МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

Кафедра «Програмна інженерія та інформаційні технології управління»

Звіт з лабораторної роботи №1

На тему: Знайомство з пакетом STATISTICA

З предмету «Математична статистика»

Виконав:

Студент групи КН-36а

Рубан Ю. Д.

Перевірив:

Голоскоков О. Є.

Харків – 2018

**Тема:** Знайомство з пакетом STATICTICA, розробленого компанією StatSoft.

**Ціль:** Познайомитись з пакетом для всестороннього статичного аналізу STATISTICA.

**Виконання роботи**

STATISTICA - це універсальна інтегрована система, призначена для статистичного аналізу та візуалізації даних, управління базами даних і розробки користувальницьких додатків, містить широкий набір процедур аналізу для застосування в наукових дослідженнях, техніці, бізнесі, а також спеціальні методи добування даних. Крім загальних статистичних і графічних засобів у системі є спеціалізовані модулі, наприклад, для проведення соціологічних або біомедичних досліджень, вирішення технічних і, що дуже важливо, промислових завдань: карти контролю якості, аналіз процесів і планування експерименту. Робота з усіма модулями відбувається в рамках єдиного програмного пакета, для якого можна вибирати один з декількох запропонованих інтерфейсів користувача. За допомогою реалізованих у системі STATISTICA потужних мов програмування, забезпечених спеціальними засобами підтримки, легко створюються закінчені користувальницькі рішення і вбудовуються в різні інші програми або обчислювальні середовища.

Професійний пакет для всебічного статистичного аналізу, розроблений компанією StatSoft. У пакеті STATISTICA реалізовані процедури для аналізу даних(data analysis), управління даними (data management), видобутку даних (data mining), візуалізації даних (data visualization).

Найсильнішою стороною пакета є графіка і засоби редагування графічних матеріалів. Представлено сотні типів графіків, матриці і піктограми. Існує можливість розробити свій дизайн графіка і додати його до меню. Засоби керування графіками містять у собі роботу одночасно з декількома графіками, зміну розмірів складних об’єктів, розширені можливості малювання з додаванням художньої перспективи і спеціальних ефектів, розбивку сторінок. Наприклад, трьохмірні графіки можна обертати, накладати один на одного, стискувати або збільшувати. Крім того можна побачити на графіках, які фрагменти там змінилися під впливом змін в одній з перемінних.

Переваги:

– широкий вибір різноманітних інструментів для статистичного аналізу даних;

– дозволяє обробити масивні об'єми даних;

– можливість проведення кластерного, дискримінантного, канонічного, дисперсійного і факторного аналізу;

– наявність функції групування даних;

– можливість побудови рядів, лінійних і нелінійних моделей;

– великий набір інструментів для дослідження кореляції між різними змінними;

– підтримка нейронних мереж;

– можливість імпорту даних з Excel документів;

– наявність блоку інструментів data mining;

– можливість побудови діаграм і 2D/3D графіків (більше 10000 типів);

– можливість категоризації за змінним

Недоліки:

– відносно висока вартість продукту для вітчизняного бізнесу;

– термін дії пробної версії програми обмежений 1 місяцем.

Підсумки та коментарі:

Програмний комплекс STATISTICA призначається для роботи з величезними масивами даних. Фактично цей інструмент підходить для багатьох галузей економіки, включаючи фінансовий сектор, фармакологію, геологорозвідку, телекомунікації, страхування, енергетику та промисловість. Дана утиліта допоможе провести аналіз даних з мінімальними трудозатрами, і вже на основі оброблених результатів побудує точний прогноз, необхідний для прийняття правильного рішення.

Система STATISTICA виробляється компанією StatSoft Inc. Перші програмні продукти - PsyhoStat-2,3 - були орієнтовані на статистичні дослідження соціологічних даних. Перший комерційний продукт - Statistical Supplement for Lotus 1-2-3, з'явився в 1985 р.

З 1985 р. почався швидкий ріст фірми. StatSoft випускає першу систему статистичного аналізу для комп'ютерів Apple Macintosh під назвою StatFast і статистичний пакет для IBM PC під назвою STATS+. У 1986 р. починається робота над основною лінією програмних продуктів фірми - інтегрованих статистичних пакетів для комплексної обробки даних.

У 1991 р. виходить перша версія системи STATISTICA/DOS, яка являє собою новий напрям розвитку статистичного програмного забезпечення. У ній реалізований так званий графічно-орієнтований підхід до аналізу даних. Цей пакет мав ряд істотних переваг перед іншими статистичними пакетами (за рахунок оптимізації вдалося домогтися підвищення швидкості обробки більш, ніж у 10 разів порівняно з іншими пакетами, пакет міг працювати фактично з необмеженим обсягом даних). У 1992 р. вийшла версія STATISTICA для Macintosh, яка швидко здобула заслужену популярність серед користувачів.

