

Рис. 1: Основні статистики і таблиці

Лекція № 2 Основні статистичні методи

Набір основних статистичних методів об'єднано в пункті меню Statistics – Basic Statistics/Tables. Серед них слід виділити (рис. 1):

- Описові статистики (Descriptive Statistics)
- Кореляційні матриці (Correlation Matrices)
- Т-тести (t-test, independent, by groups; t-test, independent, by variables; t-test, dependent samples; t-test single-sample)
- Однофакторний дисперсійний аналіз (Breakdown & one-way ANOVA)

Крім згаданих методів тут доступні методи побудови таблиць та калькулятори порівняння характеристик двох вибірок і ймовірнісний калькулятор.

1 Описові статистики

Описові статистики використовуються для визначення основних характеристик (середнього, дисперсії, максимального та мінімального значень і інших) генеральної сукупності. Визначення таких характеристик доступно з використанням Statistics – Basic Statistics/Tables – Descriptive Statistics. Вікно відповідних процедур містить закладки, що дозволяють вибрати потрібні статистики, методи досліджень та графічні представлення даних, а також деякі опції виводу інформації. Крім закладок маємо кнопки вибору змінних для аналізу (Variables), вибору випадків (Select Cases) та вагової змінної (w), а також спосіб видалення випадків з пропущеними значеннями змінних (MD deletion): Pairwise – видалається тільки той випадок, в якому є пропущене значення досліджуваної змінної, Casewise – видалається випадок, в якому є пропущене

