



ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ЕКОЛОГО-СТАТИСТИЧНИХ ДОСЛІДЖЕНЬ

- Сучасний ринок програмних продуктів для обробки великих масивів екологічної інформації.
 - Комп'ютерні технології обробки електронних таблиць Microsoft Excel.
 - Робота з банками екологічної інформації.
-

Техніка обробки зібраної інформації залежить від методики досліджень. Частина інформації можна опрацьовувати безпосередньо на об'єкті досліджень, але основну – в камеральних умовах. Сьогодні широко використовують обробку інформації на ЕОМ.

Для обробки інформації, використовують різноманітні статистичні методи – розподілу і групування, середніх і відносних величин, дисперсійний і кореляційно-регресійний, індексний та рядів динаміки, а також математичні методи моделювання і прогнозування. Статистичні методи розглядаються в другій частині.

1.3.1. Сучасний ринок програмних продуктів для обробки великих масивів екологічної інформації

Оперативна, якісна і точна обробка великих масивів статистичної інформації може бути виконана лише з використанням сучасних засобів обчислювальної техніки. Наявність потужних, надійних і разом з тим простих в експлуатації програмних продуктів статистичного аналізу звільняє дослідника від рутинних операцій, розширює сферу застосування статистичних методів в різних галузях людської діяльності, сприяє появі якісно нових можливостей статистичного аналізу і моделювання даних. Використання пакетів прикладних програм - це єдиний реальний практичний інструмент розв'язування задач багатofакторного кореляційно-регресійного та аналізу в багатовимірному просторі.

Програмне забезпечення статистичних досліджень досить розвинуте. Сучасний ринок програмних продуктів пропонує різноманітні пакети програм для статистичної обробки даних. Всесвітньо відомі статистичні пакети для комплексної обробки даних: *BMDP, SPSS, SAS, Systat, Minitab, S-Plus, Statgraphics Statistica* та інші.

Використання згаданих пакетів програм дає змогу автоматизувати процес статистичного дослідження в таких напрямках:

- створення файлів даних і таблиць;
- групування даних;
- графічний аналіз даних;
- розрахунок варіаційних характеристик вибірових сукупностей;
- побудова рядів розподілу;
- аналіз рядів динаміки і прогнозування їх майбутніх рівнів;
- кореляційно-регресійний аналіз;
- багатомірний аналіз.

З 1995 р. Світовим лідером на ринку статистичного програмного забезпечення визнається інтегрована система *Statistica* для *Windows* (версія 5.0), розроблена фірмою Stat Soft. Перша версія програми з'явилася у 1991р. для операційної системи MS-DOS і була новим напрямом розвитку статистичного програмного забезпечення. В ній реалізовано графічно-орієнтований підхід до статистичного аналізу даних, суть якого полягає в отриманні всебічного візуального представлення інформації на всіх етапах статистичної обробки даних.

Багатофункціональна, графічно орієнтована на обробку масових даних система *Statistica* відповідає основним стандартам *Windows* (динамічний обмін даними з іншими додатками, підтримка основних операцій з буфером обміну, робота в мережевому середовищі та інші).

Передусім це стандарти користувацького інтерфейсу — *MDI*, використання буфера-обміну, механізму динамічного зв'язку (*DDE*) з іншими додатками; система підтримує всі операції, реалізовані за допомогою методу *Drag-and-Drop* — «Перетягти та опустити», включаючи автозаповнення, інші.

Складніші процедури обробки даних у системі *Stratgraphics* виконує спеціалізований модуль *Data Management* — «Управління даними», а для обробки великих масивів даних або даних з довгими текстовими значеннями застосовують процедури *Megafile Manager Data* — «Менеджера мегафайлів».

Система *Stratgraphics* працює з чотирма типами документів. Це: електронна таблиця *Spreadsheet*, призначена для введення і перетворення первинних даних;

електронна таблиця *Scrollsheet* — для виведення результатів аналізу;

графік — для візуалізації результатів обробки та аналізу даних;

звіт — файл у формі *RTF* (розширений текстовий формат), в якому зберігається текстова, числова і графічна інформація.

Усі статистичні процедури системи розбито на окремі модулі, кожен з яких об'єднує групу логічно зв'язаних між собою статистичних методів і в рамках конкретної моделі забезпечує повний і всебічний аналіз закономірностей.

