

ПРАКТИЧНО-СЕМІНАРСЬКЕ ЗАНЯТТЯ 3. Побудова двовимірних статистичних моделей на матриці вихідних даних.

Третім завданням практичної роботи було визначення пар показників, які пов'язані між собою, визначення кількісної оцінки сили зв'язку між ознаками.

Двовимірний наліз – містить дві випадкові величини, утворює двовимірну модель та аналізує дві величини. В ході практичної роботи вивчаються такі двовимірні моделі як: кореляційний та коваріаційний аналізи, які виконується в програмі MINITAB 14. шляхом «Stat» → «Basic Statistics» → «Correlation».

Кореляційний аналіз – метод обробки статистичних даних, що полягає у вивченні тісноти зв'язку між змінними; при цьому порівнюються коефіцієнти кореляції між однією парою або множиною пар ознак для встановлення між ними статистичної взаємодії, що дозволяє виявити найбільш пов'язані змінні та побудувати поверхні їх залежності.

Мета кореляційного аналізу – забезпечити отримання деякої інформації про одну змінну за допомогою іншої змінної. В випадках, коли можливе досягнення мети, говорять, що змінні корелюють. В самому загальному вигляді сприйняття гіпотези про наявність кореляції означає, що зміна значення змінної А відбудеться одночасно з пропорційною зміною значення В. Мірою залежності між експериментальними наборами даних є числа – коефіцієнти зв'язку.

