

ТЕХНІЧНЕ ЗАБЕЗПЕЧЕННЯ ДІЯЛЬНОСТІ ПРАВООХОРОННИХ ОРГАНІВ

УДК 004.942:343.92

А.С. БАБІЙ,
М.М. ЗАЦЕРКЛЯНИЙ, докт. техн. наук, проф.

Національний університет внутрішніх справ

АВТОМАТИЗАЦІЯ АНАЛІЗУ СЕЗОННИХ КОЛИВАНЬ РІВНЯ ЗЛОЧИННОСТІ

Розглянута методика виявлення сезонних коливань злочинності як складової аналізу кримінальних проявів.

Основним об'єктом аналізу та оцінки в органах внутрішніх справ є кримінальна обстановка, що складається на території, яка обслуговується даним органом. Її аналіз допускає вивчення різноманітної інформації, що характеризує стан правопорушень в регіоні і результати боротьби з ними, сили і засоби органа внутрішніх справ, фактори зовнішнього середовища, що впливають як на правопорушення, так і на діяльність самого органа. Інформація, яка використовується з цією метою, різноманітна як за змістом, так і за формою. Проте, провідна роль при аналізі кримінальної обстановки належить статистичній інформації, оскільки тільки вона дозволяє встановити кількісні закономірності явищ і процесів, з якими доводиться мати справу правоохоронним органам. А пізнання кількісних закономірностей у цій сфері надає більш надійну основу для управління правоохоронною діяльністю, тому важливу роль при здійсненні аналізу кримінальної обстановки відіграють різноманітні методи опрацювання статистичної інформації.

Статистична інформація, яка спостерігається в правоохоронній діяльності, відображається у вигляді динамічного ряду послідовно розташованих у хронологічному порядку значень того чи іншого показника, а отже, аналіз кримінальної обстановки зводиться до аналізу динамічних рядів. Повноцінний аналіз динамічного ряду можна виконати тільки на основі математичних методів із використанням засобів обчислювальної техніки.

На теренах СРСР у системі органів внутрішніх справ особливо активно проблемами аналізу злочинності з використанням математичних методів стали займатися з кінця 60-х років минулого століття [1]. Для їх розв'язування у Штабі МВС СРСР був створений спеціальний відділ. Саме він на основі аналізу динаміки злочинності підготував перший в СРСР кримінологічний прогноз на п'ятирічний період, а на його основі – й перспективний план вдосконалення діяльності органів внутрішніх

справ на 1971-1975 р. Але трохи далі період бурхливого розвитку математичних методів аналізу злочинності змінився спадом.

Разом із тим, сучасний стан криміногенної ситуації в Україні потребує не тільки використання адміністративних методів впливу на рівень злочинності й викорінення корупційних діянь у правоохоронних органах, вдосконалення системи оцінок якості роботи підрозділів, але й використання новітніх досягнень у галузях математичного моделювання, інформаційних технологій, засобів обчислювальної та комунікаційної техніки.

Дана робота спрямована на розробку методики виявлення сезонних коливань злочинності, що є складовою частиною аналізу кримінальних явищ. Сезонними називаються коливання, які мають строго періодичний або близький до нього характер і завершуються упродовж одного року [2].

Розглянемо динамічний ряд злочинності $\{Y_t\}, t = \overline{1, T}$, породжуваний аддитивним випадковим процесом:

$$Y_t = U_t + V_t + \varepsilon_t, t = \overline{1, T}, \quad (1)$$

де U_t – трендова складова, яка характеризує загальну тенденцію розвитку динамічного ряду; V_t – сезонна компонента; ε_t – випадкова компонента (завжди присутня в динамічному ряді); T – кількість спостережень (для виявлення сезонності динамічного ряду злочинності рекомендується використовувати місячні дані, чи принаймні кварталні).

