - Фур'є - зображення ІХ $h(n)$, де $\hat{\omega}$ — нормована частота.

Для частотної характеристики $H\left(e^{j\hat{\omega}}\right)$ і передавальної функції $H(z)$ вірно співвідношення:

$$H\left(e^{j\hat{\omega}}\right) = H(z) \Big|_{z=e^{j\hat{\omega}}} \quad (14)$$

У створеній надбудові проведено аналіз часового ряду, що відображає кількість громадян, які вперше звернулися до лікувально-профілактичних закладів України зі скаргами на органи травлення. Вихідні дані – щомісячна кількість пацієнтів, які звернулися за проміжком з січня 2006 року по грудень 2015 року.

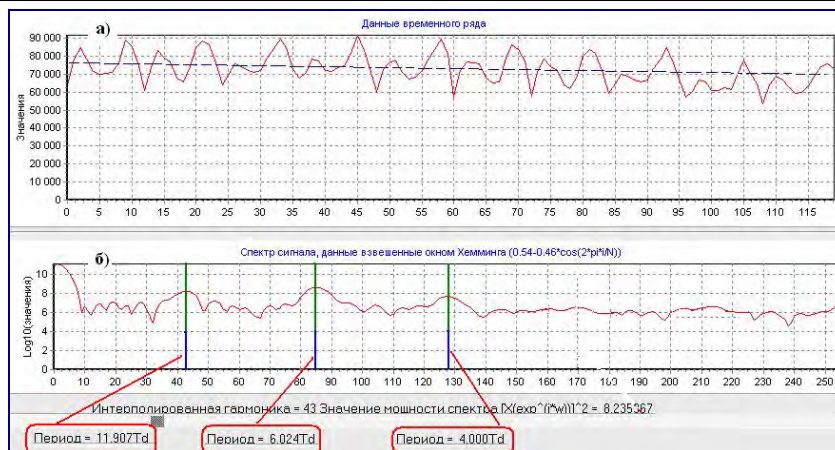


Рисунок 2 – Часовий ряд, що відображає показники за 10 років

Графічне зображення досліджуваного часового ряду наведено на рисунку 2 і дає первинне уявлення про зміну статистичних показників (кількості хворих) за минулі 10 років. Візуально на графіку можна виділити:

- довготривалу тенденцію – пунктирна лінія йде поступово вниз (рисунок 2-а);
- циклічні зміни, які зазвичай пов'язані з коливаннями фізіологічної активності організму людини (рисунок 2-а);
- сезонні коливання (рисунок 2-б);
- нерегулярні коливання, пов'язані з непередбачуваними випадковими подіями (рисунок 2-б).

На рисунку 2-б представлений спектр сигналу (згладжений вікном Хеммінга), що має гармоніки, яким відповідають значення періодів 12, 6 і 4. Виходячи з того, що часовий ряд, який аналізується, складається з щомісячних показань (тобто має крок дискретизації за часом $T_d = 1$), цифри вказують на те, що звернення громадян до медучастин має певну закономірність, яка має три різних за часом періоди: річний (12 місяців); піврічний (6 місяців) і посезонний (зима, весна, літо, осінь – 4 місяці).

Графічне зображення ряду даних після обробки методом експоненціального згладжування (у надбудові) при різних значеннях параметру α надано на рисунку 3. Результати згладжування ряду даних у вигляді фрагменту таблиці Excel та графіків залишків надані на рисунку 4.