Наприклад, у модулі **Basic Statistics/Table** — «Основні статистики і таблиці» пропонується широкий вибір методів розв'язувального статистичного аналізу:

- ♦ характеристики варіації і форми розподілу,
- ♦ групування та класифікації,
- ♦ таблиці дисперсійного аналізу *Anova*,
- ♦ всі види коефіцієнтів щільності зв'язку,
- ♦ критерії для тестування нормальності розподілу,
- ♦ критерії істотності зв'язку тощо.

Модуль **Multiple Regression** — «Множинна регресія» включає:

- ♦ вичерпний набір засобів множинної лінійної і нелінійної регресії,
- ♦ багатofакторного прогнозування,
- ♦ аналіз залишків і викидів,
- ♦ тестування гіпотез регресійного аналізу.

Модуль **Time Series/Forecasting** — «Часові ряди і прогнозування» об'єднує процедури аналізу закономірностей динаміки:

тенденцій розвитку і коливань,

- ♦ різні методи згладжування рядів,
- ♦ описування трендів,
- ♦ описування сезонної декомпозиції,
- ♦ методи авторегресійного аналізу,

методи прогновної екстраполяції.

Система **Statistica** включає модуль **Anova/Manova** — «Дисперсійний аналіз», увесь арсенал методів багатовимірної аналізу (кластерний, дискримінантний, факторний аналіз, факторне шкалювання, канонічні кореляції).

Особливе місце посідає модуль **Seopath** — «Моделювання взаємозв'язків системами структурних рівнянь».

Зазначені модулі покривають практично весь спектр сучасних методів статистичного дослідження і моделювання. *Запуск модуля* здійснюється через перемикач модулів — **Module Swither**. У кожному модулі робота починається із «Стартової панелі», де відкривається файл первинних даних, вибирається процедура обробки даних і визначаються відповідні їй параметри.

Стартова панель — основне діалогове вікно модуля. Структуру діалогу в усіх модулях уніфіковано, її можна подати схематично (рис. 1.1).

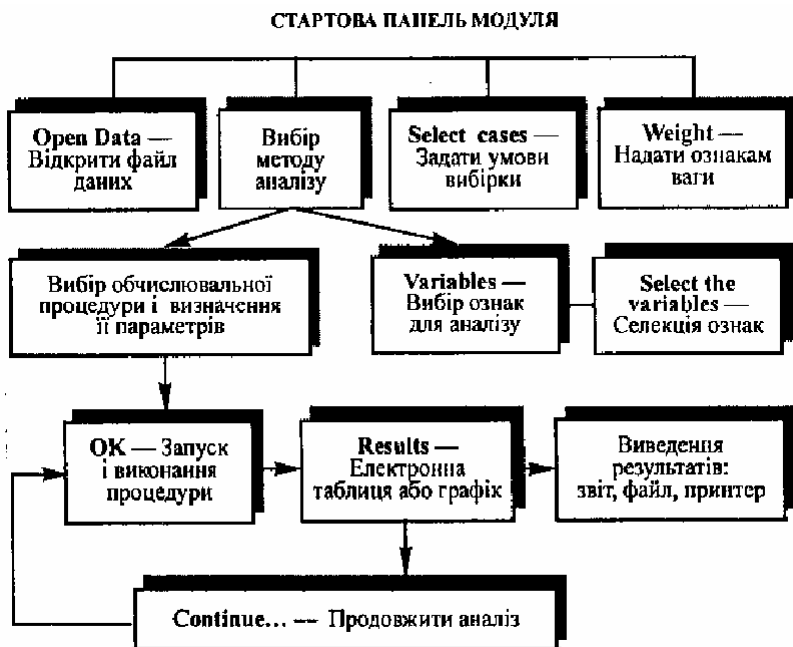


Рис. 1.1. Схема структури діалогу в модулі

У системі *Statistica* реалізовано принцип постійного логічного підказування. Якщо користувач не може визначитися щодо наступного кроку діалогу, через команду *Enter* система сама спрямує до відповідного діалогового вікна. Якщо виникають складнощі з вибором параметрів обчислювальної процедури, вони задаються системою «за умовчужанням».

