**ПРАКТИЧНА РОБОТА № 9, 10**

**ДИСПЕРСІЙНИЙ АНАЛІЗ ДАНИХ**

1. **Мета роботи та завдання**

Освоїти методи дисперсійного аналізу для статистичного дослідження даних.

Завданнями роботи є:

- Оволодіння методикою дослідження величин, що змінюються під впливом кількісних, якісних та випадкових факторів;

- Надбання навичок дослідження впливу тих чи інших факторів на мінливість середніх значень.

1. **Теоретичні відомості**

В будь-якому експерименті середні значення досліджуваних величин змінюються у зв’язку зі зміною основних факторів (кількісних та якісних), що визначають умови досліду, або випадкових факторів. Дослідження впливу тих чи інших факторів на мінливість середніх значень є задачею дисперсійного аналізу.

Дисперсійний аналіз використовує обумовлену дією незалежних факторів властивість адитивності дисперсії випадкової величини. В залежності від числа джерел дисперсії розрізняють однофакторний та багатофакторний дисперсійний аналіз.

Дисперсійний аналіз ефективний при вивченні кількох факторів. Але при класичному методі вивчення змінюють тільки один фактор, а решту залишають незмінними. При цьому для кожного фактору проводиться своя серія спостережень, що не використовується при вивченні інших. Дисперсійний аналіз полягає у відокремленні і оцінці факторів, що викликають зміну досліджуваної випадкової величини.

При цьому проводиться розклад сумарної вибіркової дисперсії на складові, обумовлені незалежними факторами. Щоб вирішити, чи дієвий вплив даного фактору, необхідно оцінити значимість відповідної вибіркової дисперсії за критерієм Фішера. Коли розрахункове значення критерію Фішера виявиться меншим табличного, то вплив досліджуваного фактору немає підстав вважати значимим. Коли ж розрахункове значення критерію Фішера виявиться більшим табличного, то цей фактор впливає на зміни середніх.

В подальшому ми вважаємо, що виконуються наступні припущення:

1) випадкові помилки спостережень мають нормальний розподіл;

2) фактори впливають тільки на зміну середніх значень, а дисперсія спостережень залишається постійною.

Фактори, що розглядаються в дисперсійному аналізі, бувають двох типів з випадковими рівнями та з фіксованими. В першому випадку мається на увазі, що вибір рівнів проходить з необмеженої сукупності можливих рівнів. Якщо рівні обираються випадковим чином, математична модель експерименту називається моделлю з випадковими рівнями факторів. Коли всі рівні фіксовані – модель з фіксованими рівнями факторів. Коли частина факторів розглядається на фіксованих рівнях, а решта обираються випадковим чином – модель змішаного типу.

Для однофакторного дисперсійного аналізу розглядається дія єдиного фактору *А* (кількісного чи якісного), котрий приймає *а* різних значень (рівнів фактора).

Таблиця 1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Групи за одним фак­тором  *А* | Окремі спостереження | | | | | | | Суми за групами | Середні за гру­пами |
| 1 | 2 | 3 | … | *j* | … | *n* |
| 1 | *x*11 | *x*12 | *x*33 | … | *x1j* | … | *x1n* |  |  |
| 2 | *x*21 | *x*22 | *x*23 | … | *x*2*j* | … | *x*2*n* |  |  |
| … | … | … | … | … | … | … | … | … | … |
| *i* | *xi*1 | *xi*2 | *xi*3 | … | *xij* | … | *xin* |  |  |
| … | … | … | … | … | … | … | … | … | … |
| *a* | *xa*1 | *xa*2 | *xa*3 | … | *xaj* | … | *xan* |  |  |
|  | | | | | | | |  |  |

При цьому *n* – це кількість спостережень для фіксованого рівня фактору, *N* – загальна кількість спостережень (*N=na*). Слід зазначити, що кількість спостережень для фіксованого рівня фактору може бути різною для різних рівнів фактору.

