

## ЛАБОРАТОРНЕ ЗАНЯТТЯ 4

### ДИСПЕРСІЙНИЙ АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

**Мета:** лабораторне заняття передбачає засвоєння методів проведення однофакторного та двофакторного дисперсійного аналізу із використанням табличного процесору MS Excel.

**Обладнання та матеріали:** ПЕОМ із встановленою ОС Microsoft Windows XP або Windows Seven, процесор таблиць Microsoft Excel 2007 або 2010, Інтернет браузер, методичні вказівки.

#### План:

1. Однофакторний дисперсійний аналіз.
2. Двофакторний дисперсійний аналіз.

#### Теоретичні відомості.



Аналізуючи біологічні особливості тварин і рослин в природі або при розведенні їх в культурі, легко переконатися в їх постійній залежності від багатьох зовнішніх і внутрішніх факторів. Так, миші варіюють в своєму зростанні і розвитку в залежності від розмірів їхніх батьків, від часу народження, від числа дитинчат у пометах, в яких вони народилися, від погодних умов, від кількості одержуваної ними їжі і т.д. Вага телят при народженні і в наступні періоди їхнього життя визначається спадковими особливостями їх матерів і батьків, умовами ембріонального розвитку, рівнем годівлі та утримання. Все це створює величезне різноманіття тварин і рослин за їх морфологічними і фізіологічними властивостями не тільки в природних умовах, але навіть при розведенні їх людиною у відносно однорідних умовах.

Така складна обумовленість біологічних властивостей організмів привела до необхідності розробки відповідних математичних методів, за допомогою яких можна було б виділити вплив окремих факторів і оцінити їх відносну роль в загальній мінливості цих властивостей. До їх числа відноситься дисперсійний аналіз, який на даний час являє собою самостійну і дуже важливу галузь біологічної статистики. Сутність його полягає у встановленні ролі окремих факторів в мінливості тієї чи іншої ознаки.

Справа в тому, що вплив тих чи інших факторів на досліджувану ознаку (або ознаки) ніколи не може бути виділено в чистому вигляді. Хоча при проведенні дослідів і намагаються зберегти умови максимально однорідними, все ж різні досліді дають дещо неоднакові результати. Це пояснюється тим, що на них впливають численні випадкові обставини, багато інших чинників, які змінюються від досліді до досліді і не піддаються контролю. Тим більше є великою роль таких додаткових неконтрольованих факторів при проведенні аналізу не в експериментальних умовах, а безпосередньо в природних умовах.

Тому виникає важливе завдання розкладання загальної мінливості ознаки

на складові частини, які з одного боку визначаються досліджуваними конкретними факторами, а з іншого – викликаються випадковими, неконтрольованими причинами. Дисперсійний аналіз дозволяє оцінювати значимість впливу окремих факторів, а також їх відносну роль у загальній мінливості.

### Хід виконання роботи.

**Завдання 1.** Для перевірки впливу гучності сигналу на швидкість реакції на швидкість реакції випадковим чином відібрали 3 групи досліджуваних. Першій групі (5 осіб) пред'являли звукових сигнал у 10 дБ, другій (6 осіб) – 30 дБ, третьої (4 особи) – 50 дБ. У досліджуваних кожної групи фіксували час реакції у мілісекундах.

Номер групи	1 (10дБ)	2 (30дБ)	3 (50дБ)
	304	272	223
	268	264	184
Результати	272	256	209
вимірювань, мс	262	269	183
	283	285	
		247	

Сформулюйте статистичні гіпотези ( $H_0$  та  $H_A$ ) за даними умовами та перевірте їх.



*Для проведення дисперсійного аналізу у MS Excel використовується процедури «Однофакторный дисперсионный анализ» та «Двухфакторный дисперсионный анализ» застосування яких передбачає наступний й алгоритм дій:*

- Створення таблиці з даними.
- Формулювання  $H_0$  та  $H_A$  гіпотез ( $H_0^A$  та  $H_0^B$  – у випадку двофакторного дисперсійного аналізу).
- Виклик надбудови «Анализ данных».
- Встановлення необхідних умов проведення статистичного тесту.
- Інтерпретація результатів статистичного тесту.
- Відхилення чи прийняття  $H_0$  гіпотези на прийнятому рівні статистичною значимості за результатами статистичного тесту.

