

дисперсію, викликану відмінністю між групами, з дисперсією, викликану внутрішньогруповою мінливістю.

Суть дисперсійного аналізу полягає в тому, що із загальної варіативної ознаки вичленяють три види варіативної :

- 1) варіативна, обумовлена дією кожного з досліджуваних чинників;
- 2) варіативна, обумовлена взаємодією досліджуваних чинників;
- 3) випадкова варіативна, обумовлена діями інших чинників(невідомих чинників або відомих чинників, вплив яких не входить в завдання дослідження).

Варіативна, обумовлена дією досліджуваних чинників і їх взаємодією, співвідноситься з випадковою варіативною. Показником цього співвідношення є F-критерій Фішера.

$$F_{\text{эмп.}A} = (\text{Варіативна, обумовлена чинником } A) / (\text{Випадкова варіативна})$$

$$F_{\text{эмп.}y} = (\text{Варіативна, обумовлена чинником } B) / (\text{Випадкова варіативна})$$

$$F_{\text{эмп.}AB} = (\text{Варіативна, обумовлена взаємодією чинників } A \text{ і } B) / (\text{Випадкова варіативна})$$

Чим більшою мірою варіативна ознаки обумовлена досліджуваними чинниками або їх взаємодією, тим вище емпіричні значення F- критерію Фішера.

**Нульова гіпотеза** в дисперсійному аналізі: середні величини досліджуваної ознаки при усіх градаціях досліджуваного чинника однакові. Наприклад: середня тривалість перебування відвідувачів в ресторані (досліджувана ознака) однакова у разі, якщо в залі звучить тиха, середня або гучна музика (гучність - передбачуваний чинник, який впливає на тривалість перебування відвідувачів в ресторані, а «тиха», «середня» і «гучна» - градації цього чинника.

**Альтернативна гіпотеза:** Середня тривалість перебування відвідувачів в ресторані розрізняється залежно від гучності звукової в залі музики.

При істинності нульової гіпотези (про рівність середніх в декількох

групах спостережень, вибраних з генеральної сукупності), оцінка дисперсії, пов'язаної з внутрішньогруповою мінливістю, має бути близької до оцінки міжгрупової дисперсії.

Метод дисперсійного аналізу стає незамінним тоді, коли ми досліджуємо одночасну дію двох або більше за чинники, оскільки він дозволяє виявити взаємодію чинників в їх впливі на одну і ту ж ознаку.

Існує декілька видів дисперсійного аналізу. Необхідний варіант вибирається з урахуванням числа чинників і наявних вибірок з генеральної сукупності.

**Однофакторний дисперсійний аналіз** служить для аналізу дисперсії за даними двох або декількох вибірок. При аналізі перевіряється нульова гіпотеза про те, що кожен приклад витягнутий з одного і того ж базового розподілу вірогідності. Альтернативна гіпотеза: базові розподіли вірогідності в усіх вибірках різні. Якщо є всього дві вибірки, застосовують функцію TTEST. Для більше, ніж двох вибірок не існує узагальнення функції TTEST, і замість цього можна скористатися моделлю однофакторного дисперсійного аналізу.

**Двохфакторний дисперсійний аналіз з повтореннями** застосовується, якщо дані можна систематизувати за двома параметрами. Наприклад, в дослідженні, спрямованому на пошук кращих засобів для схуднення, добровольцям пропонувалися три різні дієти (А, В і С) і два варіанти поведінки - регулярне зайняття фітнесом і відсутність спеціалізованих фізичних навантажень. Таким чином, для кожної з 6 можливих пар умов {дієта, фізичне навантаження} є набір спостережень за зниженням ваги добровольців. За допомогою двофакторного дисперсійного аналізу з повтореннями можна перевірити наступні гіпотези:

1. Чи витягнуті дані про зниження ваги для різних дієт з однієї генеральної сукупності незалежно від фізичного навантаження.
2. Чи витягнуті дані про зниження ваги для різного характеру фізичного навантаження з однієї генеральної сукупності незалежно від типу дієти.
3. Чи витягнуті 6 вибірок, значень, що представляють усі пари {дієта,

фізичне навантаження}, використовувані для оцінки впливи різних типів дієт (крок 1) і типу фізичного навантаження (крок 2), з однієї генеральної сукупності. Альтернативна гіпотеза припускає, що вплив конкретних пар {дієта, фізичне навантаження} перевищує вплив окремо дієти і окремо фізичного навантаження.

**Двохфакторний дисперсійний аналіз без повторення** корисний при класифікації даних за двома вимірами, як і двухфакторний дисперсійний аналіз з повторенням. Проте при цьому аналізі передбачається тільки одно спостереження для кожної пари. При цьому аналізі можна додавати перевірки в кроки 1 і 2 двухфакторні дисперсійні аналізи з повтореннями, але недостатньо даних для додавання перевірок в крок 3.

Дисперсійний аналіз дозволяє констатувати зміну ознаки, але при цьому не вказує напряму цих змін. Щоб отримати наочне уявлення про напрям змін, необхідно спеціально графічно представляти отримані дані по градації чинника.

### Постановка завдання

Директор з маркетингу хоче встановити, чи значущо розрізняються продажі в групах магазинів залежно від рівня використовуваної реклами. Результати продажів представлені в таблиці 16.1.

Таблиця 16.1

### Результати продажів залежно від рівня реклами

|   | А           | В                      | С                       | Д                      |
|---|-------------|------------------------|-------------------------|------------------------|
| 1 | Номер тижня | Рівень реклами високий | Рівень реклами середній | Рівень реклами низький |
| 2 | 1           | 8                      | 7                       | 4                      |
| 3 | 2           | 7                      | 8                       | 5                      |
| 4 | 3           | 9                      | 5                       | 3                      |
| 5 | 4           | 5                      | 4                       | 6                      |
| 6 | 5           | 6                      | 6                       | 2                      |
| 7 | 6           | 8                      | 7                       | 4                      |
| 8 | Середнє     |                        |                         |                        |

### Нульова гіпотеза:

Відмінності в продажах груп магазинів з різним рівнем реклами, є не більше вираженими, ніж випадкові відмінності усередині кожної групи. (Іншими словами: середні продажів в кожній з трьох груп не розрізняються).

### Рішення за допомогою електронних таблиць

1. Завантажити Excel.
2. Ввести дані. Розрахувати середні значення продажів для кожної групи магазинів.
3. Побудувати графік для порівняння середніх трьох груп магазинів (див. рис. 16.1).

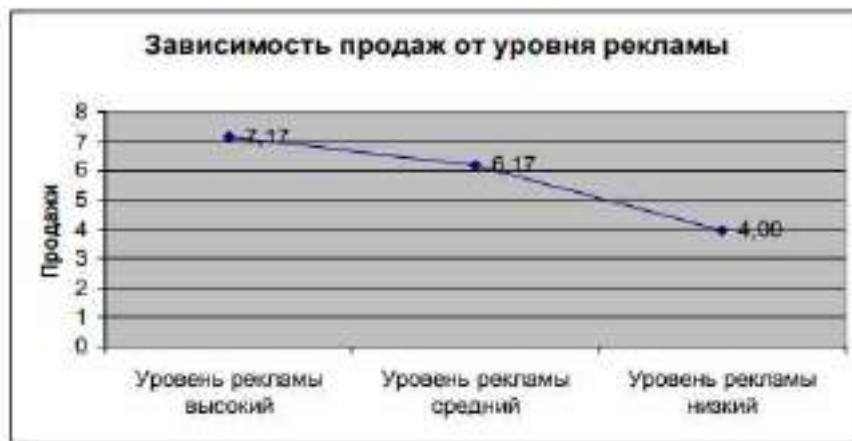


Рис. 16.1. Графік для порівняння середніх

4. Вибрати меню *Сервіс, Аналіз даних, Однофакторний ДА*.
5. У вікні «Однофакторный дисперсионный анализ» (рис. 16.2) задати необхідні параметри.

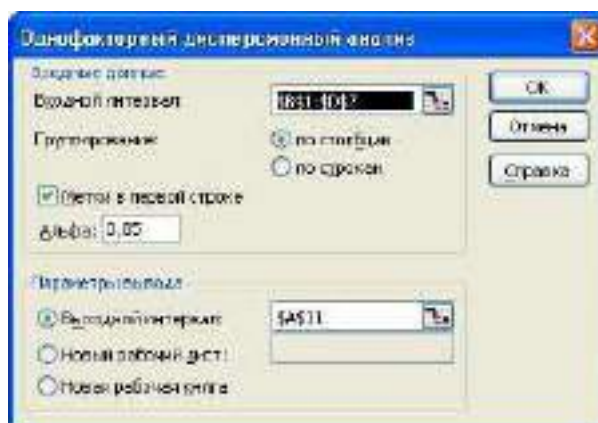


Рис. 16.2. Вікно «Однофакторный дисперсионный анализ»

6. У вікні результатів порівняти  $F$  критичне і  $F$  емпіричне (див. рис. 16.3).  
Якщо  $F$  критичне  $< F$  емпіричного, нульова гіпотеза відкидається.

| Однофакторний дисперсійний аналіз |                |             |                |                  |                     |                   |
|-----------------------------------|----------------|-------------|----------------|------------------|---------------------|-------------------|
| ПІДСУМКИ                          |                |             |                |                  |                     |                   |
| <i>Групи</i>                      | <i>Рахунок</i> | <i>Сума</i> | <i>Середнє</i> | <i>Дисперсія</i> |                     |                   |
| Рівень реклами високий            | 6              | 43          | 7,17           | 2,17             |                     |                   |
| Рівень реклами середній           | 6              | 37          | 6,17           | 2,17             |                     |                   |
| Рівень реклами низький            | 6              | 24          | 4              | 2                |                     |                   |
|                                   |                |             |                |                  |                     |                   |
| Дисперсійний аналіз               |                |             |                |                  |                     |                   |
| <i>Джерело варіації</i>           | <i>SS</i>      | <i>df</i>   | <i>MS</i>      | <i>F</i>         | <i>P - Значення</i> | <i>F критичне</i> |
| Між групами                       | 31,44444444    | 2           | 15,72          | 7,447            | 0,005671839         | 3,682             |
| Усередині груп                    | 31,66666667    | 15          | 2,11           |                  |                     |                   |
|                                   |                |             |                |                  |                     |                   |
| Разом                             | 63,11111111    | 17          |                |                  |                     |                   |

Рис. 16.3. Вікно результатів

### Рішення задачі в SPSS

1. Задати дві змінні:  $Y$  - залежна змінна - число продажів,  $F1$  - незалежна змінна – фактор (рівень реклами) з градаціями: 1 - низький, 2 - середній, 3 -

високий.

2. Ввести початкові дані (представлення даних в SPSS відрізняється від представлення в електронних таблицях і наводиться в таблиці 16.2).

Таблиця 16.2.

**Представлення початкових даних в SPSS для Windows**

| F1(рівень реклами) | Y(число продажів) |
|--------------------|-------------------|
| високий            | 8                 |
| високий            | 7                 |
| високий            | 9                 |
| високий            | 5                 |
| високий            | 6                 |
| високий            | 8                 |
| середній           | 7                 |
| середній           | 8                 |
| середній           | 5                 |
| середній           | 4                 |
| середній           | 6                 |
| середній           | 7                 |
| низький            | 4                 |
| низький            | 5                 |
| низький            | 3                 |
| низький            | 6                 |
| низький            | 2                 |
| низький            | 4                 |

3. Вибрати меню **Analyze, General Linear Model, Univariate**. Перенести змінну Y в список залежних змінних, змінну F1 - в список незалежних змінних (див. рис. 16.4).



Рис. 16.4. Вікно «Univariate»

4. Розглянути результати у вікні виведення (рис. 16.5).

Для нашого прикладу значення критерію Фишера  $F_{\text{емпирическое}}=7,447$  і значущість  $p=0,006$ , отже нульова гіпотеза відхиляється.

**Tests of Between-Subjects Effects**

Dependent Variable: Продажи

| Source          | Type III Sum of Squares | df | Mean Square | F       | Sig. |
|-----------------|-------------------------|----|-------------|---------|------|
| Corrected Model | 31,444 <sup>a</sup>     | 2  | 15,722      | 7,447   | ,006 |
| Intercept       | 600,889                 | 1  | 600,889     | 284,632 | ,000 |
| F1              | 31,444                  | 2  | 15,722      | 7,447   | ,006 |
| Error           | 31,667                  | 15 | 2,111       |         |      |
| Total           | 664,000                 | 18 |             |         |      |
| Corrected Total | 63,111                  | 17 |             |         |      |

a. R Squared = ,498 (Adjusted R Squared = ,431)

Рис. 16.5. Вікно виведення

5. Побудувати графік для порівняння середніх : **Analyze, General Linear Model, Univariate**, натиснути кнопку **Plot**, з'явиться діалогове вікно «**Profile Plots**» (див. рис. 16.6); перенести змінну F1 в рядок **Horizontal Axis**, натиснути кнопку **Add**, потім **Continue**, **OK**. На рис. 16.7 представлений графік, який має вийти в результаті цих дій.