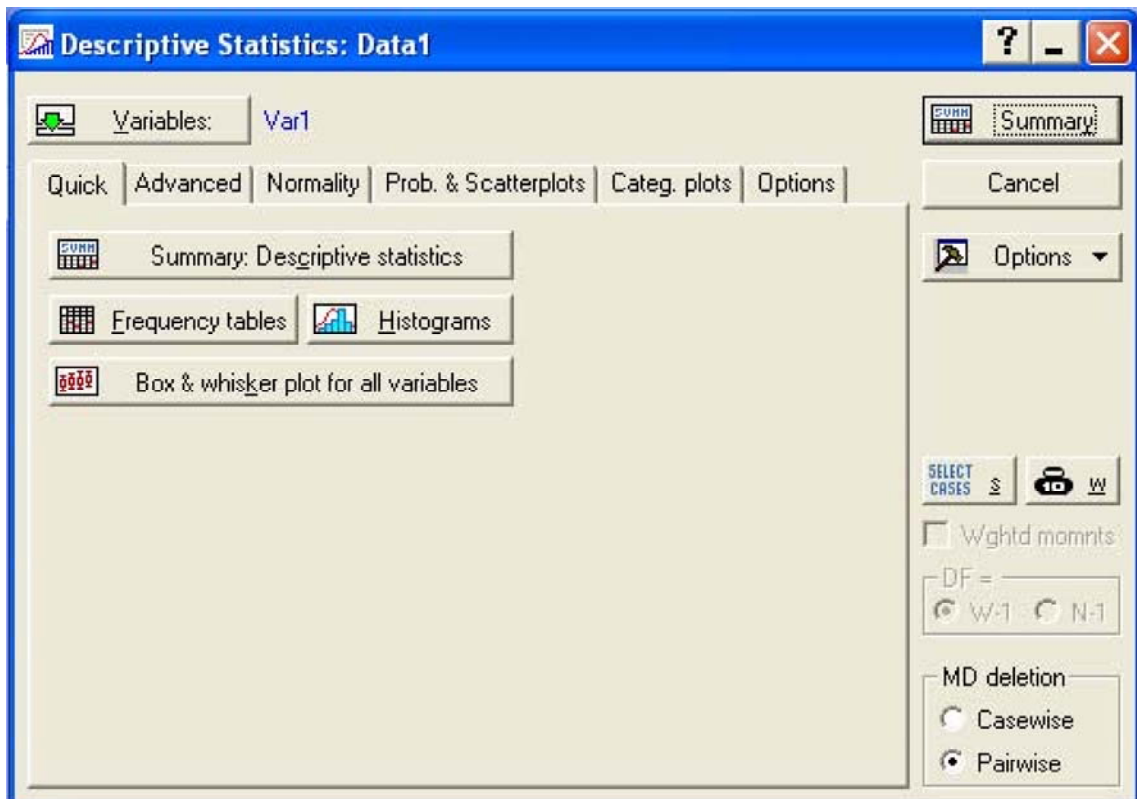


Рис. 2: Описові статистики: Quick

значення хоча б в одній вибраній змінній. Розглянемо детально згадане вікно. Перша закладка (рис. 2) (Quick) містить кнопки виводу результатів (Summary: Descriptive Statistics те ж саме, що і кнопка Summary), побудови таблиць частот (Frequency tables), гістограм (Histograms) та графіків (Box & whisker plot for all variables), що відображають середні значення і міри розкиду значень всіх вибраних змінних. Закладка Advanced (рис. 3) дозволяє вибрати статистики для обчислення. Це статистики положення – кількості (Location, valid N):

- Valid N — кількість наявних значень вибраних змінних;
- Mean — середнє значення (середнє арифметичне);
- Sum — сума значень змінних;
- Median — медіана вибірки (середина варіаційного ряду);
- Mode — мода (значення, що найчастіше зустрічається);
- Geom. mean — середнє геометричне;
- Harm. mean — середнє гармонічне;

розкиду – моментів (Variation, moments):

- Standard Deviation — стандартне відхилення (середньоквадратичне відхилення);
- Variance — дисперсія (квадрат стандартного відхилення);