Для дослідження зміни даних зазвичай відокремлюють три напрямки

- Загальна зміна всіх даних *xij* (незалежно від того, до якої групи вони відносяться) відносно загальної середньої

- Зміна середніх значень за кожною групою *xi* відносно загальної середньої ,

- Зміна всіх даних *xij* всередині кожної відносно середнього значення за групою.

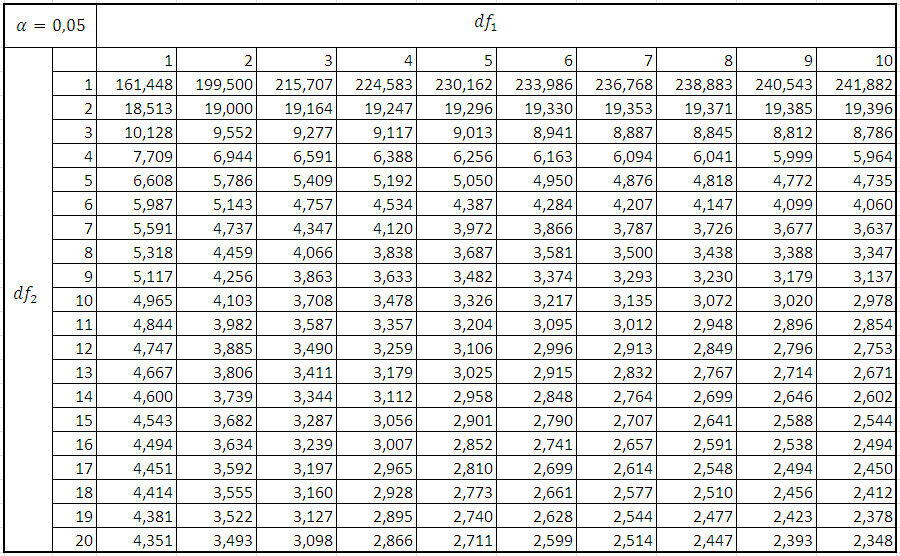
Для цього треба знайти суму квадратів відхилень, число відповідних степеней свободи та середній квадрат згідно таблиці.

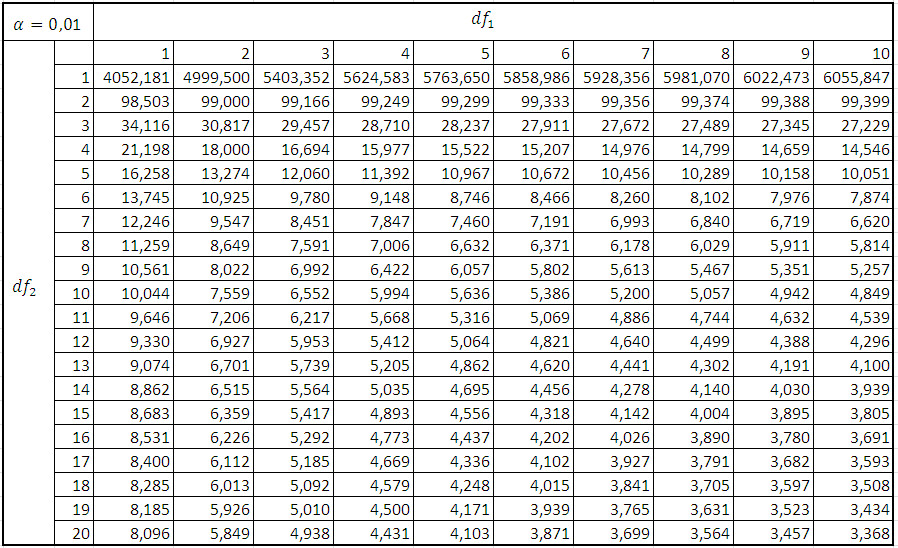
Таблиця 2

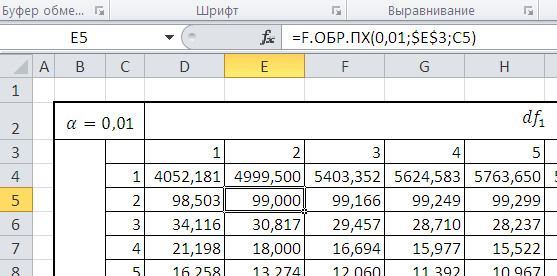
|  |  |  |
| --- | --- | --- |
| Сума квадратів *ss* | Число степеней свободи *df* | Середній квадрат *ms* |
|  |  |  |
|  |  |  |
|  |  |  |