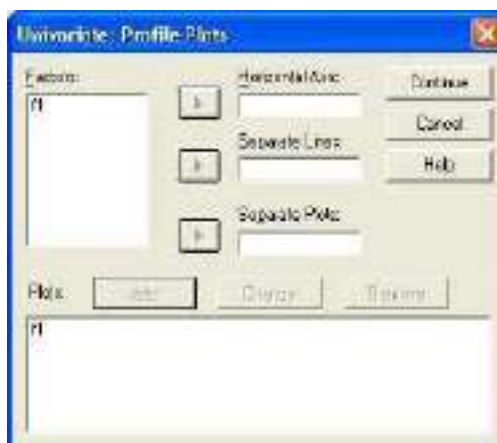


Рис. 16.6. Вікно «Univariate : Profile Plots»

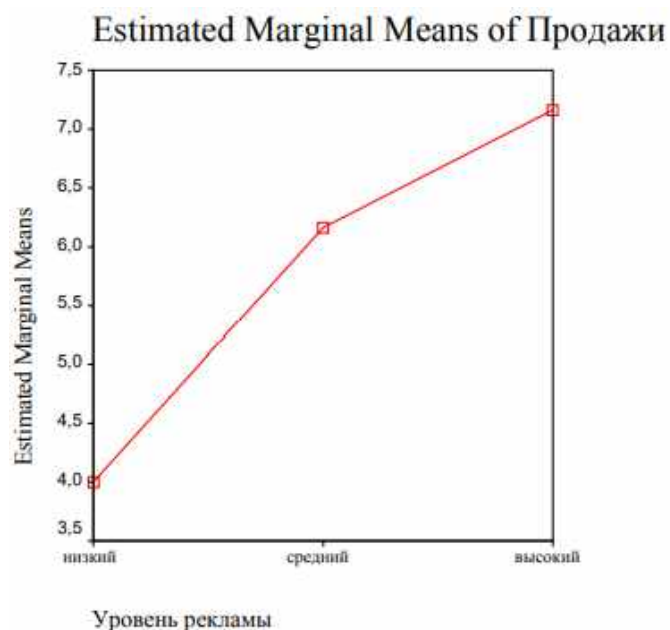


Рис. 16.7. Графік для порівняння середніх

### Завдання для самостійного виконання

1. Виконайте перевірку по F- критерію нульової гіпотези на рівні 0,01 по представлених в таблиці. 16.3 даним, які описують вагові втрати в кілограмах випробовуваними, такими, що дотримувалися різних дієт.

Таблиця 16.3

**Вагові втрати випробовуваних(кг), що дотримувалися різних дієт**

| Дієта 1 | Дієта 2 | Дієта 3 | Дієта 4 |
|---------|---------|---------|---------|
| 2,7     | 4,95    | 9,45    | 2,25    |
| 3,6     | 5,85    | 9       | 4,05    |
| 1,35    | 6,75    | 7,65    | 4,5     |
| 2,25    |         | 7,2     | 3,15    |
| 2,7     |         |         | 3,15    |

2. Для з'ясування впливу грошового стимулювання на продуктивність праці програмістів шести однорідним групам з 5 чоловік були запропоновані завдання однакової трудності. Завдання пропонувалися кожному випробовуваному незалежно від усіх інших. Групи відрізняються між собою величиною грошової винагороди за вирішувану задачу (таблиця 16.4).

Таблиця 16.4.

**Величина винагороди від меншої до більшої**

| Гр1 | Гр2 | Гр3 | Гр4 | Гр5 | Гр6 |
|-----|-----|-----|-----|-----|-----|
| 10  | 8   | 12  | 12  | 24  | 19  |
| 11  | 10  | 17  | 15  | 16  | 18  |
| 9   | 16  | 14  | 16  | 22  | 27  |
| 13  | 13  | 9   | 16  | 18  | 25  |
| 7   | 2   | 16  | 19  | 20  | 24  |

Визначте за допомогою однофакторного дисперсійного аналізу чи впливає величина винагороди на продуктивність праці програмістів?

**Вимоги до звіту**

Звіт про роботу повинен містити:

- постановку завдання, початкові дані, формулювання гіпотези, результати, висновки про істинність або хибність гіпотези;
- файл з даними.

**Контрольні запитання**

1. Призначення однофакторного дисперсійного аналізу.
2. Як робиться однофакторний дисперсійний аналіз за допомогою Excel?
3. Як робиться однофакторний дисперсійний аналіз за допомогою SPSS?
4. Дайте змістовну інтерпретацію отриманим результатам.

**ЛАБОРАТОРНЕ ЗАНЯТТЯ № 17****Тема: Двохфакторний дисперсійний аналіз**

**Мета:** навчитися використати пакет статистичного аналізу електронних таблиць Excel і можливості SPSS для Windows для двухфакторного дисперсійного аналізу.

**Постановка завдання**

Маркетологи використали три рівні реклами товарів в магазині: високий,

середній і низький. У купонного розпродажу були два рівні. Купон на 20-доларову знижку або давали потенційним покупцям (рівень в цьому випадку означали номером 1), або не давали (цей рівень означали номером 2 в таблиці 17.1 і в таблиці 17.2).

Результати експериментів з рекламою і купоном об'єднали в таблицю розміром 3x2 з шістьма комірками. Тридцять магазинів було вибрано випадковим чином, і для кожної комбінації умов експерименту (рівнів реклами і рівнів купонного розпродажу) випадковим чином узяли по п'ять магазинів, як показано в таблиці 17.2. Експеримент тривав два місяці. Визначили об'єм продажів в кожному магазині, нормалізували його, взявши до уваги сторонні чинники (розмір магазину, товарообіг і так далі) і перерахували за десятибальною шкалою. На додаток була отримана якісна оцінка відносного числа постійних покупців для кожного магазину, також з використанням десятибальної шкали. Отримані дані приведені в таблицю 17.3.

Треба перевірити, чи впливають рівень реклами і рівень купонного розпродажу на число продажів.

*Таблиця 17.1*

**Дані про продажі залежно від рівня реклами і рівня купонного розпродажу**

| Номер магазину | Рівень купонного розпродажу | Внутрішньо-магазинна реклама | Продажі | Постійні покупці |
|----------------|-----------------------------|------------------------------|---------|------------------|
| 1              | 1                           | 1                            | 10      | 9                |
| 2              | 1                           | 1                            | 9       | 10               |
| 3              | 1                           | 1                            | 10      | 8                |
| 4              | 1                           | 1                            | 8       | 4                |
| 5              | 1                           | 1                            | 9       | 6                |
| 6              | 1                           | 2                            | 8       | 8                |
| 7              | 1                           | 2                            | 8       | 4                |
| 8              | 1                           | 2                            | 7       | 10               |
| 9              | 1                           | 2                            | 9       | 6                |
| 10             | 1                           | 2                            | 6       | 9                |
| 11             | 1                           | 3                            | 5       | 8                |
| 12             | 1                           | 3                            | 7       | 9                |

|    |   |   |   |    |
|----|---|---|---|----|
| 13 | 1 | 3 | 6 | 6  |
| 14 | 1 | 3 | 4 | 10 |
| 15 | 1 | 3 | 5 | 4  |
| 16 | 2 | 1 | 8 | 10 |
| 17 | 2 | 1 | 9 | 6  |
| 18 | 2 | 1 | 7 | 8  |
| 19 | 2 | 1 | 7 | 4  |
| 20 | 2 | 1 | 6 | 9  |
| 21 | 2 | 2 | 4 | 6  |
| 22 | 2 | 2 | 5 | 8  |
| 23 | 2 | 2 | 5 | 10 |
| 24 | 2 | 2 | 6 | 4  |
| 25 | 2 | 2 | 4 | 9  |
| 26 | 2 | 3 | 2 | 4  |
| 27 | 2 | 3 | 3 | 6  |
| 28 | 2 | 3 | 2 | 10 |
| 29 | 2 | 3 | 1 | 9  |
| 30 | 2 | 3 | 2 | 8  |

### Рішення задачі за допомогою SPSS для Windows

1. У редакторові даних описати змінні: Y - залежна змінна - число продажів; F1 - незалежна змінна - фактор 1 - рівень купонного розпродажу з градаціями: 1 - купон давали, 2 - купон не давали; F2 - незалежна змінна 2 (фактор 2) - рівень реклами з градаціями: 1 - низький, 2 - середній, 3 - високий.

2. Ввести початкові дані.

3. Перенести змінні у вікні «*Univariate*» (див. рис. 17.1)



Рис. 17.1. Вікно «Univariate»

4. Для побудови графіку натиснути кнопку Plot, перенести змінну f1 в полі **Horizontal Axis**, f2 в полі **Separate Lines**. Клацнути на вимикачі **Add** (див. рис. 17.2), натиснути **Continue**, **OK**.

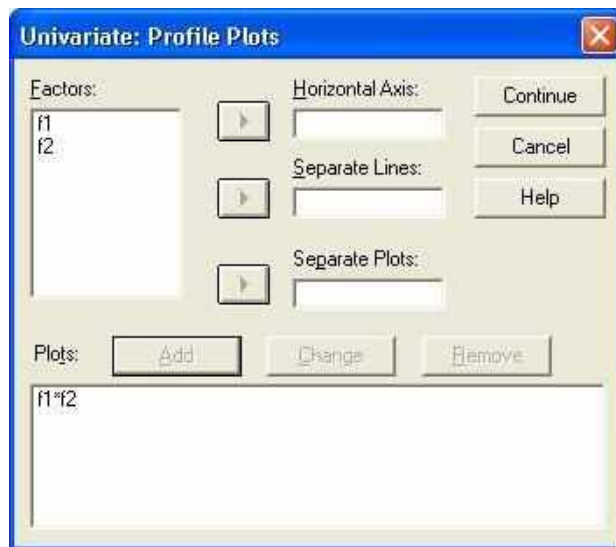


Рис. 17.2. Діалогове вікно «Univariate : Profile Plots»

5. Розглянути результати у вікні виведення (рис. 17.3 - 17.4).

### Tests of Between-Subjects Effects

Dependent Variable: продаж

| Source          | Type III Sum of Squares | df | Mean Square | F        | Sig. |
|-----------------|-------------------------|----|-------------|----------|------|
| Corrected Model | 162.667(a)              | 5  | 32.533      | 33.655   | .000 |
| Intercept       | 1104.133                | 1  | 1104.133    | 1142.207 | .000 |
| KUPON           | 53.333                  | 1  | 53.333      | 55.172   | .000 |
| RECLAMA         | 106.067                 | 2  | 53.033      | 54.862   | .000 |
| KUPON * RECLAMA | 3.267                   | 2  | 1.633       | 1.690    | .206 |
| Error           | 23.200                  | 24 | .967        |          |      |
| Total           | 1290.000                | 30 |             |          |      |
| Corrected Total | 185.867                 | 29 |             |          |      |

a. R Squared = .875 (Adjusted R Squared = .849)

Рис. 17.3. Результати виконання дисперсійного аналізу у вікні виведення

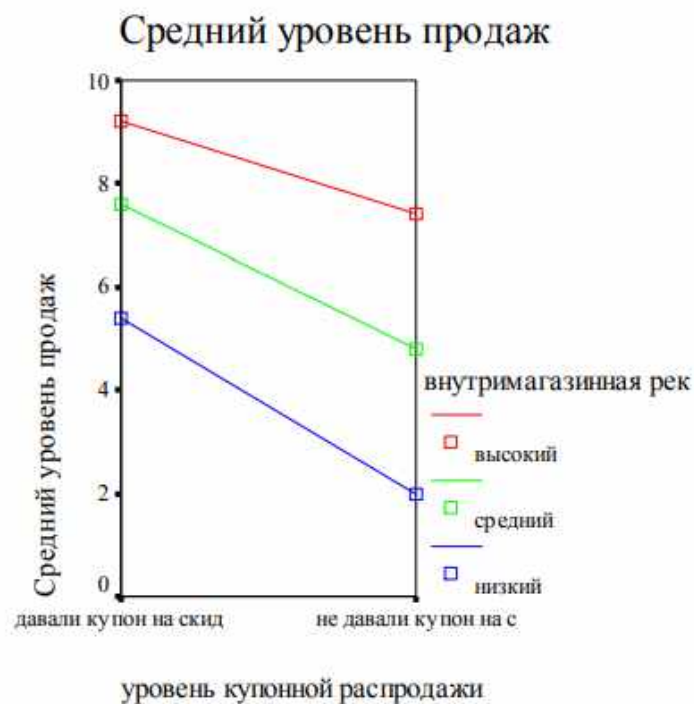


Рис. 17.4. Графік для середніх

Аналіз отриманих результатів показує, що на рівень продажів впливає рівень купонної реклами і рівень внутрішньомagaзинної реклами, а їх взаємодія виявляється не значимою.

## **Рішення задачі двохфакторного дисперсійного аналізу за допомогою електронних таблиць**

### **Постановка завдання**

В одному з досліджень вивчалася така особова риса, як домінантність дорослих чоловіків і жінок. Автори припускали, що домінантність має бути вища у людей, які були первістками у своїх сім'ях. Виявилось, що вплив кожного з двох досліджуваних факторів - статі і порядку народження незначима, а їх взаємодія - значима. У чоловіків домінантність зі збільшенням порядку народження знижується, а у жінок - навпаки підвищується. За даними рис. 17.4. перевірте за допомогою двохфакторного дисперсійного аналізу (ДА) наведене вище твердження.

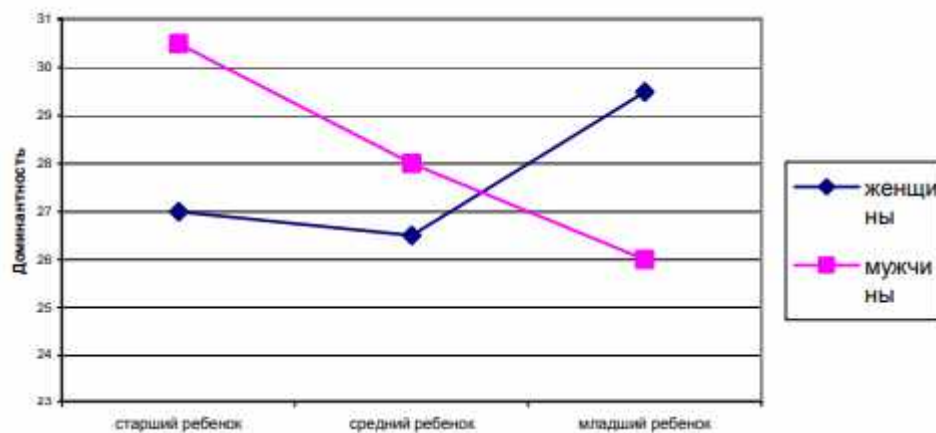
|   | В                                                                             | С          | Д             |
|---|-------------------------------------------------------------------------------|------------|---------------|
| 1 | Показники домінантності у чоловіків і жінок<br>залежно від порядку народження |            |               |
| 2 |                                                                               |            |               |
| 3 |                                                                               | A1 - жінки | A2 - чоловіки |
| 4 | B1 - старша дитина                                                            | 26         | 30            |
| 5 |                                                                               | 28         | 31            |
| 6 | B2 - середня дитина                                                           | 27         | 27            |
| 7 |                                                                               | 26         | 29            |
| 8 | B3 - молодша дитина                                                           | 29         | 25            |
| 9 |                                                                               | 30         | 27            |

Рис. 17.4. Показники домінантності у чоловіків і жінок залежно від порядку народження

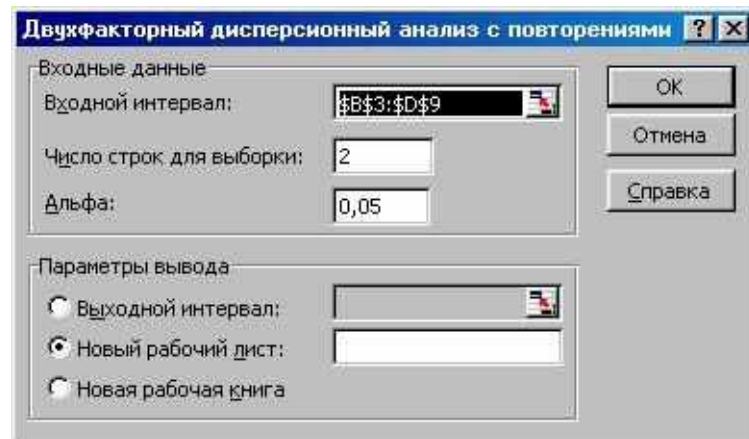
### **Хід роботи**

1. Ввести дані. Розрахувати середні значення показників домінантності для шести груп.
2. Побудувати діаграму для порівняння середніх значень показника домінантності шести груп випробовуваних.