Одним із методів виявлення сезонності, придатним для аналізу динамічних рядів злочинності, є ітераційний метод Четверикова, сутність якого полягає в такому:

1. Спочатку динамічний ряд $\{Y_t\}$ вирівнюється ковзною середньою із періодом ковзання T_0 , тобто береться $(T_0 + 1)$ членів початкового ряду, із яких

перший і останній беруться з половинною вагою: $\alpha_{-T_0/2} = \alpha_{T_0/2} = 1/2$ членів ряду, які випадають із обох його кінців, або поновлюються екстраполяцією вирівняного ряду, або в наступних стадіях робіт залишаються осторонь (рекомендуємо втрачені при згладжуванні рівні ряду не поновлювати):

$$Y_t = \frac{Y_{t+T_0/2} + Y_{t-T_0/2+1} + \dots + Y_t + \dots + Y_{t+T_2/2-1} + Y_{t+T_0/2}}{T_0} \quad (2)$$

У результаті цієї операції дістається попередня оцінка тренда

$$Y'_t = U'_t$$

і відхилення початкового динамічного ряду від вирівняного

$$l'_{ij} = l_t = Y_t - Y'_t, t = \overline{1, T},$$

або в розгорнутому вигляді:

$$l'_{ij} = Y_{ij} - Y'_{jt}, (i = \overline{1, m}; j = \overline{1, T_0}). \quad (3)$$

2. Для кожного року і обчислюється σ_i - середнє квадратичне відхилення, на яке потім діляться окремі місячні (квартальні) відхилення відповідного року:

$$l_{ij} = \frac{l'_{ij}}{\sigma_i}, \quad (4)$$

де

$$\sigma_i = \left[\frac{\sum_{j=1}^{T_0} l_{ij}^2 - (\sum_{j=1}^{T_0} l_{ij})^2 / T_0}{T_0 - 1} \right]^{1/2}. \quad (5)$$

3. З "нормованих" таким чином відхилень обчислюється в першому наближенні середня сезонна хвиля:

$$V_j^1 = \frac{\sum_{i=1}^m l_{ij}}{m}. \quad (6)$$

4. Середня сезонна хвиля множиться на середнє квадратичне відхилення кожного року і віднімається від рівнів початкового динамічного ряду:

$$U_{ij}^1 = Y_{ij} - V_j^1 \sigma_i. \quad (7)$$

5. Так, утворений ряд позбавлений сезонної компоненти. Він знову згладжується ковзною середньою (для місячних даних по п'яти або семи точках у залежності від інтенсивності дрібних коливань і тривалості більш значних), тобто виконується друга ітерація. В результаті одержується нова оцінка тренда $U_t^{(2)}$.

6. Відхилення емпіричного ряду Y_t від ряду $U_t^{(2)}$, одержаного в п.5 $l_t^{(2)} = Y_t - U_t^{(2)}$, (8) знову піддаються аналогічному опрацюванню згідно пунктів 2 і 3 для виявлення наступного наближення середньої сезонної хвилі.

7. Вилучення остаточної сезонної компоненти із динамічного ряду злочинності проводиться після множення середньої сезонної хвилі на k_i - коефіцієнт напруженості сезонної хвилі:

$$k_i = \frac{\sum_{j=1}^{T_0} l_{ij}^{(2)} \varepsilon_{ij}}{\sum_{j=1}^{T_0} \varepsilon_{ij}^2}, \quad (9)$$

де $l_{ij}^{(2)}$ - вирівняні значення ряду; ε_{ij} - випадкова компонента:

$$\varepsilon_{ij} = l_{ij}^{(2)} - V_j^{(2)}.$$

У разі потреби ітераційний процес продовжується.

Таким чином, із динамічного ряду вилучаються сезонна і трендова компоненти. Залишкова послідовність, тобто, випадкова компонента використовується для перевірки адекватності і точності проведеного аналізу.

Перевірка адекватності ґрунтується на перевірці виконання у залишковій послідовності таких чотирьох властивостей:

- випадковості коливань рівнів;
- відповідності нормальному закону розподілу;
- рівності математичного сподівання нулю;
- незалежності значень рівнів, тобто відсутності істотної автокореляції.

Якщо не виконується принаймні одна з цих умов, говорити про адекватність немає підстав. Ця перевірка здійснюється з використанням ряду статистичних критеріїв.

Одним із критеріїв перевірки випадковості коливань рівнів залишкової послідовності є критерій піків (поворотних точок). Рівень випадкової компоненти (послідовності) ε_i вважається максимумом, якщо він більший від двох поруч розміщених рівнів, тобто $\varepsilon_{i-1} < \varepsilon_i > \varepsilon_{i+1}$, і мінімумом, якщо він менший від обох сусідніх рівнів, тобто $\varepsilon_{i-1} > \varepsilon_i < \varepsilon_{i+1}$. В обох випадках ε_i вважається поворотною точкою; загальну кількість поворотних точок для залишкової послідовності ε_i позначимо через \bar{p} .

