

Теми практичних занять

№ з/п	Практ. робота	Назва теми	Кількість годин	Оцінка бали
1	ПР1	Усвідомлення геологічної і математичної суті задач моделювання.	2	4
2		Підготовка вихідних даних, формування бази даних на комп'ютері.	2	
3		Постановка задач моделювання на конкретному фактичному матеріалі	2	
4	ПР2	Побудова одновимірних статистичних моделей векторів змінних. Трансформація законів розподілу.	4	4
5	ПР3	Побудова двовимірних статистичних моделей на матриці вихідних даних. Двовимірний регресійно-кореляційний аналіз.	6	4
6	ПР4	Кластер – аналіз вихідних даних. Інтерпретація отриманих результатів.	4	4
7		Факторний аналіз вихідних даних (метод головних компонент). Інтерпретація отриманих результатів.	4	
8	ПР5	Моделювання просторових змінних методом ковзного статистичного вікна. Аналіз та інтерпретація результатів.	4	4
9		Оформлення результатів та написання звітів.	4	
		Разом	32	20

Примітка: практичні заняття 2 - 9 здаються студентами особисто у вигляді звіту (10-12 с.) і оцінюються викладачем індивідуально.

Практичне заняття III.

Усвідомлення геологічної і математичної суті задач моделювання..
Підготовка вихідних даних, формування бази даних на комп'ютері.
Постановка задач моделювання на конкретному фактичному матеріалі

Питання для обговорення

1. Суть моделювання
2. Формулювання конкретного об'єкта та предмету моделювання
3. Поняття «статистичний показник»
4. Поняття «матриці даних»
5. Вимоги до складання матриці даних

6. Відносні та абсолютні показники

Рекомендації до підготовки та проведення заняття.

1. Студенти отримують чи збирають статистичний матеріал для моделювання
2. Побудова інформаційної бази даних по статистичній сукупності.

Ресурс часу: 6 годин.

Критерії оцінювання (максимум 4 балів)

- Виконання практичної роботи – 4 балів

Практичне заняття П2.

Побудова одновимірних статистичних моделей векторів змінних.
Трансформація законів розподілу.

Питання для обговорення

1. Сутність одновимірного аналізу
2. Нормальний розподіл випадкових величин
3. Мода
4. Медіана
5. Центральний момент першого, другого, третього порядку
6. Види середнього значення

Рекомендації до підготовки та проведення заняття.

Наступним завданням практичної роботи є виявлення залежності динаміки певних показників від часу, включає методи згладжування і аналітичного виміру. Згладжування базується на процедурі визначення усередненої траєкторії розвитку показників в минулому і її продовження на майбутнє. Аналітичне вимірювання передбачає підбір математичної функції, яка найкращим чином відображає тенденцію динаміки показника в часі, і розрахунок на його основі прогнозних значень даного показника. Виконується за допомогою одновимірного аналізу, здійснюється в програмі STATISTICA

Одновимірний аналіз обробка однієї випадкової величини. В основі даного аналізу є дослідження властивостей і характеристик рядів випадкових величин, встановлення приналежності цих рядів до певного теоретичного розподілу, визначення подібності або відмінності даного ряду в порівнянні з іншими рядами.

Основними вихідними складовими одновимірного аналізу є поняття: простий статистичний ряд (сукупність), упорядкований статистичний ряд (ранжируваних), варіаційний ряд, генеральна сукупність і вибірка, обсяг вибірки, частота і ймовірність, закон розподілу, критерії однорідності та інші.

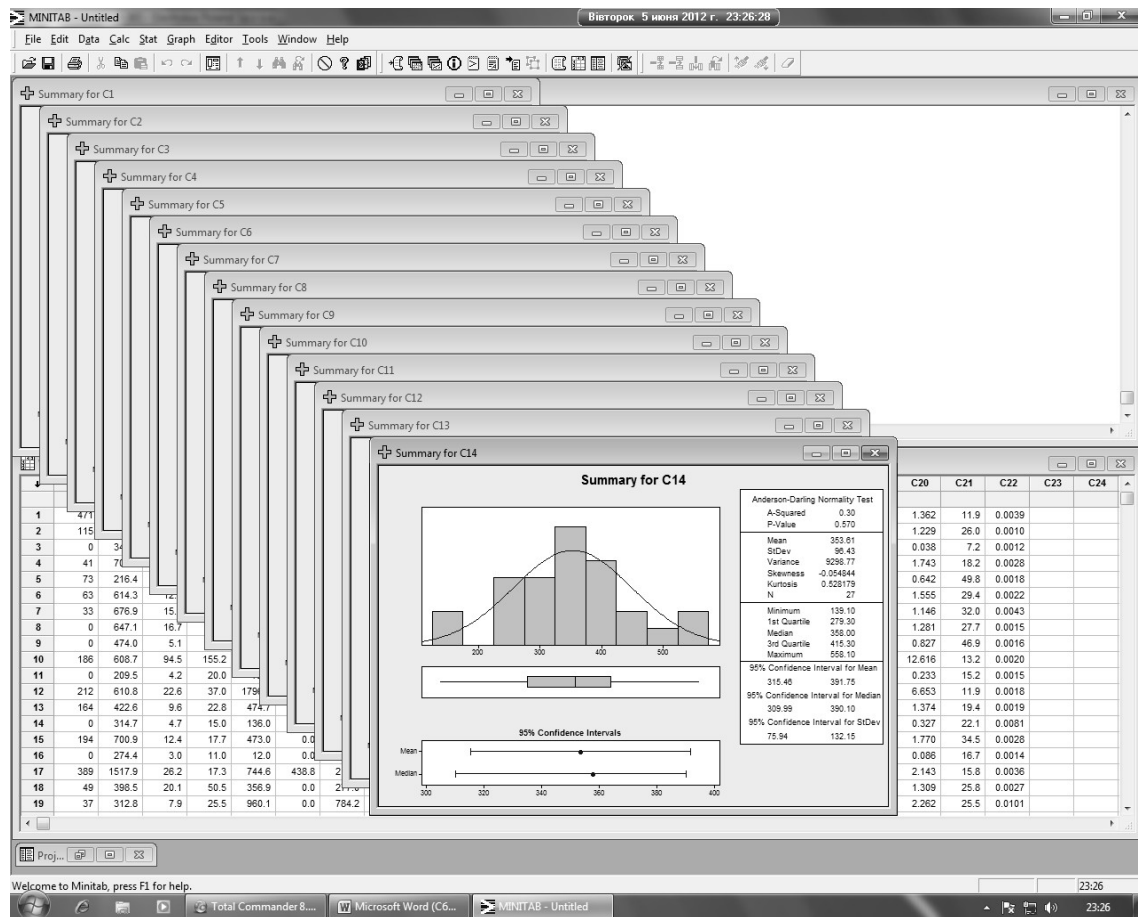


Рис. 1. Нормальний розподіл статистичного показника

Сукупність випадкових величин, отриманих при однакових умовах є результатом проведення спостереження, дослідів, експериментів і називається простою статистичною сукупністю або статистичним рядом. Статистичний ряд це первинна форма запису статистичного матеріалу у вигляді таблиці. Для наочного уявлення вихідного матеріалу за даними статистичного ряду будується графік змін значень цієї величини в часі або в просторі або в хронологічній послідовності.

Генеральна сукупність – це нескінченне або кінцеве число елементів або компонентів, що складаються з якісно однорідних показників. Будь-яку частину генеральної сукупності, відібрану за певними правилами, яка характеризує генеральну сукупність, називають статистичної вибіркою. Вибіркову сукупність (вибірку) створюють зазвичай для полегшення обробки інформації.

