

гибкость реконфигурируемых устройств обуславливают перспективность их применения для данных целей. Способность ПЛИС быстро изменять внутреннюю структуру позволят создавать многофункциональные системы, ориентированные на решение широкого круга задач информационной безопасности локальных компьютерных сетей.

1. *Коростиль Ю.М., Давиденко А.Н., Гильгурт С.Я.* Анализ угроз и опасностей в компьютерных системах на предмет защиты цифровыми реконфигурируемыми устройствами // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ НАН України. – Київ, 2010. – Вип. 54. – С. 9–16.
2. *Гильгурт С.Я.* Обзор современных реконфигурируемых унифицированных вычислителей // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ НАН України. – Вип. 49. – Київ: 2008. – С. 17–24.
3. *Касперский Е.В.* Компьютерное зловредство. – СПб.: Питер, 2007. – 208 с.
4. *Гильгурт С.Я.* Особенности применения реконфигурируемых вычислителей для аппаратной защиты информационных систем // Зб. наук. пр. ІПМЕ НАН України. – Вип. 38. – Київ: 2007. – С. 36–41.
5. *Зима В.М., Молдовян А.А., Молдовян Н.А.* Безопасность глобальных сетевых технологий. – 2-е изд. – СПб.: БХВ-Петербург, 2003. – 368 с.
6. *Конеев И.Р., Беляев А.В.* Информационная безопасность предприятия. – СПб.: БХВ-Петербург, 2003. – 752 с.
7. *Столингс В.* Основы защиты сетей. Приложения и стандарты: Пер с англ. – М.: Изд. дом "Вильямс", 2002. – 432 с.
8. *Щеглов А.Ю.* Защита компьютерной информации от несанкционированного доступа. – СПб.: Наука и техника, 2004. – 384 с.
9. *Лукацкий А.В.* Обнаружение атак. – СПб.: БХВ-Петербург, 2001. – 624 с.
10. *Baker Z.K., Prasanna V.K.* A Methodology for the Synthesis of Efficient Intrusion Detection Systems on FPGAs // Proceedings of the 12th Annual IEEE Symposium on Field-Programmable Custom Computing Machines. Washington: IEEE Computer Society, 2004. P. 135–144.

Поступила 3.03.2010р.

УДК 519.711

В.О. Артемчук

ОБЧИСЛЕННЯ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК ВИБІРКИ В ІНФОРМАЦІЙНО-АНАЛІТИЧНІЙ СИСТЕМІ ЕКОЛОГО- ЕНЕРГЕТИЧНОГО МОНІТОРИНГУ

Вступ

В статтях [1, 2, 3] обґрунтовано актуальність, поставлено та розв'язано задачу збереження даних еколого-енергетичного моніторингу для їх обробки,

© В.О. Артемчук

21

аналізу та інтеграції створеної бази до програмного додатку, написаного на мові програмування Borland C++ Builder 6.0 з використанням технології ADO (ActiveX Data Object).

Таким чином, створену базу даних Microsoft Access, було поєднано з розробленим програмним додатком та розпочато розвиток даного програмного продукту в „Аналітико-інформаційну систему еколого-енергетичного моніторингу”, в якій першочерговою операцією є мультикритеріальна вибірка даних з бази та її графічне представлення, про що докладніше було описано в роботах [3, 4]. Наступним етапом розробки „Аналітико-інформаційної системи еколого-енергетичного моніторингу” є обчислення статистичних характеристик отриманих вибірок та розробки засобів програмної реалізації цих характеристик.

У даній роботі розглянуті основні аспекти проектування, створення алгоритму та програмної реалізації обчислення статистичних характеристик мультикритеріальної вибірки. Обґрунтовано вибір інструментів для вирішення поставленої задачі. Наведено приклади роботи створених програмних модулів.