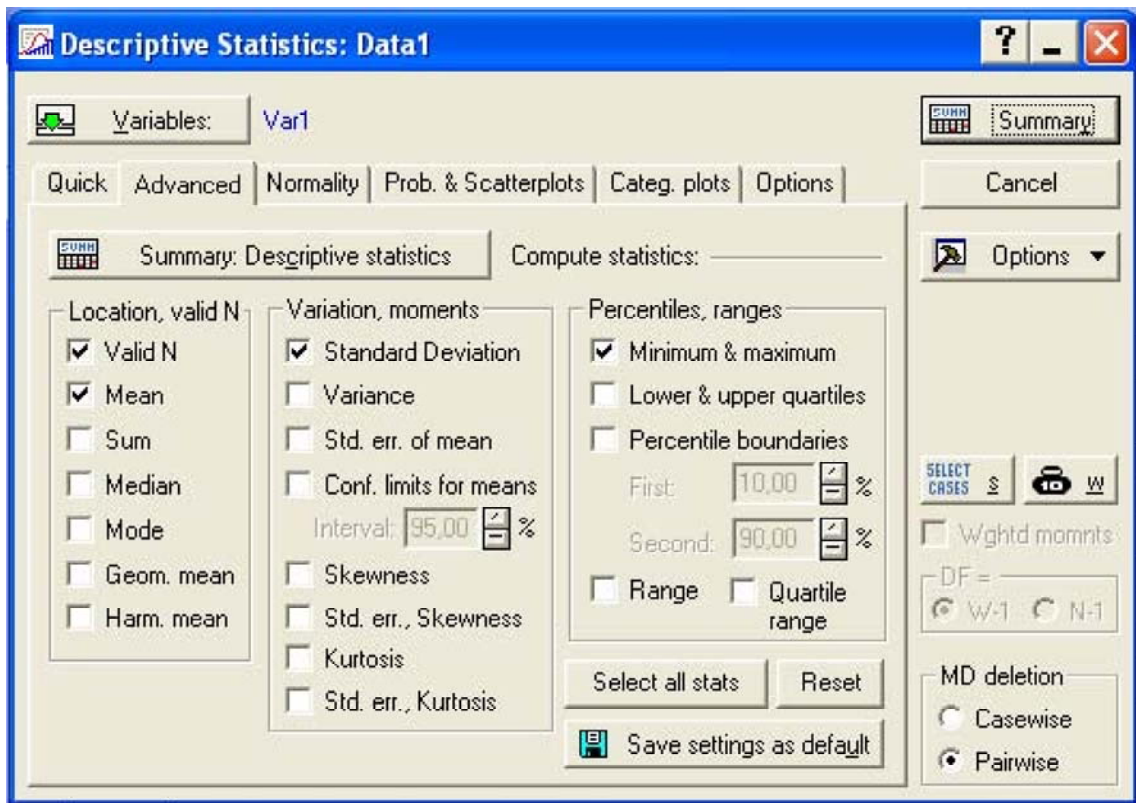


Рис. 3: Описові статистики: Advanced

- Std. err. of mean — стандартна помилка середнього (стандартне відхилення поділене на корінь квадратний з об'єму вибірки);
- Conf. limits for means — межі довірчого інтервалу для середнього значення (є можливість вибрати рівень надійності — Interval: %);
- Skewness — вибірковий коефіцієнт асиметрії (характеризує несиметричність розподілу);
- Std. err., skewness — стандартна помилка коефіцієнта асиметрії;
- Kurtosis — вибірковий коефіцієнт ексцесу (характеризує гостровершинність розподілу);
- Std. err., kurtosis — стандартна помилка коефіцієнта ексцесу;

та порядку (Percentiles, ranges):

- Minimum & maximum — найбільше та найменше вибіркове значення;
- Lower & upper quartiles — 25%-ний та 75%-ний квантилі (вбіркове значення, або середнє арифметичне двох сусідніх значень, яке є границею 25%-ної чи 75%-ної частини варіаційного ряду);
- Percentile boundaries — значення квантилів із заданими порядками (в %);
- Range — розмах варіації (різниця між найбільшим та найменшим вибірковими значеннями);
- Quartile range — різниця між верхнім (75%-ним) та нижнім (25%-ним) квантилями.