Важливою характеристикою системи є наявність засобів всебічної графічної підтримки процесу обробки даних і візуалізації результатів аналізу. Графічні можливості й засоби системи унікальні. Вона включає сотні різних типів користувацьких і спеціальних статистичних графіків, доступних у будь-якому модулі й на будь-якому етапі статистичної обробки даних. Інструменти компонування складної графічної інформації з текстовою і числовою інформацією розкладаються у кожному модулі.

Використання сучасних комп'ютерних технологій обробки даних, інтерактивний спосіб взаємодії з системою перетворюють статистичний аналіз, моделювання та прогнозування в захоплююче

дослідження закономірностей навколишнього світу. Завдяки різноманітним формам організації діалогу, максимально простій із звичиними для статистики термінами мові спілкування, наявності контекстно-залежної довідкової системи, мові програмування STATISTICA BASIC пакет є ефективним інструментом проведення статистичного дослідження як для користувача-початківця, так і для професіонала.

1.3.2. Комп'ютерні технології обробки електронних таблиць Microsoft Excel

Необхідно зазначити, що спеціалізовані пакети, до яких належать Statistica, - це відносно дорогі, а отже, і не завжди доступні програмні продукти. Переважну більшість задач статистичного аналізу можна досить ефективно розв'язувати з використанням програми обробки електронних таблиць Microsoft Excel. Спектр доступних стандартних статистичних функцій останніх версій Microsoft Excel майже не поступається спеціальним програмам обробки статистичних даних (понад 70 функцій). Для автоматизації розрахунків статистичних показників і параметрів підручника, можна скористатися наведеними далі статистичними функціями:

ФРАСП (х, ступінь вільності в чисельнику, ступінь вільності в знаменнику).

Результат: F - розподіл ймовірності.

ВЕРОЯТНОСТЬ (інтервал числових значень х, множина ймовірностей появи значень х, нижня границя, верхня границя).

Результат: ймовірність того, що числові значення знаходяться в заданих границях.

ДИСП (число 1, число 2, ...).

Результат: дисперсія вибіркової сукупності.

ДИСПР (число 1, число 2, ...).

Результат: дисперсія генеральної сукупності.

ДОВЕРИТ (рівень значимості, стандартне відхилення, розмір вибірки).

Результат: довірчий інтервал для генеральної сукупності.

КВАДРОТКЛ (число 1, число 2,...).

Результат: сума квадратів відхилень даних від середнього значення.

КОРЕЛ (масив у, масив х).

Результат: коефіцієнт кореляції між масивами значень змінних.

ЛИНЕЙН (масив значень у, масив значень х, константа, статистика). *Результат:* розрахунок параметрів рівняння лінійної регресії: $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m$

Якщо константа набуває значення ЛОЖЬ, то розраховуються параметри рівняння без вільного члена, якщо - ИСТИНА, то будується модель з вільним членом. У разі потреби у додатковій регресійній статистиці необхідно надати останньому аргументу функції значення ИСТИНА.

Результати роботи функції ЛИНЕЙН видаються у вигляді такого масиву інформації:

Таблиця 1.2

b_m	b_{m-1}	...	b_1	b_0
σ_{bm}	σ_{bm-1}	...	σ_{b1}	σ_{b0}
R^2	σ_y			
$F_{розр}$	$V = n - m - 1$			
$\sum (\hat{y} - \bar{y})^2$	$\sum (y - \bar{y})^2$			

Перед виконанням функції область результатів потрібно виділити, результати з'являться на екрані після послідовного натиску клавіш

$F2 \Rightarrow Ctrl + Shift + Enter$.

АГРФПРИБЛ (масив у, масив х, константа, статистика).

Результат: масив значень, що характеризує експоненційну регресійну криву $y = b_0b_1$ (константа опущена або набуває значення ЛОЖЬ) або $y = b^x$ (константа набуває значення ИСТИНА).

Статистика - це логічне значення, яке показує, чи потрібно видавати додаткову статистичну інформацію. Якщо статистика набуває значення ИСТИНА, то стандартна функція розраховує додаткову статистику.

МЕДИАНА (число 1, число 2). *Результат:* медіана масиву значень.