Далі необхідно порівняти з табличним значенням розрахункове значення критерію Фішера



****Таблиця критерію Фішера для рівня значущості 0,05

Таблиця критерію Фішера для рівня значущості 0,01



Також табличне значення критерію Фішера можна обчислити за допомогою вбудованої функції *Ecxel* *F.ОБР.ПХ* (див. Рис.1).

Рисунок 1 – Приклад застосування функції *F.ОБР.ПХ*

Коли розрахункове значення критерію Фішера виявиться меншим табличного, то вплив досліджуваного фактору немає підстав вважати значимим. Коли ж розрахункове значення критерію Фішера виявиться більшим табличного, то цей фактор впливає на зміни середніх.

**3. Методика виконання роботи**

* 1. **Варіанти завдань**

**Варіант 1**

Для рослини *Nymphea* 4 рази на добу визначали вміст каротиноїдів:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Години | Визначення | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 18 | 1,42 | 1,30 | 1,68 | 1,59 | 1,49 | 1,62 | 1,36 | 1,26 | 1,58 | 1,66 |
| 24 | 1,45 | 1,38 | 1,49 | 1,71 | 1,54 | 1,57 | 1,34 | 1,32 | ,66 | 1,39 |
| 6 | 1,48 | 1,42 | 1,58 | 1,67 | 1,50 | 1,80 | 1,35 | 1,36 | 1,67 | 1,49 |
| 12 | 1,43 | 1,38 | 1,47 | 1,33 | 1,22 | 1.35 | 1,10 | 1,08 | 1,34 | 1,11 |

**Варіант 2**

Для рослини *Syringa Emodi* в різні години доби вивчали вміст каротиноїдів:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Години | Визначення | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 18 | 0,60 | 0,64 | 0,69 | 0,52 | 0,65 | 0,52 | 0,58 | 0,63 | 0,69 | 0,48 |
| 24 | 0,61 | 0,72 | 0,72 | 0,58 | 0,46 | 0,52 | 0,72 | 0,59 | 0,71 | 0,79 |
| 6 | 0,58 | 0,59 | 0,66 | 0,46 | 0,47 | 0,56 | 0,70 | 0,60 | 0,60 | 0,64 |
| 12 | 0,70 | 0,57 | 0,67 | 0,87 | 0,52 | 0,66 | — | 0,59 | 0,65 | 0,66 |

**Варіант 3**

Для рослини *Abutilon* в різні години доби вивчали дані про вміст хлорофілу (у мг/дм3):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Години | Визначення | | | |
| 1 | 2 | 3 | 4 |
| 15 | 3,06 | 2,88 | 2,83 | 2,41 |
| 18 | 3,20 | 2,97 | 2,50 | 3,03 |
| 21 | 1,82 | 1,73 | 1,33 | 2,25 |
| 24 | 1,67 | 1,26 | 1,52 | 1,36 |
| 6 | 2,76 | 1,26 | 1,46 | 1,32 |
| 9 | 2,78 | 2,70 | 2,49 | 1,66 |
| 12 | 2,41 | 3,22 | 1,90 | 2,00 |

**Варіант 4**

Для рослини *Solánum lycopérsicum* в різні години доби вивчали дані про вміст хлорофілу в умовних одиницях:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Години | Визначення | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 18  24  6  12 | 0,220,230,250,24 | 0,270,24⎯ 0,28 | 0,220,240,190,25 | 0,230,260,240,27 | 0,200,230,200,22 | 0,280,25⎯ 0,27 | 0,320,260,300,32 | 0,280,280,270,30 | 0,290,300,200,29 | 0,270,230,260,27 |