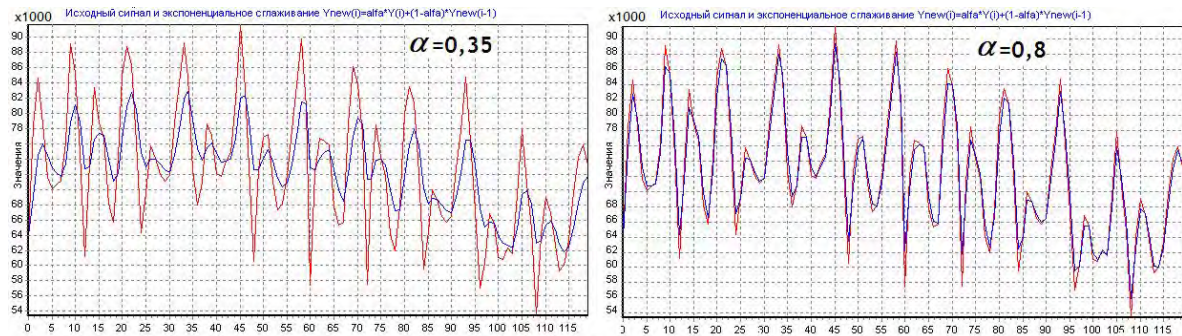


Рисунок 3 – Метод экспоненціального згладжування

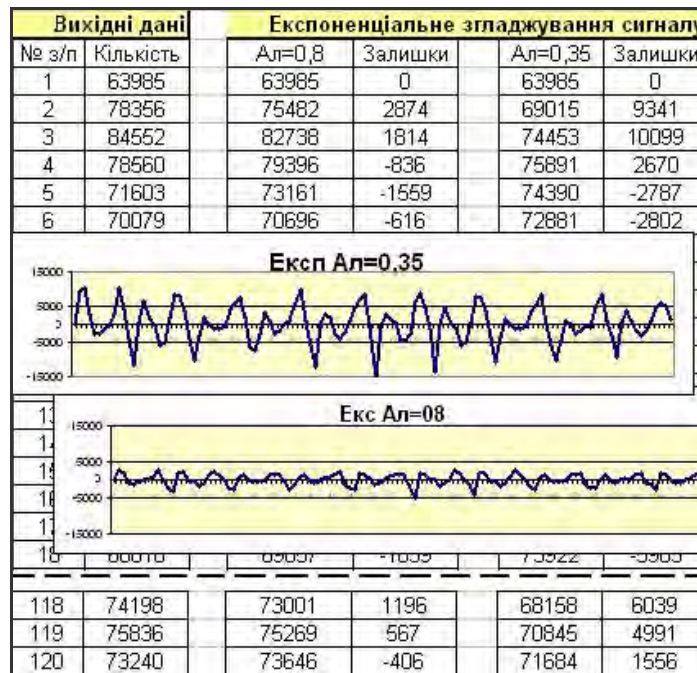


Рисунок 4 – Результати експоненціального згладжування на листі Excel

Достовірність будь-яких результатів виконується шляхом порівняння. Аналогічну обробку даних, щодо звернення громадян у медичні установи, було здійснено у додатку ITSM2000 [8].

На рисунку 5 показані види екранів додатку ITSM2000 та надбудови Excel для алгоритму нелінійного перетворення Бокса-Кокса. Результати у графічному вигляді схожі. Порівняльний аналіз по рисункам є якісним.

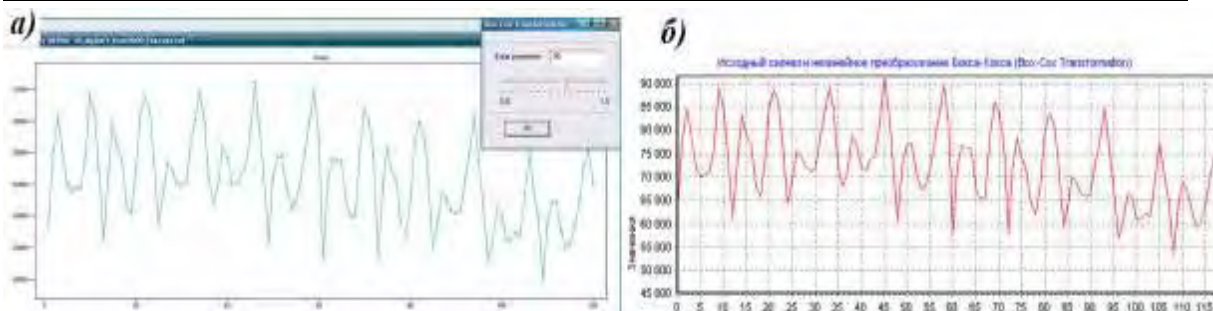


Рисунок 5 – Нелінійне перетворення Бокса-Кокса,
а)- в ITSM 2000, б)- в надбудові

На рисунку 6 зображена періодограма, яка зроблена у програмі в ITSM 2000. Пік в області деякої частоти ω_0 вказує на те, що в спектральному розкладанні автокореляційної функції присутня відповідна гармонійна компонента. Чим вище і різкіше виділено пік, тим більша частина потужності зосереджена близько частоти ω_0 і тим більшу роль відіграє ця частота в описі відповідного випадкового процесу [9].