В 1994 р. виходить версія STATISTICA для Windows. Вона відразу ж займає лідируюче положення серед статистичних пакетів. В результаті порівняльного тестування вона отримує перше місце в наступних наукових та комп'ютерних виданнях Medical Software Reviews(2/9), INSIGHT(May, 1994), MacWELT - німецьке видання журналу MacWORLD (Жовтень 1994), c't Magazine(німецький аналог журналу BYTE, March 1994), WINDOWS Magazine (Листопад 1994) де STATISTICA 4.5 порівнювалася з статистичними пакетами BMDP 1.0, SPSS 6.1, Statgraphics 1.0, Systat 5.01. У 1995 р. STATISTICA була включена в число 100 кращих програмних продуктів (WINDOWS Magazine, лютий 1995 р.).

В кінці 1995 р. вийшла версія STATISTICA 5.0. Від попередньої версії вона відрізняється більш зручним інтерфейсом, повну сумісність з Windows 95. Вона включає в себе потужні засоби роботи з даними, багаті графічні можливості і велику кількість методів і процедур статистичного аналізу. STATISTICA5.0 повністю задовольняє основним стандартам середовища Windows. Це насамперед стандарти користувальницького інтерфейсу MDI , використання технологій DDE - Динамічного обміну даними з інших додатків, OLE - Зв'язування і впровадження об'єктів, підтримка основних операцій з буфером обміну і ін Наявність і підтримка основних Windows стандартів дозволяє вам зробити STATISTICA одним з інструментів повсякденної роботи і використовувати її поряд з іншими додатками Windows, обмінюватися даними з якими не представляє ніякої праці.Скачати StatSoft STATISTICA 8.0.550 Enterprise Single User

На відміну від попередніх версій у неї включений внутрішній мова програмування Statistica BASIC, який дозволяє користувачеві розширювати можливості системи. Користувач може додати власну панель інструментів з тим чи іншим методом статистичного аналізу. Можливість дописувати (нарощувати) систему за допомогою вбудованої мови програмування (з якого можна викликати і будь-яку зовнішню бібліотеку DLL) є безсумнівною перевагою пакета.

У 1996-98 роках з'явилися нові випуски пакета – STATISTICA 5.1, 5.1-97 і 5.1-98. В STATISTICA систему додані нові спеціалізовані модулі, враховані всі нові формати Windows і MS Office, зроблені різні доповнення і поліпшення. Завдяки найбільш сучасним і зручним аналітичним інструментам компанія зайняла лідируючі позиції серед виробників статистичного програмного забезпечення. Дві основні лінійки продуктів StatSoft:STATISTICA Enterprise-wide Data Analysis System (SEDAS) і STATISTICA Enterprise-wide SPC System (SEWSS) не тільки не мають аналогів у всьому світі, але і повністю відповідають вимогам замовника. SEWSS, що включає всебічний набір досконалих SPC-аналізів (інструментів регулювання статистичних процесів та контролю якості), зараз успішно проводить мониторниг якості і контроль виробничого процесу на рівні всього підприємства, як найбільш актуальне виробниче обладнання в світі.

У 1999-2000 рр. StatSoft представила поліпшені і спрощені програмні продукти, зокрема STATISTICA 5.5, включає такі найбільш повні інструменти аналізу, як GLM (Загальні лінійні моделі), GSR (Загальна ступінчаста регресія), GLZ (Узагальнені лінійні моделі), and PLS (Приватні найменші квадрати).

У 2001-2003 рр. презентація STATISTICA 6.0 ознаменувала появу нового покоління статистичних програмних продуктів. STATISTICA 6.0 - це високотехнологічний продукт, заснований на COM архітектурі, що володіє унікальними функціональними і корисними можливостями. STATISTICA 6.0 отримала найкращі відгуки серед клієнтів як ніяка інша версія за всю 17-летнюю2000 практику StatSoft.

На даний момент пакет STATISTICA має більш 500 тис. зареєстрованих користувачів у всьому світі і найбільш динамічно розвивається пакетом на ринку статистичного програмного забезпечення. Є версії системи німецькою, французькою, японською, іспанською, польською та іншими мовами. Користувачами системи є найбільші університети, дослідницькі центри, компанії, банки всього світу, державні установи.