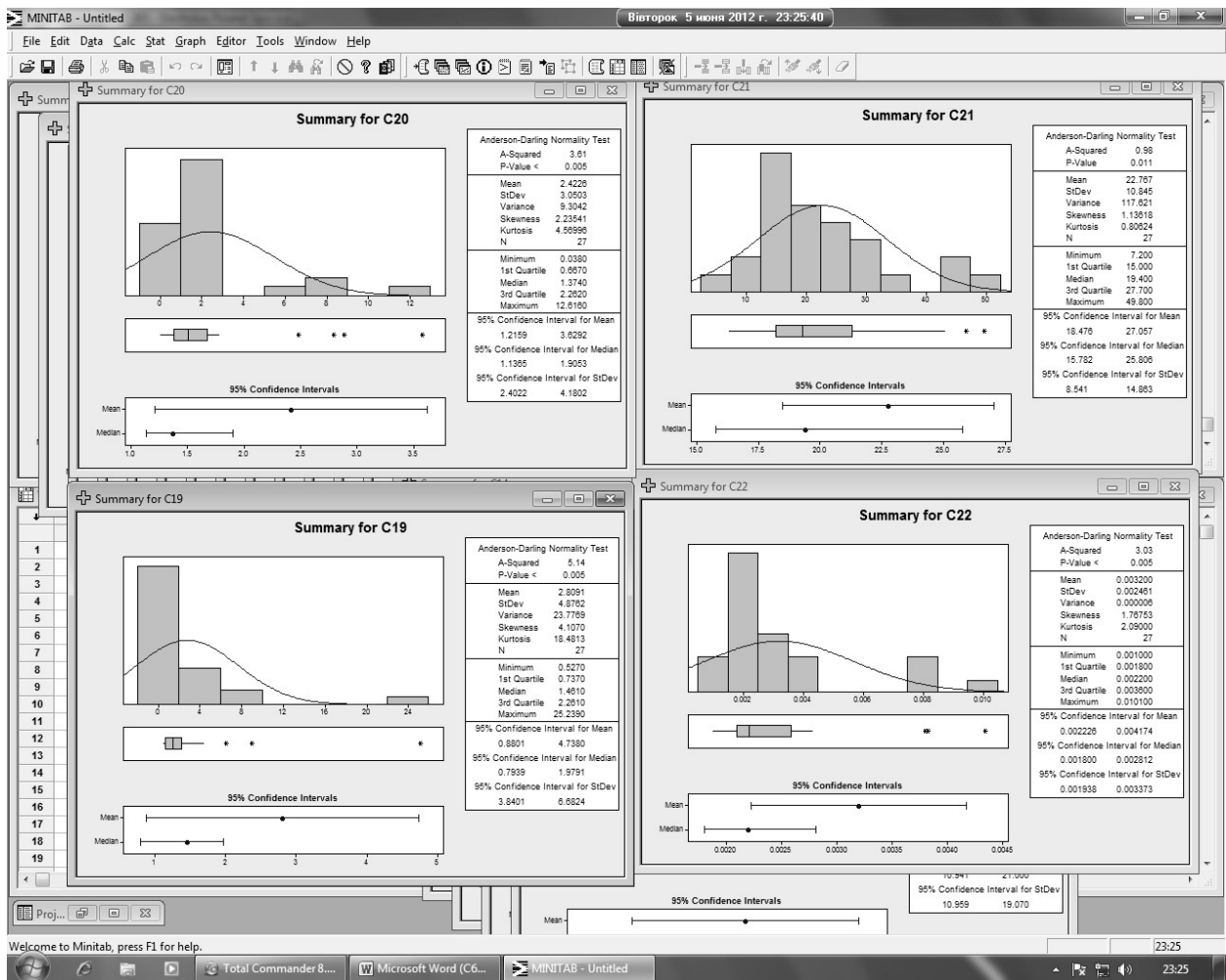


Рис. 2. Приклад одновимірного аналізу статистичних показників

Одним з найбільш складних і важливих питань одновимірного статистичного аналізу є визначення кількості спостережень в дослідженнях для отримання надійного уявлення про характер мінливості ознаки, що вивчається, в генеральній сукупності. Звичайно оптимальний обсяг вибірки пропорційний ступеню мінливості ознаки. Якщо ознака сильно змінюється, то кількість вимірювань слід збільшити. Величину вибіркової сукупності при виконанні географічних досліджень можна визначити двома способами: по таблиці досить великих чисел; розрахунковим способом. В обох випадках кількість спостережень (чисельність або обсяг вибірки) визначається, виходячи з довірчої ймовірності.

Систематизація та впорядкування даних, що представляють статистичну сукупність. Приведення їх у певну систему і характеристика цієї системи. Будь-яка статистична сукупність характеризується обсягом ряду і абсолютної частотою повторення однакової ознаки. Частота \square - це число, яке показує скільки разів зустрічається дана ознака в досліджуваній сукупності. Закон зміни частоти \square це і є закон розподілу. Існує три способи подання закону розподілу: табличний, графічний і аналітичний. Результати обробки вихідних даних, як правило, спочатку завжди оформляються у вигляді статистичних таблиць. Така форма дозволяє надати матеріалу

зручність, компактність і раціональність.

Спосіб, який дозволяє в наочній формі отримати уявлення про закономірності розподілу, називається графічним зображенням варіаційного ряду. Існує кілька способів графічного зображення рядів розподілу. При цьому зазвичай використовують дані таблиці емпіричного розподілу. При графічному зображенні рядів розподілу на горизонтальній осі відкладаються значення інтервалів або спостережень значення випадкової величини, а на вертикальній осі \square частоти. Для наочного уявлення варіаційного ряду частіше використовують графічні зображення у вигляді гістограми, полігону, кумуляти, кривою концентрації Лоренса і інші. Гістограма наочно показує розподіл досліджуваних величин, через що подібний спосіб вже часто використовується для ілюстрації особливостей статистичного розподілу. Варіаційний ряд при цьому зображується у вигляді стовпчиків, кордони між якими проходять по координатах, відповідних кордонів між класами. При цьому підстава стовпчиків по ширині дорівнює величині інтервалу ознаки, а висота пропорційна частоті окремих класів. Розглянемо основні статистичні характеристики варіаційних рядів. Одним з основних параметрів статистичного ряду є середнє значення ознаки або центр, щодо якого розподіляються члени сукупності. Значення середніх при вивченні різного роду закономірностей географічних явищ, процесів і об'єктів дуже велике. Вони дозволяють: визначити загальну тенденцію розвитку явищ; оцінювати значення окремої величини шляхом порівняння її з середньою, визначити наявність зв'язку між явищами за допомогою аналізу середніх двох або декількох ознак по територіях або часових проміжків.

При обробці даних досліджень як найважливіших характеристик варіаційного ряду застосовуються різні середні значення. Середні величини прийнято розділяти на прості і зважені. Якщо середні значення обчислюються по безпосередньому переліку значень ознаки в кожній одиниці сукупності, то такі середні називаються простими (невваженими). Якщо середні обчислюються по варіаційному ряду з урахуванням статистичної ваги кожного варіанту, то їх називають зваженими.

Середні величини бувають різного роду. Якщо вид середньої невідомий, то мається на увазі середня арифметична величина.

Середні величини частіше використовують статечні і структурні (порядкові) середні. Статечні середні, в свою чергу, поділяються на середні арифметичні, середні гармонійні, середні квадратичні, середні кубічні та інші. Структурні середні \square це мода, медіана, квартили, децили і ін.

Медіаною називається середнє (серединне) значення ознаки ранжированного варіаційного ряду, тобто значення, рівновіддалене від початку і кінця, перебудованого в зростаючому або спадному порядку.

Модою називається ймовірна (що частіше зустрічається) в даному статистичному ряду величина. Мода є найбільшою ординатою кривої розподілу при одновершинній розподілі. У загальному випадку крива розподілу може мати кілька вершин і, відповідно, вона буде мати кілька мод.

Вважається, що нормальний розподіл є однією з емпірично перевірених

істин і його положення розглядаються як один з фундаментальних законів природи. Точна форма нормального розподілу «колоколоподібна крива» \square визначається тільки двома параметрами: середнім і стандартним відхиленням. Для нормального розподілу характерно, що 68% всіх значень знаходяться в межах ± 1 стандартне відхилення від середнього, а діапазон ± 2 стандартних відхилення включає 95% значень. Про прийняття за основу від нормального розподілу свідчать коефіцієнти асиметрії та ексцесу. Коефіцієнт асиметрії показує відхилення розподілу від симетричного (нормальний розподіл \square абсолютно симетричне, відповідно коефіцієнт асиметрії дорівнює нулю). Позитивне значення коефіцієнта асиметрії свідчить про наявність розподілу з «довгим правим хвостом», негативне з «довгим лівим хвостом». Коефіцієнт ексцесу показує «гостроту піку» розподілу (для нормального розподілу коефіцієнт ексцесу також дорівнює нулю). Якщо коефіцієнт ексцесу є позитивним, пік є загостреним, якщо є негативним \square пік закруглений.

В завдання практичної роботи входить приведення емпіричного розподілу випадкової величини до нормального шляхом нелінійних перетворень.

Ресурс часу 4 години.

Критерії оцінювання (максимум 4 бали)

- Виконання практичної роботи – 4 балів

Практичне заняття ПЗ.

Побудова двовимірних статистичних моделей на матриці вихідних даних. Двовимірний регресійно-кореляційний аналіз.

Равданням практичної роботи ПЗ є визначення пар показників, пов'язаних між собою, визначення кількісної оцінки сили зв'язку між ознаками. Двовимірна модель містить дві випадкові величини, і аналізує дві величини. В ході практичної роботи вивчаються такі двовимірні моделі як: кореляційний і регресійний аналізи.

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

Мета кореляційного аналізу \square забезпечити отримання деякої інформації про однієї змінної за допомогою іншої змінної. У випадках, коли можливе досягнення мети, кажуть, що змінні корелюють. У найзагальнішому вигляді сприйняття гіпотези про наявність кореляції означає, що зміна значення змінної. А станеться одночасно з пропорційною зміною значення В. Мірою залежності між експериментальними наборами даних є числа \square коефіцієнти зв'язку.

Головні завдання кореляційного аналізу:

- 1) оцінка за вибірковими даними коефіцієнтів кореляції;

- 2) перевірка значущості вибірових коефіцієнтів кореляції або кореляційного відношення;
- 3) оцінка близькості виявленої зв'язку до лінійної;
- 4) побудова довірчого інтервалу для коефіцієнтів кореляції.

Визначення сили й напрямку взаємозв'язку між змінними є однією з важливих проблем аналізу даних. У загальному випадку для цього застосовують поняття кореляції. Кореляція є залежністю двох випадкових величин. При цьому, зміна однієї або декількох цих величин приводить до систематичного зміни іншої або інших величин.

Математичної мірою кореляції двох випадкових величин служить коефіцієнт кореляції.