## Зміна показників домінантності залежно від порядку народження чоловіків і жінок



3. Вибрати меню *Дані, Аналіз даних, Двухфакторный ДА*.
4. У вікні дисперсійного аналізу задати необхідні параметри.



5. Розглянути результати.

6. У вікні результатів порівняти три значення  $F$  критичного і три значення  $F$  емпіричного. Визначити, чи впливає на домінантність дитини чинник статі, чинник порядку народження і їх взаємодія.

| Источник вариации | SS    | df    | MS    | F     | P-Значение | F критическое |
|-------------------|-------|-------|-------|-------|------------|---------------|
| Выборка           | 4,67  | 2,00  | 2,33  | 1,87  | 0,23       | 5,14          |
| Столбцы           | 0,75  | 1,00  | 0,75  | 0,60  | 0,47       | 5,99          |
| Взаимодействие    | 26,00 | 2,00  | 13,00 | 10,40 | 0,01       | 5,14          |
| Внутри            | 7,50  | 6,00  | 1,25  |       |            |               |
| Итого             | 38,92 | 11,00 |       |       |            |               |

Якщо  $F$  критичне  $< F$  емпіричного, нульова гіпотеза відкидається.

### Завдання для самостійної виконання:

1. В одному з експериментів по психології програмування вивчався вплив коментарів в лістингу на легкість модифікації програми. Студенти початкового курсу були розділені на дві групи по 6 чоловік в кожній. В якості тесту було запропоновано дві форми 26-рядкової програми на Фортрані, оперуючої інформацією про оцінки студентів. Перша форма містила єдиний блок коментарів високого рівня на початку програми, що повідомляє про її призначення. Друга форма програми не містила коментарів високого рівня, але мала 19 однорядкових коментарів низького рівня. Вивчивши програму і ті, і інші студенти повинні були провести три її модифікації, на які давалося 30 хвилин. Три завдання на модифікацію оцінювалися за 10-бальною шкалою. Результати експерименту приведені в таблиці.

#### Оцінки продуктивності в експерименті з коментуванням програми

| Коментарі     | Оцінка модифікації |          |          |
|---------------|--------------------|----------|----------|
|               | 1-а мод.           | 2-а мод. | 3-я мод. |
| Високорівневі | 6                  | 8        | 3        |
|               | 7                  | 10       | 4        |
|               | 5                  | 8        | 4        |
|               | 7                  | 7        | 5        |
|               | 7                  | 9        | 5        |
|               | 7                  | 9        | 4        |
| Низькорівневі | 6                  | 6        | 1        |
|               | 6                  | 7        | 2        |
|               | 7                  | 8        | 3        |
|               | 5                  | 6        | 2        |
|               | 6                  | 6        | 2        |
|               | 5                  | 5        | 1        |

Чи впливає на результати тесту (продуктивність праці студентів) рівень коментування програми, вид її модифікації або їх взаємодія?

### **Вимоги до звіту**

Звіт про роботу повинен містити:

- постановку завдання, початкові дані, формулювання гіпотези, результати, висновки про істинність або хибність гіпотез;
- файли з даними.

### **Контрольні запитання**

1. Яка мета двохфакторного дисперсійного аналізу?
2. Як робиться двохфакторний дисперсійний аналіз за допомогою Excel?
3. Як інтерпретуються його результати?
4. Як задаються дані для дисперсійного аналізу за допомогою SPSS? Як інтерпретуються результати?
5. Наведіть свій приклад завдання для двохфакторного дисперсійного аналізу і вирішіть її.

## **ЛАБОРАТОРНЕ ЗАНЯТТЯ № 18**

**Тема:** *Дискримінантний аналіз у SPSS для Windows*

**Мета:** навчитися використати можливості SPSS для аналізу дискримінанта даних і інтерпретувати його результати

### **Теоретичні відомості**

За допомогою дискримінантного аналізу на підставі деяких ознак (незалежних змінних) об'єкт може бути зарахований до однієї з двох або декількох груп (число груп визначається числом категорій залежної змінної). У двовимірній аналізі дискримінанта об'єкти відносяться до однієї з двох груп, наприклад, що купили або не купили даних продукт. А незалежними змінними в цьому випадку виступають вік, доход покупців, та ін. показники.

Аналіз дискримінанта використовується для аналізу даних у тому випадку, коли залежна змінна категоріальна, а предиктори (незалежні змінні)

— інтервальні. В результаті аналізу дискримінанта будується так звана функція дискримінанта

$d = b_1x_1 + b_2x_2 + \dots + b_nx_n + a$ , де  $x_1$  і  $x_n$  — значення змінних, що відповідають даним випадкам,  $b_1 - b_n$ ,  $a$  - коефіцієнти, які і належить оцінити за допомогою аналізу дискримінанта. Коефіцієнти підбираються так, щоб по значеннях функції дискримінанта можна було з максимальною чіткістю провести розділення по групах.

Процедура **дискримінантного аналізу** складається з п'яти кроків. Перший крок — формулювання проблеми, вимагає **визначення** цілей, залежною і незалежних змінних. Вибірку ділять на дві частини. Аналізовану вибірку використовують для обчислення функції дискримінанта; **перевірочну** — для перевірки достовірності моделі. Другий **крок** — визначення функції, включає виведення такої лінійної комбінації предикторів (функцій **дискримінантів**), щоб групи максимально можливо розрізнялися між собою значеннями предикторів.

Визначення статистичної значущості є третім кроком. Вона включає перевірку нульової гіпотези про те, що в сукупності середні усіх функцій дискримінантів в усіх групах рівні між собою. Якщо нульову гіпотезу відхиляють, то має сенс інтерпретувати результати.

Четвертий крок — інтерпретація дискримінантних вагів або **коефіцієнтів** аналогічна такій же стадії в множинному регресійному аналізі.

П'ятий крок — перевірка достовірності. Вона включає розробку класифікаційної матриці. Дискримінантні ваги, визначені за допомогою аналізованої вибірки, множать на значення незалежних змінних в перевірочній вибірці, щоб отримати показники дискримінантів для випадків в цій вибірці. Потім випадки розподіляють по групах, виходячи з показників дискримінантів і відповідного правила ухвалення рішення. Визначають відсоток вірно класифікованих випадків і порівнюють його з відсотком випадків, яке можна чекати на основі класифікації методом випадкового вибору.

Для оцінки коефіцієнтів існує два відомі підходи. Прямий метод включає оцінку функції дискримінанта при одночасному введенні усіх предикторів. Альтернативний йому покроковий метод включає послідовне введення передбачених змінних, виходячи з їх здатності дискримінувати групи.

### Постановка завдання

Маркетологи вивчали змінні, що впливають на сімейний відпочинок. Сім'ям, які відпочивали на курорті в останні два роки, присвоєний код 1; тим же, які не відвідали курорт за вказаний період часу, присвоєний код 2.

Обидві вибірки (як аналізована, так і перевірна) збалансовані з точки зору відвідуваності курорту. Аналізована вибірка містить 15 сімей кожної категорії, а перевірна — по 6 сімей кожної категорії.

Крім того, отримані дані про дохід, відношення до подорожі, значення, що надається сімейному відпочинку, розміру сім'ї і віку глави сім'ї. Усі дані наведені в таблиці 18.1.

*Таблиця 18.1.*

### Чинники, що впливають на сімейний відпочинок

| Номер | Відвідування курорту | Щорічний дохід | Відношення до подорожі | Значення, що надається сімейному відпочинку | Розмір сім'ї | Вік глави сім'ї |
|-------|----------------------|----------------|------------------------|---------------------------------------------|--------------|-----------------|
| 1     | 1                    | 50,2           | 5                      | 8                                           | 3            | 43              |
| 2     | 1                    | 70,3           | 5                      | 7                                           | 4            | 61              |
| 3     | 1                    | 52,9           | 7                      | 5                                           | 6            | 52              |
| 4     | 1                    | 46,5           | 7                      | 5                                           | 5            | 36              |
| 5     | 1                    | 52,7           | 6                      | 6                                           | 6            | 55              |
| 6     | 1                    | 75             | 8                      | 7                                           | 5            | 68              |
| 7     | 1                    | 46,2           | 5                      | 3                                           | 3            | 62              |
| 8     | 1                    | 57             | 2                      | 4                                           | 6            | 51              |

|    |   |      |   |   |   |    |
|----|---|------|---|---|---|----|
| 9  | 1 | 64,1 | 7 | 5 | 4 | 57 |
| 10 | 1 | 68,1 | 7 | 6 | 5 | 45 |
| 11 | 1 | 73,4 | 6 | 7 | 5 | 44 |
| 12 | 1 | 7,9  | 5 | 8 | 4 | 64 |
| 13 | 1 | 56,2 | 1 | 8 | 6 | 54 |
| 14 | 1 | 49,3 | 4 | 2 | 3 | 56 |
| 15 | 1 | 62   | 5 | 6 | 2 | 58 |
| 16 | 2 | 32,1 | 5 | 4 | 3 | 58 |
| 17 | 2 | 36,2 | 4 | 3 | 2 | 55 |
| 18 | 2 | 43,2 | 2 | 5 | 2 | 57 |
| 19 | 2 | 50,4 | 5 | 2 | 4 | 37 |
| 20 | 2 | 44,1 | 6 | 6 | 3 | 42 |
| 21 | 2 | 38,3 | 6 | 6 | 2 | 45 |
| 22 | 2 | 55   | 1 | 2 | 3 | 57 |
| 23 | 2 | 45,1 | 3 | 5 | 3 | 51 |
| 24 | 2 | 35   | 6 | 4 | 5 | 64 |
| 25 | 2 | 37,3 | 2 | 7 | 4 | 54 |
| 26 | 2 | 41,8 | 5 | 3 | 3 | 56 |
| 27 | 2 | 57,1 | 8 | 3 | 2 | 36 |
| 28 | 2 | 33,4 | 6 | 8 | 2 | 50 |
| 29 | 2 | 37,5 | 3 | 2 | 3 | 48 |
| 30 | 2 | 41,3 | 3 | 3 | 2 | 42 |

### Рішення задачі за допомогою SPSS

#### Хід роботи

1. Описати змінні і ввести дані згідно таблиці 18.1.
2. Вибрати Меню *Analyze, Classify, Discriminant, Analysis*.
3. У вікні, що відкрилося, перенести змінну **відпочинок** (відвідування курорту) у вікно групуючої змінної, а інші змінні в список незалежних змінних

або предикторів.

4. Після клацання по вимикачу **Define Range...** (*Визначити проміжок*) ввести мінімальне і максимальне значення змінної **відпочинок**: 1 і 2 (рис. 18.1).

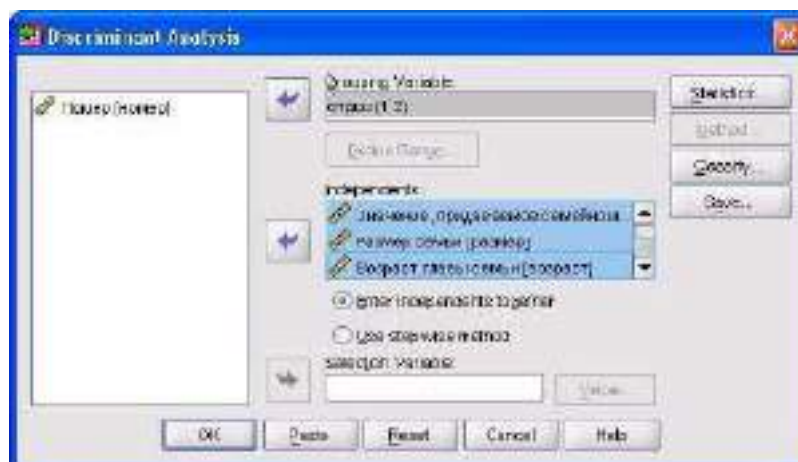


Рис. 18.1. Діалогове вікно «Discriminant Analysis»

5. Спершу залишити встановлений за умовчанням метод: **Enter independents together** (Одночасний облік усіх незалежних змінних), при якому в аналізі одночасно братимуть участь усі незалежні змінні.

6. Натиснути кнопку **Statistics** і активувати опції: **Means** (Середні значення), **Univariate ANOVAs** (Одновимірні тести ANOVA), **Unstandardized Function Coefficients** (Нестандартизовані коефіцієнти функції) і **Within-group Correlation Matrice** (Кореляційна матриця усередині групи) (рис. 18.2).



Рис. 18.2. Діалогове вікно «Discriminant Analysis Statistics»

7. Натиснути кнопку **Classify** і зробити додатково запит на виведення діаграм по окремих групах (**Separate groups Plots**), результатів для окремих спостережень (**Casewise results**) і зведеної таблиці (**Summary table**) (рис. 18.3.)

