



УДК 31:[004.6:004.94]

## ІНФОРМАЦІЙНІ СИСТЕМИ І ТЕХНОЛОГІЇ АНАЛІЗУ СТАТИСТИЧНОЇ ІНФОРМАЦІЇ

Наталія Міщук

*Львівський національний університет імені Івана Франка  
79008 м.Львів, проспект Свободи, 18  
E-mail: mishchuk\_n@ukr.net*

*В даному дослідженні розглянуті передумови створення єдиної статистичної інформаційної системи, визначені принципи створення, функціонування та динамічного розвитку статистичної інформаційної системи підприємства, здійснено огляд пакетів прикладних програм для обробки статистичної інформації.*

*Ключові слова: інформаційне суспільство, статистична інформаційна система, інформаційні технології в статистиці, програмні продукти для аналізу даних, візуалізації даних, статистичного моделювання та прогнозного аналізу.*

**Вступ.** Динамічний розвиток інформаційного суспільства та інформаційної економіки зокрема, виступає однією з домінуючих тенденцій загальносвітового розвитку. На основі достовірної, точної, своєчасної та повної інформації про стан, тенденції та фактори впливу на розвиток повинні формуватись державна політика та управління цим процесом, оскільки від їх обґрунтованості значною мірою залежать результати для громадян, суспільства та держави.

Побудова орієнтованого на інтереси людей, відкритого для всіх і спрямованого на розвиток інформаційного суспільства є одним з головних пріоритетів і для України. У такому суспільстві кожен міг би створювати і накопичувати інформацію та знання, мати до них вільний доступ, користуватися і обмінюватися ними, щоб надати можливість повною мірою реалізувати потенціал кожній людині, сприяючи суспільному і особистому розвитку та підвищуючи якість життя.

На сьогодні розвиток інформаційного суспільства, поширення інформаційних технологій в усі сфери життєдіяльності людини та суспільства стали нормою подальшої еволюції цивілізації. Не стала виключенням і державна статистика як одна з найважливіших міжгалузевих ланок у системі управління економікою будь-якої країни.

**Основний виклад матеріалу.** Для вдосконалення державного управління і прийняття виважених та обґрунтованих рішень щодо забезпечення сталого розвитку національної економіки та соціальної сфери, залучення населення до обговорення актуальних питань і завдань подальшого розвитку країни необхідна якісна та своєчасна статистична інформація про економічне, соціальне та екологічне становище України та її регіонів. У зв'язку з цим забезпечення доступності ключових статистичних даних для широкого загалу та прозорості статистичної методології є



необхідною умовою для формування громадянського суспільства та подальшої міжнародної інтеграції нашої держави.

Особливої уваги потребує подальший динамічний розвиток інформаційних технологій для опрацювання статистичної інформації на базі інформаційно-обчислювальної мережі Держкомстату України.

Функціонування інформаційних технологій в статистиці орієнтоване на можливість їх адаптації до мінливих структур управління економікою, до потреб отримання оперативних і надійних статистичних даних.

Застосування сучасної обчислювальної техніки та найновіших інформаційних технологій в органах державної статистики є об'єктивною необхідністю у зв'язку з величезною кількістю об'єктів статистичного спостереження, великих обсягів вирішуваних завдань і оброблюваної інформації при постійно зростаючих вимогах до оперативності, достовірності та аналітичності звітних даних.

Інший важливий напрям удосконалення автоматизації статистичних робіт пов'язаний із створенням великих методологічно сумісних між собою інтегрованих фондів статистичних даних, що охоплюють значні періоди часу, та організацією швидкого і зручного доступу до зберігаються в них даними.

Створення таких фондів дозволяє автоматизувати ведення динамічних рядів статистики і використовувати наявну в них інформацію для вирішення різноманітних аналітичних завдань та інформаційно-довідкового обслуговування керівних органів.

Розглянуті передумови автоматизації статистичних робіт у системі Держкомстату України реалізуються створенням статистичної інформаційної системи (СІС).