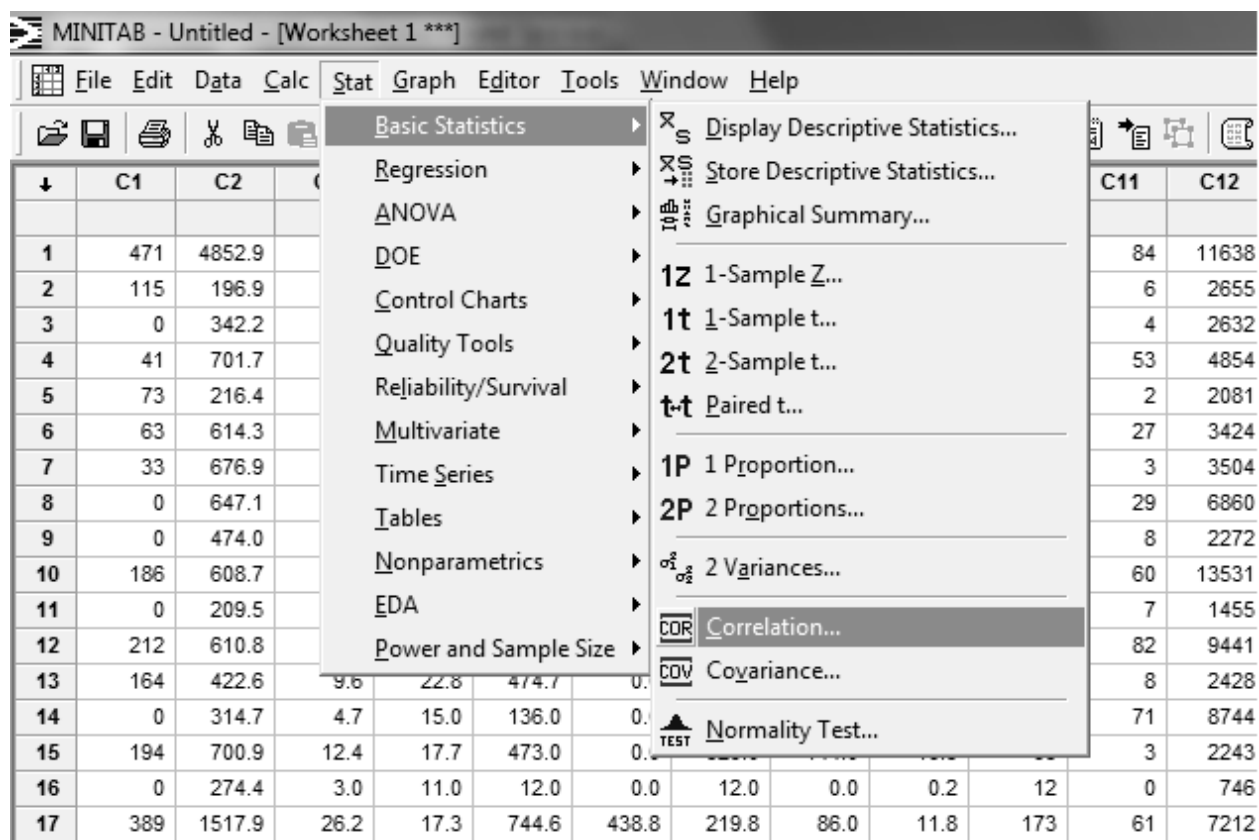


Рис. 3. Алгоритм вибору кореляційного аналізу в програмі MINITAB 1.

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

Коефіцієнт кореляції
тіснота зв'язку

- > ± 0.91 Дуже сильна
- ± 0.71-0.90 Сильна

± 0.51-0.70 Відносна
 ± 0.21-0.50 Слабка
 <± 0.20 Дуже слабка

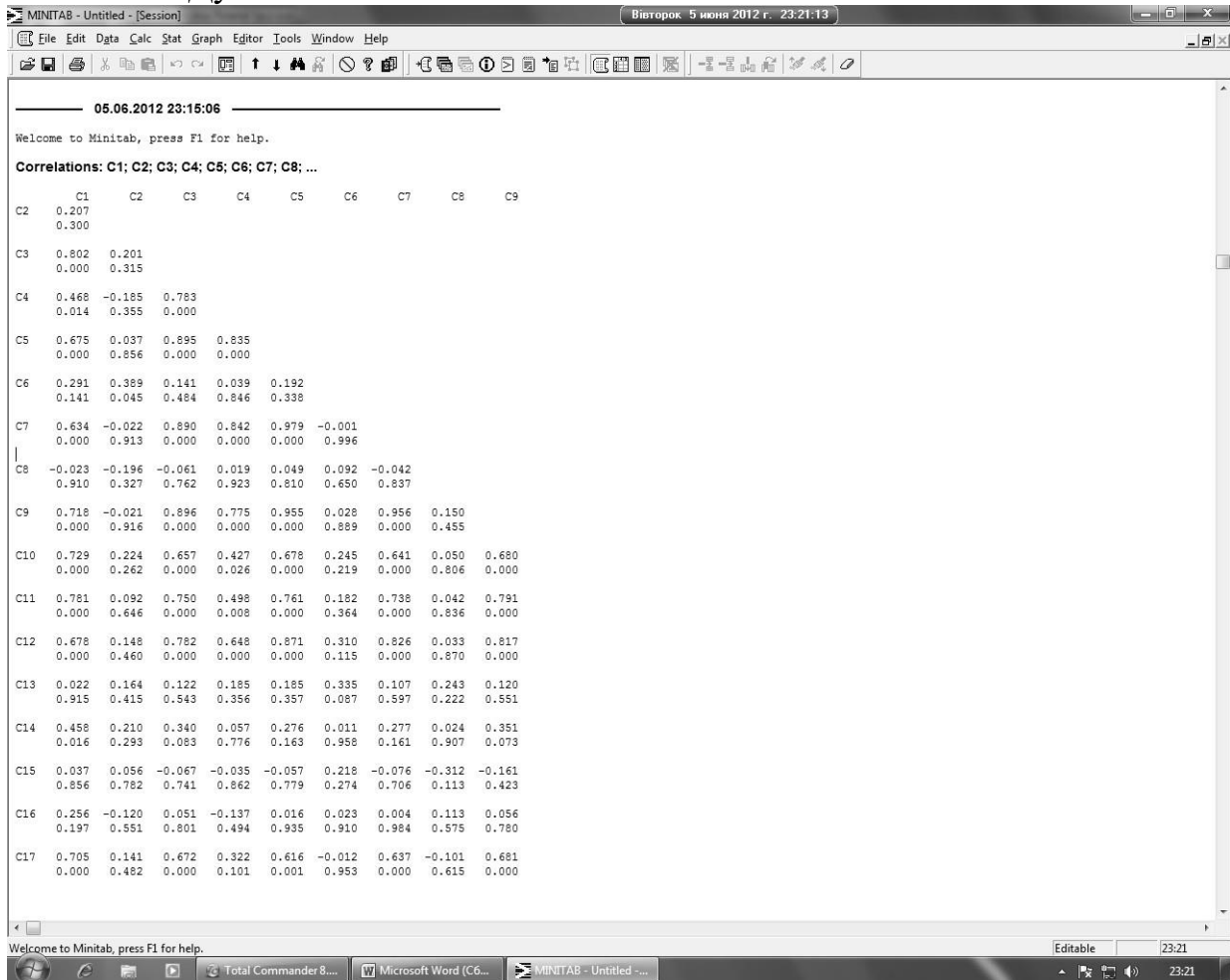


Рис. 4. Кореляційна матриця

Коефіцієнт кореляції, а в загальному випадку кореляційна функція, дозволяє встановити ступінь взаємозв'язку між змінними. Кореляція може бути лінійної або нелінійної залежно від типу залежності, яка фактично існує між змінними. Досить часто на практиці розглядають тільки лінійну кореляцію (взаємозв'язок), але більш глибокий аналіз вимагає 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 разі максимальної тісноти зв'язку між показниками на діаграмі розсіювання їх залежність буде представлена прямою лінією. Іноді зустрічається таке явище, як псевдокореляція, тобто кореляція, обумовлена впливом інших показників, які залишилися поза увагою дослідника. Значимість коефіцієнта кореляції залежить від довжини динамічних рядів. У великих динамічних рядах навіть слабкі залежності будуть значущими, в той час, як в незначних \square навіть дуже сильні залежності не є статистично надійними. Тому говорять про надійність кореляційної залежності, пов'язаної з репрезентативністю вихідних динамічних рядів і свідчить про те, наскільки

ймовірно, що виявлена залежність, знову буде виявлена при збільшенні періоду ретроспекції або екстраполяції на майбутнє. Надійність виявлених залежностей оцінюється за допомогою стандартної статистичної заходи p -рівня статистичного рівня значимості. Даний показник знаходиться в зворотній залежності до надійності результату: чим вище p -рівень, тим нижче статистична надійність виявленої залежності, і навпаки.

Ресурс часу – 6 годин

Критерії оцінювання (максимум 4 бали)

- Виконання практичної роботи – 4 бали

Практичне заняття П4.

Кластер – аналіз вихідних даних. Інтерпретація отриманих результатів. Факторний аналіз вихідних даних (метод головних компонент). Інтерпретація отриманих результатів.

Завданням практичної роботи є групування об'єктів і виявлення значущих чинників геологічного процесу.

Багатовимірним аналізом називають сукупність різних методів, призначених для вивчення багатовимірних явищ. У багатовимірному просторі досліджувані об'єкти розташовуються, як правило, не рівномірно, а утворюють певні скупчення. Ці скупчення можна розглядати як класи об'єктів. Причому в різних просторах виділяються різні класи. У багатовимірному просторі класифікація більш обґрунтована. Але це означає, що при безмежному зростанні числа ознак в тій же пропорції зростає точність класифікації.

Існує багато математичних методів і прийомів, які на основі інформації закладеної в матриці даних, дозволяють об'єктивно класифікувати економіко-географічні об'єкти (в тому числі регіоналізувати їх). Такі методи і прийоми об'єднані в багатомірний аналіз, розглядається у вузькому і широкому сенсах. У вузькому сенсі це аналіз матриці даних, має два і більше стовпців, в широкому сенсі це аналіз матриці даних, коли кількість стовпців не обмежена знизу. В цьому випадку багатомірний аналіз включає в себе і одновимірний. Отже, одновимірний аналіз можна розглядати як окремих випадок багатомірного, а багатомірний як узагальнення одновимірного. Часто в процесі обробки матриці даних доводиться їх згортати, тобто зводити багатомірний простір до n - k -мірному, і навіть до одновимірного. У цьому випадку багато ознак замінюються декількома, але такими, кожна з яких синтезує інформацію відповідної групи ознак або є найбільш репрезентативною.

Кластерний аналіз передбачає групування об'єктів за подібністю певних характеристик. Термін вперше введений Трайеном / Tryon /, 1939 р. Це один з методів класифікації, що передбачає поділ вихідної сукупності об'єктів на кластери (класи, групи). Кластер це група об'єктів, що мають схожі тенденції або особливості розвитку. З математико-статистичної точки зору, кожен кластер повинен мати такі властивості: щільність об'єктів в

межах кластера повинен бути більша за густину поза ним, можливість відокремлення від інших кластерів і т.д. Складність завдань кластерного аналізу полягає в тому, що реальні геологічні об'єкти є багатовимірними, тобто описуються не одним, а певною сукупністю параметрів, тому об'єднання в групи здійснюється в просторі багатьох вимірів. Згідно з критерієм об'єднання регіонів в кластери є мінімум відстані в просторі показників, які їх описують. Звідси поняття відстані між регіонами в просторі даних.

В ході практичної роботи виконується проведений кластерний аналіз, в програмі MINITAB «Stat» → «Multivariate» → «Cluster Obrevations».

Існують наступні групи методів кластерного аналізу:

- 1) ієрархічні методи;
- 2) ітеративні методи;
- 3) факторні методи;
- 4) методи згуцень;
- 5) методи, які використовують теорію графів.

До поширених в економіці відносять ієрархічні і ітеративні.