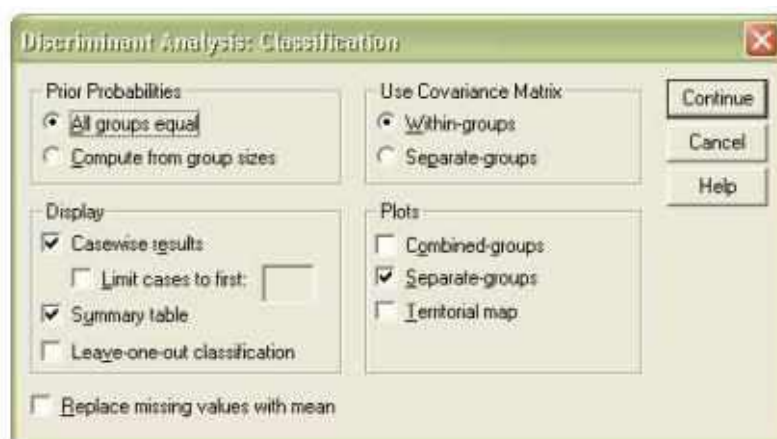


Рис. 18.3. Діалогове вікно «Discriminant Analysis : Classification»

8. За допомогою вимикача **Save...** зберегти значення функції дискримінанта в додатковій змінній (Discriminant Scores).

9. Розглянути і проаналізувати результати у вікні виведення.

У таблиці «Статистика груп» (рис. 18.4) наводяться середні значення і стандартні відхилення для усіх ознак в кожній групі окремо і сумарні показники.

| Group Statistics  |                                        |         |                |                    |          |
|-------------------|----------------------------------------|---------|----------------|--------------------|----------|
| Посещение курорта |                                        | Mean    | Std. Deviation | Valid N (listwise) |          |
|                   |                                        |         |                | Unweighted         | Weighted |
| 1                 | Ежегодный доход                        | 59.7200 | 10.16951       | 15                 | 15.000   |
|                   | Отношение к путешествию                | 5.3333  | 1.91485        | 15                 | 15.000   |
|                   | Значение, придаваемое семейному отдыху | 5.8000  | 1.82052        | 15                 | 15.000   |
|                   | Размер семьи                           | 4.4667  | 1.30201        | 15                 | 15.000   |
|                   | Возраст главы семьи                    | 53.7333 | 8.77062        | 15                 | 15.000   |
| 2                 | Ежегодный доход                        | 41.8467 | 7.51588        | 15                 | 15.000   |
|                   | Отношение к путешествию                | 4.3333  | 1.95180        | 15                 | 15.000   |
|                   | Значение, придаваемое семейному отдыху | 4.2000  | 1.89737        | 15                 | 15.000   |
|                   | Размер семьи                           | 2.8667  | .91548         | 15                 | 15.000   |
|                   | Возраст главы семьи                    | 50.1333 | 8.27101        | 15                 | 15.000   |
| Total             | Ежегодный доход                        | 50.7833 | 12.64178       | 30                 | 30.000   |
|                   | Отношение к путешествию                | 4.8333  | 1.96668        | 30                 | 30.000   |
|                   | Значение, придаваемое семейному отдыху | 5.0000  | 2.00000        | 30                 | 30.000   |
|                   | Размер семьи                           | 3.6667  | 1.37297        | 30                 | 30.000   |
|                   | Возраст главы семьи                    | 51.9333 | 8.57395        | 30                 | 30.000   |

Рис. 18.4. Статистика групп

У наступній таблиці (рис. 18.5) наводяться результати тесту про те, наскільки значимо розрізняються між собою змінні в обох групах. Для цього наводяться значення тестової величини Лямбда Уилкса («Wilks-Lambda») і

застосовується простий дисперсійний аналіз. Для усіх змінних (окрім віку і відношення до подорожі) виходить значима відмінність між обома групами.

| Tests of Equality of Group Means       |               |        |     |     |      |
|----------------------------------------|---------------|--------|-----|-----|------|
|                                        | Wilks' Lambda | F      | df1 | df2 | Sig. |
| Ежегодный доход                        | .483          | 29.966 | 1   | 28  | .000 |
| Отношение к путешествию                | .933          | 2.006  | 1   | 28  | .168 |
| Значение, придаваемое семейному отдыху | .834          | 5.554  | 1   | 28  | .026 |
| Размер семьи                           | .649          | 15.158 | 1   | 28  | .001 |
| Возраст главы семьи                    | .954          | 1.338  | 1   | 28  | .257 |

Рис. 18.5. Тест на рівність групових середніх

Потім обчислюється кореляційна матриця (рис. 18.6).

| Pooled Within-Groups Matrices |                                        |                 |                         |                                        |              |                     |
|-------------------------------|----------------------------------------|-----------------|-------------------------|----------------------------------------|--------------|---------------------|
|                               |                                        | Ежегодный доход | Отношение к путешествию | Значение, придаваемое семейному отдыху | Размер семьи | Возраст главы семьи |
| Correlation                   | Ежегодный доход                        | 1.000           | .135                    | .112                                   | .017         | .010                |
|                               | Отношение к путешествию                | .135            | 1.000                   | .089                                   | -.044        | -.213               |
|                               | Значение, придаваемое семейному отдыху | .112            | .089                    | 1.000                                  | .048         | .045                |
|                               | Размер семьи                           | .017            | -.044                   | .048                                   | 1.000        | -.007               |
|                               | Возраст главы семьи                    | .010            | -.213                   | .045                                   | -.007        | 1.000               |

Рис. 18.6. Кореляційна матриця

На наступному кроці обчислюються і аналізуються коефіцієнти дискримінантної функції.

| Eigenvalues |                    |               |              |                       |
|-------------|--------------------|---------------|--------------|-----------------------|
| Function    | Eigenvalue         | % of Variance | Cumulative % | Canonical Correlation |
| 1           | 1.751 <sup>a</sup> | 100.0         | 100.0        | .798                  |

a. First 1 canonical discriminant functions were used in the analysis

| Wilks' Lambda       |               |            |    |      |
|---------------------|---------------|------------|----|------|
| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
| 1                   | .364          | 25.803     | 5  | .000 |

Рис. 18.7. Власні значення і лямбда Уїлкоксона

При допомозі Лямбда Уїлкса (рис. 18.7) робиться тест на те, чи значимо в обох групах відрізняються один від одного середні значення функції дискримінанта; значення  $p < 0,05$  вказує на дуже значиму відмінність.

Значення, що виводиться під ім'ям «Eigenvalue» (Власне значення), відповідає відношенню суми квадратів між групами до суми квадратів усередині груп. Великі власні значення вказують на вдало підібрані дискримінантні функції.

Наступна таблиця дає уявлення про те, як сильно окремі змінні, що застосовуються у дискримінантній функції, корелюють із стандартизованими значеннями цієї дискримінантної функції. При цьому кореляційні коефіцієнти були розраховані в обох групах окремо і потім усереднені.

Далі наводяться самі коефіцієнти дискримінантної функції (рис. 18.8).

**Canonical Discriminant Function Coefficients**

|                                        | Function |
|----------------------------------------|----------|
|                                        | 1        |
| Ежегодный доход                        | .081     |
| Отношение к путешествию                | .077     |
| Значение, придаваемое семейному отдыху | .112     |
| Размер семьи                           | .481     |
| Возраст главы семьи                    | .022     |
| (Constant)                             | -7.944   |

Unstandardized coefficients

Рис. 18.8. Нестандартизовані коефіцієнти

Наприкінці наводиться класифікаційна таблиця (рис. 18.9) з вказівкою досягнутої точності прогнозування. Значення цієї точності рівне 90%, що є гарним результатом.

**Classification Results<sup>a</sup>**

|          |       | Predicted Group Membership |       | Total |
|----------|-------|----------------------------|-------|-------|
|          |       | 1                          | 2     |       |
| Original | Count | Посещение курорта          |       |       |
|          |       | 1                          | 2     |       |
|          | 1     | 12                         | 3     | 15    |
|          | 2     | 0                          | 15    | 15    |
|          | %     |                            |       |       |
|          | 1     | 80.0                       | 20.0  | 100.0 |
|          | 2     | .0                         | 100.0 | 100.0 |

<sup>a</sup>. 90.0% of original grouped cases correctly classified.

Рис. 18.9. Результати класифікації

### **Вимоги до звіту**

- файли з початковими даними і результатами розрахунків;
- відповіді на контрольні питання.

### **Контрольні запитання**

1. Призначення аналізу дискримінанта.
2. Як він здійснюється в SPSS?
3. Назвіть етапи виконання аналізу дискримінанта.
4. Як розділити загальну вибірку для цілей аналізу і перевірки?
5. Що таке коефіцієнт  $\lambda$  Уїлкса? Для яких цілей його використовують?
6. Що таке класифікаційна матриця?
7. Як визначають статистичну значущість аналізу дискримінанта?
8. Чим відрізняється покроковий метод дискримінанта від прямого?
9. Дайте змістовну інтерпретацію отриманим результатам.

## **ЛАБОРАТОРНЕ ЗАНЯТТЯ № 19**

### **Тема: Ієрархічний кластерний аналіз у SPSS для Windows**

**Мета:** навчитися робити ієрархічний кластерний аналіз і описувати отримані кластери.

### **Теоретичні відомості**

**Кластерний аналіз** (таксономія, автокласифікація, розпізнавання образів) можна назвати методом конденсації інформації. Завданням кластерного аналізу є таке представлення багатовимірного масиву інформації в стислому виді, щоб втрата інформації не була надмірною. Він дозволяє об'єднати величезну кількість об'єктів в невелике число однорідних груп.

**Приклад.** Нехай ми маємо матрицю даних, яка включає характеристики 22-х людей (об'єктів) за двома кількісними ознаками: стаж і зарплата. Відкладаючи значення ознак по осях координат, ми можемо зображувати усі

об'єкти на площині у вигляді точок: X - стаж, Y - зарплата.

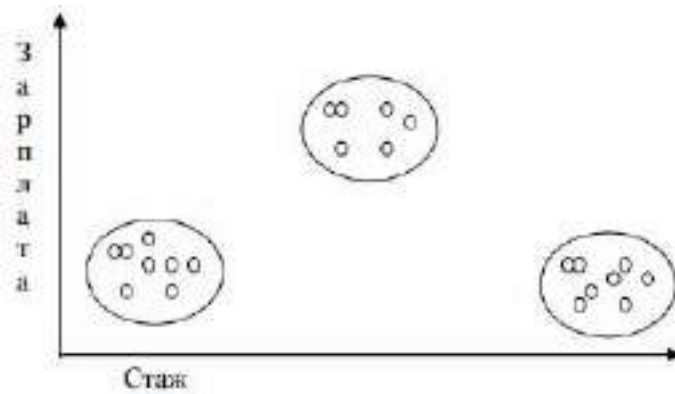


Рис. 19.1. Графічне представлення кластерів

Як видно з рис. 19.1, усі об'єкти можна розбити на три групи так, що усередині однієї групи знаходяться об'єкти з близькими значеннями ознак. Перша група - це люди з невеликим стажем роботи і низькою зарплатою, друга - з середнім стажем і середньою зарплатою і третя група - люди з великим стажем і низькою зарплатою.

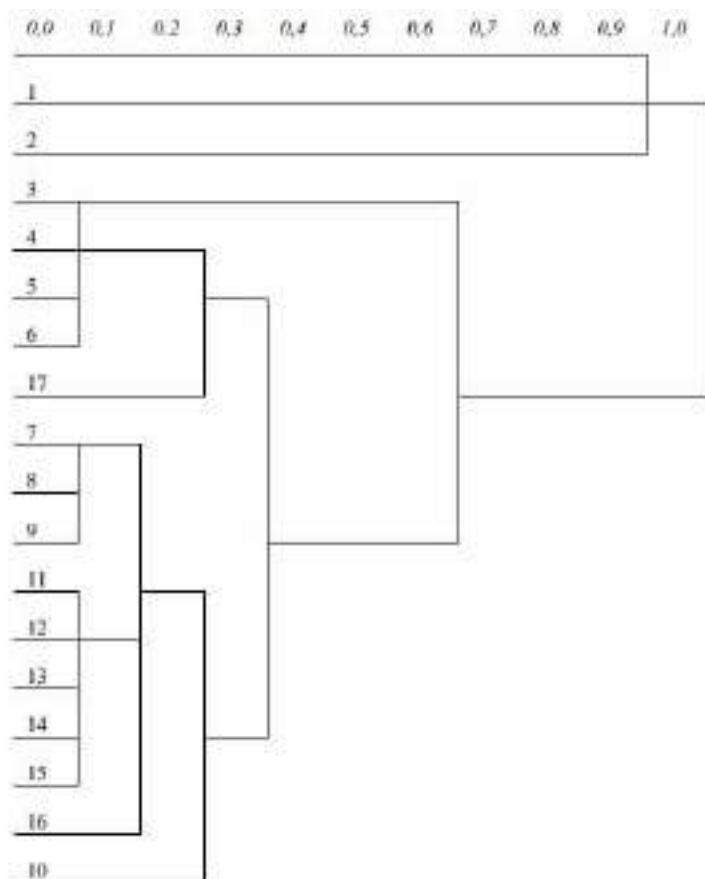
Величезна кількість близьких між собою точок називається **кластером** або **таксоном** і при інтерпретації результатів розглядається як деякий тип.

Одним з різновидів кластерного аналізу є **ієрархічний** кластерний аналіз. Перед початком ієрархічної кластеризації усі об'єкти вважаються окремими кластерами, які в ході алгоритму об'єднуються. На першому кроці вибирається пара найближчих кластерів, які об'єднуються в один кластер. В результаті число кластерів стає  $N-1$ . Процедура повторюється, поки усі класи не об'єднуються. На будь-якому етапі об'єднання можна перервати, отримавши потрібне число кластерів. Результат роботи алгоритму визначають способи обчислення відстані між об'єктами і визначення близькості між кластерами. Найбільш споживаною мірою виміру відстані є відстань **Евклідова відстань** між об'єктами.

Процес агрегації даних може бути графічно представлений за допомогою дендрограми (інші назви: «**дендограма-дерево**» «**діаграма-дерево**»).

**Дендограма** - це графічне зображення процедури і результатів процесу послідовної кластеризації, який здійснюється в термінах матриці відстаней або

У дендрограмі-дереві об'єкти розташовуються вертикально ліворуч (номери об'єктів на рис. 19.2 позначені звичайним шрифтом), а результати кластеризації – справа. Значення відстаней або схожості, що відповідають побудові нових кластерів, зображаються у вигляді горизонтального ряду цифр, розташованого поверх дендрограми (на рис. 19.2 – цифри, набрані курсивом).