Методи дослідження

Відповідно до поставленої задачі було проведено дослідження найбільш широковживаних статистичних характеристик, що можуть бути обраховані після формування мультикритеріальної вибірки, формул для їх обчислення та алгоритмів програмування в Borland C++ Builder 6.0. Проведено перевірку отриманих результатів на практиці, тобто в реальному режимі роботи програми з реальними даними.

Вирішення задачі

До головних статистичних характеристик, що можуть бути обраховані за результатами вибірки можна віднести:

1. Середнє значення:

$$\bar{x} = \frac{1}{n} \sum x_i \quad (1)$$

2. Максимальне значення:

$$x_{\max} = \max \{x_i\} \quad (2)$$

3. Мінімальне значення:

$$x_{\min} = \min \{x_i\} \quad (3)$$

4. Середнє лінійне відхилення:

$$\bar{\ell} = \frac{\sum |x_i - \bar{x}|}{n} \quad (4)$$

5. Дисперсію:

$$\sigma^2 = \frac{1}{n} \sum x_i^2 - (\bar{x})^2 \quad (5)$$

Загальна дисперсія визначається як середня арифметична з квадратів відхилень кожного значення ознаки від їх загальної середньої величини. Дана дисперсія характеризує варіацію досліджуваної ознаки за рахунок впливу всіх чинників. Вище була наведена спрощена формула (5) для її обрахунку.

6. Середнє квадратичне відхилення являє собою корінь квадратичний з дисперсії:

$$\sigma = \sqrt{\frac{1}{n} \sum x_i^2 - (\bar{x})^2} \quad (6)$$

Середнє квадратичне відхилення називають стандартним відхиленням. Воно як і середнє лінійне відхилення, є іменованою величиною. Середнє квадратичне відхилення використовують при оцінці тісноти зв'язку між явищами, при обчисленні помилок вибіркового спостереження, дослідженні рядів розподілу та ін.

Для нормального або близького до нормального розподілу між середнім квадратичним і лінійним відхиленнями встановлено таке співвідношення:

$$\sigma = 1,25\bar{\ell} \quad (7)$$

Середнє квадратичне відхилення не завжди зручне для використання, тому що воно не дозволяє порівнювати між собою середні квадратичні відхилення у варіаційних рядах, варіанти яких виражені у різних одиниць виміру.

7. Асиметрію:

$$A = \frac{\sum (x_i - \bar{x})^3}{n\sigma^3} \quad (8)$$

8. Екссес:

$$E = \frac{\sum (x_i - \bar{x})^4}{n\sigma^4} - 3 \quad (9)$$

9. Лінійний коефіцієнт варіації:

$$v_{\ell} = \frac{\bar{\ell}}{x} \quad (10)$$

10. Середньоквадратичний коефіцієнт варіації:

$$v_{\sigma} = \frac{\sigma}{x} \quad (11)$$

11. Суму:

$$Sum = \sum x_i \quad (12)$$

12. Суму квадратів:

$$Sum^2 = \sum x_i^2 \quad (13)$$

13. Розмах варіації являє собою різницю між найбільшим і найменшим значенням ознаки:

$$R = x_{\max} - x_{\min} \quad (14)$$

Розмах варіації простий для обчислення, але він відображає лише крайні значення ознаки і не дає уяви про ступінь варіації усередині сукупності.

14. Модальне значення (моду). Модою називається величина ознаки (варіанта), яка найчастіше зустрічається в даній сукупності.

15. Медіанне значення (медіану). Медіана – це варіанта, яка займає середнє положення в рангованому варіаційному ряду.

Якщо вибірка здійснювалася за двома ознаками (x та y) одночасно, то можна додатково обрахувати:

16. Для вимірювання тісноти зв'язку і визначення його напрямку при лінійній залежності використовують лінійний коефіцієнт кореляції, який визначається за формулою:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} \quad (15)$$

17. Коефіцієнт детермінації даного випадку можна визначити за формулою:

$$R^2 = \left(\frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} \right)^2 \quad (16)$$

Коефіцієнт детермінації показує ступінь варіації ознаки під впливом чинника покладеного в основу групування.