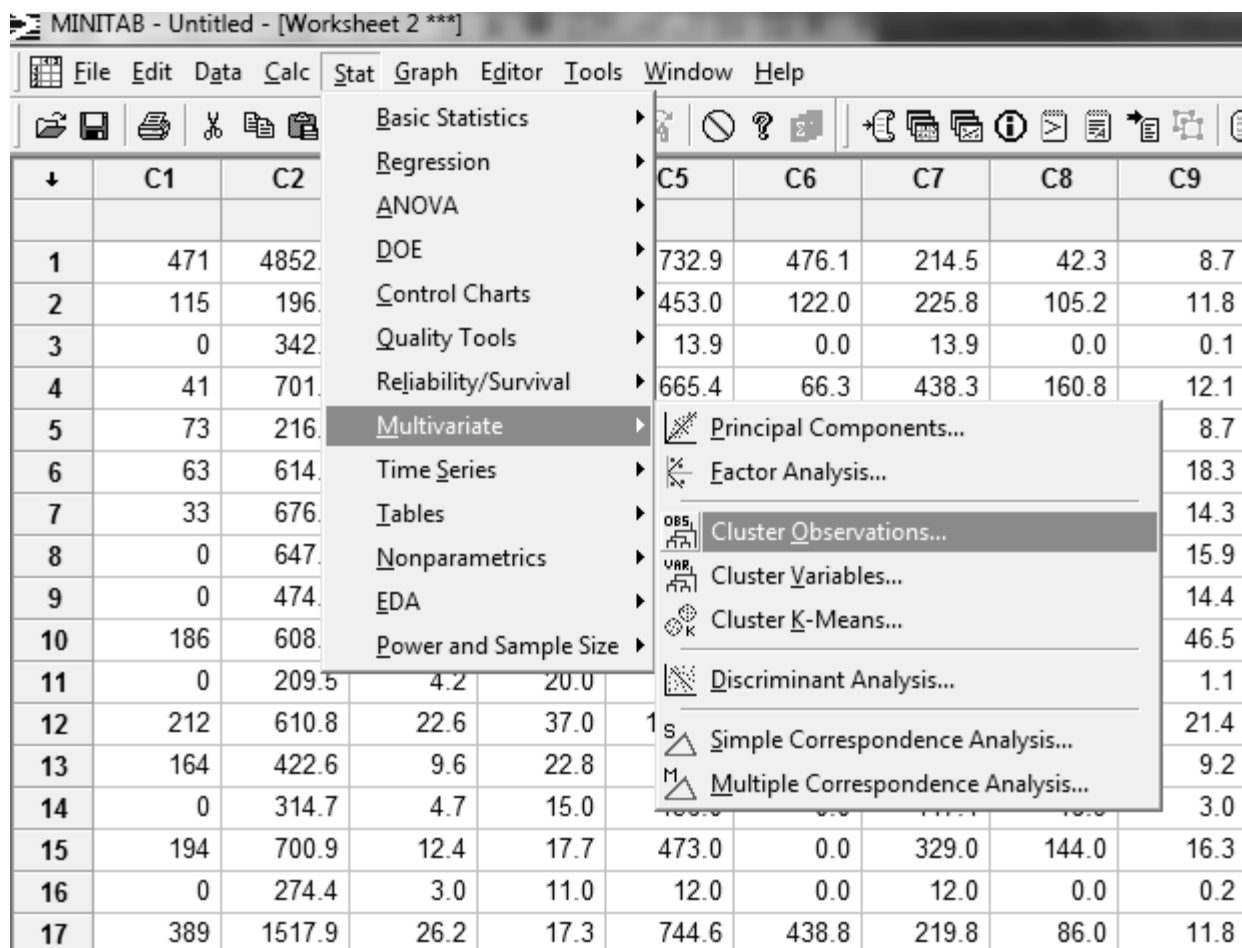


Рис. 5. Алгоритм вибору кластерного аналізу в програмі MINITAB 14

Для проведення класифікації необхідно ввести поняття подібності об'єктів по спостережуваних змінних. У кожен кластер (клас, таксон) повинні потрапити об'єкти, що мають подібні характеристики. Вибір відстані між

об'єктами є вузловим моментом дослідження, від нього багато в чому залежить остаточний варіант розбиття об'єктів на класи при даному алгоритмі розбиття.

Алгоритм ієрархічного агломеративного кластерного аналізу можна представити у вигляді послідовності процедур:

- 1) нормуються вихідні дані;
- 2) розраховується матриця відстаней або матриця заходів подібності;
- 3) знаходиться пара найближчих кластерів, за обраним алгоритмом об'єднуються ці два кластери. Нового кластеру присвоюється менший з номерів об'єднуються кластерів;
- 4) процедури 2, 3 і 4 повторюються до тих пір, поки всі об'єкти не будуть об'єднані в один кластер або до досягнення заданого "порога" подібності.

Для розрахунку відстаней між об'єктами у багатовимірному просторі передбачається здійснення процедури нормалізації даних.

У кластерному аналізі для кількісної оцінки подібності вводиться поняття метрики. Подібність або відмінність між класифікованими об'єктами встановлюється в залежності від метричного відстані між ними. Якщо кожен об'єкт описується k ознаками, то він може бути представлений як точка в k -вимірному просторі, і схожість з іншими об'єктами буде визначатися як відповідне відстань. У кластерному аналізі використовуються різні заходи відстані між об'єктами.

Найбільш поширеними є такі види відстаней:

- Евклідова відстань (Euclidean distances), розраховується по теоремі Піфагора: відстані по кожній з координат вводяться в квадрат, а потім з їх суми визначається корінь квадратний
- Манхеттенська відстань (відстань міських кварталів, city-block / Manhattan / distances), розраховується як сума різниць по кожній з координат;
- відстань Чебишева (Chebyshev distance metric), регіони визначають як "різні", якщо вони розрізняються за якоюсь однією координаті;
- відсоток незгоди (percent disagreement), використовується, коли вихідні дані не мають кількісного вираження.

Методи кластеризації діляться на дві великі групи : агломеративні (від англ. Agglomerate - скупчення) і дивізійні (від англ. Division - поділ).

Агломеративні методи передбачають послідовне об'єднання найвизначніших регіонів на основі розрахованих відстаней між ними. Процедура кластеризації така. На першому кроці, кожен регіон утворює окремий кластер, далі в новий кластер об'єднуються два регіони, ступінь подібності яких є найбільшою. На останньому кроці усі регіони об'єднуються в один кластер.

Міра подібності кластерів і регіонів визначається наступними способами:

- одиничної зв'язку (single linkage, «метод найближчого сусіда»): мінімум найменших відстаней до будь-якого одного регіону в кластері;
- повної зв'язку (complete linkage, «метод найвіддаленіших сусідів»):

мінімум найбільших відстаней до будь-якого одного регіону в кластері;

- «середньої» зв'язку: мінімум середньоарифметичного значення відстаней до всіх регіонів в кластері;

- центроїдного мінімум відстаней до центрів тяжкості кластерів.

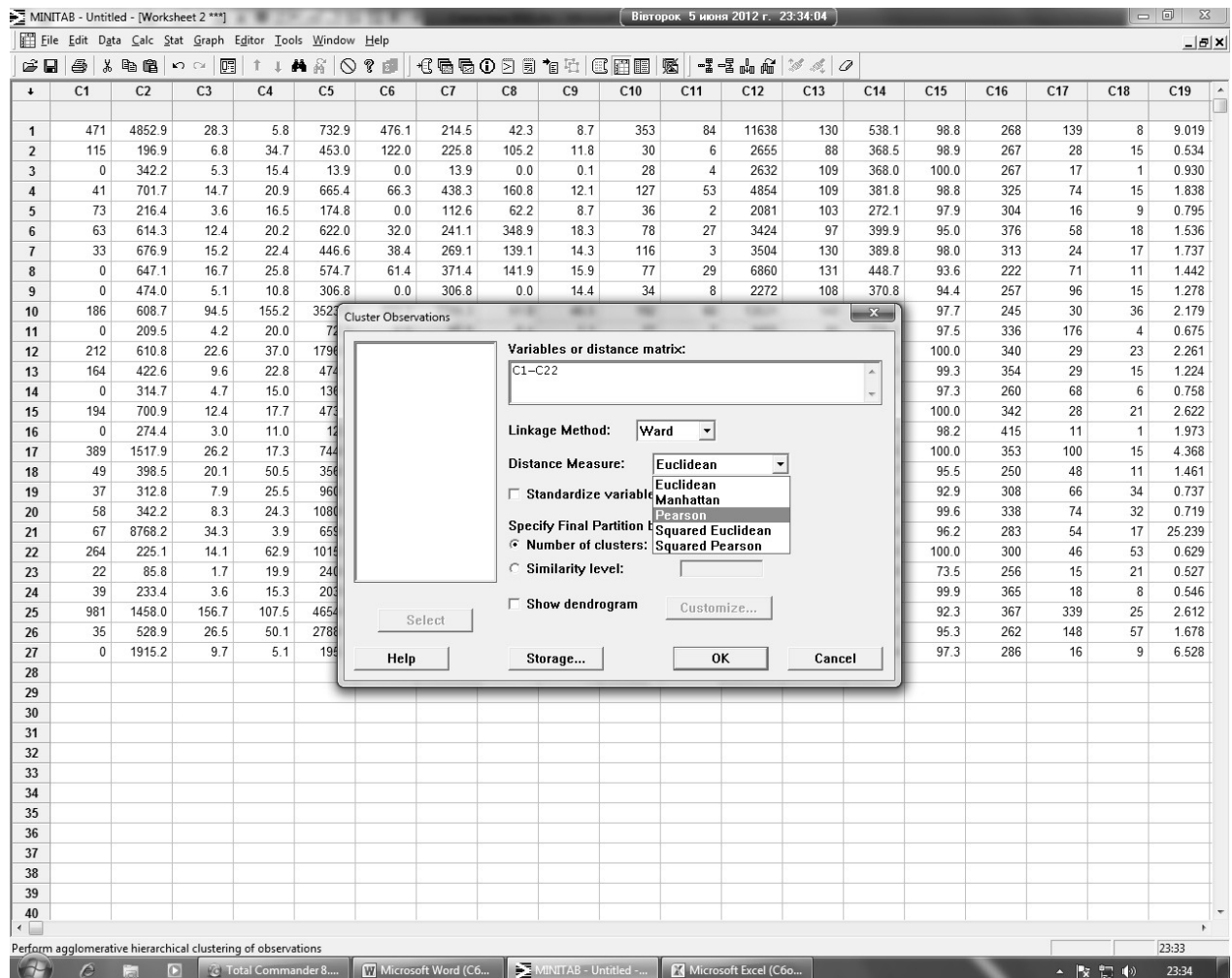


Рис. 6. Вибір відстані при виконанні кластерного аналізу

Центр тяжіння кластера визначається як середнє по кожному параметру.