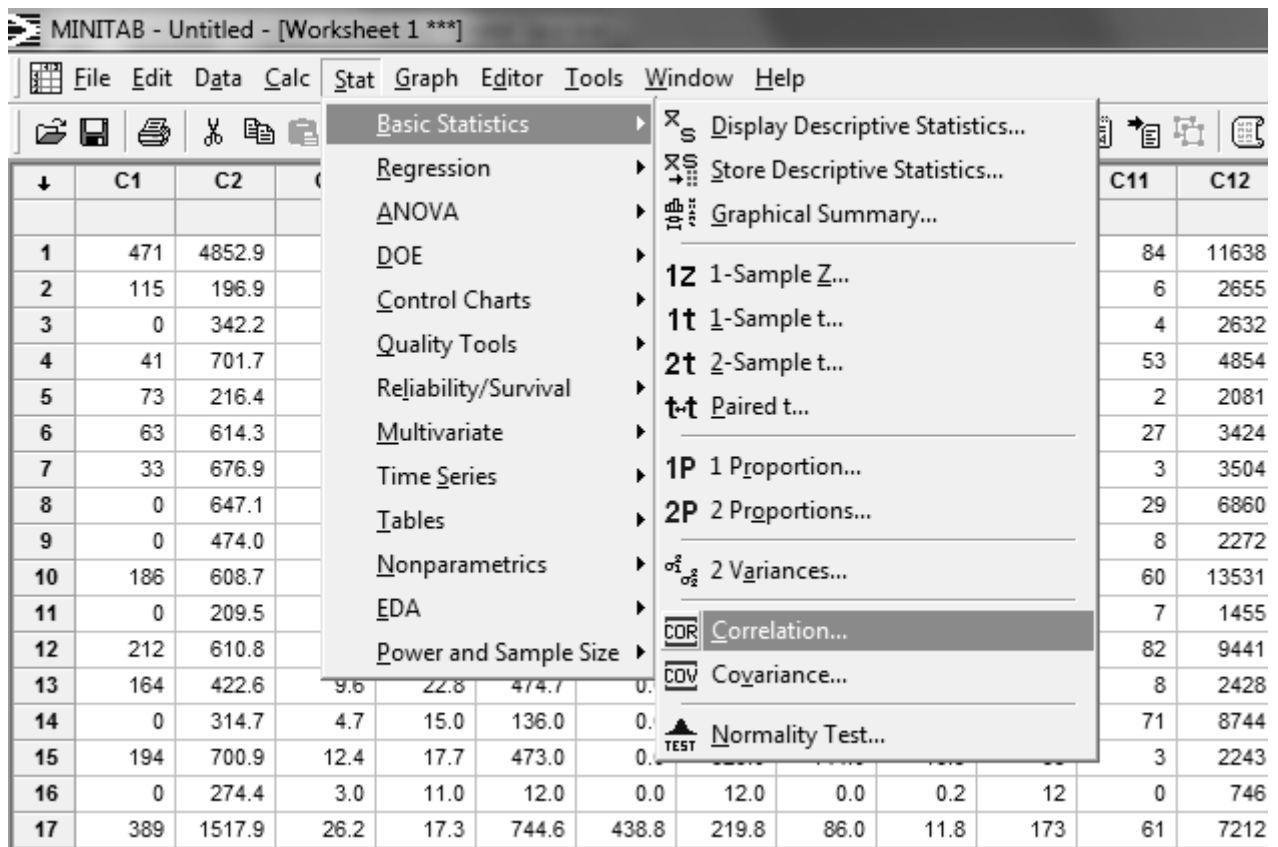


Рис. 3. Алгоритм вибору кореляційного аналізу в програмі MINITAB 1.

Головні завдання кореляційного аналізу:

- 1) оцінка за вибірковими даними коефіцієнтів кореляції;
- 2) перевірка значущості вибіркових коефіцієнтів кореляції або кореляційного відношення;
- 3) оцінка близькості виявленого зв'язку до лінійного;
- 4) побудова довірчого інтервалу для коефіцієнтів кореляції.

Визначення сили та напрямку взаємозв'язку між змінними є однією з важливих проблем аналізу даних. В загальному випадку для цього застосовують поняття кореляції. Кореляція є залежністю двох випадкових величин. При цьому, зміна однієї або кількох цих величин призводить до систематичної зміни іншої або інших величин. Математичною мірою кореляції двох випадкових величин слугує коефіцієнт кореляції.