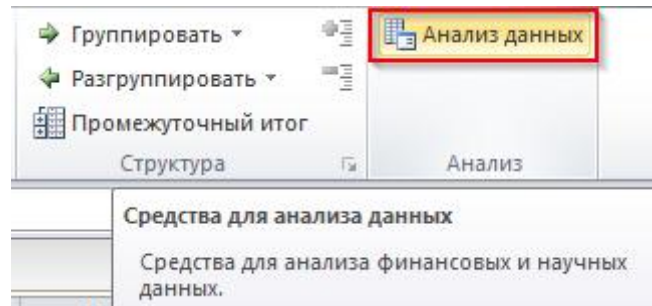
1.1. Сформулюємо нульову ( $H_0$ ) та альтернативну ( $H_A$ ) гіпотези:

**$H_0$  – Гучність сигналу впливає на швидкість реакції.**

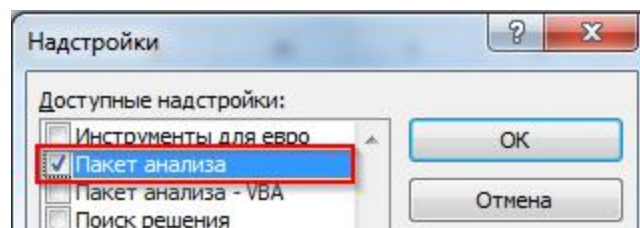
### ***$N_A$ – Гучність сигналу не впливає на швидкість реакції.***

Для перевірки вищезазначених припущень використаємо однофакторний дисперсійний аналіз. Для цього уведіть дані у відповідну таблицю (починаючи із комірки A1), таким чином, щоб у одному стовпчику опинились дані, які відповідають одному значенню досліджуваного фактора, а стовпчики розташовувались у порядку зростання (зменшення) фактора, що досліджується.

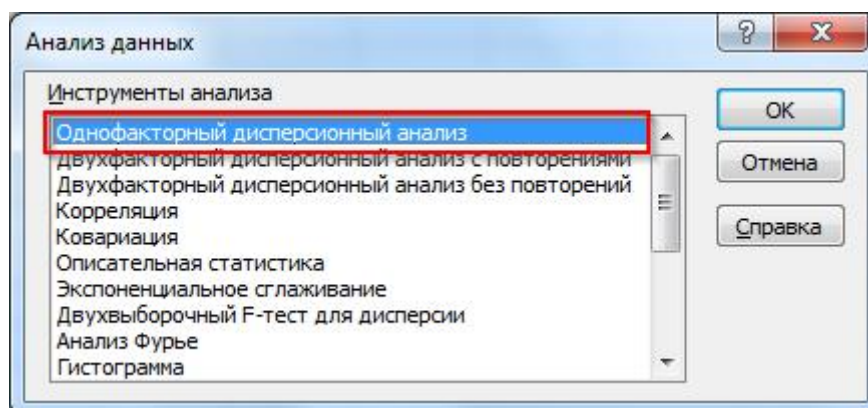
1.2. Далі, на стрічці MS Excel активуйте вкладку «Данные» та натисніть на кнопку «Анализ».