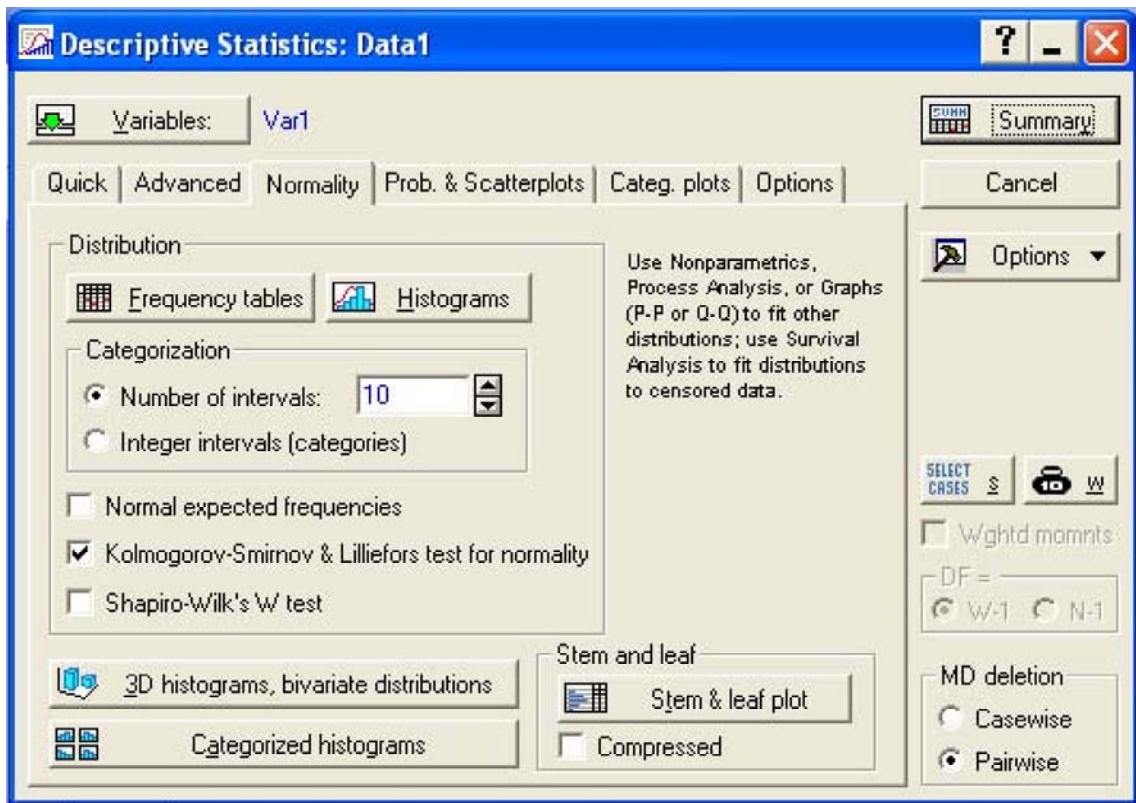


Рис. 4: Описові статистики: Normality

Тут же маємо кнопки вибору всіх статистик (Select all stats), відновлення налаштувань за замовчуванням (Reset) та збереження поточних налаштувань для використання за замовчуванням (Save settings as default).

Закладка Normality (рис. 4) містить процедури перевірки нормальності розподілу (Distribution) вибраних змінних. Якщо вибрати потрібну кількість інтервалів розбиття (Number of intervals) варіаційного ряду, то стають доступними критерії Колмогорова-Смірнова і Ліллієфорса (Kolmogorov-Smirnov & Lilliefors test for normality) та Шапіро-Уїлкса (Shapiro-Wilk's W test) перевірки гіпотези про нормальність. Значення цих критеріїв та максимальний рівень значущості p , на якому гіпотеза про нормальність підтверджується відображаються на гістограмі (створюється кнопкою Histograms) чи таблиці частот (створюється кнопкою Frequency tables). Опція Normal expected frequencies додає в таблицю частот прогнозовані значення частот нормального розподілу з оціненими за вибіркою параметрами. Крім того на закладці розміщені кнопки побудови тривимірних гістограм для двовимірних розподілів (3D histograms, bivariate distributions) та категоризованих гістограм (Categorized histograms). Особливим способом відображення частот є таблиця Stem and leaf.

Графічні процедури опису даних зібрані на закладках Prob. & Scatterplots і Categ. plots (рис. 5, 6). Ліворуч на першій з них маємо групу кнопок побудови графіків (двовимірних (з регресійним аналізом) — 2D scatterplot, тривимірних — 3D scatterplot, категоризованих — Categorized scatterplot, поверхонь — Surface plot, матричних — Scatterplot matrix). Справа розміщені кнопки побудови ймовірнісних графіків (Normal..., Half normal..., Detrended normal probability plot), які відображають графічний метод перевірки гіпотези про вид розподілу змінної. Наступна закладка містить кнопки для створення категоризованих графіків — однотипних графіків, побудованих за окремими групами (категоріями) значень. Категорії визначаються значеннями однієї чи кількох номінальних змінних. Новими типами графіків тут є Categorized box &