Результат кластеризації візуалізується у вигляді дендрограми кластеризації (tree diagram, дерево об'єднання), на одній осі якої відкладаються регіони, на другий \square відстані об'єднання (linkage distance). Дивізійні методи передбачають поетапне (ітераційне) поділ регіонів на задану кількість кластерів. Поширеним серед дивізійних є метод k-середніх (kmeans clustering), який вирішує завдання виділення заздалегідь заданої кількості кластерів, які максимально відрізняються, тобто знаходяться на найбільших відстанях один від одного.

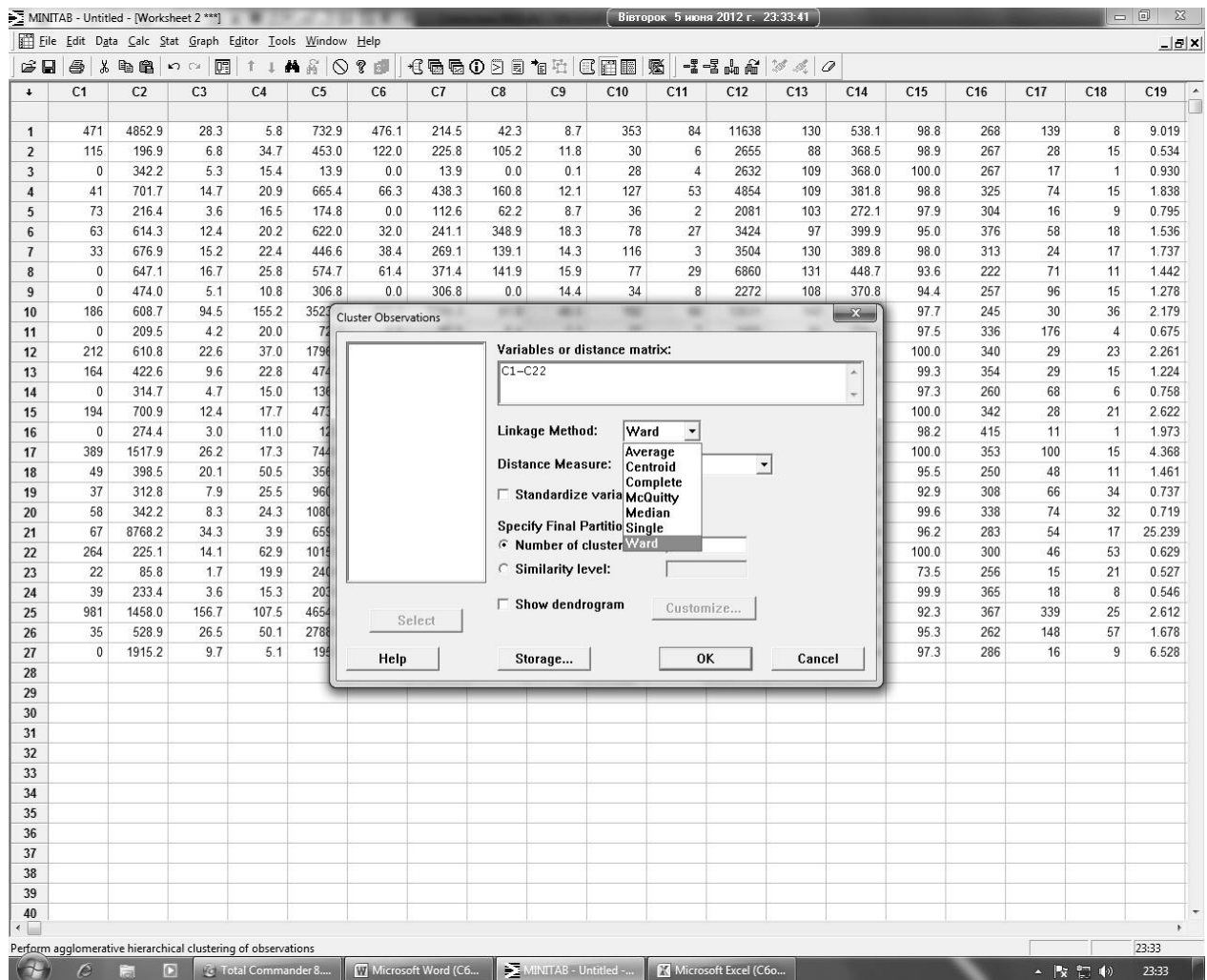


Рис. 7. Вибір методу при виконанні кластерного аналізу

Процедура кластеризації така. На першому кроці задається деякий випадкове поділ даних на задану кількість кластерів (k), розраховуються центри тяжіння кластерів. Далі здійснюється переміщення регіонів: кожен регіон відноситься до того кластеру, відстань до центра ваги якого мінімальна, розраховуються центри тяжкості нових кластерів. Ця процедура повторюється, поки не буде знайдена стабільна конфігурація, тобто склад кластерів перестане змінюватися.

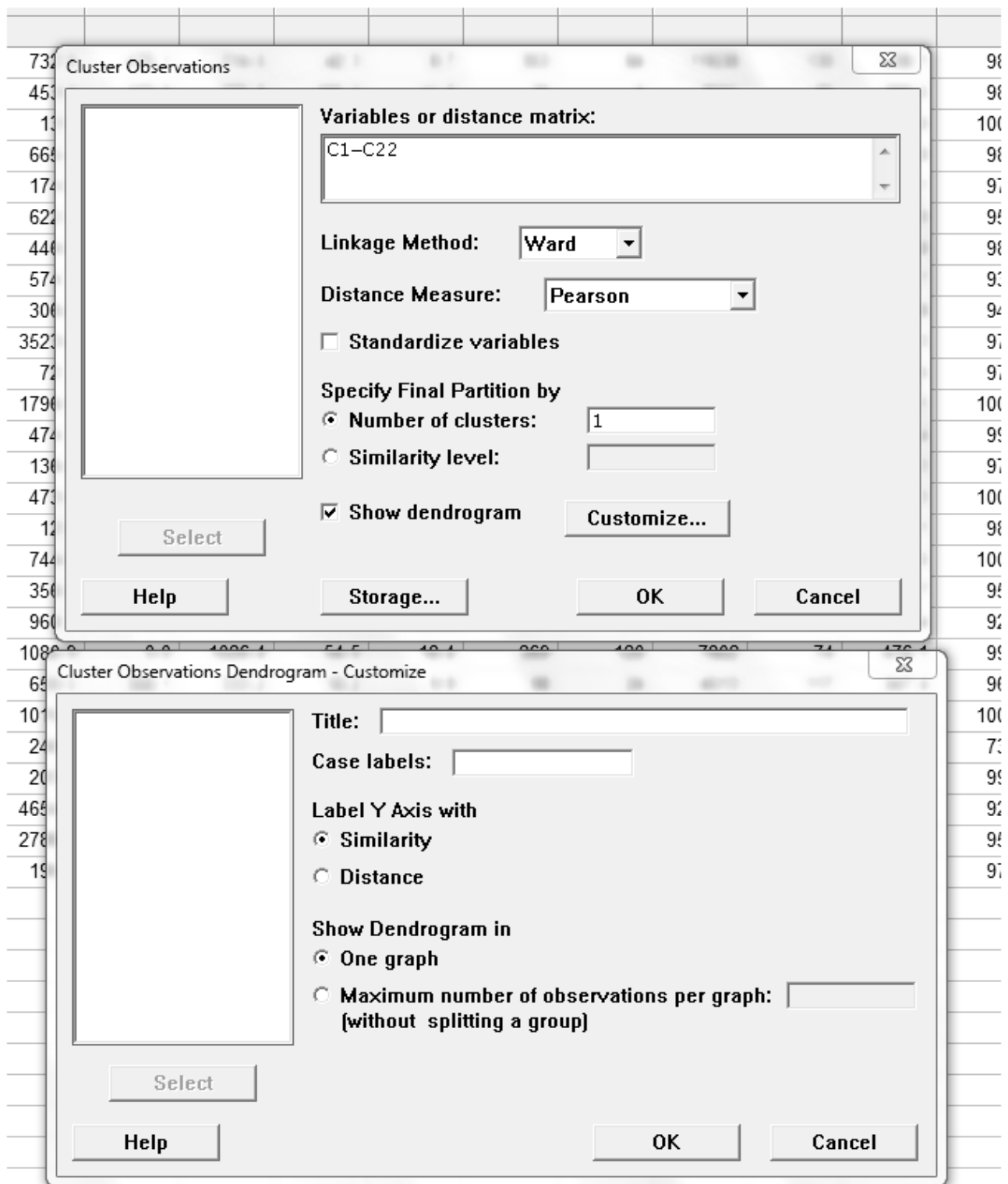


Рис. 8. Вибір основних параметрів при виконанні кластерного аналізу

Для візуалізації результатів будуються графіки, що представляють собою проєкції будь-якої пари показників на площину, на яких точки, що відносяться до одного кластеру, окантовуються. Одним з найбільш важливих і складних питань при кластеризації є вибір оптимальної кількості кластерів. Зазвичай згідно вихідної гіпотези визначається початкова кількість кластерів, а потім змінюючи його, емпіричним шляхом вибирають остаточний варіант

кластеризації. Для того, щоб оцінити, наскільки виділені кластери відрізняються, розраховують середні значення базисних показників для кожного кластера.