**Варіант 5**

Для рослини *Solánum lycopérsicum* в різні години доби вивчали дані про вміст хлорофілу (умовні одиниці):

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Години | Визначення | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 18 | 0,087 | 0,106 | 0,091 | 0,097 | 0,077 | 0,103 | 0,118 | 0,104 | 0,099 | 0,108 |
| 24 | 0,086 | 0,092 | 0,095 | 0,100 | 0,086 | 0,096 | 0,101 | 0,113 | 0,113 | 0,092 |
| 6 | 0,093 | — | 0,091 | 0,089 | 0,081 | — | 0,118 | 0,106 | 0,088 | 0,096 |
| 12 | 0,088 | 0,091 | 0,101 | 0,113 | 0,079 | 0,097 | 0,114 | 0,107 | 0,103 | 0,093 |

**Варіант 6**

Для рослини *Nymphea* 4 рази на добу визначали вміст каротиноїдів:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Години | Визначення | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 18 | 1,42 | 1,8 | 1,68 | 1,59 | 1,49 | 1,62 | 1,36 | 1,76 | 1,58 | 1,85 |
| 24 | 1,55 | 1,54 | 1,55 | 1,71 | 1,62 | 1,57 | 1,58 | 1,32 | 1,66 | 1,49 |
| 6 | 1,48 | 1,42 | 1,58 | 1,5 | 1,5 | 1,1 | 1,35 | 1,36 | 1,67 | 1,49 |
| 12 | 1,43 | 1,65 | 1,47 | 1,59 | 1,54 | 1,35 | 1,5 | 1,46 | 1,54 | 1,24 |

**Варіант 7**

Для рослини *Syringa Emodi* в різні години доби вивчали вміст каротиноїдів:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Години | Визначення | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 18 | 0,6 | 0,64 | 0,59 | 0,52 | 0,65 | 0,52 | 0,58 | 0,63 | 0,54 | 0,48 |
| 24 | 0,61 | 0,72 | 0,72 | 0,58 | 0,78 | 0,81 | 0,72 | 0,59 | 0,71 | 0,79 |
| 6 | 0,58 | 0,59 | 0,66 | 0,68 | 0,69 | 0,71 | 0,7 | 0,6 | 0,6 | 0,64 |
| 12 | 0,5 | 0,57 | 0,54 | 0,47 | 0,52 | 0,66 | 0,54 | 0,59 | 0,65 | 0,66 |

**Варіант 8**

Для рослини *Abutilon* в різні години доби вивчали дані про вміст хлорофілу (у мг/дм3):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Години | Визначення | | | |
| 1 | 2 | 3 | 4 |
| 15 | 3,06 | 2,88 | 1,98 | 1,56 |
| 18 | 3,2 | 2,97 | 2,5 | 2,44 |
| 21 | 1,82 | 1,73 | 1,33 | 1,78 |
| 24 | 1,67 | 1,88 | 1,97 | 1,65 |
| 6 | 2,65 | 1,58 | 1,95 | 2,1 |
| 9 | 2,4 | 2,68 | 2,49 | 1,89 |
| 12 | 2,41 | 3,22 | 1,9 | 2,56 |

**Варіант 9**

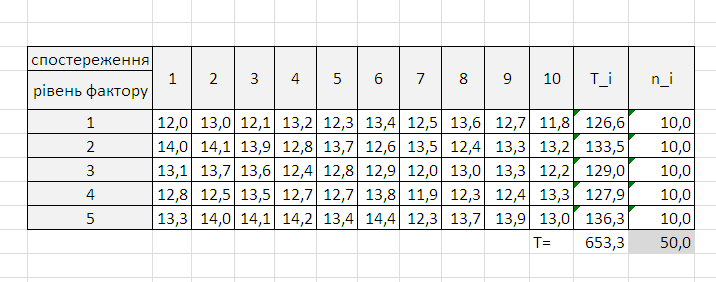
Для рослини *Solánum lycopérsicum* в різні години доби вивчали дані про вміст хлорофілу (умовні одиниці):