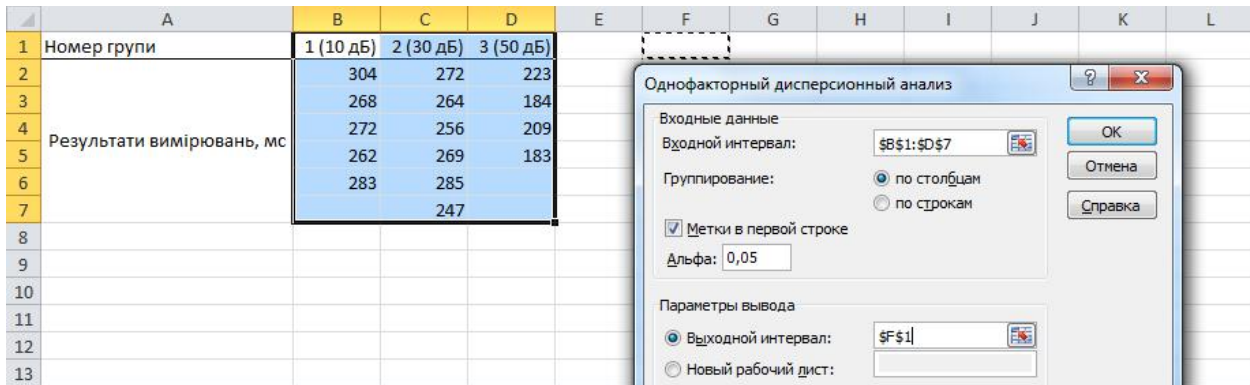
*Якщо ця вкладка відсутня, зробіть наступні кроки: вкладка «Файл» на стрічці, пункт «Параметры» та у діалоговому вікні, що з'явиться оберіть пункт «Надстройки» і далі – «Пакет анализа» та натисніть кнопку «Перейти». У діалоговому вікні, що з'явиться зробіть відмітку на пункті «Пакет анализа» та натисніть кнопку «ОК»*



1.3. У списку «Инструменты анализа» діалогового вікна «Анализ данных» оберіть процедуру «Однофакторный дисперсионный анализ», після чого натисніть кнопку «ОК»;



1.4. У діалоговому вікні, що з'явилося задайте вхідний інтервал (комірки B1:D7). У розділі «Группирование» встановити перемикач у положення «**по столбцам**». Встановіть позначку «**Метки в первой строке**» та вкажіть вихідний діапазон, тобто те місце в якому будуть виводитись результати аналізу, для чого встановіть перемикач у положення «**Выходной интервал**», клацніть ЛКМ по кнопці із зображенням таблиці у діалоговому вікні та оберіть комірку F1. Натисніть кнопку «**ОК**».



У результаті цих дій вихідний діапазон буде включати в себе результати дисперсійного аналізу: середні значення, дисперсії, критерій Фішера та інші показники.

Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	Счет	Сумма	Среднее	Дисперсия		
1 (10 дБ)	5	1389	277,8	273,2		
2 (30 дБ)	6	1593	265,5	173,9		
3 (50 дБ)	4	799	199,75	384,9166667		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	15341,88	2	7670,941667	29,53154425	0,0000232	3,885293835
Внутри групп	3117,05	12	259,7541667			
Итого	18458,93	14				

1.5. **Інтерпретація результатів.** Вплив фактора, що досліджується, визначається за величиною значимості критерію Фішера, яка знаходиться у таблиці «Дисперсионный анализ» на перетині рядка «Между группами» и столбца «P-Значение». У випадках, коли P-Значение < 0,05, критерій Фішера є значимим, і вплив досліджуваного фактору можна вважати доведеним на визначеному рівні довірчої вірогідності (у нашому випадку – 0,95).

**Завдання 2.** Досліджували процес росту бактерій у рідкому культуральному середовищі при трьох різних значеннях рН та чотирьох різних значеннях температури. Усі дослідні зразки знаходились у 12 окремих

пробірках, ступінь росту визначалась шляхом вимірювання оптичної щільності культури (табл. 4.1).

Таблиця 4.1 – Оптична щільність культури бактерій за умов культивування при різних значеннях температури та рН

Температура, °С	рН 5,5	рН 6,5	рН 7,5
<b>25</b>	<b>10</b>	<b>19</b>	<b>40</b>
<b>30</b>	<b>15</b>	<b>25</b>	<b>45</b>
<b>35</b>	<b>20</b>	<b>30</b>	<b>55</b>
<b>40</b>	<b>15</b>	<b>22</b>	<b>40</b>

Визначити, чи впливає температура культивування та рН культурального середовища на інтенсивність росту бактерій?