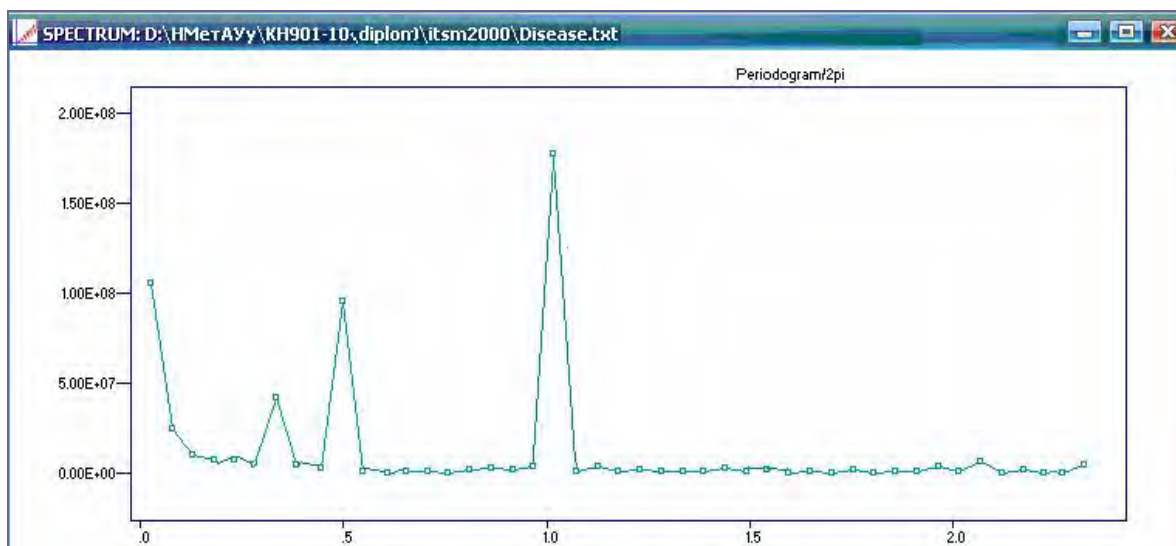


Рисунок 6 – Періодограма в ITSM 2000

При розгляді часових рядів за одиницю часу береться один місяць (одне спостереження за місяць). На періодограмі (рисунок 4.36) по осі X аргумент відповідає наступним значенням: 0,25 – 3 місяці, 0,5 – 6 місяців, 1,0 – 12 місяців. Наявність піків в області значень $\approx 0,33$, 0,5 і 1,0 свідчить про існування в досліджуваному ряду трьох циклів:

- 4-х місячного ($0,33 \cdot 12$ місяців = 4 місяця);
- піврічного циклу ($0,5 \cdot 12$ місяців = 6 місяців);

– річного циклу ($1,0 \cdot 12$ місяців = 12 місяців).

Спектр сигналу, отриманий після обробки часового ряду у надбудові (рисунок 7), показує наявність трьох явно виражених циклів з періодами 4, 6 та 12 місяців. Отже, спектральний аналіз у надбудові і в програмі ITSM 2000 тотожні.



Рисунок 7 – Спектр сигналу в «Надбудові»

Кількісний аналіз порівняння результатів згладжування, реалізованих у створеній надбудові алгоритмів, представлений на рисунку 8.