Рішення про число кластерів приймають з теоретичних і практичних міркувань. У ієрархічній кластеризації важливим критерієм ухвалення рішення про число кластерів є відстані, при яких відбувається об'єднання кластерів.

Результатом роботи алгоритмів таксономії зазвичай є розбиття величезної кількості об'єктів на групи (кластери) в просторі ознак, заданих дослідником, а

також розрахунок деяких узагальнених характеристик кожного кластера (центр, середні, заходи варіації).

Існують методи, які дозволяють проводити класифікацію в просторі ознак, вимірюваних за допомогою різного типу шкал: номінальних, порядкових, рангових.

### Постановка завдання

18 претендентів пройшли 10 різних тестів у відділі кадрів підприємства. Максимальна оцінка, яку можна було отримати по тесту - 10 балів. Список тестів був наступним:

1. Пам'ять на числа.
2. Математичні завдання.
3. Винахідливість при прямому діалозі.
4. Тест на складання алгоритмів.
5. Упевненість під час виступу.
6. Командний дух.
7. Винахідливість.
8. Співпраця.
9. Визнання в колективі.
10. Сила переконання.

Розподіліть претендентів в кластери, виходячи з результатів тестування.

### Хід роботи

1. Опишіть змінні і введіть дані відповідно до наступної таблиці (для змінної з прізвищем задайте тип **String**).

Таблиця 19.1

### Результати тестування претендентів

| № | Претендент | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | T10 |
|---|------------|----|----|----|----|----|----|----|----|----|-----|
| 1 | Volker R   | 10 | 10 | 9  | 10 | 10 | 10 | 9  | 10 | 10 | 9   |
| 2 | Sigrid K   | 10 | 10 | 4  | 10 | 5  | 5  | 4  | 5  | 4  | 3   |
| 3 | Elmar M    | 5  | 4  | 10 | 5  | 10 | 4  | 10 | 5  | 3  | 10  |

|    |            |    |    |    |    |    |    |    |    |    |    |
|----|------------|----|----|----|----|----|----|----|----|----|----|
| 4  | Peter B    | 10 | 10 | 9  | 10 | 10 | 10 | 9  | 10 | 10 | 9  |
| 5  | Otto R     | 4  | 3  | 5  | 4  | 3  | 10 | 4  | 10 | 10 | 5  |
| 6  | Elke M     | 10 | 10 | 4  | 10 | 5  | 4  | 3  | 4  | 5  | 5  |
| 7  | Sarah K    | 4  | 4  | 5  | 5  | 4  | 10 | 5  | 10 | 10 | 6  |
| 8  | Peter T    | 4  | 5  | 3  | 4  | 5  | 10 | 4  | 10 | 10 | 4  |
| 9  | Gudrun M   | 4  | 5  | 10 | 4  | 10 | 5  | 10 | 4  | 3  | 10 |
| 10 | Siglinde P | 10 | 10 | 4  | 10 | 5  | 4  | 4  | 5  | 4  | 4  |
| 11 | Werner W   | 4  | 5  | 10 | 5  | 10 | 4  | 10 | 4  | 5  | 10 |
| 12 | Achim Z    | 10 | 10 | 9  | 10 | 10 | 9  | 9  | 10 | 10 | 10 |
| 13 | Dieter K   | 6  | 5  | 4  | 3  | 5  | 10 | 5  | 10 | 10 | 5  |
| 14 | Boris P    | 4  | 5  | 10 | 4  | 10 | 5  | 10 | 3  | 4  | 10 |
| 15 | Silke W    | 10 | 10 | 9  | 10 | 10 | 9  | 10 | 9  | 10 | 10 |
| 16 | Clara T    | 6  | 5  | 3  | 4  | 4  | 10 | 4  | 10 | 10 | 5  |
| 17 | Manfred K  | 10 | 10 | 5  | 10 | 4  | 5  | 4  | 3  | 4  | 5  |
| 18 | Richard M  | 4  | 5  | 10 | 4  | 10 | 4  | 10 | 4  | 4  | 10 |

2. Виберіть меню *Analyze, Classify, Hierarchical Cluster* (Ієрархічний кластерний аналіз) (рис. 19.3)



Рис. 19.3. Діалогове вікно «Hierarchical Cluster Analysis»

3. Помістіть змінні **t1** - **t10** в полі тестованих змінних, а змінну ім'я претендента(**name**) - використайте для позначення випадків.

Оглядова таблиця порядку злиття повинна виглядати таким чином (див.

рис. 19.4).

| Agglomeration Schedule |                  |           |              |                             |           |            |
|------------------------|------------------|-----------|--------------|-----------------------------|-----------|------------|
| Stage                  | Cluster Combined |           | Coefficients | Stage Cluster First Appears |           | Next Stage |
|                        | Cluster 1        | Cluster 2 |              | Cluster 1                   | Cluster 2 |            |
| 1                      | 1                | 4         | ,000         | 0                           | 0         | 6          |
| 2                      | 14               | 18        | 2,000        | 0                           | 0         | 4          |
| 3                      | 12               | 15        | 2,000        | 0                           | 0         | 6          |
| 4                      | 9                | 14        | 2,000        | 0                           | 2         | 8          |
| 5                      | 2                | 10        | 2,000        | 0                           | 0         | 13         |
| 6                      | 1                | 12        | 3,000        | 1                           | 3         | 15         |
| 7                      | 13               | 16        | 4,000        | 0                           | 0         | 12         |
| 8                      | 9                | 11        | 4,000        | 4                           | 0         | 11         |
| 9                      | 5                | 7         | 5,000        | 0                           | 0         | 14         |
| 10                     | 6                | 17        | 6,000        | 0                           | 0         | 13         |
| 11                     | 3                | 9         | 6,500        | 0                           | 8         | 15         |
| 12                     | 8                | 13        | 7,000        | 0                           | 7         | 14         |
| 13                     | 2                | 6         | 7,500        | 5                           | 10        | 16         |
| 14                     | 5                | 8         | 12,833       | 9                           | 12        | 16         |
| 15                     | 1                | 3         | 194,000      | 6                           | 11        | 17         |
| 16                     | 2                | 5         | 198,500      | 13                          | 14        | 17         |
| 17                     | 1                | 2         | 219,407      | 15                          | 16        | 0          |

Рис. 19.4. Таблиця порядку злиття кластерів

Оскільки значний стрибок в показнику *Coefficient* спостерігається після 14-го кроку, то для даних, які складаються з 18.1-ти випадків, оптимальним є рішення, що складається з  $18.1 - 14 = 4$  кластера.

4. Знову активуйте вікно «**Hierarchical Cluster Analysis**» (Ієрархічний кластерний аналіз) і, клацнувши на кнопці **Statistics**, вкажіть в розділі приналежність до кластера *Single solution i 4 clusters* (рис. 19.5).

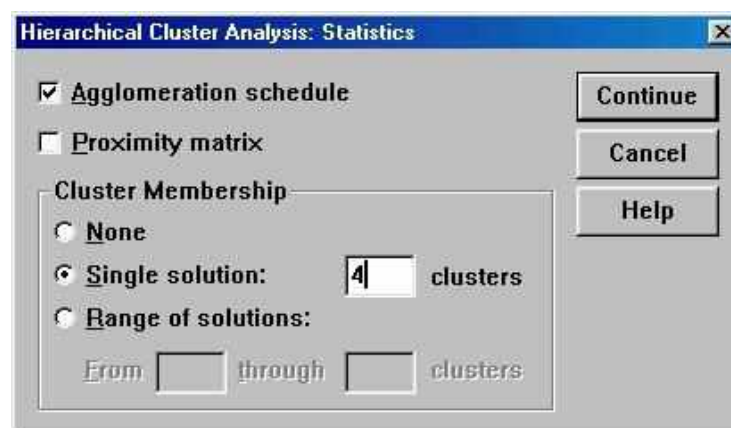


Рис. 19.5. Діалогове вікно «Hierarchical Cluster Analysis : Statistics»

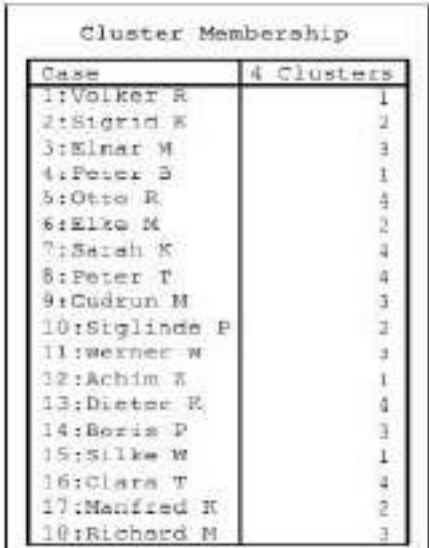
5. Пройдіть вимикач **Save**. Тепер для кожного випадку виводитиметься і

інформація про приналежність до кластера.

6. Отримайте таблицю про приналежність до кластерів (див. рис. 19.6).

Розібратися в значенні кластерів допоможуть кластерні профілі. Вони є середніми значеннями змінних, які включені в аналіз, розподілені по кластерній приналежності.

7. Виберіть в меню **Analyze, Compare Means, Means**. Змінним t1- t10 присвойте статус залежних, а змінній «приналежність до кластера» - незалежною і почніть розрахунок (див. рис. 19.7). Результати розрахунку приведені на рис. 19.8.



| Case           | 4 Clusters |
|----------------|------------|
| 1:Volker R     | 1          |
| 2:Sigrid K     | 2          |
| 3:Elmar W      | 3          |
| 4:Peter B      | 1          |
| 5:Otto R       | 4          |
| 6:Elke M       | 2          |
| 7:Sarah K      | 4          |
| 8:Peter T      | 4          |
| 9: Gudrun M    | 3          |
| 10:Stiglinde P | 2          |
| 11:Werner W    | 3          |
| 12:Achim T     | 1          |
| 13:Dieter K    | 4          |
| 14: Boris P    | 3          |
| 15:Silke W     | 1          |
| 16:Clara T     | 4          |
| 17:Manfred H   | 2          |
| 18:Richard M   | 3          |

Рис. 19.6. Приналежність до кластерів

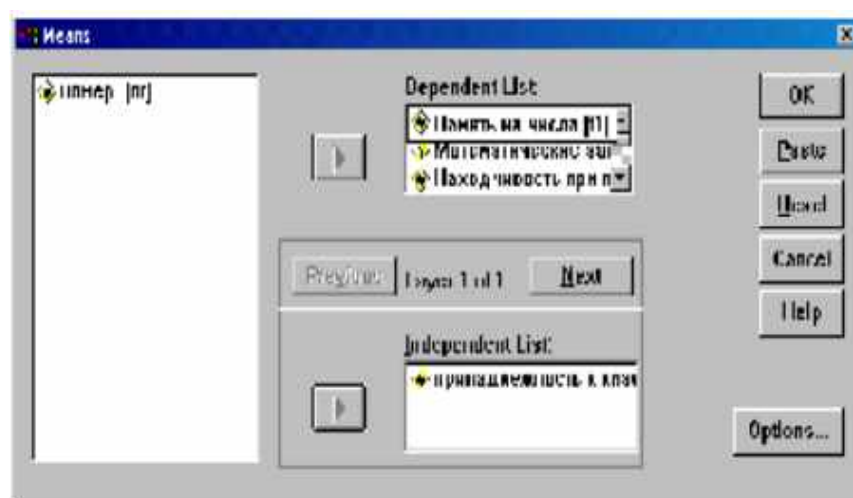


Рис. 19.7. Обчислення кластерних центрів

8. Дайте змістовний опис кластерам (рис. 19.8).

| Report                           |                           |       |       |       |       |
|----------------------------------|---------------------------|-------|-------|-------|-------|
| Mean                             |                           |       |       |       |       |
|                                  | принадлежность к кластеру |       |       |       |       |
|                                  | 1                         | 2     | 3     | 4     | Total |
| Память на числа                  | 10,00                     | 10,00 | 4,20  | 4,80  | 6,94  |
| Математические задачи            | 10,00                     | 10,00 | 4,80  | 4,40  | 7,00  |
| Находчивость при прямом диалоге  | 9,00                      | 4,25  | 10,00 | 4,00  | 6,83  |
| Тест на составление алгоритмов   | 10,00                     | 10,00 | 4,40  | 4,00  | 6,78  |
| Уверенность во время выступления | 10,00                     | 4,75  | 10,00 | 4,20  | 7,23  |
| Командный дух                    | 9,50                      | 4,50  | 4,40  | 10,00 | 7,11  |
| Находчивость                     | 9,25                      | 3,75  | 10,00 | 4,40  | 6,89  |
| Сотрудничество                   | 9,75                      | 4,25  | 4,00  | 10,00 | 7,00  |
| Признание в коллективе           | 10,00                     | 4,25  | 3,80  | 10,00 | 7,00  |
| Сила убеждения                   | 9,50                      | 4,25  | 10,00 | 5,00  | 7,22  |

Рис. 19.8. Середні значення усіх змінних по кластерах

### Завдання для самостійного вирішення:

Створіть файл пиво.sav, який містить деякі дані про 17-ть сортів пива (таблиця 19.2).

Таблиця 19.2

### Інформація про 17-ть сортів пива

| Марка пива  | Країна-виробник | Витрати | Калорії | % алкоголю |
|-------------|-----------------|---------|---------|------------|
| Budweiser   | 1               | 0,43    | 144     | 4,7        |
| Lowenbrau   | 1               | 0,48    | 157     | 4,9        |
| Michelob    | 1               | 0,50    | 162     | 5,0        |
| Kronenbourg | 3               | 0,73    | 170     | 5,2        |
| Heineken    | 4               | 0,77    | 152     | 5,0        |
| Schmidts    | 1               | 0,30    | 147     | 4,7        |
| Pabst       | 1               | 0,38    | 152     | 4,9        |
| Miller      | 1               | 0,43    | 99      | 4,3        |
| Budweiser   | 1               | 0,44    | 113     | 3,7        |
| Coors       | 1               | 0,46    | 102     | 4,1        |
| Dos         | 5               | 0,70    | 145     | 4,5        |

|         |   |      |     |     |
|---------|---|------|-----|-----|
| Becks   | 2 | 0,76 | 150 | 4,7 |
| Rolling | 1 | 0,36 | 144 | 4,7 |
| Pabst   | 1 | 0,38 | 68  | 2,3 |
| Tuborg  | 1 | 0,43 | 155 | 5,0 |
| Olympia | 1 | 0,46 | 72  | 2,9 |
| Schlitz | 1 | 0,47 | 97  | 4,2 |

- змінна «виробник» вказує на країну-виробника пива, де США закодовані за допомогою одиниці;
- витрати наведені в доларах США для місткості, що дорівнює 12-ти унціям для рідини (приблизно одна третина літра);
- калорійність вказана для однакової кількості пива;
- вміст алкоголю наводиться у відсотках.