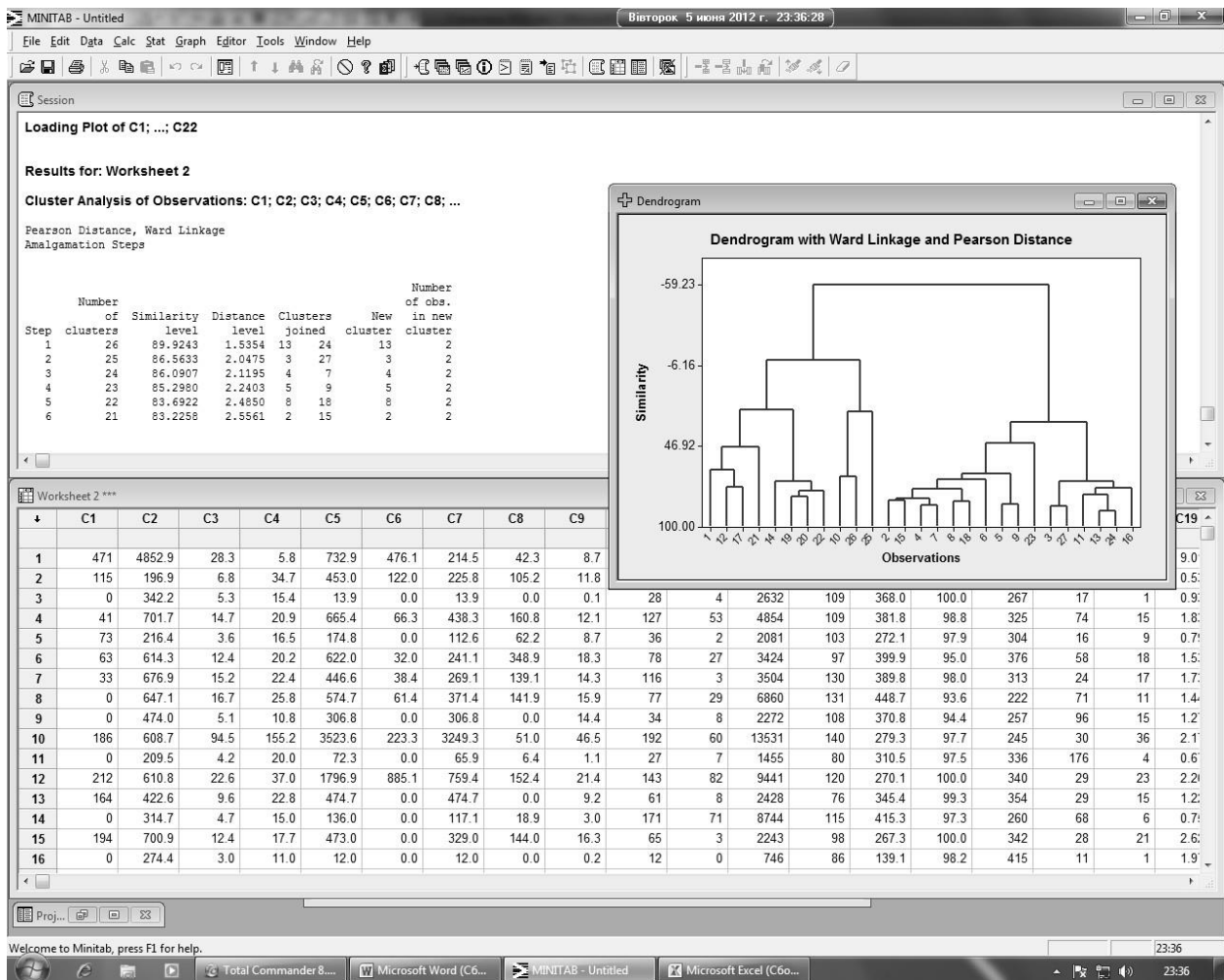


Рис. 9. Приклад результату кластерного аналізу

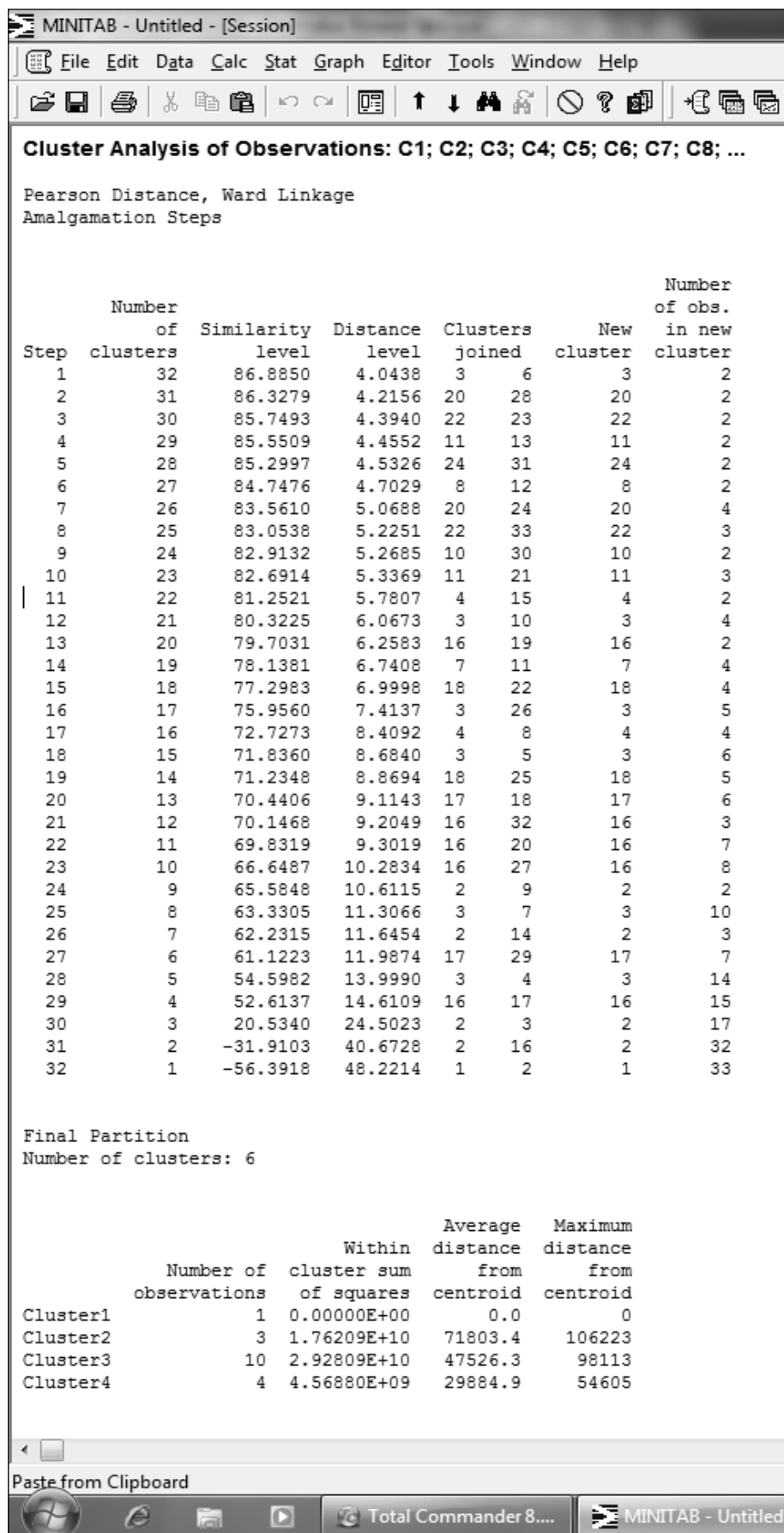


Рис. 10. Приклад протоколу кластерного аналізу

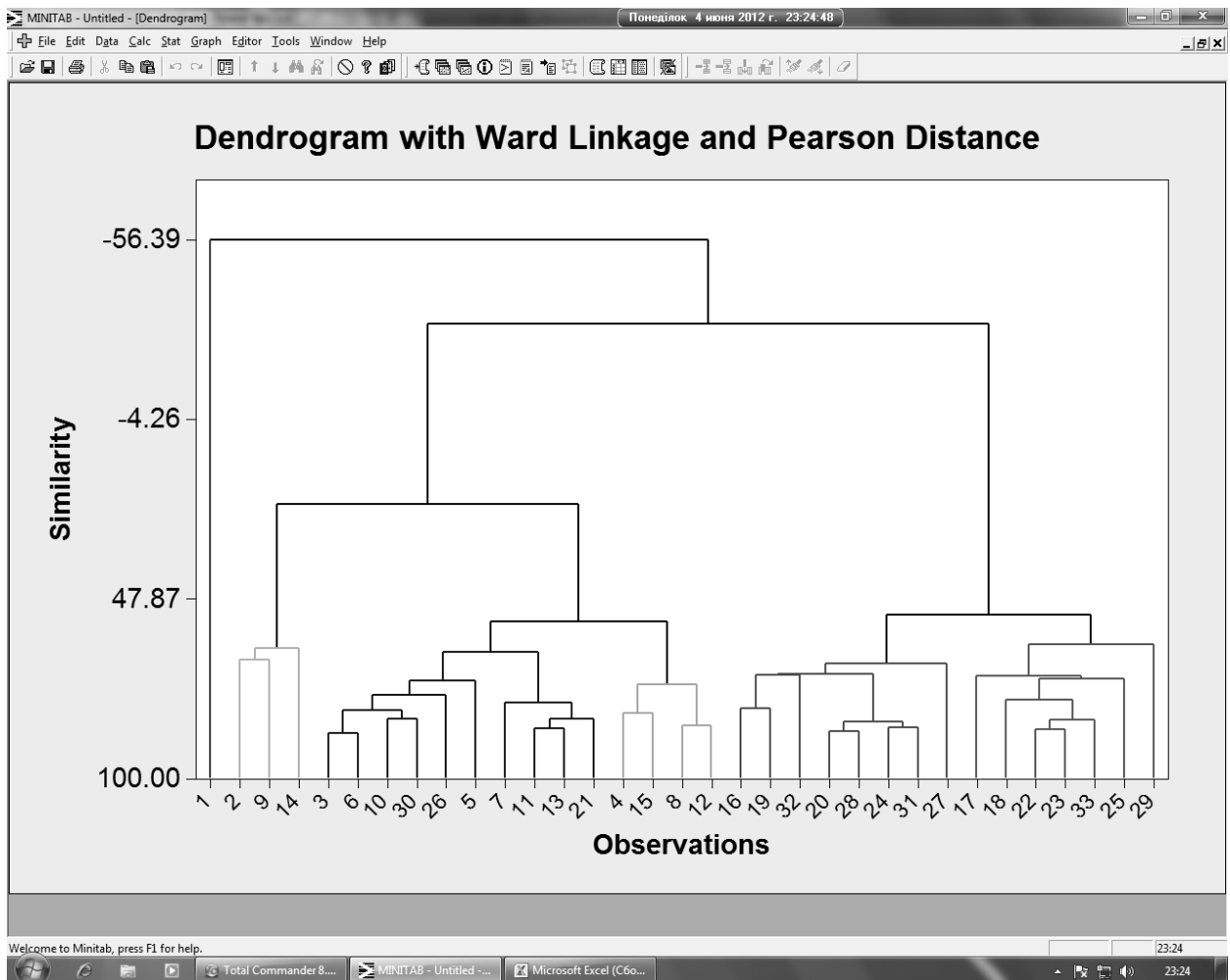


Рис. 11. Приклад дендрограми кластерного аналізу

Кластер-аналіз застосовується при наявності багатовимірних сукупностей статистичних показників, сутність його полягає в об'єднанні їх в групи (кластери) за принципом мінімального розміру в багатовимірному просторі. Залежно від кількості об'єктів кластеризація і угруповання виконуються послідовно в кілька кроків таким чином, щоб на останньому кроці в одну загальну групу потрапили всі об'єкти. На перших кроках класифікації формуються найбільш однорідні групи по об'єктах, які мають найбільшу схожість. Поступово, «послаблюючи» критерій щодо подібності об'єктів, об'єднується все більшу кількість об'єктів. З кожним кроком до кластерів вищого порядку включаються цілі групи об'єктів, які все сильніше розрізняються між собою. На останньому кроці всі об'єкти об'єднуються в один кластер. В кінці процедури отримують групи неоднорідні. В результаті кластерного аналізу отримують багаторівневу ієрархічну класифікацію, яка відображає найбільш суттєві особливості взаємини між об'єктами. Таким чином, отримані кластери \square це група об'єктів, які мають подібні особливості розвитку. Такий аналіз проводиться для територіальних об'єктів по ряду показників, для подальшого угруповання районів і виявлення стійких груп.

Факторний аналіз це багатовимірний статистичний метод, який використовується для вивчення дієчих факторів на основі вивчення