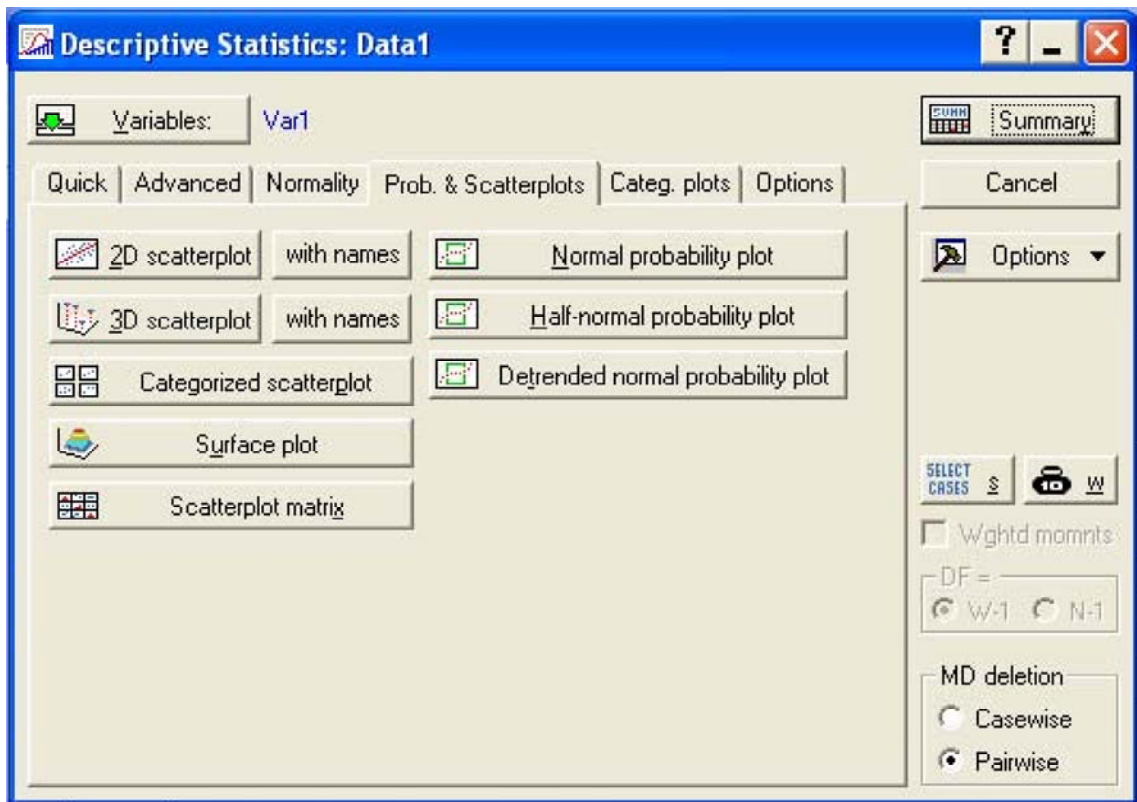


Рис. 5: Описові статистики: Prob. & Scatterplots

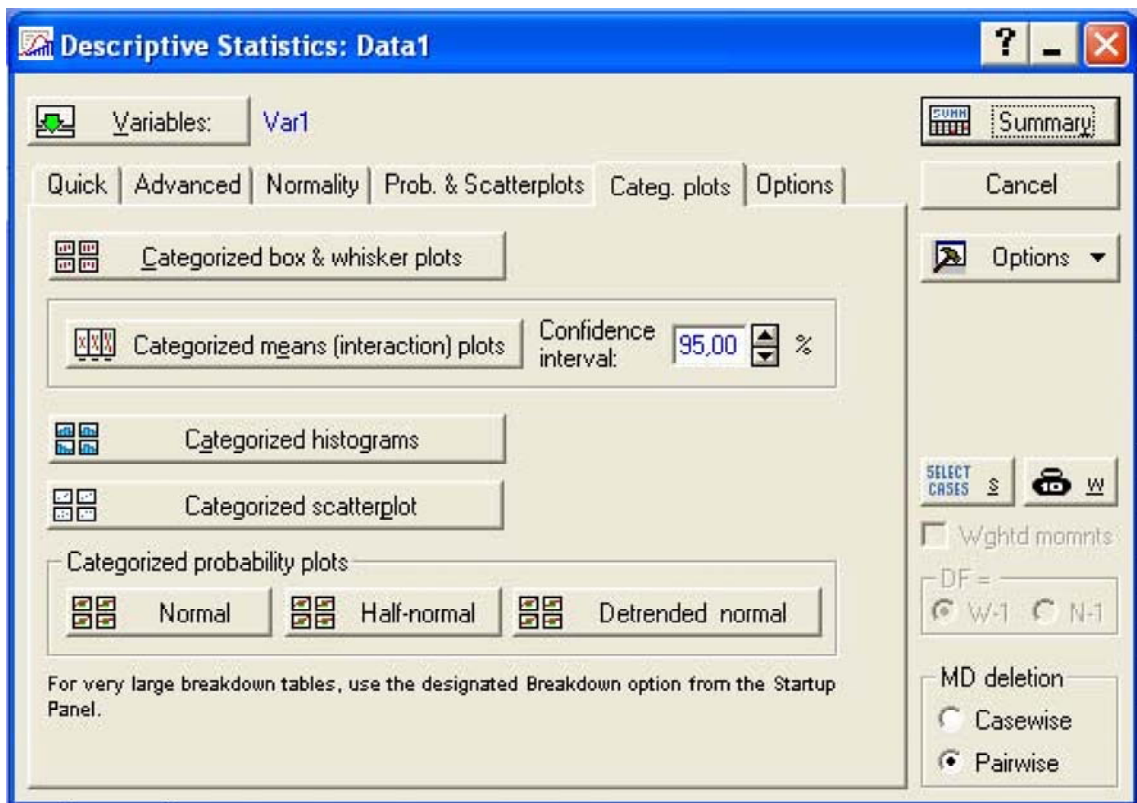


Рис. 6: Описові статистики: Categ. plots

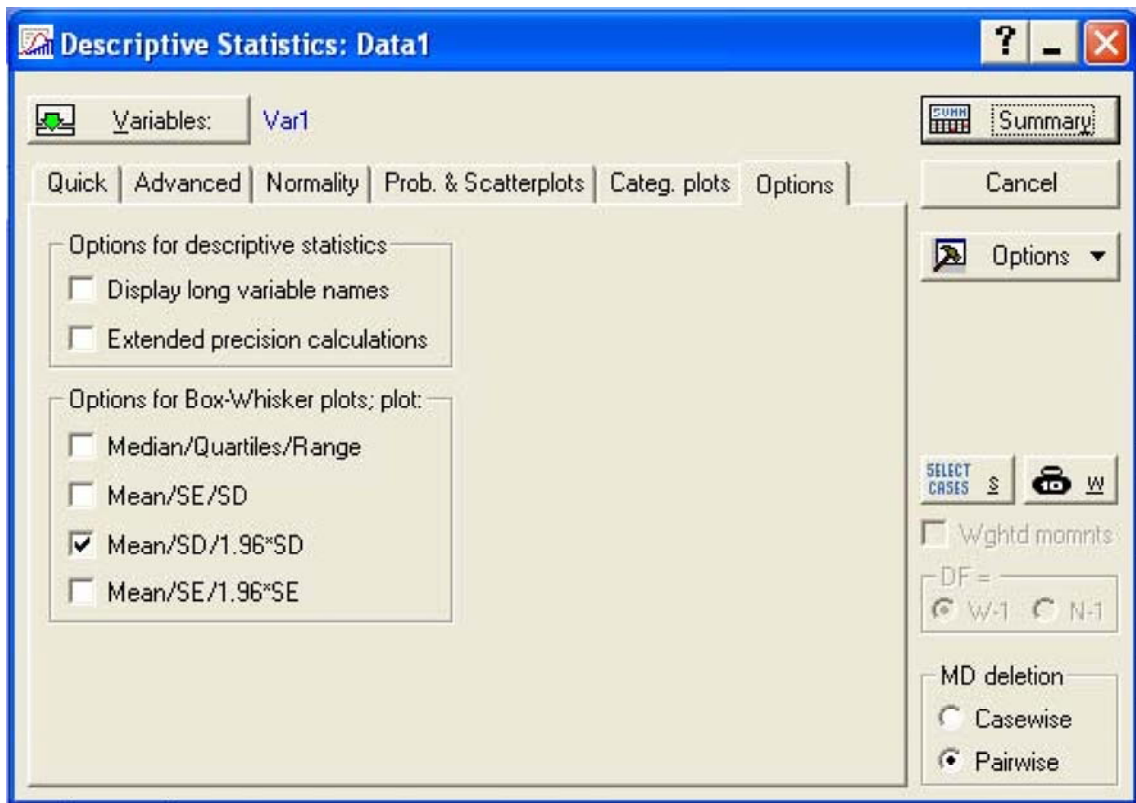


Рис. 7: Описові статистики: Options

whisker plots, який відображає оцінки середніх значень та мір розкиду значень за категоріями, та Categorizer means (interactions) plots, на якому зображуються точкові та інтервальні оцінки середніх значень за категоріями.

На останній закладці — Options — можемо (рис. 7) встановити опції для виводу результатів обчислень (Options for descriptive statistics) та для графіків типу "вусатих коробок" (Options for Box-Whisker plots). В першому випадку маємо можливість разом з назвою змінної виводити коментарі (чи формули) до змінної (Display long variable names) та використовувати обчислення підвищеної точності (Extended precisions calculations).