18. Рівняння лінійної регресії:

$$y = r \cdot \frac{\sigma_y}{\sigma_x} (x - \bar{x}) + \bar{y} \quad (17)$$

та

$$x = r \cdot \frac{\sigma_x}{\sigma_y} (y - \bar{y}) + \bar{x} \quad (18)$$

Для обчислення вищезгаданих статистичних характеристик були написані два класи-шаблони (див. рис. 1).

Для відображення результатів розрахунків у проект було додано спеціальну форму FormStatistics (див. рис. 2), яка на даний момент має чотири закладки (по одній для кожного виду вибірки: вибірка даних екологічних спостережень, вибірка неекологічних даних, вибірка визначення залежностей між екологічними та неекологічними даними за районами по місяцях, вибірка визначення залежностей між екологічними та неекологічними даними за районами по роках). Кожна закладка форми містить дві кнопки „Застосувати” для визначення статистичних характеристик вибірки та „Відмінити” для виходу з даної форми та повернення до попередньої, а також власне таблицю, в якій будуть відображатися обраховані статистичні характеристики вибірки.

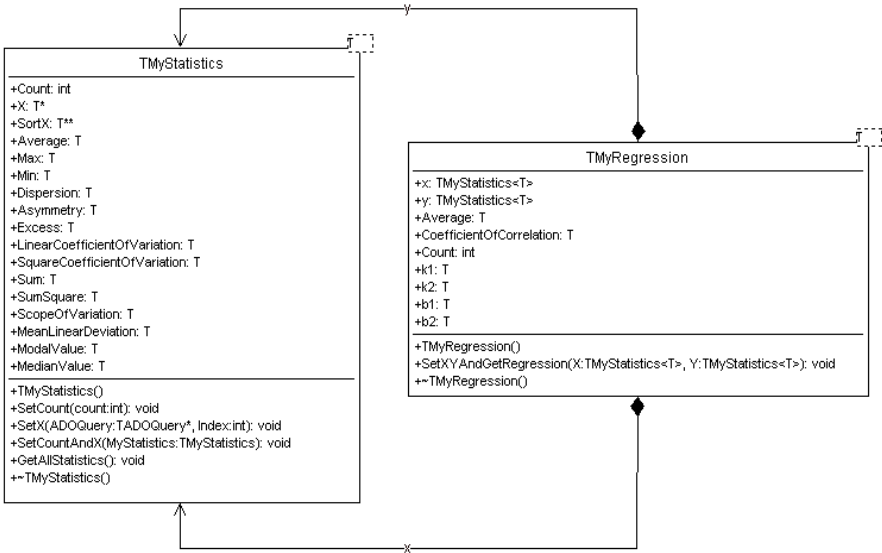


Рис. 1. UML діаграма класів для обчислення статистичних характеристик вибірки

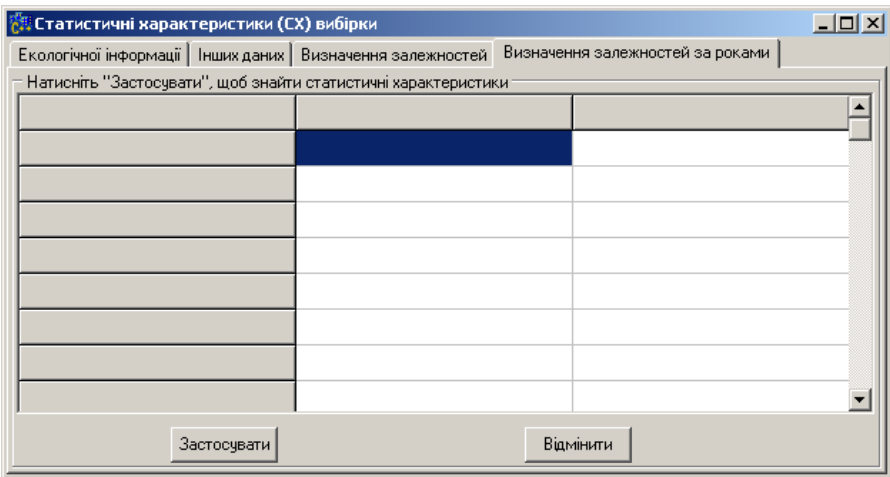


Рис. 2. Загальний вигляд FormStatistics

Оскільки статистичні характеристики обчислюються саме для сформованих вибірок, то і виклик відповідної підпрограми і форми можна здійснити лише після отримання вибірки шляхом натиснення кнопки „Статистика” (див. рис. 3) на формі FormSelect (Формування вибірки).

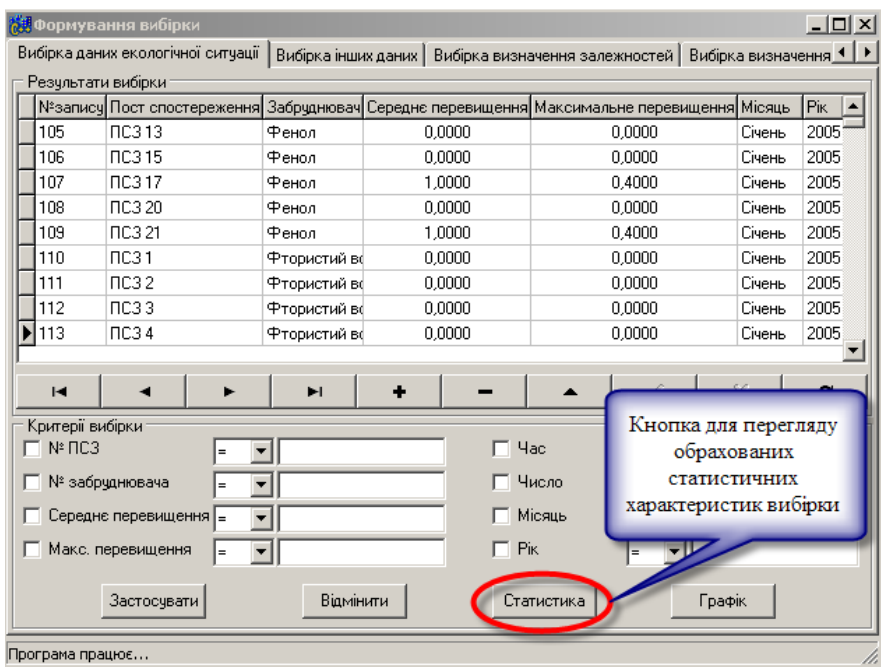
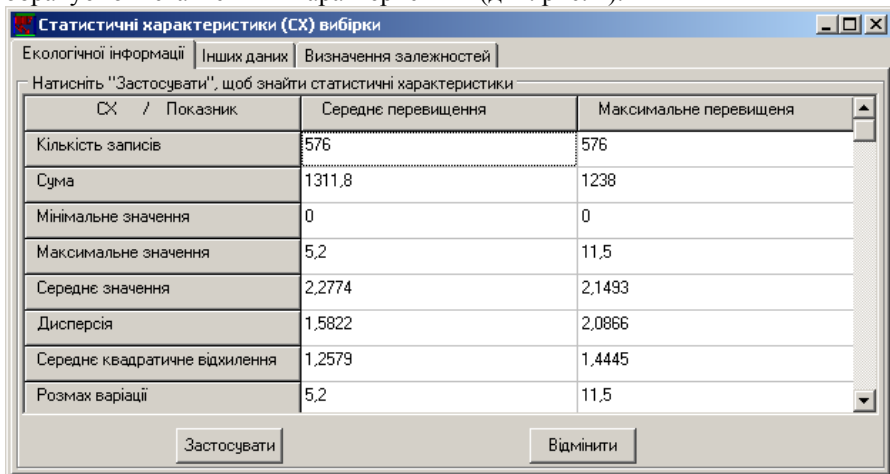