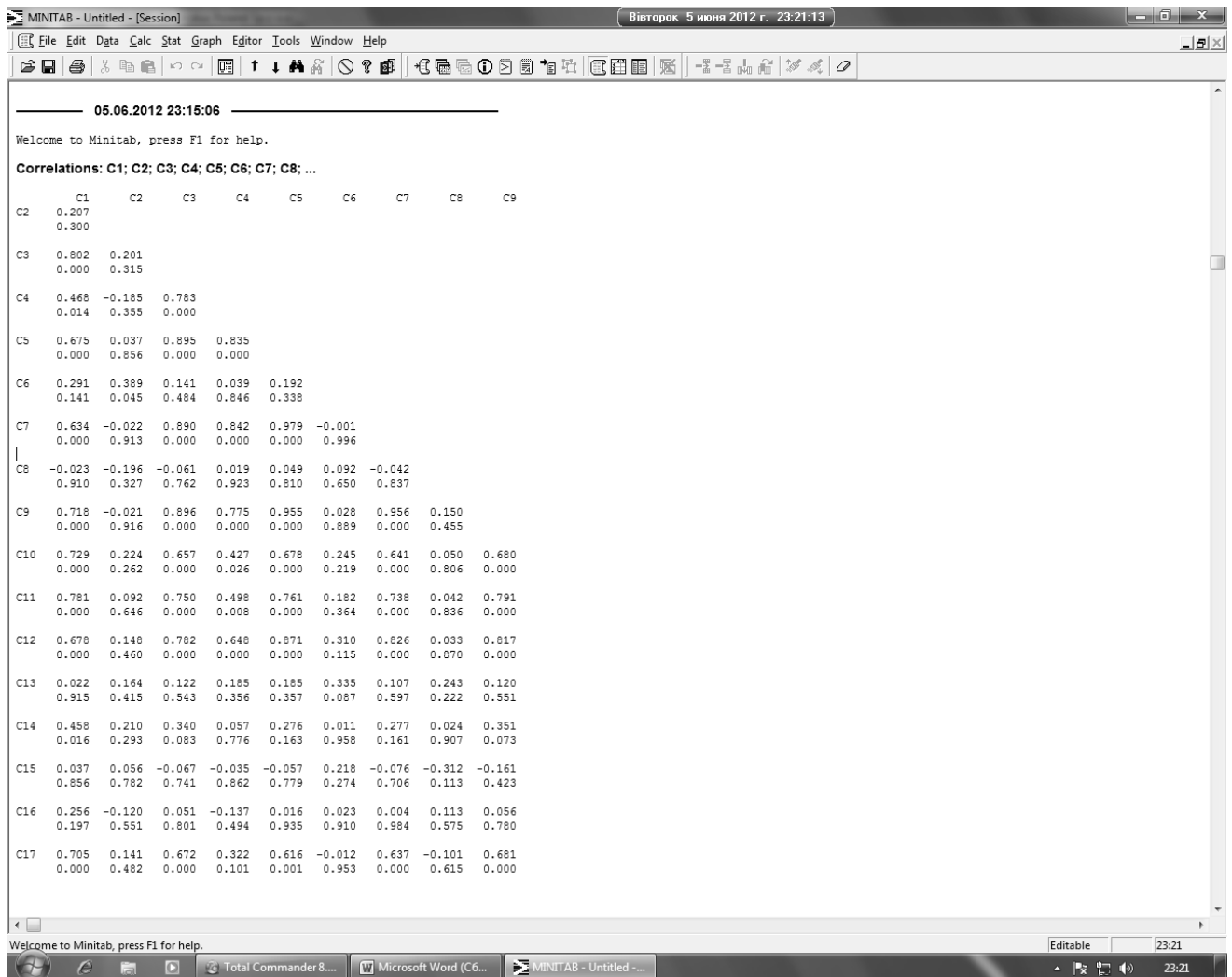


Рис. 4. Кореляційна матриця

Коефіцієнт кореляції приймає значення в межах від -1 до +1. Знак «мінус» означає обернену залежність, «+» – пряму. Якщо коефіцієнт дорівнює нулю, то лінійний зв'язок між динамічними рядами є відсутнім, а якщо одиниці - наявна функціональна залежність. Досить часто помилково вважається, що значення коефіцієнта кореляції менше 0.2 свідчить про відсутність зв'язку. Насправді він свідчить про наявність дуже слабкого зв'язку, але зв'язок між показниками є.

Коефіцієнт
кореляції
> ± 0.91
± 0.71-0.90
± 0.51-0.70
± 0.21-0.50
< ± 0.20

Тіснота зв'язку
Дуже сильна
Сильна
Помірна
Слабка
Дуже слабка

Коефіцієнт кореляції, а в загальному випадку кореляційна функція, дозволяють встановити ступінь взаємозв'язку між змінними. Кореляція може бути лінійною або нелінійною в залежності від типу залежності, яка фактично існує між змінними. Досить часто на практиці розглядають тільки лінійну кореляцію (взаємозв'язок), але більш глибокий аналіз потребує використання для дослідження процесів нелінійних залежностей. Складну нелінійну залежність можна спростити, але знати про її існування необхідно для того, щоб побудувати адекватну модель процесу.

У випадку максимальної тісноти зв'язку між показниками на діаграмі розсіяння їх залежність буде представлена прямою лінією. Іноколи зустрічаються і з таким явищем як псевдокореляція, тобто кореляція, що обумовлена впливом інших показників, які залишилися поза увагою дослідника. Значимість коефіцієнту кореляції залежить від довжини динамічних рядів. У великих динамічних рядах навіть слабкі залежності будуть значимими, в той час, як у незначних – навіть дуже сильні залежності не є статистично надійними. Тому ведуть мову про надійність кореляційної залежності, яка пов'язана із репрезентативністю вихідних динамічних рядів і свідчить про те, наскільки ймовірно, що виявлена залежність, знову буде виявлена при збільшенні періоду ретроспекції чи екстраполяції на майбутнє. Надійність виявлених залежностей оцінюється за допомогою стандартної статистичної міри – p -рівня – статистичного рівня значимості. Даний показник знаходиться у оберненій залежності до надійності результату: чим вищим є p -рівень, тим нижчою є статистична надійність виявленої залежності, і навпаки.

Коваріація – в теорії ймовірностей та математичній статистиці, числова характеристика залежності випадкових величин. Сутність коваріації полягає в тому, що вона виникає внаслідок невизначеності результату перемножування двох сукупностей чисел.

Побудова коваріаційної матриці досить подібна з побудовою матриці кореляції, виконується в програмі MINITAB 14, шляхом «Stat» → «Basic Statistics» → «Covariance». В порівнянні з кореляційним зв'язком коваріаційний не вказує на тісноту зв'язку між певними групами факторів він лише вказує на результат перемножування двох сукупностей чисел.