МОДА (число 1, число 2, ...) *Результат:* мода масиву значень.

НОРМАЛИЗАЦИЯ (х, середнє значення, стандартне відхилення).

Результат: нормалізоване значення із заданим середнім і стандартним відхиленням.

НОРМРАСП (середнє значення, стандартне відхилення, константа).

Результат: інтегральна або диференціальна нормальна функція розподілу.

ПРЕДСКАЗ (x^{np} , масив значень у, масив значень х). *Результат:* значення функції в точці x^{np} , передбачене на основі рівняння лінійної

регресії.

РАНГ (число, масив даних, порядок). *Результат*: ранг числа в масиві даних. Якщо порядок опущений або дорівнює 0, то ранг обчислюється виходячи із ранжування в напрямку спадання значень, інакше - в напрямку зростання.

СГЕОМ (число 1, число 2). *Результат*: середнє геометричне масиву додатних аргументів.

СРЗНАЧ (число 1, число 2). *Результат*: середня арифметична аргументів.

СРОТКЛ (число 1, число2, ...). *Результат*: середня абсолютних значень відхилень даних від середньої.

СТАНДОТКЛОН (число 1, число 2). *Результат*: оцінка стандартного відхилення вибірових даних.

СТАНДОТКЛОНП (число 1, число 2, ...). *Результат*: оцінка стандартного відхилення даних генеральної сукупності.

СТЬЮДРАСП (x, ступінь вільності, константа). *Результат*: t-розподіл Стюдента. Якщо константа дорівнює 1, то видається значення одностороннього розподілу, якщо - 2, то видається значення двостороннього розподілу.

ТТЕСТ (масив 1, масив 2, константа, тип). *Результат*: ймовірність того, що вибірки проведені із генеральних сукупностей з рівними середніми. Якщо константа дорівнює 1, то використовують односторонній розподіл, якщо - 2, то використовують двосторонній розподіл. Останній аргумент характеризує тип тестування: 1 - парний; 2 - двовибірковий з рівними дисперсіями; 3 - двовибірковий з нерівними дисперсіями.

Додаткову інформацію про статистичні функції можна отримати, скориставшись довідниковою підсистемою Microsoft Excel.

Крім арсеналу обчислювальних засобів Microsoft Excel, можна використати широкий спектр засобів графічного представлення даних. У Microsoft Excel передбачені засоби побудови графіків функцій, роботи з усіма можливим типами діаграм, нанесення даних на вбудовані географічні карти, створення різноманітних графічних об'єктів.

Враховуючи можливості роботи з базами даних, Microsoft Excel можна вважати дешевим і зручним інструментом обробки статистичних даних. Таким чином наявність комп'ютерних засобів (Statistica, Excel) дозволяє достатньо легко будувати багатофакторні лінійні регресійні моделі особам без базової математичної освіти.

1.3.3. Робота з банками екологічної інформації

Інформаційне забезпечення. Основу інформаційного забезпечення статистичного дослідження становлять дані статистичної звітності, які містяться у:

- первісних документах статистичної звітності,
- регіональних статистичних бюлетенях і статистичних щорічниках Держкомстату України,
- офіційних матеріалах Міністерства економіки України, Міністерства фінансів України, Міністерства праці та соціального захисту України,
- матеріалах Національного банку України, Державного митного комітету України, Міждержавного статистичного комітету СНД, Українсько-Європейського центру з питань законодавства тощо.

Додатковими інформаційними джерелами можуть служити матеріали спеціально організованих статистичних спостережень, аналітичні матеріали (бізнес-огляди):

❖ *українських НДІ*

- Інститут економічного прогнозування НАН України,
- Міжнародний центр перспективних досліджень,
- Інститут економіки НАН України,
- Інститут регіональних досліджень НАН України,
- Інвестиційна компанія ДІКОМ та ін.

❖ *зарубіжних організацій*

- Міжнародний статистичний інститут,
- Світовий банк,
- Міжнародна організація праці,
- Міжнародний банк реконструкції та розвитку,
- Міжнародний валютний фонд,
- Міжнародна фінансова корпорація,
- Міжнародна асоціація розвитку,
- Статистична комісія ООН,
- Комісія з питань народонаселення Економічної та Соціальної ради ООН,
- Організація економічного співробітництва і розвитку,
- Конференція європейських статистиків та ін.