Основна мета розробки та впровадження СІС полягає в реалізації можливостей, які виникають внаслідок розвитку економіко-математичних методів впровадження інтерактивних технологій, сучасної телекомунікативної техніки, для повного і якісного задоволення потреб управління економікою у статистичній інформації на державному, галузевому та регіональному рівнях з найменшими витратами людського капіталу та мінімальними матеріальними витратами.

Визначальними принципами створення та функціонування СІС є: системність розробки її елементів; єдність організаційної структури; єдність інформаційної бази; єдність технічної бази; інтегрована обробка даних; типізація проектних рішень; розвиток системи.

Автоматизація статистичної діяльності на підприємствах і в організаціях є актуальним завданням щодо вдосконалення функціонування СІС, вирішення якого дозволить отримати необхідну статистичну інформацію для управління, а також істотно скоротити витрати на підготовку вихідних даних в органах державної статистики.

В процесі реформування економіки України змінилася роль підприємства, а також умови створення та функціонування підприємств. Ці зміни, в свою чергу, висунули на перше місце не лише управління виробничою діяльністю, а й завдання стратегічного управління. Виникли нові підходи до визначення цілей, завдань, технологічних рішень і вимоги до розробки інформаційної системи підприємства.

Статистичний облік на підприємстві та технологія його машинної реалізації переорієнтується, насамперед, на внутрішні потреби керівництва для виконання ним функцій управління в умовах інформаційної економіки.



Зазначені можливості автоматизованого рішення задач на підприємствах можуть бути реалізовані при створенні статистичної інформаційної системи підприємства (СІСП).

Створення СІСП повинно базуватися на двох основних принципах: інтеграції в інформаційну систему управління підприємством та інтеграції із зовнішнім інформаційним простором.

Процес створення СІСП повинен включати наступні роботи:

- визначення інформаційної потреби керівництва підприємством в статистичній інформації;
- визначення цілей і завдань функціонування СІСП в рамках інформаційної системи управління підприємством;
- визначення складу статистичних показників і методів їх розрахунку;
- визначення складу математико-статистичних методів, що забезпечують підтримку прийняття рішень на основі статистичної інформації;
- розробку принципів створення і функціонування СІСП.

Зазначимо, що найбільш важливими відмінними рисами СІСП є поглиблена аналітичність, використання математико-статистичних методів, висока оперативність та варіантність розрахунків. При цьому основний акцент робиться не тільки на отримання вихідної інформації, але й на її багатостороннє дослідження з отриманням варіантних розрахунків, а також на прогностичну орієнтацію і активність системи з отриманням випереджаючої інформації.

Впровадження СІСП може здійснюватися в кілька етапів. На перших етапах можлива орієнтація на програмні засоби, які впроваджені у державній СІС, оскільки вони спроектовані з урахуванням можливості їх адаптації, коригування та розвитку, а також пройшли тривалу перевірку.

На наступних етапах необхідна інтеграція ППП з комплексом програмно-технологічних засобів системи управління підприємством. Її суть полягає у створенні єдиної інформаційної бази первинних даних, яка буде джерелом отримання результатуючої інформації для виконання всіх функцій управління підприємством.

Побудова такої системи передбачає подальшу формалізацію процедур перетворення даних, прикладом якої може служити СТАРТ-технологія. Інформаційною базою її функціонування повинні стати первинні оперативні дані, що відображають динаміку функціонування підприємства.

На перших етапах впровадження СІСП пакети прикладних програм для проведення аналітичних досліджень є більш універсальними в порівнянні з пакетами, що реалізують регламентні завдання. Саме тому, в умовах інформаційної економіки вони можуть бути більш успішно використані на рівні підприємств і організацій.

Програмне забезпечення статистичних досліджень достатньо розвинуте. Серед відомих статистичних пакетів прикладних програм статистичної обробки інформації, які пройшли багаторічну перевірку можна виділити наступні пакети STATISTICA, SPSS, SAS, MathCAD, MathLab, Statgraphics, Stata та інші.