Для випадкової вибірки математичне очікування числа точок повороту \bar{p} і дисперсія σ_p^2 відшукуються за формулами:

$$\bar{p} = \frac{2}{3}(T - 2); \quad \sigma_p^2 = \frac{16T - 29}{90}. \quad (10)$$

Критерієм випадковості з 5 %-м рівнем значимості, тобто, з довірчою ймовірністю 95 %, є нерівність

$$p > \left[\bar{p} - 1,96 \sqrt{\sigma_p^2} \right], \quad (11)$$

де квадратні дужки означають цілу частину числа. Якщо ця нерівність не виконується, говорити про адекватність не доводиться.

Перевірка відповідності розподілу випадкової компоненти нормальному закону виконується лише наближено за допомогою дослідження показників асиметрії (γ_1) і ексцесу (γ_2), оскільки часові ряди, як

правило, не достатньо великі. При нормальному розподілі показники асиметрії та ексцесу деякої генеральної сукупності дорівнюють нулю. Вважаємо, що відхилення від тренда є вибіркою з генеральної сукупності, тому можна знайти тільки вибіркові характеристики асиметрії й ексцесу та їх похибки:

$$\bar{\gamma}_1 = \frac{\frac{1}{T} \sum_{i=1}^T \varepsilon_i^2}{\sqrt{\left(\frac{1}{T} \sum_{i=1}^T \varepsilon_i^2\right)^3}}; \quad \sigma_{\bar{\gamma}_1} = \sqrt{\frac{6(T-2)}{(T+1)(T+3)}};$$

$$\bar{\gamma}_2 = \frac{\frac{1}{T} \sum_{i=1}^T \varepsilon_i^4}{\left(\frac{1}{T} \sum_{i=1}^T \varepsilon_i^2\right)^2} - 3; \quad \sigma_{\bar{\gamma}_2} = \sqrt{\frac{24T(T-2)(T-3)}{(T+1)^2(T+3)(T+5)}} \quad (12)$$

У цих формулах $\bar{\gamma}_1$ - вибіркова характеристика асиметрії; $\bar{\gamma}_2$ - вибіркова характеристика ексцесу; $\sigma_{\bar{\gamma}_1}$ і $\sigma_{\bar{\gamma}_2}$ - відповідні середні квадратичні відхилення (похибки).

Якщо одночасно виконуються такі нерівності:

$$|\bar{\gamma}_1| < 1,5 < \sigma_{\bar{\gamma}_1}; \quad \left| \bar{\gamma}_2 + \frac{6}{T+1} \right| < 1,5\sqrt{2} \sigma_{\bar{\gamma}_2}, \quad (13)$$

то гіпотеза про нормальний характер розподілу випадкової компоненти приймається.

Якщо виконується принаймні одна нерівність

$$|\bar{\gamma}_1| \geq 2 \sigma_{\bar{\gamma}_1}; \quad \left| \bar{\gamma}_2 + \frac{6}{T+1} \right| < 2 \sigma_{\bar{\gamma}_2}, \quad (14)$$

то говорити про адекватність немає підстав. Інші випадки вимагають додаткової перевірки за допомогою більш складних критеріїв [3].

Перевірка рівності математичного сподівання випадкової компоненти нулю, якщо вона розподілена за нормальним законом, здійснюється на основі t-критерію Стюдента. Розрахункове значення цього критерію знаходиться за формулою:

$$t = \frac{\bar{\varepsilon} - 0}{S_{\varepsilon}} \sqrt{T}, \quad (15)$$

де $\bar{\varepsilon}$ - середнє арифметичне значення рівнів залишкової послідовності ε ; S_{ε} - стандартне (середнє квадратичне) відхилення для цієї послідовності.

Якщо розрахункове значення t менше табличного значення t_{α} статистики Стюдента із заданим рівнем значимості α і числом степенів свободи $T-1$, то гіпотеза про рівність нулю математичного очікування випадкової послідовності приймається; у противному разі ця гіпотеза відхиляється і модель вважається неадекватною.