Інформацію про екологічні ситуації на окремих територіях чи об'єктах, про вплив антропогенної діяльності на стан довкілля та здоров'я людей можна отримати за даними екологічної експертизи, екологічний стан на окремих об'єктах описується в екологічних паспортах підприємства.

Банки екологічної інформації. Банки екологічної статистичної інформації – це вторинна накопичена інформація, певним чином упорядкована чи опрацьована. Найчастіше така інформація подається у вигляді статистичних збірників, щорічників.

Статистичний щорічник України представляє собою накопичену інформацію за ряд років, тобто є по суті банком статистичної інформації про соціально-економічне становище держави. Він складається з таких розділів: національні рахунки, фінанси і кредит, ціни і тарифи, матеріально-енергетичні ресурси, промисловість, інвестиційна та будівельна промисловість, транспорт і зв'язок, торгівля і послуги, зовнішньоекономічна діяльність, структурні зміни в економіці, населення, зайнятість населення, доходи населення, освіта, наука та інформатика, культура і відпочинок, медичне обслуговування, правопорушення, природні ресурси та охорона навколишнього середовища, міжнародні зіставлення. Структура статистичного щорічника побудована таким чином, щоби повною мірою відобразити багатогранність соціально-економічних явищ і процесів, які відбуваються в державі. Особлива увага приділяється показникам ринкового перетворення економіки, структурних змін, змін форм власності, зайнятості населення тощо. Інформація про розвиток економіки України міститься у розділах, які укладено за галузевою ознакою. Співпраця Держкомстату України з міжнародними статистичними організаціями дала змогу навести міжнародні зіставлення показників соціально-економічного становища держав Співдружності, а також міжнародні порівняння по окремих країнах світу.

Статистичні збірники «Довкілля України», «Довкілля Житомирщини» та аналогічні збірники інших регіонів України представляють собою накопичену інформацію за ряд років, тобто є по суті банком статистичної інформації про екологічне становище держави та різних її областей. Він складається з таких розділів:

❖ **охорона атмосферного повітря**, що складається з таких таблиць:

«Викиди шкідливих речовин в атмосферне повітря, всього»,

«Викиди шкідливих речовин в атмосферне повітря від автотранспорту»,

«Викиди шкідливих речовин в атмосферне повітря стаціонарними джерелами забруднення»,

«Викиди шкідливих речовин в атмосферне повітря від стаціонарних джерел забруднення у розрахунку на квадратний кілометр»,

«Викиди шкідливих речовин в атмосферне повітря від стаціонарних джерел забруднення у розрахунку на одну особу»,

«Групування регіонів за рівнем викидів шкідливих речовин в атмосферне повітря від стаціонарних джерел забруднення у розрахунку на одну особу»,

«Кількість підприємств, що мають стаціонарні джерела забруднення»,

«Частка викидів шкідливих речовин від автотранспорту у загальному обсязі викидів»,

«Обсяги викидів діоксиду вуглецю від стаціонарних джерел забруднення»;

❖ **охорона та використання водних ресурсів**, що складається з таких таблиць:

«Забір води із природних водних об'єктів»,

«Забір води із природних водних об'єктів у розрахунку на одну особу»,

«Забір води із підземних водних об'єктів»,

«Використання свіжої води»,

«Використання свіжої води на господарсько-питні потреби»,

«Використання свіжої води на господарсько-питні потреби у розрахунку на одну особу»,

«Групування регіонів за рівнем використання свіжої води на господарсько-питні потреби у розрахунку на одну особу»,

«Використання свіжої води на виробничі потреби»,

«Використання свіжої води на сільськогосподарські потреби»,

«Економія забору води за рахунок оборотного та повторного водопостачання»,

«Загальне водовідведення»,

«Водовідведення в поверхневі водойми»,

«Скидання неочищених та недостатньо-очищених стічних вод у поверхневі водні об'єкти»,

«Скидання нормативно-очищених стічних вод у поверхневі водні об'єкти»,

«Скидання нормативно-чистих стічних вод у поверхневі водні об'єкти»,

«Обсяг оборотної та послідовно (повторно) використаної води»;

❖ **земельні ресурси та їх охорона**, що складається з таких таблиць:

«Загальна земельна площа та її розподіл»,

«Розподіл сільськогосподарських угідь»,

«Структура земельного фонду регіону» - зведена таблиця,

«Розподіл загальної земельної площі та сільськогосподарських угідь по землевласниках і землекористувачах» - зведена таблиця,

«Зведені показники внесення добрив на всіх землях сільськогосподарськими підприємствами».