Пакет «SPSS» є одним з найпотужніших універсальних і зручних в експлуатації статистичних пакетів зарубіжного виробництва, який реалізує наступні можливості:

- виконує близько 100 процедур статистичної обробки в базовому модулі, а сімейство додаткових модулів являє собою фактично весь статистичний інструментарій;
- здійснює доступ до даних, що знаходяться практично в будь-якому місці, включаючи можливість легко і швидко з'єднувати кілька баз даних;



- дає можливість за допомогою вбудованої мови скриптів змінювати як інтерфейс, так і різні процедури.

Пакет «SPSS» дозволяє організувати повний цикл досліджень для статистичного аналізу та прогнозування даних на основі широкого набору сучасних методів прикладної статистики

Система SAS – це могутній комплекс з більш як двадцятьма різними програмними продуктами, об'єднаними один з одним «засобами доставки інформації». SAS відрізняє неперевершена потужність щодо набору статистичних алгоритмів. Система надає користувачеві можливість приєднання його оригінальних алгоритмів.

Для оброблення статистичної інформації широко використовується інтегрована система статистичного аналізу й оброблення даних STATISTIKA. Основними компонентами системи STATISTIKA є: електронні таблиці для введення вхідних даних, а також спеціальні таблиці введення числових результатів аналізу; потужна графічна система для візуалізації даних і результатів статистичного аналізу; набір спеціалізованих статистичних модулів, у яких зібрано групи логічно зв'язаних між собою статистичних процедур; спеціальний інструментарій для підготовки звітів; убудовані мови програмування SCL і STATISTICA BASIC, які дають змогу користувачеві розширити стандартні можливості системи.

MATLAB – це високорівнева мова та інтерактивне середовище для програмування, чисельних розрахунків і візуалізації результатів. За допомогою MATLAB можна аналізувати дані, розробляти алгоритми, створювати моделі та додатки.

MATLAB дозволяє користувачеві отримати доступ до даних з файлів, інших програм, баз даних, зовнішніх пристроїв. Існує можливість зчитувати дані з файлів таких популярних форматів як Microsoft Excel, текстових або двійкових файлів, зображень, аудіо та відео файлів, наукових форматів (NetCDF і HDF). Функції введення-виведення дозволяють працювати з файлами даних будь-яких форматів.

Компанія MathWork для аналізу і моделювання даних за допомогою статистики і машинного навчання пропонує продукт Statistics and Machine Learning Toolbox із сімейства продуктів MATLAB.

Statistics and Machine Learning Toolbox™ надає функції і додатки для опису, аналізу і моделювання даних з використанням статистичних даних та машинного навчання. Зокрема, користувач має можливість використовувати описові статистики та графіки для дослідження та аналізу даних, застосовувати ймовірнісний розподіл до даних, генерувати випадкові числа для моделювання методом Монте-Карло, а також виконувати перевірки гіпотез. Регресії та класифікації алгоритмів дозволяють зробити висновки з даних і будувати прогностичні моделі.

Для аналізу багатовимірних даних, Statistics and Machine Learning Toolbox дозволяє визначити ключові змінні або функції, які впливають на модель з послідовним відбором ознак, покроковою регресією, аналізом головних компонент, регуляризації та інших методів зниження розмірності. Панель інструментів забезпечує контрольовані і неконтрольовані алгоритми машинного навчання, у тому числі підтримка векторних машин (SVM), посиленням і згрупованим деревом рішень, k-найближчих сусідів, K-засобів, K-методів, ієрархічної кластеризації, змішаних моделей Гауса і прихованих моделей Маркова.

Особливостями даного продукту є [3]:



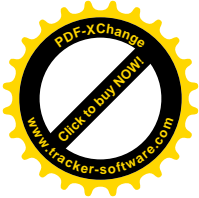
- описова статистика і візуалізація – чисельне дослідження даних та генерація зведених статистичних даних, у тому числі визначення центральної тенденції, дисперсії, форм і співвідношення; обчислення сумарної статистики для вибірки даних, яка містить відсутні (NaN) значення; візуалізація даних за допомогою одновимірного, двовимірного і багатовимірного простору; групування даних, використовуючи методи кластерного аналізу, такі як ієрархічної кластеризації та K-середньої кластеризації;
- розподіл ймовірностей – визначення частоти даних моделей, генерування випадкової вибірки, оцінка ймовірність функції розподілу, розрахунок зведених статистичні показників (середнє і медіана);
- перевірка і тестування гіпотез – надає параметричні і непараметричні тести гіпотез,
- кластерний аналіз;
- дисперсійний аналіз (ANOVA) надає можливість здійснювати аналіз дисперсії і коваріації, N-дисперсійний аналіз, багатовимірний аналіз відхилень, множинні порівняння для даних з декількома відповідями, моделювання з повторними вимірами;
- регресія – використовуються лінійні, узагальнені лінійні, нелінійні та непараметричні методи для контрольованого навчання;
- класифікація – являє собою тип контрольованого машинного навчання, в якому алгоритм «дізнається», як класифікувати нові спостереження за прикладом помічених даних. Для вивчення моделей класифікації в інтерактивному режимі, використовується додаток Classification Learner;
- зниження розмірності – факторний аналіз, послідовний підбір функцій;
- промислова статистика (дизайн експериментів (DOE); аналіз даних на весь термін служби; статистичне управління процесом) – інструменти для проектування експериментів, аналізу надійності і використовуваності дані, контроль за якістю процесу та спостереження даних;
- прискорення статистичних розрахунків – дозволяє використовувати паралельні або розподілені обчислення для прискорення певних статистичних обчислень.

Потужним пакетом для обробки статистичних і графічних даних вважаємо і програму Stata, оскільки це є повний, інтегрованим статистичний пакет програмного забезпечення, який забезпечує все необхідне для аналізу даних, управління даними, і графіки як спосіб візуалізації даних.

Stata містить сотні статистичних інструментів від стандартних методів таких як базові таблиці і зведення, лінійна регресія, кластерний аналіз тощо, до потужних методів – багаторівневі моделі, структурне моделювання, бінарні розрахунки, Байєсівський аналіз, точні статистичні дані тощо.

Невід'ємною частиною Stata є мова програмування матриць Mata – це інтерактивне середовище для маніпулювання матрицями і повна середовище для розробки скомпільованого і оптимізованого коду. Mata включає в себе спеціальні функції для обробки панельних даних, виконує операції на дійсними чи комплексними матрицями, забезпечує повну підтримку об'єктно-орієнтованого програмування, і повністю інтегрована з кожним елементом Stata.

Stata дозволяє легко генерувати якісні публікації та чітко оформлені графіки.



Stata використовується для досліджень більш ніж 30 років, надає все необхідне для аналізу даних, управління даними, і графіки у різноманітних сфера діяльності, а саме біостатистика, економіка, освіта, епідеміологія, фінанси, бізнес, і маркетингу, медицина, політологія, суспільна охорона здоров'я, державна політика, соціологія.