2 Кореляційні матриці

Визначення оцінок коефіцієнтів кореляції Пірсона може бути виконане за допомогою пунктів Statistics – Basic Statistics/Tables – Correlation Matrices. В цьому вікні розміщені кнопки вибору типу матриці кореляцій: квадратна кореляційна матриця (One variable list), яка складається з оцінок коефіцієнтів кореляцій та їх характеристик для кожної пари змінних з одного списку; прямокутна (Two list (rect. matrix)), яка містить кореляції кожної змінної з одного списку з кожною змінною з іншого списку (списки можуть перетинатися). На закладках цього вікна розміщені кнопки виводу результатів (рис. 8, 9) та опції змісту результатів (рис. 10).

Зупинимось тільки на опціях представлення кореляційної матриці. На закладці Options маємо можливість вибрати спосіб подання результатів — Display format for correlation matrices. Опція Display simple matrix (highlight p's) дозволяє виводити матрицю коефіцієнтів кореляції та виділяти кольором значимі коефіцієнти, Display r, p-levels, and N's крім коефіцієнтів кореляції виводить рівні значущості та об'єми вибірок, Display detailed table of results виводить оцінки середнього, стандартного відхилення кожної змінної, оцінки коефіцієнтів кореляції кожних двох

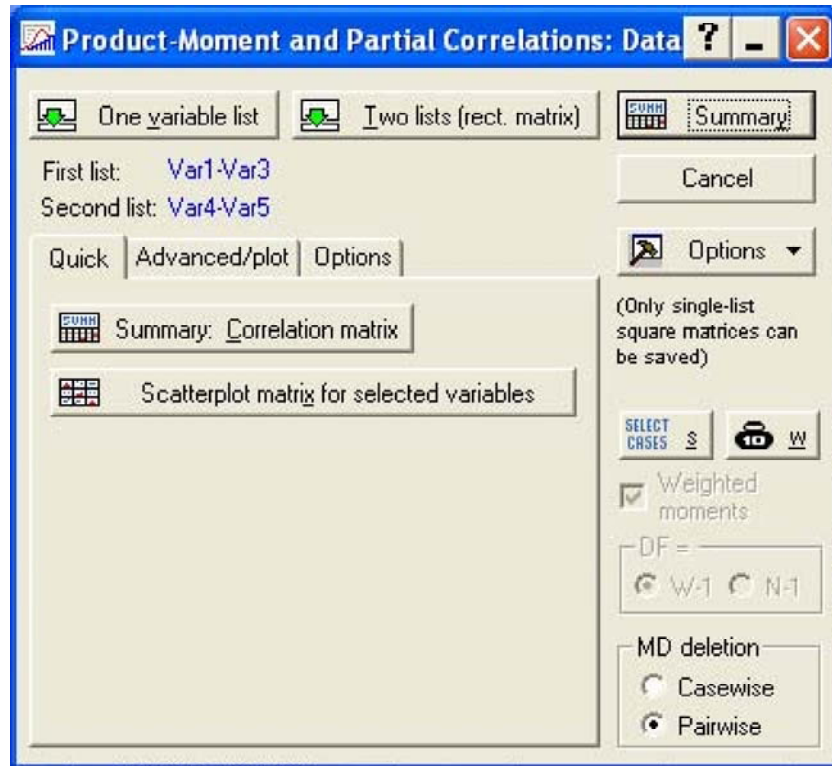


Рис. 8: Кореляційні матриці: Quick

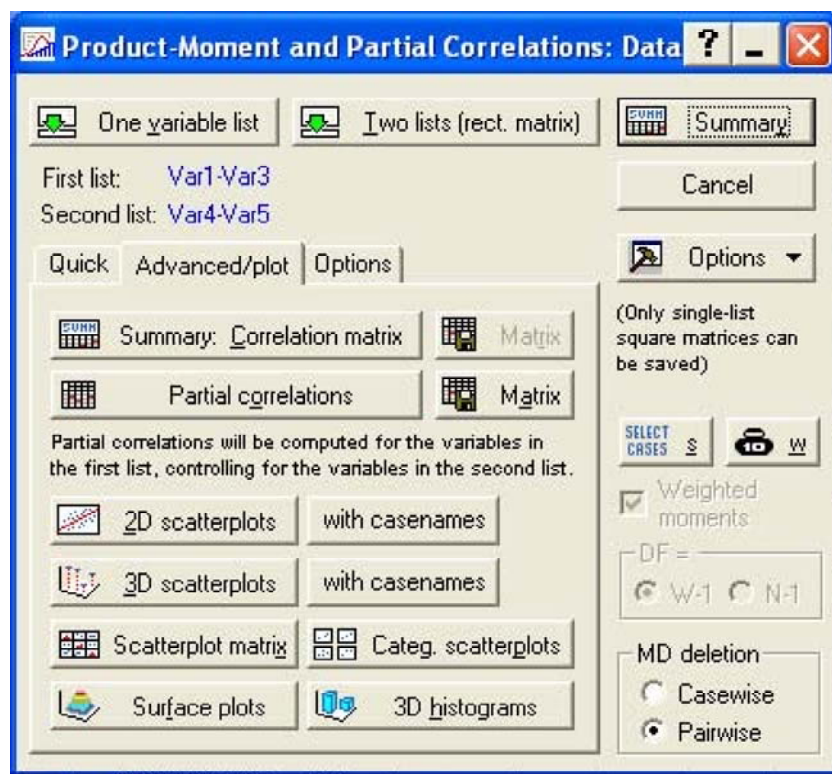


Рис. 9: Кореляційні матриці: Advanced/plot

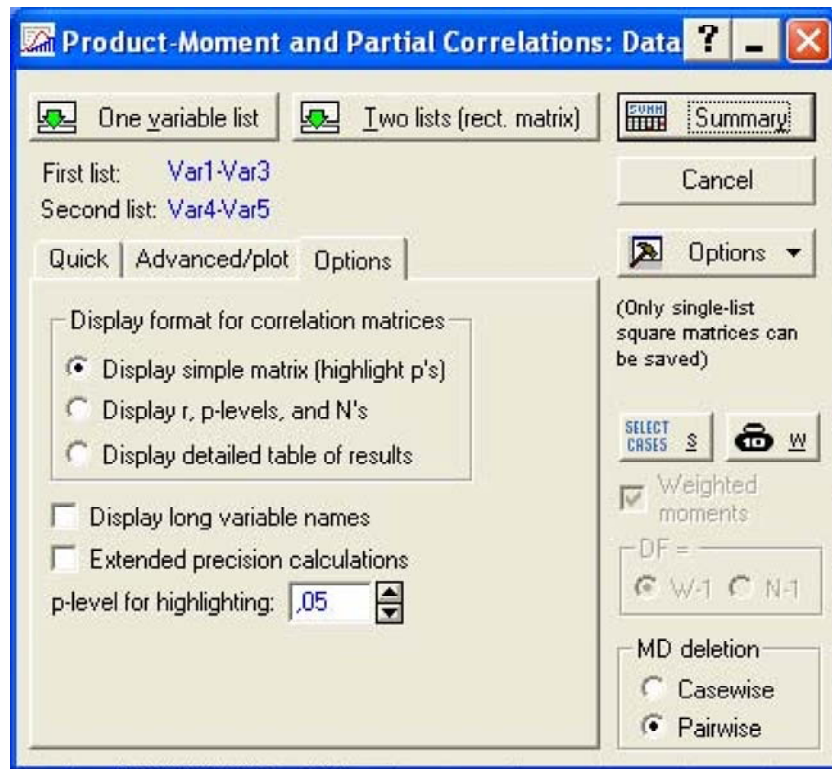


Рис. 10: Кореляційні матриці: Options

змінних, їх квадрати, значення статистики критерію щодо значимості коефіцієнта кореляції, його значимість, оцінки параметрів лінійної регресії, що задають залежність однієї змінної від іншої. Крім того на закладці є можливість встановити граничне значення рівня значущості (p-level for highlighting) при яких коефіцієнти кореляції будуть позначатись як значимі.

- 3 Т-тести перевірки гіпотез про рівність середніх та стандартних відхилень
- 4 Однофакторний дисперсійний аналіз