Визначена задача полягає у дослідження впливу двох факторів (температури та рН) на зміну значень випадкової величини  $x$  (інтенсивність росту бактеріальної культури). Різні значення факторів називають рівнями факторів. Для кожного набору рівнів факторів проводиться одне вимірювання, отже необхідно застосовувати двофакторний дисперсійний аналіз без повторень (у іншому випадку потрібна процедура двофакторного дисперсійного аналізу з повтореннями).



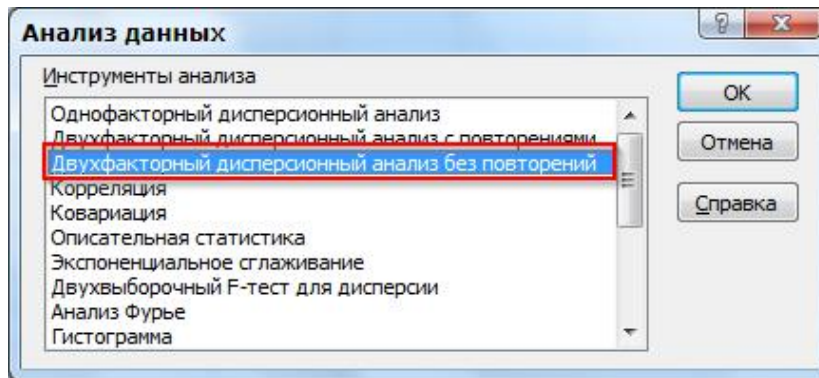
Дані для задач однофакторного дисперсійного аналізу зазвичай формуються у прямокутну таблицю, кожний рядок якої – дані про зміну випадкової величини за рівнями першого фактора ( $A$ ), стовпчики – дані про зміну випадкової величини за рівнями другого фактора ( $B$ ). Задача двофакторного дисперсійного аналізу у тому, щоб встановити, чи є фактори (кожний окремо або обидва разом) причиною мінливості випадкової величини. Далі формулюються дві нульових гіпотези, для кожного фактора своя, про те, що математичні очікування випадкових величин за дії різних рівнів фактора однакові. Якщо нульова гіпотеза не відхиляється, то відповідний фактор не впливає на мінливість, а у протилежному випадку – впливає.

2.1. Сформулюйте нульові гіпотези для вищезазначеної задачі. Останні можуть бути сформульовані наступним чином  $H_0^A$ : «Температура не впливає на інтенсивність росту бактеріальної культури»,  $H_0^B$  «рН не впливає на інтенсивність росту бактеріальної культури»; альтернативні гіпотези  $H_0^A$  – «Температура впливає на інтенсивність росту бактеріальної культури»,  $H_0^B$  – «рН впливає на інтенсивність росту бактеріальної культури».

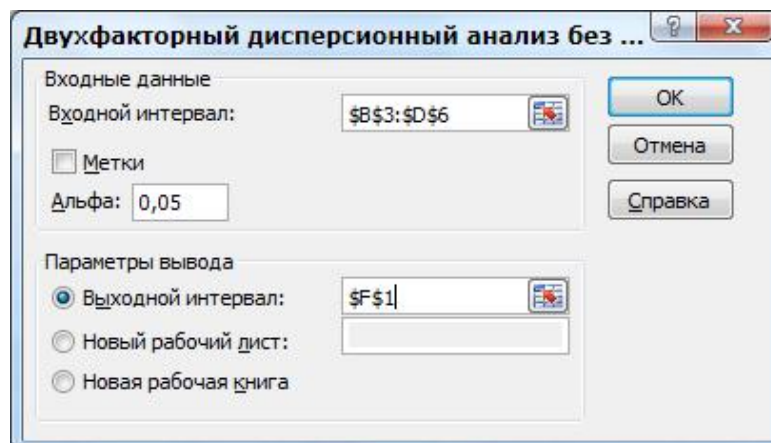
2.2. Створіть таблицю із даними.