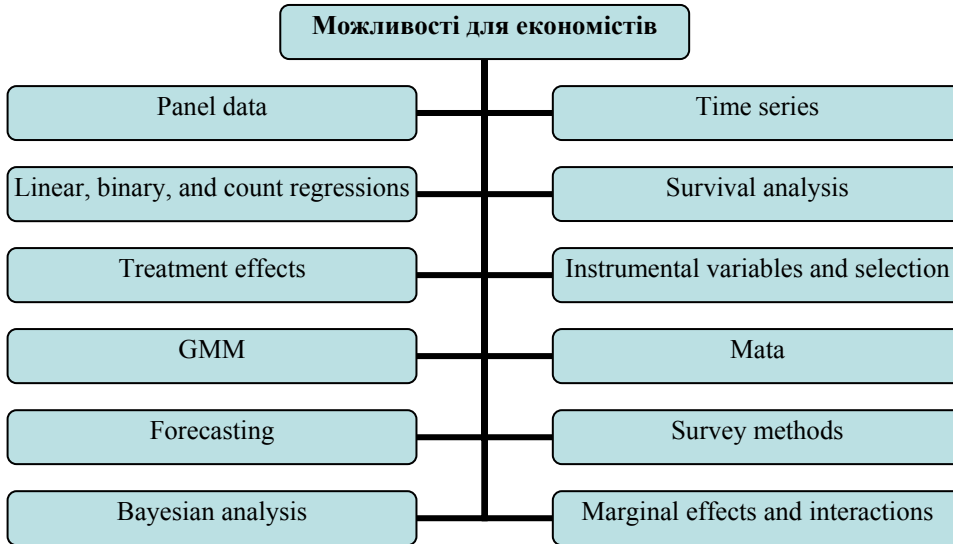


Рис. 1. Структура можливості програми Stata для економічних досліджень

Продукт Stata володіє широкими можливостями, точністю розрахунків, розширюваністю і відтворюваністю. Крім того, Stata дозволяє візуалізувати всі статистичні дані, побудувати графіки та надає інструменти управління даними, які необхідні для вирішення широкого спектру економічних питань.

Компанія Statpoint Technologies, Inc пропонує кілька продуктів для статистики та аналізу даних. Кожен користувач може вибрати продукт, який найкращим чином відповідає його потребам.

Таблиця 1

**Порівняння продуктів**

<b>Продукт</b>	<b>Цільова аудиторія</b>	<b>Тип продукту</b>
Statgraphics Centurion	Аналітики даних, які потребують широкий спектр статистичних інструментів	Додаток для Windows Автономний
Statgraphics sigma express	Практики, які потребують статистичних функцій для поліпшення якості	Додаток в Excel
Statgraphics stratus	Аналітики даних, які потребують статистичної функціональності у веб-браузері	Веб-додаток
Statgraphics Web Services	Розробники веб-додатків	Веб-сервіс
Statbeans	Розробники додатків JAVA	Статистична бібліотека



Statgraphics є потужним, але інтуїтивно зрозумілий продуктом для аналізу даних, візуалізації даних, статистичного моделювання та прогнозного аналізу. Це комплексна програма має в наявності понад 230 процедур від сумарної статистики до потужних статистичних моделей.

**Висновок.** Створення, експлуатація та розвиток нових інформаційних технологій передбачають участь висококваліфікованих фахівців, у тому числі і економістів-статистиків, які володіють основами проектування та функціонування таких технологій.

Створення СІСП повинно базуватися на двох основних принципах: інтеграції в інформаційну систему управління підприємством та інтеграції із зовнішнім інформаційним простором.

Для проведення аналітичних досліджень в умовах сучасної інформаційної економіки універсальними програмними засобами є статистичні пакети прикладних програм, і тому вони можуть успішно використовуватись на рівні підприємств і організацій.

- 
1. Орленко Н.С. Інформаційні системи і технології в статистиці [Текст]: навч.-метод. посіб. для самост. вивч. дисципліни / Н.С. Орленко; Державний вищий навчальний заклад "Київський національний економічний ун-т ім. Вадима Гетьмана". – К.: КНЕУ, 2008. – 282 с.
  2. Ушакова І. О. Інформаційні системи і технології в статистиці / Ушакова І.О. – Харків : Вид. ХДЕУ, 2006. – 164 с.
  3. <http://www.mathworks.com/products/statistics>
  4. <http://www.stata.com/>
  5. <http://www.statgraphics.com/>

## INFORMATION SYSTEMS AND TECHNOLOGY ANALYSIS OF STATISTICAL INFORMATION

Nataliia Mishchuk

*Ivan Franko National University of Lviv,  
Prospekt Svobody 18, UA – 79008, Ukraine*

There are considered prerequisites for creating a single statistical informationsystem, defined principles of creation, functioning and dynamic development of the statistical information system, implemented an overview of software packages for processing of statistical information in this study.

Keywords: communicatory society, statistical information system, information technologies in statistics, software programs for data analysis, data visualization, statistical modeling and predictive analytics.





## ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ АНАЛИЗА СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ

Наталія Мищук

*Львовский национальный университет имени Ивана Франка  
79008 м. Львов, проспект Свободы, 18*

В данном исследовании рассмотрены предпосылки создания единой статистической информационной системы, определены принципы создания, функционирования и динамичного развития статистической информационной системы предприятия, осуществлен обзор пакетов прикладных программ для обработки статистической информации.

Ключевые слова: информационное общество, статистическая информационная система, информационные технологии в статистике, программные продукты для анализа данных, визуализации данных, статистического моделирования и прогнозного анализа.