**Основні можливості пакету STATISTICA**

Розглянемо можливості пакету на прикладі, припустимо, що нам дані статистичні дані з опиту населення, який включає в себе такі питання: стать, вік, сімейний стан, частота перегляду спортивних матчів та частота заняттям спортом. Необхідно побудувати графік перегляду спортивних матчів, частоту заняття спортом, відношення сімейного стану серед населення, а також знайти середнє значення віку опитуваних.

Спочатку відкриємо програму STATISTICA(інтерфейс програми показаний на рисунку 1) і створимо новий проект, натиснувши на кнопку new на панелі управління. Після цього відкриється вікно (рисунок 2), на якому вибрати number of variables = 5, після чого натиснути кнопку ОК, матимемо наступне(рисунок 3)

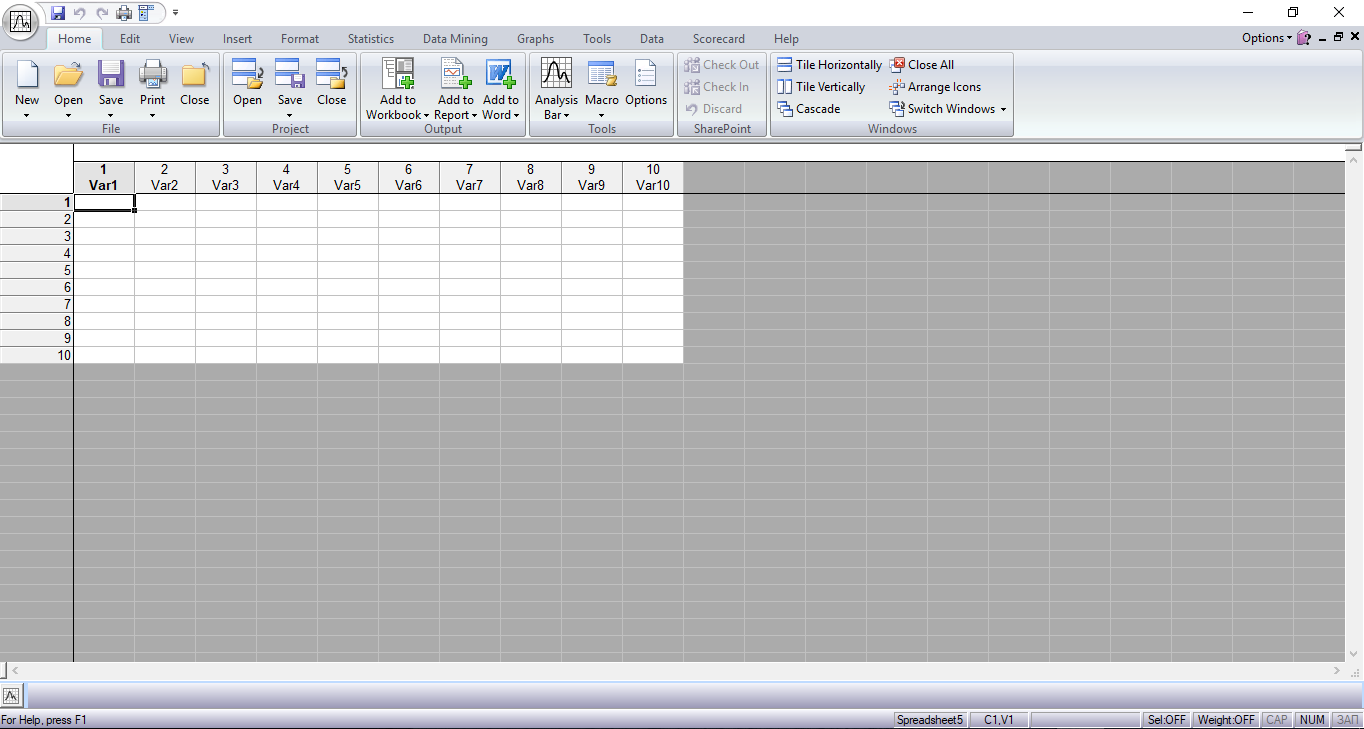


Рисунок 1 – Головний інтерфейс програми

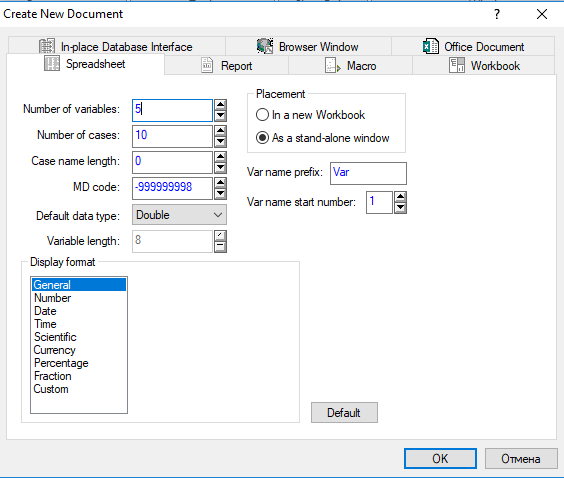


Рисунок 2 – Вікно для створення нового файлу

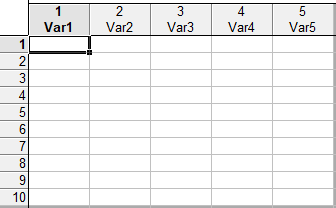


Рисунок 3 – Нова таблиця

Далі необхідно перейменувати стовпці «var1», «var2», «var3», «var4» та «var5» на «Стать», «Вік», «Сімейний стан», «Частота перегляду» та «Заняття спортом». Для цього натиснемо пкм на заголовок стовпця і із контекстного меню виберемо пункт variable specs у вікні якого в полі name введемо необхідні імена(рисунок 4).