	A	B	C	D	E
1		Оптическая щільність, ум.од.			
2	Температура, °C	pH 5,5	pH 6,5	pH 7,5	
3	25	10	19	40	
4	30	15	25	45	
5	35	20	30	55	
6	40	15	22	40	

2.3. На вкладці «Данные» стрічки MS Excel оберіть панель «Анализ» та натисніть на кнопку «Анализ данных». У діалоговому вікні «Анализ данных» оберіть пункт «Двухфакторный дисперсионный анализ данных без повторений» та натисніть на кнопку «ОК».



2.4. У діалоговому вікні, що з'явилося вкажіть вхідний інтервал (комірки B3:D6), вихідний інтервал (комірку F1) та натисніть на кнопку «ОК».



Результати аналізу будуть виведені у вказаний вами діапазон:



	F	G	H	I	J	K	L	M
1	Двухфакторный дисперсионный анализ без повторений							
2								
3	<b>ИТОГИ</b>							
		<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>			
4	Строка 1	3	69	23	237			
5	Строка 2	3	85	28,33333	233,3333333			
6	Строка 3	3	105	35	325			
7	Строка 4	3	77	25,66667	166,3333333			
8								
9	Столбец 1	4	60	15	16,6666667			
10	Столбец 2	4	96	24	22			
11	Столбец 3	4	180	45	50			
12								
13								
14	Дисперсионный анализ							
15		<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критическое</i>	
16	Строки	238,6666667	3	79,55556	17,46341463	0,002280032	4,757062663	
17	Столбцы	1896	2	948	208,097561	0,000002870	5,14325285	
18	Погрешность	27,33333333	6	4,555556				
19								
20	Итого	2162	11					

2.5. Інтерпретація результатів, полягає у оцінці отриманих Р-значень. У нашому випадку воно  $< 0,05$ , як для дії фактора «Температура» (рядки) так і для фактора «рН» (стовпчики). Окрім цього, обраховане значення F-критерію (17,46341463 та 208,097561) більше, ніж F-критичне (4,757062663 та 5,14325285 відповідно). Отже із 95% довірчою імовірністю ми можемо прийняти гіпотези  $H_A^A$  та  $H_A^B$  і стверджувати, що **«Як температура так і рН впливають на інтенсивність росту бактеріальної культури»**.

### Завдання для самостійного виконання.

1. Досліджували відмінність продуктивності відтворення одного і того ж навчального матеріалу у трьох групах випробовуваних (по 5 осіб), які розрізнялися умовами пред'явлення цього матеріалу для запам'ятовування. Результати обстеження наведені в таблиці.

Обстежувана особа, № з/п	Умо ва 1	Умо ва 2	Умо ва 3
1	5	8	11
2	4	7	9
3	3	6	7
4	6	9	10
5	7	5	8

Визначити, чи впливають умови пред'явлення навчального матеріалу на продуктивність його відтворення?

2. Досліджували активність ензиму манозо-6-фосфат ізомерази (MPI) та MPI генотипи у ракоподібних – бокоплавів *Platorchestia platensis*. Визначити, чи впливає генотип та стать бокоплавів на активність манозо-6-фосфат ізомерази.

Генотип	Самиці	Самці
FF	2,838	1,884
	4,216	2,283
	2,889	4,939
	4,198	3,486
FS	3,550	2,396
	4,556	2,956
	3,087	3,105
	1,943	2,649
SS	3,620	2,801
	3,079	3,421
	3,586	4,275
	2,669	3,110

Визначити, чи впливає а) генотип та б) стать бокоплавів на активність манозо-6-фосфат ізомерази?

#### Контрольні питання.

1. Сформулюйте завдання дисперсійного аналізу.
2. Наведіть приклади задач однофакторного дисперсійного аналізу.
3. Поясніть, що таке рівні фактора?
4. Яку інформацію про мінливість випадкової величини містить коефіцієнт детермінації?
5. Сформулюйте нульові гіпотези двофакторного дисперсійного аналізу.
6. Які твердження можна формулювати в завданні двофакторного дисперсійного аналізу? Наведіть приклади.
7. Опишіть алгоритм виконання однофакторного дисперсійного аналізу.