Рис. 3. Загальний вигляд FormSelect з вибіркою за замовчуванням (усі записи), на якій виділено кнопку „Статистика”

Наведемо декілька прикладів роботи створеного модулю. Здійснимо вибірку максимальних та середньодобових перевищень норм забруднень двоокису азоту на пунктах спостереження м. Києва за 2005-2007 рр. та обрахуємо її статистичні характеристики (див. рис. 4).



СХ / Показник	Середнє перевищення	Максимальне перевищення
Коефіцієнт варіації квадратичний	0,55231 або 55,231%	0,67209 або 67,209%
Коефіцієнт варіації лінійний	0,26681 або 26,681%	0,31826 або 31,826%
Середнє лінійне відхилення	0,60764	0,68403
Мода	0	0
Медіана	2,5	2,1
Асиметрія	0,84443	1,4296
Екссес	-1,7572	1,7952
Коефіцієнт кореляції	0,7713	0,7713

Рис. 4. Статистичні характеристики вибірки максимальних та середньодобових перевищень норм забруднень двоокису азоту на пунктах спостереження м. Києва за 2005-2007 рр.

З отриманих результатів слід відмітити, що за 2005-2007 рр. максимальні та середньодобові перевищення норм забруднень двоокису азоту на пунктах спостереження м. Києва в середньому були відповідно рівні 2,1493 та 2,2774, а подекуди досягали значень 11,5 та 5,2 відповідно. Також відмітимо доволі тісний зв'язок (лінійний коефіцієнт кореляції рівний 0,7713) між максимальними та середньодобовими перевищеннями норм забруднень двоокису азоту на пунктах спостереження м. Києва за 2005-2007 рр.

Розглянемо ще один приклад. Здійснимо вибірку визначення залежності між максимальними перевищеннями двоокису азоту в 2005 р. та вродженими вадами розвитку немовлят в 2005 р. у м. Києві та обрахуємо їх статистичні характеристики (див. рис. 5).

З обрахованих в даному прикладі статистичних характеристик звернемо увагу на лінійний коефіцієнт кореляції, який показує, що між досліджуваними показниками існує досить тісний прямо пропорційний зв'язок, і коефіцієнт детермінації, який показує, що в 2005 році в м. Києві вроджені вади розвитку немовлят на 19,031% залежали від значень максимального перевищення двоокису азоту в 2005 р.

Результати роботи

Результатом проведеної роботи є розширення додатку (програми) на мові програмування Borland C++ Builder 6.0.

Таким чином створений програмний продукт, тепер може не лише здійснювати мультикритиреальні вибірки та графічно їх представляти, а й обчислювати основні статистичні характеристики отриманих вибірок.

Статистичні характеристики (СХ) вибірки		
Екологічної інформації Інших даних Визначення залежностей Визначення залежностей за роками		
Натисніть "Застосувати", щоб знайти статистичні характеристики		
СХ / Показник	Серед./Макс. перевищення	Значення показника
Середнє лінійне відхилення	0	32,667
Мода	1,4556	123,42
Медіана	1,9417	103,71
Асиметрія	0	1,5158
Екссес	-3	-0,27942
Коефіцієнт кореляції	0,43624	0,43624
Коефіцієнт детермінації	0,19031 або 19,031%	0,19031 або 19,031%
Рівняння регресії	$x = 0,0036316 * y + 1,5677$	$y = 52,404 * x + 4,2248$

Рис. 5. Деякі статистичні характеристики вибірки визначення залежності між максимальними перевищеннями двоокису азоту в 2005 р. та вродженими вадами розвитку немовлят в 2005 р. у м. Києві

Для подальшого розвитку „Аналітико-інформаційної системи еколого-енергетичного моніторингу” планується розробка блоку медичних даних, модулю картографічного моделювання, функцій багатовимірного кореляційно-регресійного аналізу.