взаємозв'язків між значеннями змінних. Він дозволяє визначити групи факторів впливу і досліджувати значення кожної змінної в дії фактора в різний час.

В ході практичної роботи виконується факторний аналіз методом головних компонент, в MINITAB 14 шляхом «Stat» → «Multivariate» → «Prinsipal Components». Факторний аналіз дозволяє оцінити чинники розвитку.

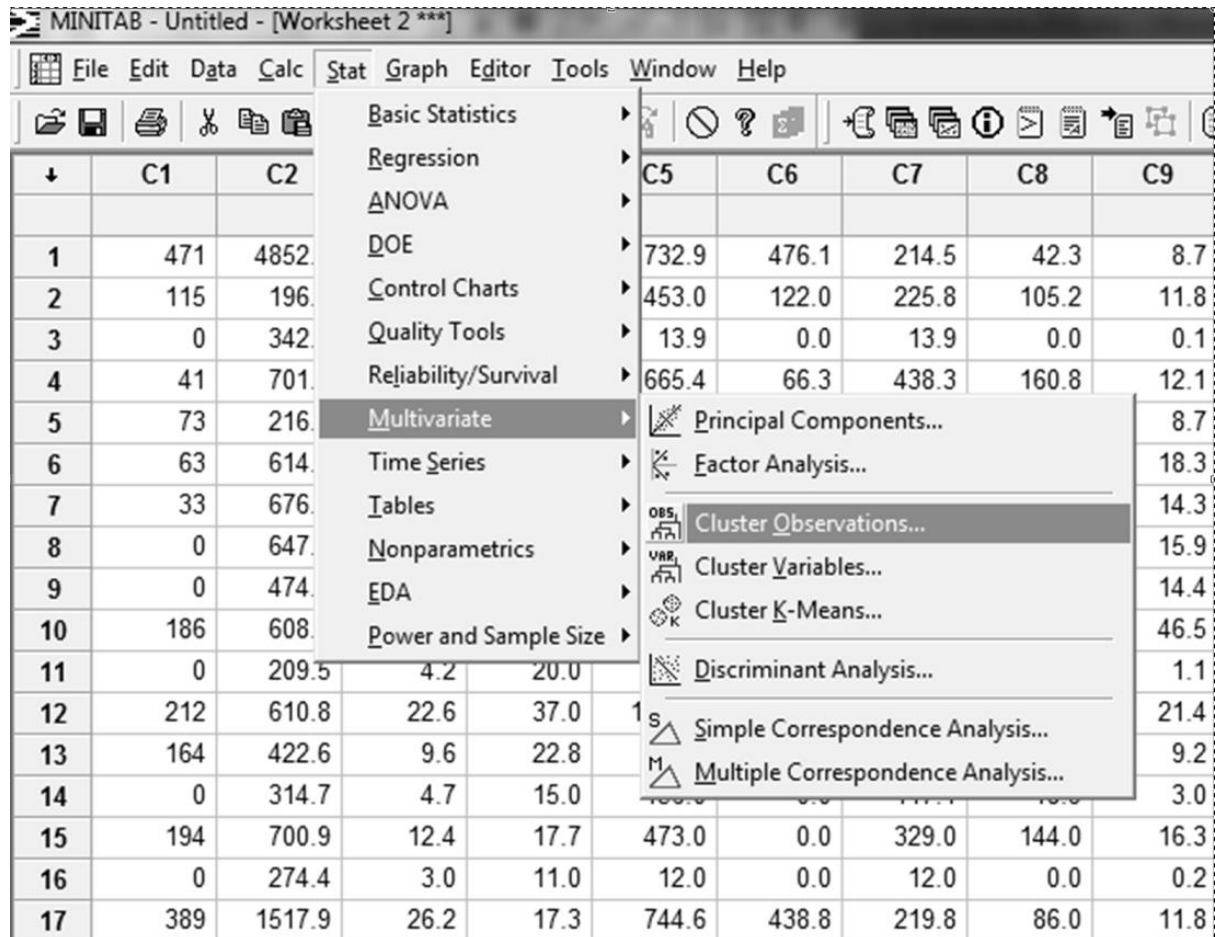


Рис. 12. Алгоритм вибору факторного аналізу в програмі MINITAB 14

В основі побудови моделей факторного аналізу лежить твердження про те, що безліч взаємопов'язаних показників, що характеризують певний процес, можна представити меншою кількістю гіпотетичних змінних факторів і безліччю незалежних залишків.

Інформаційною базою розрахунків служать ряди показників геологічного розвитку в розрізі геологічних об'єктів регіону. Набір показників є суб'єктивним; основним принципом вибору показників є достовірність і порівнянність. Після обрання оптимальної кількості факторів, виконується інтерпретація результатів, враховуючи показники, які сформували виявлені чинники

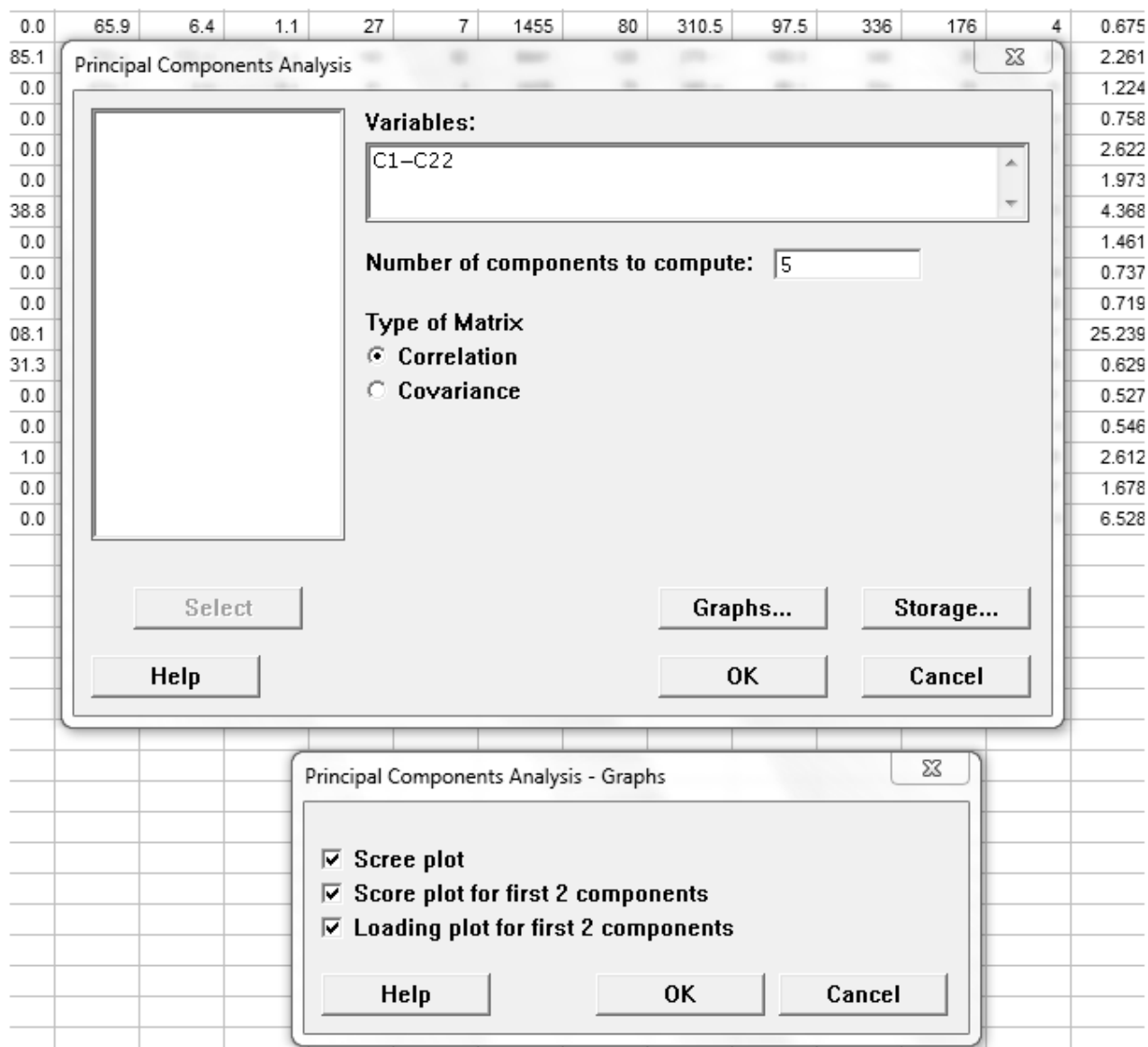


Рис. 13. Вибір вихідних даних і елементів виконання факторного аналізу

Зазвичай фактори, що впливають на розвиток процесів, характеризуються не одним, а безліччю показників, що мають між собою певний зв'язок (корелюють). Тому виникає потреба в їх об'єднанні в певну кількість груп за подібністю впливу, в результаті чого визначаються на перший погляд "приховані" (латентні) фактори, кількість яких, зрозуміло, менше кількості вихідних показників. Тобто чинники ідентифікуються.