Перевірка незалежності значень рівнів випадкової компоненти, тобто перевірка відсутності істотної автокореляції в залишковій послідовності може здійснюватися за рядом критеріїв, зокрема за d-

критерієм Дарбіна-Уотсона [4]. Його розрахункове значення визначається за формулою

$$d = \frac{\sum_{i=2}^T (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^T \varepsilon_i^2}. \quad (16)$$

Зауважимо, що розрахункове значення критерію Дарбіна-Уотсона в інтервалі від 2 до 4 свідчить про від'ємний зв'язок; у цьому випадку його треба перетворити за формулою $d' = 4 - d$ і надалі користуватися значенням d' .

Розрахункове значення критерію d (або d') порівнюється з верхнім d_2 і нижнім d_1 критичними значеннями статистики Дарбіна-Уотсона. Якщо розрахункове значення критерію d більше верхнього табличного значення d_2 , то гіпотеза про незалежність рівнів залишкової послідовності, тобто, про відсутність у ній автокореляції, приймається. Якщо значення d менше від нижнього табличного значення d_1 , то ця гіпотеза відхиляється і модель вважається неадекватною. Якщо значення d міститься між значеннями d_1 і d_2 , включаючи і ці значення, то вважається, що немає достатніх підстав зробити той чи інший висновок, і потрібні подальші дослідження.

Наявність автокореляції є сигналом, що трендову компоненту вилучено не повністю [5]. Існування автокореляції може також свідчити про систематичні похибки вимірювання, або, що більш ймовірно в соціальних процесах, існування неврахованих факторів і зв'язків, що потребують дослідження.

Статистична інформація для досліджень сезонних коливань отримується із спеціалізованих баз даних за допомогою запитів (зокрема, з інтегрованої бази даних, яка зараз впроваджується в діяльність підрозділів ОВС і яка побудована з використанням СУБД «Oracle»). Сам метод реалізований стандартними засобами табличного редактора Microsoft Excel і мови Visual Basic for Application (рис. 1).

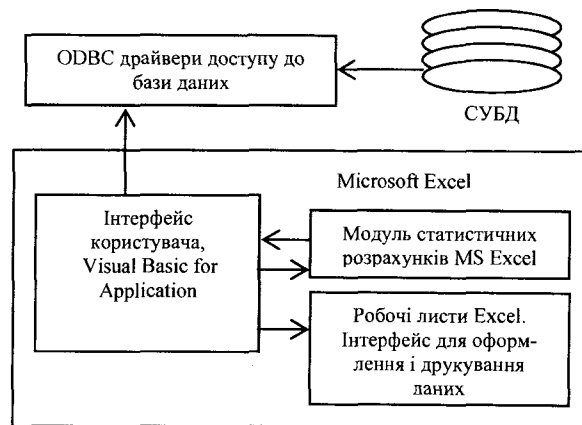


Рисунок 1 - Схема програмних компонентів системи

Розглянемо динамічний ряд кількості зареєстрованих

ДТП за період з 1996 по 2003 рік (у табл. 1 дані приведені в розрахунку на 10 тис. населення).

Таблиця 1 – Динаміка зареєстрованих ДТП за 1996-2003 рр.

| Рік | Місяць | | | | | | | | | | | |
|------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 |
| 1996 | 0,176 | 0,172 | 0,152 | 0,231 | 0,338 | 0,341 | 0,472 | 0,490 | 0,486 | 0,507 | 0,403 | 0,286 |
| 1997 | 0,197 | 0,241 | 0,214 | 0,234 | 0,259 | 0,293 | 0,314 | 0,434 | 0,462 | 0,507 | 0,376 | 0,179 |
| 1998 | 0,186 | 0,166 | 0,231 | 0,231 | 0,272 | 0,334 | 0,362 | 0,386 | 0,476 | 0,517 | 0,334 | 0,252 |
| 1999 | 0,162 | 0,197 | 0,241 | 0,303 | 0,359 | 0,366 | 0,307 | 0,417 | 0,355 | 0,455 | 0,324 | 0,266 |
| 2000 | 0,221 | 0,152 | 0,200 | 0,248 | 0,224 | 0,262 | 0,317 | 0,355 | 0,355 | 0,338 | 0,421 | 0,303 |
| 2001 | 0,186 | 0,172 | 0,186 | 0,217 | 0,252 | 0,290 | 0,334 | 0,324 | 0,334 | 0,314 | 0,472 | 0,186 |
| 2002 | 0,124 | 0,190 | 0,221 | 0,193 | 0,203 | 0,207 | 0,238 | 0,276 | 0,238 | 0,424 | 0,400 | 0,245 |
| 2003 | 0,162 | 0,083 | 0,134 | 0,200 | 0,293 | 0,234 | 0,283 | 0,417 | 0,390 | 0,472 | 0,483 | 0,345 |