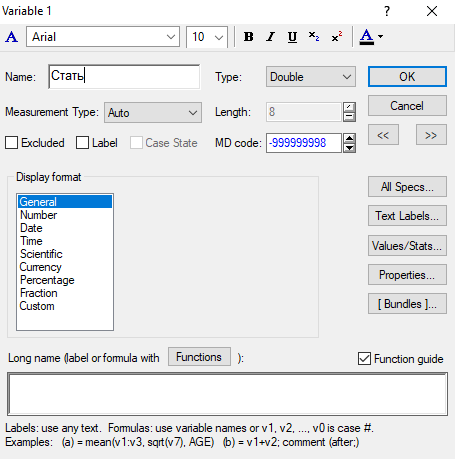


Рисунок 4 – Вікно параметрів змінної

Для того, щоб ввести всі необхідні дані нам потрібно створити ще декілька рядків. Для цього натиснемо пкм по номерам стовпчиків і виберемо пункт add cases, у цьому вікні, у полі how many написати 5, insert after – 10, матимемо наступне (рисунок 5)

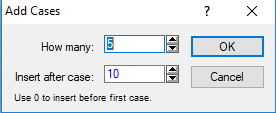


Рисунок 5 – Вікно добавлення стовпця

Для вводу значень для змінної «Стать» натиснемо пкм на заголовок колонки і виберемо variable specs->text lables. У цьому вікні введемо значення, як на рисунку 6

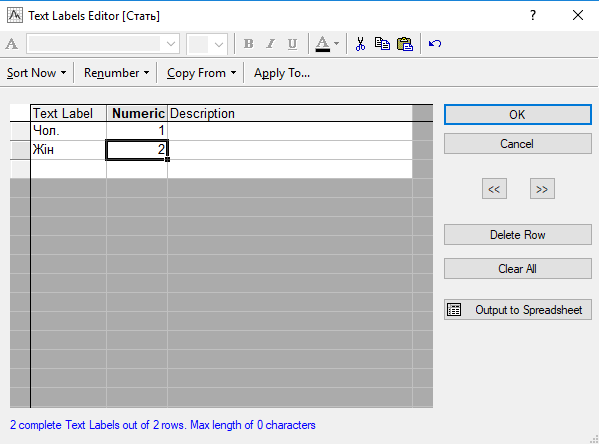


Рисунок 6 – Значення, які приймає змінна «Стать»

Для змінних «Сімейний стан», «Частота перегляду», «Заняття спортом» зробимо аналогічно. Дані введені у таблицю показані на рисунку 7.

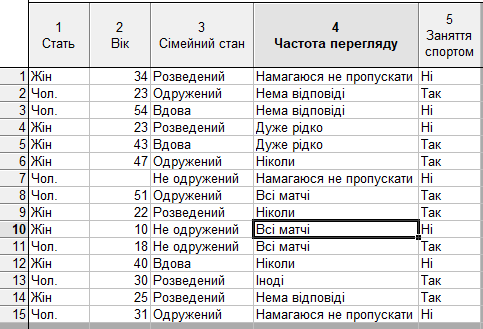


Рисунок 7 – Данні, введені до таблиці

Відобразимо графік перегляду спортивних матчів. Для цього перейдемо до вкладки Graphs на панелі інструментів. Виберемо пункт 2D->pie charts. У діалоговому вікні виберемо необхідну змінну, натиснувши кнопку variables – змінну 4.та натиснемо кнопку ОК. Результат цієї операції на рисунку 8.