Застосування факторного аналізу передбачає:

- по-перше, скорочення кількості показників (змінних), що на мові математичної статистики називається редукцією даних;
- по-друге, визначення структури взаємозв'язків між показниками (змінними), тобто класифікацію показників, інтерпретацію факторів.

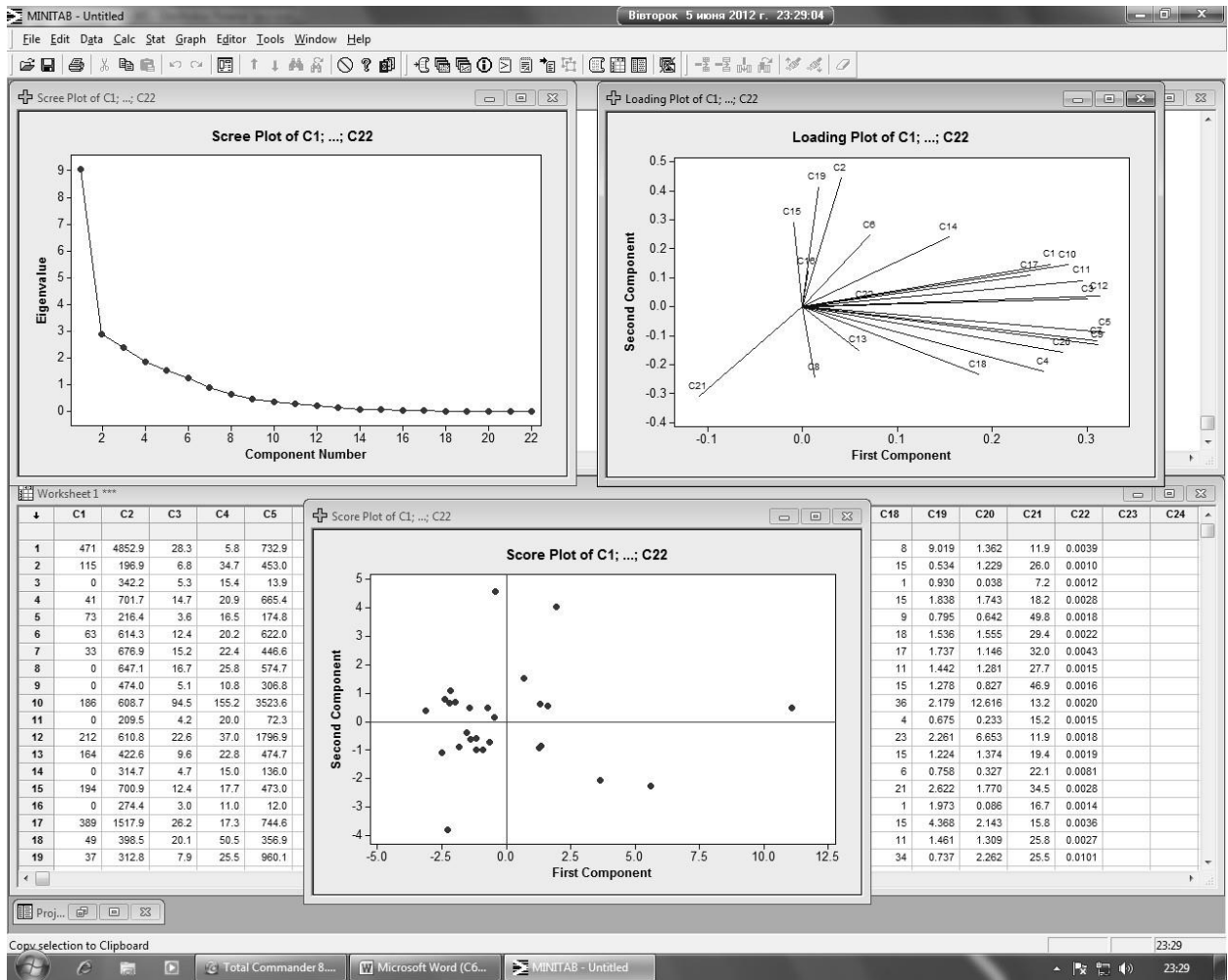


Рис. 14. Приклад результату факторного аналізу

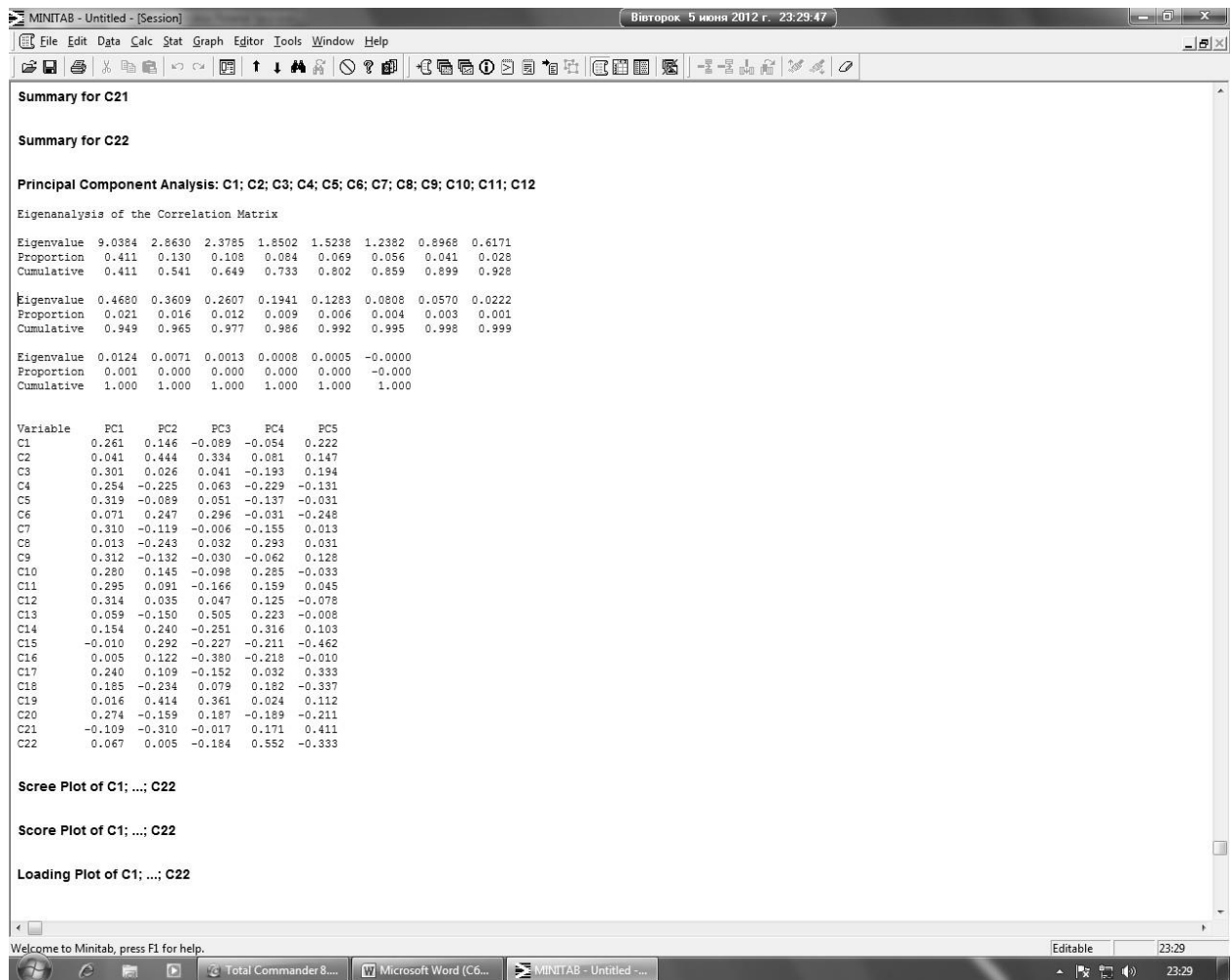


Рис. 15. Приклад протоколу факторного аналізу

З кожного виду аналізу виконується інтерпретація отриманих результатів.

Ресурс часу – 8 годин.

Критерії оцінювання (максимум 4 бали)

- Виконання практичної роботи – 4 бали

Практичне заняття П5.

Моделювання просторових змінних методом ковзного статистичного вікна. Аналіз та інтерпретація результатів.

Завданням практичної роботи є побудова статистичної поверхні поля параметр у геологічного об'єкту методом локального середнього.

У зв'язку з індивідуальним характером роботи її загальний опис недоцільний.

Ресурс часу – 6 годин.

Критерії оцінювання (максимум 4 бали)

- Виконання практичної роботи – 4 бали

Результати виконання практичних робіт оформляються у вигляді загального звіту (10-12 с.) і оцінюються індивідуально. Ресурс часу 6 годин. Максимальна оцінка 20 балів