При використанні розглянутої методики ми виявили в цьому ряду сезонну компоненту (графік рис.2), аналізуючи яку можна зробити висновок, що найменша кількість ДТП трапляється в лютому, а найбільша в жовтні. Це свідчить, що в більшості випадків ДТП скоєні з вини водія, який порушив правила дорожнього руху. Загальновідомим є факт, що взимку значно зменшується кількість водіїв на дорогах, що пояснює значне зменшення кількості ДТП. Восени, окрім збільшення кількості автомобілів, підвищується кількість ДТП, учасниками яких є неповнолітні, що пов'язано з початком занять у школах.

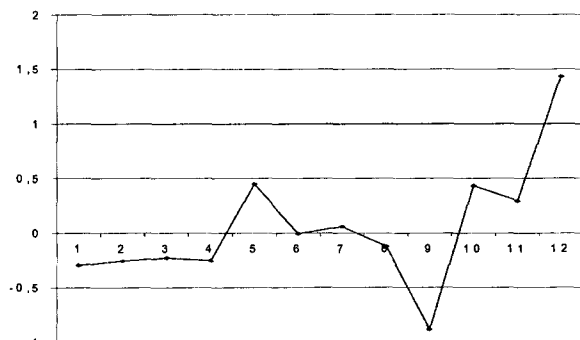


Рисунок 2 - Сезонна компонента динамічного ряду зареєстрованих ДТП

У деяких динамічних рядах сезонна компонента відсутня, хоча існують коливання, які повторюються з року в рік, але не вказують на її існування. Це підтверджено аналогічними розрахунками для виділення сезонної компо-

ненти по іншому виду злочинів - пограбуванням (рис.3). Для визначення чинників, що викликають ці коливання, потрібні додаткові дослідження.

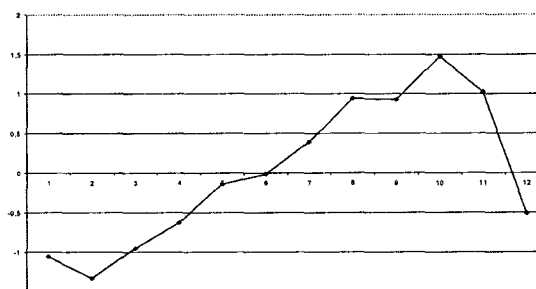


Рисунок 3 - Сезонна компонента зареєстрованих пограбувань

ЛІТЕРАТУРА

1. Вицин С.И. Моделирование в криминологии. -М.: Высш. шк. МВД СССР, 1973. -230 с.
2. Иншаков С.М. Криминология: Учебник. -М.: Юриспруденция, 2000. -426 с.
3. Арнольд В.И. "Жесткие" и "мягкие" математические модели. -М.: МЦНМО, 2000. -32 с.
4. Толстова Ю.Н. Анализ социологических данных. -М.: Научный мир, 2000. -350 с.
5. Робертс Ф.С. Дискретные математические модели с приложениями к социальным и экологическим задачам. -М.: Наука, 1986. -494 с.

Надійшла до редколегії 29.04.2005

БАБИЙ А.С., ЗАЦЕРКЛЯНИЙ Н.М. АВТОМАТИЗАЦІЯ АНАЛІЗА СЕЗОННИХ КОЛЕБАНИЙ УРОВНЯ ПРЕСТУПНОСТІ
Рассмотрена методика определения сезонных колебаний преступности как составной анализа уголовных проявлений.

BABY A.S., ZATSERKLJANYJ N.M. AUTOMATION OF THE ANALYSIS OF SEASONAL FLUCTUATIONS OF A CRIME RATE
The technique of definition of seasonal fluctuations of criminality as compound the analysis of criminal displays is considered.