A	B	C	D	E	F	G	H
№ з/п	Сигнал, disease	Надбудова	ITSM 2000	Δ	Надбудова	ITSM 2000	Δ
		Експон. згладжування, α=0.8			Перетворення Бокса-Кокса, λ=0.84		
1	63985	63985,39	63985,39	0,00	12965,48	12965,48	0,00E+00
2	78356	75482,10	75482,10	0,00	15371,25	15371,25	0,00E+00
3	84552	82738,12	82738,12	0,00	16386,03	16386,03	0,00E+00
4	78560	79395,84	79395,84	0,00	15404,86	15404,86	0,00E+00
5	71603	73161,33	73161,33	0,00	14250,34	14250,34	1,20E-13
6	70079	70695,65	70695,65	0,00	13995,19	13995,19	0,00E+00
7	70645	70655,10	70655,10	0,00	14090,04	14090,04	1,70E-13
8	71039	70962,19	70962,19	0,00	14156,03	14156,03	0,00E+00
9	76173	75130,91	75130,91	0,00	15010,65	15010,65	0,00E+00
10	89123	86324,19	86324,19	0,00	17126,95	17126,95	3,80E-13
11	85314	85515,92	85515,92	0,00	16509,95	16509,95	0,00E+00
12	74650	76822,88	76822,88	0,00	14758,05	14758,05	0,00E+00
116	63324	62671,74	62671,74	0,00	12852,86	12852,86	0,00E+00
117	69601	68215,43	68215,43	0,00	13914,97	13914,97	4,80E-13
118	74198	73001,19	73001,19	0,00	14682,94	14682,94	0,00E+00
119	75836	75269,15	75269,15	0,00	14954,85	14954,85	0,00E+00
120	73240	73645,72	73645,72	0,00	14523,56	14523,56	0,00E+00

Рисунок 8 – Результати порівняльного аналізу функцій надбудови та пакету ITSM 2000

Висновки

Виконано розробку надбудови, яка може бути використана для обробки часових рядів в електронних таблицях MS Excel. Надбудова дозволяє виконати згладжування часового ряду та провести спектра-

льний аналіз даних. Передбачено виведення отриманих результатів в лист Ms Excel.

Проведено дослідження часового ряду кількості пацієнтів з вадами органів травлення. Спектральний аналіз виявив наявність циклічності в досліджуваному ряді, про що кількісно свідчать такі показники:

- період=4Td вказує на наявність чотиримісячного циклу;
- період=6Td вказує на наявність шестимісячного циклу;
- період=12Td вказує на те, що характер кривої, яка відображає річну кількість громадян, що звернулися до медустанов, тотожний для всіх десяти інтервалів. Отримані результати підтверджуються експертними оцінками.

Підтвердження функціональності створеної надбудови в Excel було проведено шляхом порівняння результатів, отриманих при аналогічній обробці досліджуваного ряду в програмі ITSM 2000. Порівняння результатів показало їх ідентичність. Доведено, що модулі надбудови працюють якісно.

ЛІТЕРАТУРА

1. Гайдышев И.П. Решение научных и инженерных задач средствами Excel, VBA и C/C++ / И.П. Гайдышев – СПб.: БХВ – Петербург, 2004. – 512 с.
2. Дорош Н. Л., Храпач Ю. О. Результаты разработки надбудовы MS Excel для обработки часовых рядов // Системні технології. Регіональний міжвузівський збірник наукових праць. – Випуск 2 (97). – Дніпропетровськ, 2015. – С. 8-12.
3. Архангельский А.Я. Программирование в C++Builder 6 / А.Я. Архангельский – М.: издательство «Бином», 2008. – 1152 с.
4. Отнес Р. Прикладной анализ временных рядов. Основные методы / Р. Отнес, Л. Эноксон. Пер. с англ. под ред. В. И. Хохлова. – М.: Мир, 1982. – 428 с.
5. Дженкинс Г. Спектральный анализ и его приложения. Выпуск 1 / Г. Дженкинс, Д. Ваттс – М.: Мир, 1971. – 462 с.
6. Солонина А.И. Цифровая обработка сигналов и MATLAB: учеб. пособие / А. И. Солонина, Д. М. Клионский, Т. В. Меркучева, С. Н. Перов. – СПб.: БХВ-Петербург, 2013. – 512 с.
7. Лайонс Р. Цифровая обработка сигналов / Р.Лайонс – М.:Бином, 2006. – 656 с.
8. ITSM 2000 Professional Version 6.0, developed by Peter J. Brockwell and Richard A. Davis, B&D Enterprises, Inc., Copyright 1999.
9. Хемминг Р.В. Цифровые фильтры. / Р.В. Хемминг, пер. с англ. под ред. А.М. Трахтмана. – М.: Сов. радио, 1980.- 224 с.