Висновки

1. Обґрунтовано актуальність задачі розширення функціональності програмного додатку „Аналітико-інформаційної системи еколого-енергетичного моніторингу” щодо створення можливості обчислення головних статистичних характеристик вибірок.
2. Проведено дослідження статистичних характеристик, що можуть бути обраховані після формування мультикритеріальної вибірки та обрано п'ятнадцять найбільш використовуваних з них.
3. Досліджено формули для обчислення цих характеристик та розроблено алгоритми їх програмування в Borland C++ Builder 6.0.
4. Проведено перевірку отриманих результатів на практиці з реальними даними.

1. *Артемчук В.О.* Інтеграція бази даних еколого-енергетичного моніторингу в програмний додаток / *В.О. Артемчук* // Збірник наукових праць Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова. – К.: ІПМЕ ім. Г.Є. Пухова НАН України, 2009. – № 51. – С. 66-73.

2. *Артемчук В.О.* Створення бази даних для інформаційно-аналітичної системи еколого-енергетичного моніторингу / *В.О. Артемчук* // Матеріали XXVII Щорічної науково-технічної конференції „Моделювання”, 15-16 січня 2009 р.: тези допов. – К.: ІПМЕ ім. Г.Є. Пухова НАН України, 2009. – С. 48–50.

3. *Каменева І.П.* База даних еколого-енергетичного моніторингу: проектування та створення / *І.П. Каменева, В.О. Артемчук* // Збірник наукових праць Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова. – К.: ПІМЕ ім. Г.Є. Пухова НАН України, 2009. – № 50. – С. 66-72.
4. *Яцишин А.В.* Формування вибірки з бази даних еколого-енергетичного моніторингу / *А.В. Яцишин, В.О. Артемчук* // Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту : Матеріали Міжнародної наукової конференції, 18-22 травня 2009 р. - Херсон: ХНТУ, 2009. – Т.2. – Ч.2. – С. 114–117.

Поступила 10.03.2010р.

УДК 681.3

С.В. Василенко

МОДЕЛЬ ЗАБЕЗПЕЧЕННЯ ЦІЛІСНОСТІ ІНФОРМАЦІЙНИХ ОБ'ЄКТІВ ІЗ ЗАСТОСУВАННЯМ УЗАГАЛЬНЕНОГО ЗАВАДОСТІЙКОГО КОДУ УМОВНИХ ЛИШКІВ

Summary: Explored possibilities of the use in the networks of telecommunication of model of providing of integrity of information holding object on the base of the generalized antigambling code of conditional tailings. In the article of such realizable analysis of his possibilities.

Вступ

Відповідно до термінології нормативних документів Державної служби спеціальних телекомунікаційних систем і захисту інформації України [1] під цілісністю інформації розуміється її властивість, яка полягає у тому, що інформація не може бути модифікована неавторизованим користувачем або процесом. Іншими словами, під цілісністю інформації розуміється відсутність в ній будь-яких викривлень (модифікацій), які не були санкціоновані її власником, не залежно від причин або джерел виникнення таких викривлень.

Викривлення інформації, тобто порушення її цілісності, можливі на будь-якому етапі її циркуляції у обчислювальних мережах: при зберіганні, передачі або обробці. Причини таких викривлень можуть бути випадковими або навмисними. У свою чергу, випадкові викривлення можуть бути як природними, пов'язаними з дією природних чинників, так і штучними. Випадкові штучні викривлення пов'язані з діяльністю людей – з випадковими помилками персоналу. Навмисні викривлення завжди пов'язані з умисними діями порушників. І ті, і інші дії мають своїм наслідком викривлення того або іншого числа символів в цифровому представленні інформації, незалежно від використовуваної системи числення або форми