Розподіліть 17 сортів пива в кластери, виходячи з двох змінних: «калорії» і «витрати». (Вказівка: використайте ієрархічний кластерний аналіз, значення змінних - стандартизуйте - вкладка **Method, Z-оцінки**). Як визначити оптимальне число кластерів? Охарактеризуйте отримані кластери.

### **Вимоги до звіту**

Звіт повинен містити:

- файли з даними і результатами розрахунків;
- відповіді на контрольні питання.

### **Контрольні запитання**

1. Призначення кластерного аналізу. У чому відмінність кластерного аналізу від факторного аналізу?
2. Назвіть приклади використання кластерного аналізу в маркетингу, соціології, психології.
3. Які методи кластерного аналізу ви знаєте? Як він здійснюється в SPSS?
4. Як визначити оптимальне число кластерів при ієрархічному аналізі?

5. Що є найбільш поширеною мірою схожості в кластерному аналізі?
6. Що показує дендрограма і як її побудувати?

## ЛАБОРАТОРНЕ ЗАНЯТТЯ № 20

**Тема:** *Факторний аналіз у SPSS для Windows*

**Мета:** навчитися проводити факторний аналіз у SPSS для Windows та інтерпретувати його результати.

### Теоретичні відомості

Факторний аналіз, як і кластерний, є засобом конденсації інформації. Але, на відміну кластерного аналізу, завдання якого – об'єднати безліч об'єктів у невелике число однорідних груп, завдання факторного аналізу – знайти максимально взаємопов'язані групи ознак. Ці групи є нові комплексні змінні, що називаються **факторами**. Число факторів істотно менше, ніж кількість ознак, що вимірюються.

Основне призначення **факторного аналізу** – упорядкування хаотичності досліджуваного явища і такої простої структури, яка досить точно відображає і відтворює реально існуючі залежності, тобто, відображає сутність цього явища.

Вихідним для факторного аналізу є положення про взаємозв'язок між ознаками (змінними), які підлягають аналізу. В якості кількісної міри зв'язку між парами змінних використовується коефіцієнт кореляції.

Результатом факторного аналізу є так звана **факторна матриця** (інші назви – **матриця факторних навантажень** чи **матриця факторного відображення**).

Загальний вигляд факторної матриці (без числових значень) наведено у табл. 20.1.

Кожен фактор характеризується стовпцем, а кожна змінна рядком матриці факторного відображення.

## Загальний вигляд матриці факторних навантажень (факторне відображення)\*

| Ознаки | Фактори |       |       |       |       |       |
|--------|---------|-------|-------|-------|-------|-------|
|        | $F_1$   | $F_2$ | $F_3$ | $F_4$ | $F_5$ | $F_6$ |
| $P_1$  |         | *     |       |       |       |       |
| $P_2$  |         | *     | *     |       |       |       |
| $P_3$  | *       |       | *     |       |       | *     |
| $P_5$  | *       |       |       |       |       |       |
| $P_6$  | *       |       |       |       | *     |       |
| ...    | *       |       |       |       |       |       |
| $P_m$  | *       |       |       |       |       |       |

— знаком «\*» позначені комірки, в яких вміщені навантаження, які суттєво відрізняються від нуля.

На рис. 9.1 представлено класифікацію факторів, що виходять в результаті факторного аналізу.

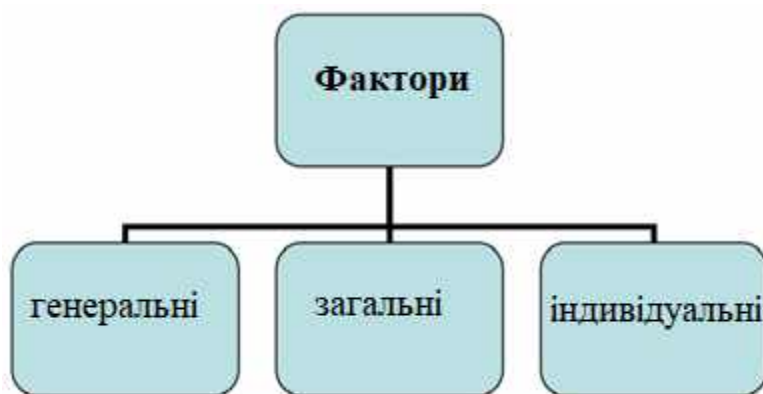


Рис. 9.1. Види факторів

Фактор називається **генеральним**, якщо він має від усіх змінних навантаження, які значно відмінні від нуля.

Фактор називається **загальним**, якщо хоча б два його навантаження відрізняються від нуля.

**Індивідуальним** називається фактор, який має ненульове навантаження лише від однієї змінної.

Змінні також можна розрізняти за кількістю високих навантажень. Кількість високих навантажень змінної на фактори називається її **складністю**. Так, наприклад, якщо змінна значно навантажує три фактори, її складність – 3.

Проблема факторів немає однозначного рішення, але з нескінченної безлічі можливих рішень потрібно вибрати лише одне. Такий вибір здійснюється шляхом обертання конфігурації векторів (змінних) навколо точки їх початку. При геометричній інтерпретації факторів є координатні осі, куди натягнуто простір загальних факторів – простір найменшої розмірності, у якому можна уявити змінні у вигляді векторів.

Обертання факторів, як і всі попередні процедури факторного аналізу, здійснюється в комп'ютері автоматично. Але, отримавши факторну матрицю після обертання, дослідник має сам вирішити – які фактори слід залишити для подальшого аналізу та інтерпретації. Як правило, для подальшого аналізу та інтерпретації залишають ті фактори, на які припадає понад 5% повної дисперсії та абсолютні значення дисперсії яких не менше одиниці.

Перш ніж робити остаточні висновки за результатами факторного аналізу, необхідно звернути увагу на такі моменти:

- Чи з'являється той чи інший фактор під час повторних вимірів однієї чи кількох аналізованих змінних?
- Чи не відображає фактор відомий чи тривіальний зв'язок між змінними? (Якщо відображено тривіальний зв'язок, фактор нецікавий для дослідника);
- Чи не є цей фактор результатом неоднорідності даних?
- Чи відображають виявлені фактори сутність явища, що вивчається?
- Чи з'являються виявлені чинники у разі повторної вибірки з тієї самої генеральної сукупності?

### **Постановка завдання**

У таблиці 20.2 містяться відповіді тридцяти респондентів на шість запитань щодо вибору зубної пасти. Респондентам пропонувалося оцінити

ступінь своєї згоди з наведеними нижче твердженнями за 7-бальною шкалою (1 – «абсолютно не згоден», 7 – «повністю згоден»):

1. Важливо мати зубну пасту, яка запобігає розвитку карієсу.
2. Мені подобається зубна паста, яка надає зубам білизни.
3. Зубна паста має зміцнювати ясна.
4. Я віддаю перевагу зубній пасті, яка освіжає дихання.
5. Запобігання псуванню зубів не є важливою перевагою даної зубної пасти.
6. Найбільш важливою причиною покупки певного виду зубної пасти є її здатність надати зубу привабливого вигляду.

*Таблиця 20.2*

**Відповіді 30 респондентів на питання про зубну пасту**

| №<br>респондента | Номер запитання |    |    |    |    |    |
|------------------|-----------------|----|----|----|----|----|
|                  | v1              | v2 | v3 | v4 | v5 | v6 |
| 1                | 7               | 3  | 6  | 4  | 2  | 4  |
| 2                | 1               | 3  | 2  | 4  | 5  | 4  |
| 3                | 6               | 2  | 7  | 4  | 1  | 3  |
| 4                | 4               | 5  | 4  | 6  | 2  | 5  |
| 5                | 1               | 2  | 2  | 3  | 6  | 2  |
| 6                | 6               | 3  | 6  | 4  | 2  | 4  |
| 7                | 5               | 3  | 6  | 3  | 4  | 3  |
| 8                | 6               | 4  | 7  | 4  | 1  | 4  |
| 9                | 3               | 4  | 2  | 3  | 6  | 3  |
| 10               | 2               | 6  | 2  | 6  | 7  | 6  |
| 11               | 6               | 4  | 7  | 3  | 2  | 3  |
| 12               | 2               | 3  | 1  | 4  | 5  | 4  |
| 13               | 7               | 2  | 6  | 4  | 1  | 3  |
| 14               | 4               | 6  | 4  | 5  | 3  | 6  |
| 15               | 1               | 3  | 2  | 2  | 6  | 4  |
| 16               | 6               | 4  | 6  | 3  | 3  | 4  |

|    |   |   |   |   |   |   |
|----|---|---|---|---|---|---|
| 17 | 5 | 3 | 6 | 3 | 3 | 4 |
| 18 | 7 | 3 | 7 | 4 | 1 | 4 |
| 19 | 2 | 4 | 3 | 3 | 6 | 3 |
| 20 | 3 | 5 | 3 | 6 | 4 | 6 |
| 21 | 1 | 3 | 2 | 3 | 5 | 3 |
| 22 | 5 | 4 | 5 | 4 | 2 | 4 |
| 23 | 2 | 2 | 1 | 5 | 4 | 4 |
| 24 | 4 | 6 | 4 | 6 | 4 | 7 |
| 25 | 6 | 5 | 4 | 2 | 1 | 4 |
| 26 | 3 | 5 | 4 | 6 | 4 | 7 |
| 27 | 4 | 4 | 7 | 2 | 2 | 5 |
| 28 | 3 | 6 | 2 | 6 | 4 | 3 |
| 29 | 4 | 7 | 3 | 7 | 2 | 7 |
| 30 | 2 | 3 | 2 | 4 | 7 | 2 |

### Хід роботи

1. Визначте змінні в SPSS для Windows та введіть дані відповідно до таблиці 20.2.

2. Виберіть меню **Analyze, Data reduction, Factor**. Перенесіть змінні v1-v6 у поле вибраних змінних (див. рис. 20.2).

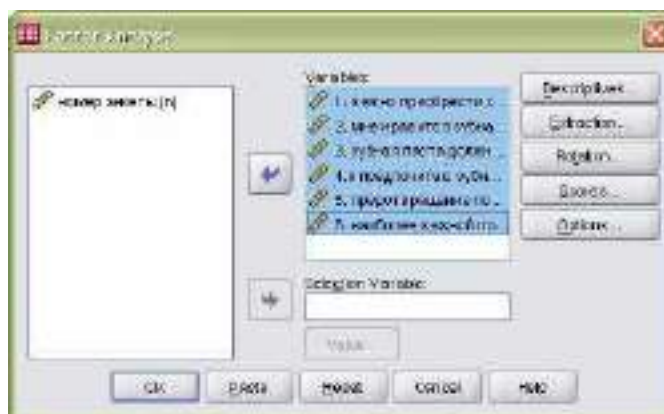


Рис. 20.2. Вікно «Factor Analysis». Результати перенесення шести змінних у полі вибраних змінних (Variables)

3. Натисніть кнопку **Descriptives**. Відкриється діалогове вікно **Factor Analysis: Descriptives** (Факторний аналіз: Описові статистики) (див. рис. 20.3):

- Залишіть встановлену за замовчуванням опцію виводу **Initial solution** (Первинне рішення).

- Виберіть опції **Coefficients** (для виведення кореляційної матриці) та **KMO and Bartlett's test of sphericity** (для перевірки придатності факторного аналізу для зібраних даних).

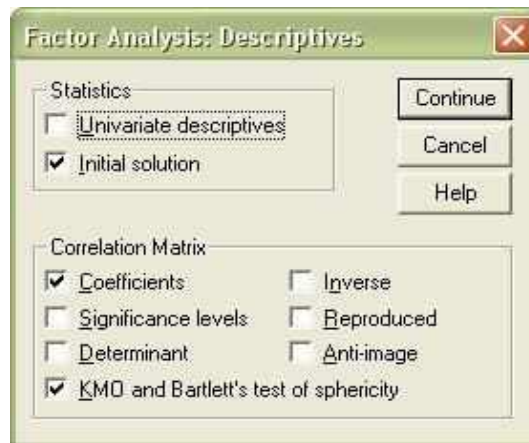


Рис. 20.3. Вікно Factor Analysis: Descriptives

4. Натисніть кнопку **Extraction**. З'явиться діалогове вікно **Factor Analysis: Extraction** (див. рис. 20.4)

– Залиште встановлення **Principal components** (Аналіз основних компонентів).

– Кількість факторів свідомо обмежимо двома. Клацніть на опції **Number of factors** (кількість факторів) і введіть число 2.

– Клацніть на відповідній опції деактивуйте виведення невідновлених значень факторів **Unrotated factor solution**.

– Активуйте опцію **Scree plot** (точкова діаграма). Точкова діаграма графічно представляє власні значення чинників, упорядковані за величиною.

Діалогове вікно **Factor Analysis: Extraction** (Факторний аналіз: Вилучення) має тепер виглядати так, як представлено на рис. 20.4.



Рис. 20.4 Вікно «Factor Analysis: Extraction»

5. Клацніть на кнопці **Rotation** (Обертання) та виберіть метод варімакс (**Varimax**) у діалоговому вікні «**Factor Analysis: Rotation**» (див. рис. 20.5).



Рис. 20.5. Вікно «Factor Analysis: Rotation»

6. Натисніть кнопку **Scores** та активуйте **Save as variables** (Зберегти як змінні), щоб розраховані значення факторів зберегти як додаткові змінні (див. рис. 20.6).