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Години | Визначення | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 18 | 0,075 | 0,089 | 0,091 | 0,074 | 0,085 | 0,093 | 0,092 | 0,078 |  | 0,077 |
| 24 | 0,086 | 0,092 | 0,095 | 0,089 | 0,086 | 0,096 | 0,079 | 0,076 | 0,095 | 0,091 |
| 6 | 0,093 | 0,097 | 0,091 | 0,089 | 0,081 | 0,097 | 0,095 | 0,096 | 0,088 | 0,084 |
| 12 | 0,122 | 0,11 | 0,101 | 0,113 | 0,015 | 0,1 | 0,114 | 0,107 | 0,103 | 0,099 |

* 1. **Хід роботи**

Для даних зі свого варіанту необхідно проаналізувати чи впливає вказаний фактор на мінливість даних.

Для цього треба побудувати таблицю з вхідними даними та сумами рядів. Для підрахунку даних для фіксованого рівня фактору можна використати функцію *СЧЕТ* (див. Рис.2)

Рисунок 2 – Таблиця з вхідними даними, сумами рядів та кількістю спостережень.

Далі розрахувати суми квадратів відхилень, кількість відповідних степеней свободи та середні квадрати згідно таблиці 2. Додати до таблиці розрахункові та критичні значення критерію Фішера (див. Рис.3) та зробити висновок щодо впливу фактора на мінливість даних

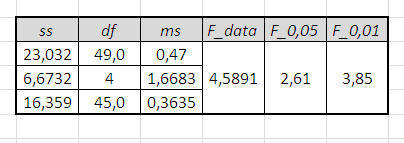


Рисунок 3 – Вихідна таблиця дисперсійного аналізу.

**Вимоги до змісту та оформлення звіту**

Звіт повинен бути продемонстрований як на паперовому носії, що містить розподіл даних за рівнями фактору та розрахунок критерію Фішера, так і в електронній формі у вигляді файлу із заповненими таблицями.

Висновок за результатами дисперсійного аналізу даних повинен містити наступні положення:

- яка інформація досліджувалася і якими методами;

- чи впливає вказаний фактор на мінливість даних;

- які прогнози і рекомендації можна зробити.

Титульний аркуш звіту повинен містити всю інформацію, необхідну для однозначної ідентифікації авторів і роботи. Для цього на титульному аркуші вказують назву дисципліни, тему і номер роботи, варіант завдання, номер групи, прізвища та ініціали студентів, посаду, прізвище та ініціали викладача та інше.

Порядок виконання робот:

1. Ознайомтесь з описом наступних функцій Excel:

*F.ОБР.ПХ*.

2. Скомпонуйте дані згідно рівнів фактору, що впливає.

3. Розрахуйте суми значень даних для кожного рівня фактору.

4. Розрахуйте кількість спостережень для кожного рівня фактору

5. Знайдіть розрахункове значення критерію Фішера.

6. Знайдіть критичні значення критерію Фішера для рівнів значущості 0,05 та 0,01.

7. Порівняйте значення критеріїв Фішера, зробіть висновок.

10. Оформіть звіт відповідно до вимог.

Контрольні питання

1. Що вивчає дисперсійний аналіз?

2. Що таке критерій Фішера?

3. З яких компонентів складається фактичне відхилення випадкової величини від середньої арифметичної?

4. Що таке рівні (градації) факторів?

5. Яка різниця між фіксованими і випадковими градаціями факторів?

6. Що таке середній квадрат (дисперсія)?

7. Напишіть робочі формули дисперсійного аналізу.

8. Які параметри оцінюють середні квадрати?

Критерії результативності виконання роботи

Лабораторна робота вважається виконаною в тому випадку, якщо студент:

1. Виконав всі зазначені завдання, дотримуючись порядку виконання роботи;

2. Освоїв методику виконання типових завдань і здатний продемонструвати роботу програми;

3. Представив звіт, що містить таблиці в вхідними та вихідними даними;

4. Відповів на всі контрольні і додаткові питання.