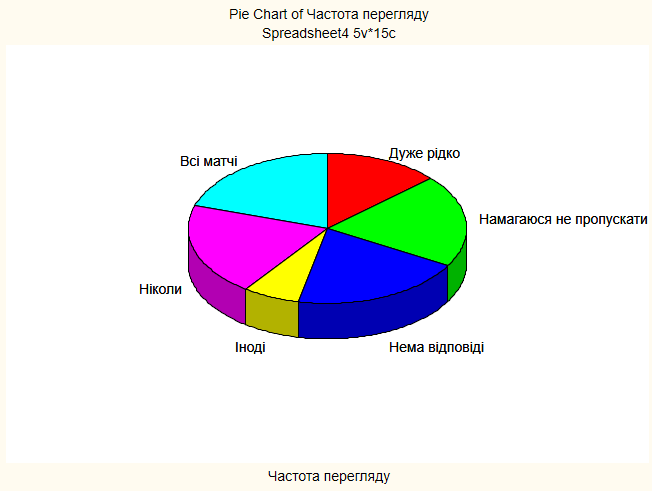


Рисунок 8 – Графік частоти перегляду спортивних матчів

Далі потрібно показати частоту заняттям спортом, для цього скористуємося графіком-гістограмою. Аналогічно у пункті Graphs->2D натиснемо на histograms і у діалоговому вікні виберемо змінну «Заняття спортом» потім перейдемо до пункту Advanced і у пункті Fill type натиснемо off, тепер можна натиснути на кнопку ОК. Результат показаний на рисунку 9.

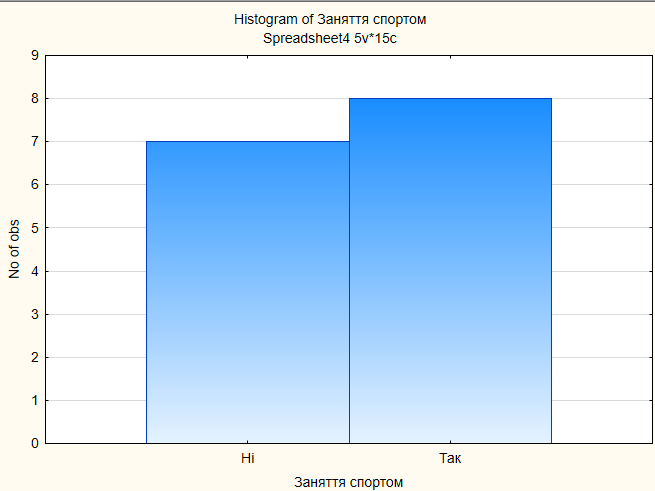


Рисунок 9 – Частота заняття спортом

Графік відношення сімейного стану будується аналогічно графіку частоти перегляду спортивних матчів. Графік показаний на рисунку 10.

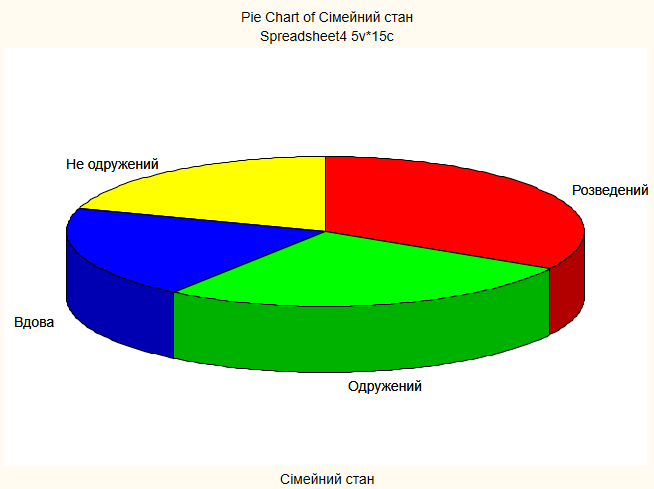


Рисунок 10 – Графік відношення сімейного стану серед населення

Тепер необхідно знайти середнє значення віку опитаних людей. Для цього виділимо змінну «Вік», натиснувши пкм виберемо пункт statistics of block data->block columns->means. Отримане значення показане на рисунку 11.



Рисунок 11 – Середнє значення віку серед опитуваних

**Висновки**

У даній лабораторній роботі було проведене знайомство з пакетом STATISTICA та було ознайомлено з основними функціями даного пакету.