Рис. 20.6. Вікно «Factor Analysis: Factor Scores»

7. Натисніть кнопку **Options...**, щоб організувати виведення коефіцієнтів, відсортованих за розміром (див. рис. 20.7).

- Активувати опцію **Sorted by size** (відсортована за розміром).
- Заборонити виведення малих факторних навантажень. Для цього активувати опцію **Suppress absolute values less than:** (не виводити абсолютні значення менші, ніж:), і ввести граничне значення 0,40.



Рис. 20.7. Вікно «Factor Analysis: Options»

8. Розгляньте результати у вікні виводу. Вони містять:

- кореляційну матрицю (рис. 20.8);
- значення критерію Бартлетта (див. рис. 20.9);
- % дисперсії, що пояснюється моделлю (рис. 20.11 та 20.12);
- діаграму «кам'яний осип» (рис. 20.10);
- повернену матрицю факторних навантажень (рис. 20.13).

Як видно із рис. 20.8 відносно високе значення коефіцієнтів кореляції спостерігається між v1 (запобігання карієсу), v3 (зміцнення ясен) і v5 (запобігання псуванню зубів). Можна очікувати, що ці змінні корелюють з тим самим набором факторів.

Аналогічно відносно високі кореляції спостерігаються між v2 (відбілювання зубів), v4 (свіже дихання) і v6 привабливість зовнішнього вигляду зубів). Також можна очікувати, що ці змінні корелюють з одними і тими самими факторами.

## Кореляційна матриця

| Correlation Matrix                                                                  |                                                                           |                                                              |                                         |                                                          |                                                                                     |                                                                                   |
|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------|--------------------------------------------------------------|-----------------------------------------|----------------------------------------------------------|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
|                                                                                     | 1. важно приобрести зубную пасту, которая предотвращает развитие кариеса; | 2. мне нравится зубная паста, которая придает зубам белизну; | 3. зубная паста должна укреплять десна; | 4. я предпочитаю зубную пасту, которая освежает дыхание; | 5. предотвращение порчи зубов не является важным преимуществом данной зубной пасты; | 6. наиболее важной причиной покупки данной зубной пасты является способность зуб- |
| Correlation                                                                         | 1.000                                                                     | -.041                                                        | .873                                    | -.086                                                    | -.858                                                                               | .004                                                                              |
| 1. важно приобрести зубную пасту, которая предотвращает развитие кариеса;           |                                                                           |                                                              |                                         |                                                          |                                                                                     |                                                                                   |
| 2. мне нравится зубная паста, которая придает зубам белизну;                        | -.041                                                                     | 1.000                                                        | -.143                                   | .590                                                     | -.007                                                                               | .713                                                                              |
| 3. зубная паста должна укреплять десна;                                             | .873                                                                      | -.143                                                        | 1.000                                   | -.248                                                    | -.778                                                                               | -.018                                                                             |
| 4. я предпочитаю зубную пасту, которая освежает дыхание;                            | -.086                                                                     | .590                                                         | -.248                                   | 1.000                                                    | -.007                                                                               | .640                                                                              |
| 5. предотвращение порчи зубов не является важным преимуществом данной зубной пасты; | -.858                                                                     | -.007                                                        | -.778                                   | -.007                                                    | 1.000                                                                               | -.136                                                                             |
| 6. наиболее важной причиной покупки данной зубной пасты является способность зуб-   | .004                                                                      | .713                                                         | -.018                                   | .640                                                     | -.136                                                                               | 1.000                                                                             |

Рис. 20.8. Фрагмент вікна виведення. Кореляційна матриця

## Критерій сферичності Бартлетта

| KMO and Bartlett's Test                          |                    |         |
|--------------------------------------------------|--------------------|---------|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. |                    | .655    |
| Bartlett's Test of Sphericity                    | Approx. Chi-Square | 115.879 |
|                                                  | df                 | 15      |
|                                                  | Sig.               | .000    |

Рис. 20.9. Фрагмент вікна виведення. Критерій сферичності Бартлетта

Як видно із рис. 20.9, наближене значення статистики  $\chi^2$  (Approx. Chi-Square) дорівнює 115,879 з 15-ма ступенями свободи, вона є значущою при рівні  $<0,05$ .

Значення статистики **КМО** (Kaiser-Meyer-Olkin Measure of Sampling Adequacy), що дорівнює 0,665, також велике ( $>0,5$ ). Отже, факторний аналіз можна прийняти як метод аналізу кореляційної матриці, наведеної на рис. 20.8.

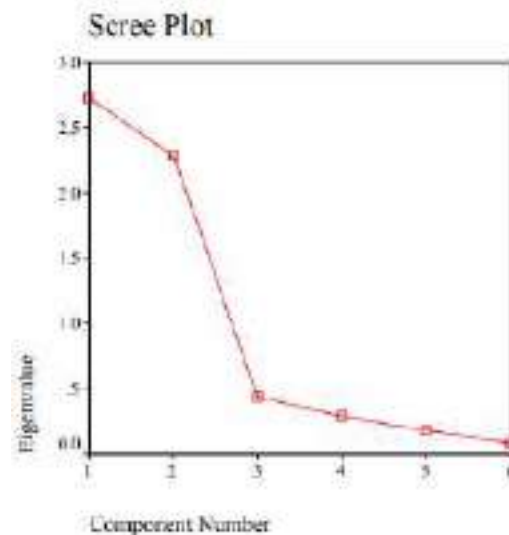


Рис. 20.10. Фрагмент вікна виведення. Діаграма «Кам’яний осип»

«Кам’янистий осип» (рис. 20.10) є графік залежності власних значень факторів від їх номерів у порядку виділення.

Зазвичай графік має чіткий розрив між крутою частиною кривою, де факторам властиві великі власні значення, та плавною хвостовою частиною (осипом). Досвід показує, що точка, з якої починається осип, вказує на дійсну кількість факторів. У нашому випадку осип починається правіше третього фактора.

На рис. 20.11 представлений фрагмент вікна виводу, у якому показані так звані **спільності** (communalities). Спільність – частка дисперсії окремої змінної, що вона ділить коїться з іншими змінними.

На рис. 20.12 представлений фрагмент вікна виведення з сумарним відсотком факторної моделі дисперсії, що пояснюється. Як видно з представленої на рис. 20.12 таблиці, є два власні значення, великі одиниці. Це свідчить про необхідність вилучення двох факторів.

Наступний показник показує відсоток повної дисперсії, який приписується кожному фактору. У нашому випадку два фактори пояснюють 83% дисперсії (що є більш ніж задовільним), а три – більше 90% (див. колонку Cumulative % на рис. 20.12).

| Communalities                                                             |         |            |
|---------------------------------------------------------------------------|---------|------------|
|                                                                           | Initial | Extraction |
| 1. важко придбати зубну пасту, яка передотращає розвиток карієсу;         | 1.000   | .926       |
| 2. мені подобається зубна паста, яка надає зубам білизну;                 | 1.000   | .769       |
| 3. зубна паста повинна укріплювати десни;                                 | 1.000   | .894       |
| 4. я віддаю перевагу зубній пасті, яка освіжає дихання;                   | 1.000   | .727       |
| 5. запобігання порчи зубів не є важливим перевагою даної зубної пасту;    | 1.000   | .877       |
| 6. найбільш важливою причиною покупки даної зубної пасту є здатність зуб- | 1.000   | .817       |

Extraction Method: Principal Component Analysis.

Рис. 20.11. Фрагмент вікна виведення. Спільності

| Total Variance Explained |                     |               |              |                                     |               |              |
|--------------------------|---------------------|---------------|--------------|-------------------------------------|---------------|--------------|
| Component                | Initial Eigenvalues |               |              | Extraction Sums of Squared Loadings |               |              |
|                          | Total               | % of Variance | Cumulative % | Total                               | % of Variance | Cumulative % |
| 1                        | 2.725               | 45.424        | 45.424       | 2.725                               | 45.424        | 45.424       |
| 2                        | 2.284               | 38.072        | 83.496       | 2.284                               | 38.072        | 83.496       |
| 3                        | .434                | 7.241         | 90.737       |                                     |               |              |
| 4                        | .289                | 4.821         | 95.557       |                                     |               |              |
| 5                        | .183                | 3.057         | 98.614       |                                     |               |              |
| 6                        | .083                | 1.388         | 100.000      |                                     |               |              |

Extraction Method: Principal Component Analysis.

Рис. 20.12. Фрагмент вікна виведення. Сумарний відсоток дисперсії, що пояснюється

На рис. 20.13 представлений фрагмент вікна виведення з повернутою матрицею факторних навантажень. Як видно з представленої на рис. 20.13 таблиці, перша, третя та п'ята змінні мають високі навантаження за першим фактором. Друга, четверта та шоста – по другому. Факторами інтерпретують з розуміння тих змінних, які його навантажують. Тому перший фактор можна інтерпретувати як здоров'я зубів, а другий - зовнішній вигляд зубів.

Таким чином, можна зробити висновок, що при виборі зубної пасти покупці керуються двома основними факторами: як паста впливає на здоров'я зубів і як вона впливає на їхній зовнішній вигляд.

## Повернена матриця факторних навантажень

| Rotated Component Matrix <sup>a</sup>                                               |           |      |
|-------------------------------------------------------------------------------------|-----------|------|
|                                                                                     | Component |      |
|                                                                                     | 1         | 2    |
| 1. важно приобрести зубную пасту, которая предотвращает развитие кариеса;           | ,962      |      |
| 3. зубная паста должна укреплять десна;                                             | ,935      |      |
| 5. предотвращение порчи зубов не является важным преимуществом данной зубной пасты; | ,932      |      |
| 6. наиболее важной причиной покупки данной зубной пасты является способность зуб-   |           | ,900 |
| 2. мне нравится зубная паста, которая придает зубам белизну;                        |           | ,876 |
| 4.я предпочитаю зубную пасту, которая освежает дыхание;                             |           | ,846 |

Extraction Method: Principal Component Analysis.  
 Rotation Method: Varimax with Kaiser Normalization.  
 a. Rotation converged in 3 iterations.

Рис. 20.13. Фрагмент вікна виведення. Повернена матриця факторних навантажень

### Завдання для самостійного виконання

У дослідженні взаємозв'язку між поведінкою в сім'ї та поведінкою при покупці отримані дані за семибальною шкалою (1 – «не згоден», 7 – «згоден») за наступними відповідями про стиль життя:

- Я краще спокійно провів би вечір удома, ніж пішов на вечірку.
- Я завжди перевіряю ціни, навіть на товар із невеликою ціною.
- Магазины мене цікавлять більше, ніж фільми.
- Я не купую товари, що рекламують на білбордах.
- Я – домосід.
- Я заощаджую, використовуючи купони для купівлі товарів.
- Компанії даремно витрачають великі гроші на рекламу.

Дані, отримані на вибірці з 25 респондентів, наведені в таблиці 20.3.

**Відповіді респондентів на питання про поведінку в сім'ї та поведінку при покупці**

| <b>№</b> | <b>v1</b> | <b>v2</b> | <b>v3</b> | <b>V4</b> | <b>v5</b> | <b>v6</b> | <b>v7</b> |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1        | 6         | 2         | 1         | 6         | 5         | 3         | 5         |
| 2        | 5         | 7         | 5         | 6         | 6         | 6         | 4         |
| 3        | 5         | 3         | 4         | 5         | 6         | 6         | 7         |
| 4        | 3         | 2         | 2         | 5         | 1         | 3         | 2         |
| 5        | 4         | 2         | 3         | 2         | 2         | 1         | 3         |
| 6        | 2         | 6         | 2         | 4         | 3         | 7         | 5         |
| 7        | 1         | 3         | 3         | 6         | 2         | 5         | 7         |
| 8        | 3         | 5         | 1         | 4         | 2         | 5         | 6         |
| 9        | 7         | 3         | 5         | 3         | 5         | 2         | 4         |
| 10       | 6         | 3         | 3         | 4         | 4         | 6         | 6         |
| 11       | 6         | 6         | 2         | 6         | 4         | 4         | 7         |
| 12       | 3         | 2         | 2         | 7         | 6         | 1         | 6         |
| 13       | 5         | 7         | 6         | 2         | 2         | 6         | 1         |
| 14       | 6         | 3         | 5         | 5         | 7         | 2         | 3         |
| 15       | 3         | 2         | 4         | 3         | 2         | 6         | 5         |
| 16       | 2         | 7         | 5         | 1         | 4         | 5         | 2         |
| 17       | 3         | 2         | 2         | 7         | 2         | 4         | 6         |
| 18       | 6         | 4         | 5         | 4         | 7         | 3         | 3         |
| 19       | 7         | 2         | 6         | 2         | 5         | 2         | 1         |
| 20       | 5         | 6         | 6         | 3         | 4         | 5         | 3         |
| 21       | 2         | 3         | 3         | 2         | 1         | 2         | 6         |
| 22       | 3         | 4         | 2         | 1         | 4         | 3         | 6         |
| 23       | 2         | 6         | 3         | 2         | 1         | 5         | 3         |
| 24       | 6         | 5         | 7         | 4         | 5         | 7         | 2         |
| 25       | 7         | 6         | 5         | 4         | 6         | 5         | 3         |

Проведіть факторний аналіз даних та проінтерпретуйте його результати.

**Вимоги до звіту**

Звіт по роботі повинен містити:

- файли з даними;
- файли з результатами;
- відповіді на контрольні питання.

### **Контрольні запитання**

1. Призначення факторного аналізу.
2. Що таке матриця факторного відображення?
3. Як факторний аналіз здійснюється у SPSS?
4. Які кнопки розташовані в головному вікні факторного аналізу та для чого вони слугують?
5. Якими способами визначається кількість факторів у моделі факторного аналізу?
6. Що показує графік «кам'янистий осип»?
7. Скільки факторів вилучено у прикладі із зубною пастою? Чому? Дайте їм змістовну інтерпретацію.

### **ЛАБОРАТОРНЕ ЗАНЯТТЯ № 21-22**

**Тема:** *Використання MS Excel для вирішення завдань методом лінійного програмування*

**Мета:** навчитися використати електронні таблиці MS Excel для вирішення завдань методом лінійного програмування.

#### **Теоретичні відомості**

**Лінійне програмування** - метод, який широко застосовують для ухвалення рішень відносно найбільш ефективного використання ресурсів (грошей, часу, техніки, робочої сили, сировини, виробничих і складських приміщень і тому подібне).