❖ **охорона та використання лісових ресурсів**, що складається з таких таблиць:

«Обсяг продукції (робіт, послуг) лісового господарства»,

«Лісовідновлення»,

«Кількість лісових пожеж»,

«Лісова площа, пройдена пожежами»,

«Збитки, заподіяні пожежами»,

«Загибель лісових насаджень»,

«Загибель лісових насаджень від несприятливих погодних умов»,

«Загибель лісових насаджень»,

«Загибель лісових насаджень від пожеж»;

❖ **небезпечні відходи**, що складається з таких таблиць:

«Наявність відходів I-III класів безпеки у спеціально відведених місцях або об'єктах та на території підприємств (на початок року)»,

«Наявність промислових токсичних відходів у сховищах організованого складування та на території підприємств»,

«Утворилось небезпечних відходів I-III класів безпеки»,

«Одержано відходів від інших підприємств I-III класів безпеки»,

«Використано відходів I-III класів безпеки»,

«Знешкоджено відходів I-III класів безпеки»,

«Передано відходів I-III класів безпеки іншим підприємствам»,

«Наявність відходів I-III класів безпеки у спеціально відведених місцях або об'єктах та на території підприємств (на кінець року)»;

❖ **економічний механізм природокористування**, що складається з таких таблиць:

«Кількість підприємств, яким пред'явлено та які сплатили збори за забруднення НПС»,

«Екологічні збори, пред'явлені підприємствам за забруднення НПС»,

«Екологічні збори, пред'явлені підприємствам за забруднення природних ресурсів»,

«Екологічні збори, сплачені підприємствами за забруднення природних ресурсів»,

«Екологічні збори, пред'явлені та сплачені підприємствами за забруднення природних ресурсів в межах встановлених лімітів»,

«Екологічні збори, пред'явлені та сплачені підприємствами за забруднення природних ресурсів понад встановлені ліміти»,

«Поточні витрати підприємств на основні природоохоронні заходи по природних ресурсах»,

«Витрати підприємств на капітальний ремонт основних виробничих фондів природоохоронного призначення».

Рекомендована література



Нормативно-правова [1-21]



Навчальна [1, 9, 10, 22, 23, 44, 49]



Словник основних термінів

Автоматизація розрахунків – використання статистичних функцій пакету програм для статистичної обробки даних.

Дані статистичні – всі факти (вся цифрова інформація), які належать до конкретної проблеми.

Інформаційне забезпечення статистичного дослідження - дані статистичної звітності, статистичних збірників, щорічників, бюлетенів тощо.

Програмне та інформаційне забезпечення статистичного дослідження – різноманітні пакети програм для статистичної обробки даних: SPSS, Systat, Minitab, S-Plus, STATGRAPHICS, STATISTICA та інші.

Функція **АГРФПРИБЛ** – визначає масив значень, що характеризує експоненційну регресійну криву.

Функція **ВЕРОЯТНОСТЬ** – визначає ймовірність того, що числові значення знаходяться в заданих границях.

Функція **ДИСП** – визначає дисперсію вибіркової сукупності.

Функція **ДИСПР** – визначає дисперсію генеральної сукупності.

Функція **ДОВЕРИТ** – визначає довірчий інтервал для генеральної сукупності.

Функція **КВАДРОТКЛ** – визначає суму квадратів відхилень даних від середнього значення.

Функція **КОРЕЛ** – визначає коефіцієнт кореляції між масивами значень змінних.

Функція **ЛИНЕЙН** – визначає розрахунок параметрів рівняння лінійної регресії.

Функція **МЕДИАНА** – визначає медіану масиву значень.

Функція **МОДА** – визначає моду масиву значень.

Функція **FRACF** – здійснює F-розподіл ймовірності.