**Характерні ознаки завдань оптимізації, які доцільно вирішувати методом лінійного програмування:**

- 1) наявність цільової функції - максимізувати, мінімізувати або

встановити рівним певному числу результат (прибуток, ціну, ті або інші витрати і тому подібне);

2) наявність певних обмежень в можливості отримати бажаний результат, сформульований як цільова функція(наприклад, фірма не в змозі відремонтувати більше тридцяти автомобілів за тиждень, в готелі можна розмістити не більше двохсот гостей і тому подібне);

3) наявність можливих альтернативних дій для вибору(наприклад, якщо фірма робить чотири види продукції, менеджерів використовують лінійне програмування для того, щоб визначити, як найраціональніше розподілити обмежені виробничі ресурси між цими видами продукції, щоб отримати максимальний прибуток);

4) цільова функція і обмеження мають бути описані лінійними рівняннями або нерівностями.

#### **Алгоритм рішення завдань лінійного програмування**

- 1) глибоко усвідомити суть виниклої проблеми;
- 2) визначити цільову функцію і обмеження;
- 3) визначити змінні, аналіз яких необхідно здійснити для ухвалення рішення;
- 4) використовуючи ці змінні, описати математично цільову функцію і обмеження;
- 5) застосувати MS Excel для вирішення завдання.

Для вирішення завдань методом лінійного програмування в MS Excel існує інструмент **«Пошук рішення»** в меню **«Сервіс»**, за допомогою якого можна вирішувати завдання з використанням до 200 змінних, кожна з яких має по два обмеження і використати додаткові до 100 обмежень.

Процедура пошуку рішення дозволяє знайти оптимальне значення формули, що міститься в комірці, яка називається цільовою. Ця процедура працює з групою комірок, прямо або побічно пов'язаних з формулою в цільовій комірці. Щоб отримати по формулі, що міститься в цільовій комірці, заданий

результат, процедура змінює значення у комірках, які на це впливають.

### Постановка завдання

Невелика кондитерська фірма прагне отримати максимальний прибуток, випускаючи два види продукції - торти і тістечка. Один торт дає фірмі 6 гривень прибутку, одно тістечко - 1 гривню 50 копійок. На виготовлення одного торта необхідно 4 години, причому потужності фірми такі, що паралельно можуть випікатися 8 тортів. За цей же час можна приготувати 50 тістечок. Таким чином, в середньому на виготовлення одного торта витрачається 0,5 людино-годин ( $4/8$ ), на виготовлення одного тістечка - 0,08 людино-годин ( $4/50$ ). Кадровий ресурс фірми - 48 людино-годин в добу. Попит на продукцію фірми - не більше 50-ти тортів і не більше за 420-ти тістечка за добу.

**Цільова функція** для такого прикладу має наступний вигляд:

$$6 \cdot X_1 + 1,5 \cdot X_2 = \max \quad (21.1)$$

де  $X_1$  - кількість тортів,  $X_2$  - кількість тістечок

**Обмеження :**

$$0,5 \cdot X_1 + 0,08 \cdot X_2 \leq 48 \quad (21.2)$$

$$X_1 \leq 50 \quad (21.3)$$

$$X_2 \leq 420 \quad (21.4)$$

### Хід роботи

1. Запустити MS Excel.

2. Ввести дані відповідно до рис. 21.1. Зверніть увагу, що в комірки B5 і C5 вводяться коефіцієнти рівняння цільової функції; у комірки B8 і C8 - коефіцієнти нерівності, що виражає перше обмеження (нерівність 21.2); у комірки B6 і C6 - коефіцієнти нерівності, що виражає друге обмеження (нерівність 21.3); у комірки B7 і C7 - коефіцієнти нерівності, що виражає третє обмеження (нерівність 21.4).

|    | A                                                        | B           | C             | D |
|----|----------------------------------------------------------|-------------|---------------|---|
| 1  | Расчет путей максимальной прибыли для кондитерской фирмы |             |               |   |
| 2  |                                                          |             |               |   |
| 3  | Факторы, которые учитываются                             | Вид изделия |               |   |
| 4  |                                                          | торт (X1)   | Пирожное (X2) |   |
| 5  |                                                          | прибыль     | 1,5           |   |
| 6  |                                                          | спрос       | 0             |   |
| 7  |                                                          | 0           | 1             |   |
| 8  | ресурсы (кадры и время)                                  | 0,5         | 0,08          |   |
| 9  |                                                          |             |               |   |
| 10 | Результат                                                |             |               |   |

Рис. 21.1. Введення в MS Excel даних завдання

3. Комірки B10 і C10 - «впливаючі комірки», в них буде отриманий шуканий результат - X1 - кількість тортів (у комірці B10) і X2 - кількість тістечок (комірка C10), випуск яких принесе найбільший прибуток в заданих умовах. Перед початком рішення задачі в кожній з цих комірок необхідно ввести число «1» (див. рис. 21.2).

4. Комірку D5 зробіть цільовою. У неї необхідно ввести цільову функцію (21.1). Процедура пошуку рішення полягає в тому, шукається максимальне, мінімальне або визначене значення формули в цільовій комірці. Щоб по формулі, що міститься в цільовій комірці, отримати заданий результат, процедура змінює значення у впливаючих комірках (див. рис. 21.2).

|    | A                                                        | B           | C             | D                      | E  | F   |
|----|----------------------------------------------------------|-------------|---------------|------------------------|----|-----|
| 1  | Расчет путей максимальной прибыли для кондитерской фирмы |             |               |                        |    |     |
| 2  |                                                          |             |               |                        |    |     |
| 3  | Факторы, которые учитываются                             | Вид изделия |               |                        |    |     |
| 4  |                                                          | торт (X1)   | Пирожное (X2) |                        |    |     |
| 5  |                                                          | прибыль     | 1,5           | =B5*\$B\$10+C5*\$C\$10 |    |     |
| 6  |                                                          | спрос       | 0             | =B6*\$B\$10+C6*\$C\$10 | <= | 60  |
| 7  |                                                          | 0           | 1             | =B7*\$B\$10+C7*\$C\$10 | <= | 420 |
| 8  | ресурсы (кадры и время)                                  | 0,5         | 0,08          | =B8*\$B\$10+C8*\$C\$10 | <= | 48  |
| 9  |                                                          |             |               |                        |    |     |
| 10 | Результат                                                | 1           | 1             |                        |    |     |

Рис. 21.2. Введення в MS Excel формул для вирішення завдання

5. У комірки E6:E8 поставити знаки, що відповідають обмеженням в

умовах завдання (у нашому випадку усі три умови звучали, як «не більше», але взагалі можна використати «=», «≤», «≥»).

6. У комірки F6:F8 ввести ті значення, які представлені в правих частинах нерівностей 21.2-21.4 (див. рис. 21.2). Тепер електронна таблиця готова до застосування процедури «Пошук рішення».

7. Зайти в меню «Сервіс» і вибрати опцію «Пошук рішення» (якщо вона не представлена в меню «Сервіс», увійти до «Надбудов» і встановити мітку біля пункту «Пошук рішення»). Відкриється діалогове вікно «Пошук рішення», представлене на рис. 21.3.

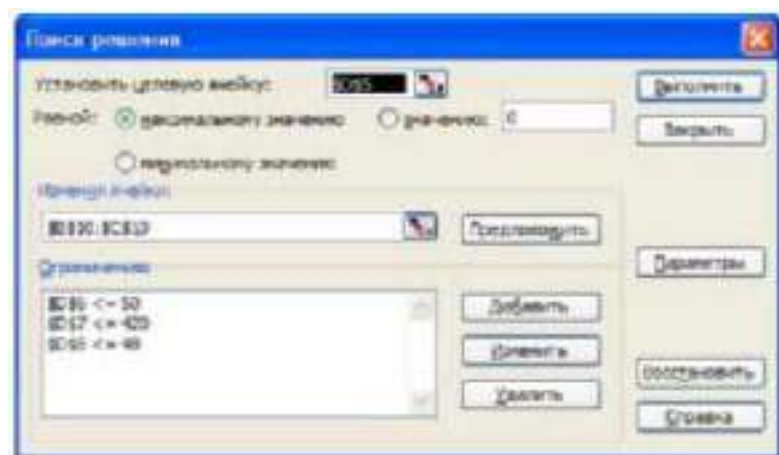


Рис. 21.3. Діалогове вікно «Пошук рішення»

8. Встановити параметри:

- а) адреса цільової комірки (для нашого завдання - це комірка D5);
- б) те, чому має бути рівне отримане в цільовій комірці значення - максимуму, мінімуму або певному значенню (у нашому - випадку - ми хочемо зробити прибуток максимальним, тобто максимізувати цільову функцію);
- в) адреси впливаючих комірок - опція «изменяя ячейки»;
- г) обмеження (після натиснення кнопки «Добавить» з'являється діалогове вікно «Добавление ограничения», представлене на рис. 21.4; після введення кожного обмеження натискають кнопку «додати»; після того, як введені усі обмеження, натискають кнопку «ОК»).

9. Після натиснення кнопки «ОК» знову відкриється діалогове вікно «Пошук рішення», представлене на рис. 21.3. Далі необхідно запустити

процедуру «Пошук рішення», натиснувши кнопку «Виконати».



Рис. 21.4. Діалогове вікно «Додавання обмеження»

**10.** Результати розрахунків представлені в комірках B10 і C10 (відповідно, кількість тортів і тістечок, яку доцільно випускати в добу) і D5 (максимальний прибуток в гривнях) (рис. 21.5).

|    | A                                                        | B           | C             | D      | E   | F |
|----|----------------------------------------------------------|-------------|---------------|--------|-----|---|
| 1  | Расчет путей максимальной прибыли для кондитерской фирмы |             |               |        |     |   |
| 2  |                                                          |             |               |        |     |   |
| 3  | Факторы, которые учитываются                             | Вид изделия |               |        |     |   |
| 4  |                                                          | торт (X1)   | Пирожное (X2) |        |     |   |
| 5  | прибыль                                                  | 6           | 1,5           | 803    |     |   |
| 6  | спрос                                                    | 1           | 0             | 29 <=  | 50  |   |
| 7  |                                                          | 0           | 1             | 420 <= | 420 |   |
| 8  | ресурсы (кадры и время)                                  | 0,5         | 0,08          | 48 <=  | 48  |   |
| 9  |                                                          |             |               |        |     |   |
| 10 | Результат                                                | 29          | 420           |        |     |   |

Рис. 21.5. Результати розрахунків

### Завдання для самостійного виконання

**1.** Керівник приватного ВНЗ формує учбовий план двоступінчатою магістрів. Кожен, що навчається на першому ступені (бакалаврат) повинен прослухати не менше 35-ти дисциплін і на другому ступені навчання (магістр) не менше 15-ти дисциплін. Крім того, в контракті кожного гарантовано, що в сумі він має право прослухати не більше 60-ти дисциплін. Кожен курс першого ступеня навчання обходиться ВНЗ у 50000 гривень (зарплата викладачів і обслуговуючого персоналу, методична література, устаткування, оренда приміщень і тому подібне), кожен курс другого ступеня навчання - в 75000 гривень. Скільки дисциплін треба включити в учбовий план на кожному зі східців навчання, щоб мінімізувати витрати ВНЗ? Вирішити завдання,

використовуючи інструмент «Пошук рішення» MS Excel.

*Підказка:*

*нехай  $X_1$  - шукане число дисциплін на першому ступені навчання;*

*$X_2$  - шукане число дисциплін на другому ступені навчання.*

*Цільова функція:  $50000 * X_1 + 75000 * X_2 = \min \$$*

*обмеження :*

*$X_1 \geq 35; \quad X_2 \geq 15; \quad X_1 + X_2 \leq 60 .$*

2. Вирішити завдання 1, змінивши умови таким чином: вартість однієї дисципліни на першому ступені навчання збільшилася на 1000 гривень, вартість однієї дисципліни на другому ступені навчання не змінилася, мінімальна кількість дисциплін, які вивчаються на першому ступені навчання, стала рівною сорока, а на другій - не змінилося.

### **Вимоги до звіту**

Звіт повинен містити:

- відповіді на контрольні питання;
- файл з результатами розрахунків.

### **Контрольні запитання**

1. Які характерні ознаки завдань оптимізації, які доцільно вирішувати методом лінійного програмування?
2. Що таке цільова функція?
3. Що таке обмеження?
4. За допомогою якого інструменту MS Excel вирішуються завдання оптимізації?
5. Опишіть призначення полів діалогового вікна «Пошук рішення».
6. Опишіть призначення полів діалогового вікна «додавання обмежень».
7. Дайте змістовну інтерпретацію результатів рішення завдань 1 і 2.

## **ЗАВДАННЯ, ЩО ВІНОСЯТЬСЯ НА САМОСТІЙНЕ ОПРАЦЮВАННЯ**

1. Вимоги до забезпечення якості та надійності вибірки у психологічному експерименті
2. Типи вимірювальних шкал.
3. Табличні та графічні методи представлення експериментальних даних.
4. Квантілі розподілу: квартилі, процентилі.
5. Міри мінливості значень ознак: розмах, дисперсія, стандартне відхилення.
6. Міри характеру розподілу: асиметрія та ексцес.
7. Завдання статистичної перевірки гіпотез в психологічних дослідженнях.
8. Статистичні критерії та їх види.
9. Вибір статистичних критеріїв для вирішення різних типів дослідницьких задач.
10. Алгоритм перевірки статистичних гіпотез.
11. Рівні статистичної значущості.
12. Статистична похибка та її види.
13. Статистичне рішення та змістовий висновок стосовно гіпотези
14. Порівняння дисперсій за F-критерієм Фішера.
15. Порівняння середніх за критерієм t-Стюдента. Непараметричні методи порівняння вибірок.
16. Порівняння двох незалежних вибірок за U-критерієм Манна-Уїтні.
17. Порівняння двох залежних вибірок за T-критерієм Вілкоксона та G-критерієм знаків.
18. Порівняння розподілу ознаки за  $\chi^2$ -критерієм Пірсона.
19. Особливі випадки використання критерію  $\chi^2$ .
20. Види коефіцієнтів кореляції й умови їх застосування у психологічних дослідженнях.