

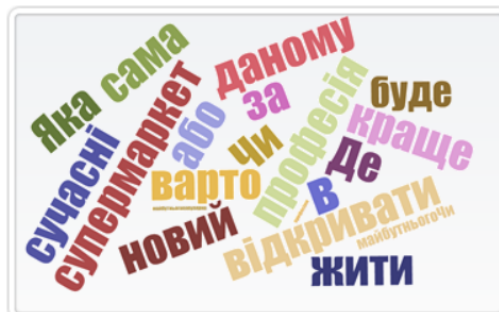
Урок 6. Основи статистичного аналізу даних

Вивчення нового матеріалу

Слайд
№ 1

Скільки буде коштувати проїзд в громадському транспорті?
Чи варто відкривати новий супермаркет в даному мікрорайоні?
Де вважають за краще жити сучасні українці?
Яка найпопулярніша професія майбутнього?

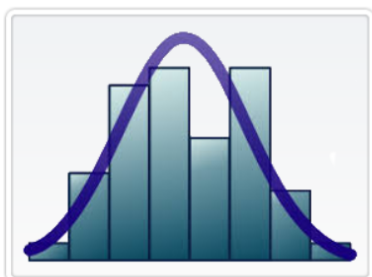
- Відповісти на ці та подібні питання можна за допомогою статистичного аналізу даних.
- Табличний процесор має потужні засоби статистичної обробки даних, що дають змогу аналізувати економічні, медикобіологічні та інші дані.



З початковими відомостями про статистику ви вже ознайомилися в курсі алгебри 9 класу.

Слайд
№ 2

Статистика — наука про методи збору, обробки, аналізу та інтерпретації даних, що характеризують масові явища і процеси, тобто явища і процеси, що стосуються не окремих об'єктів, а їх сукупностей.



Статистичні дані — сукупність чисел, які дають кількісну характеристику об'єктів та явищ, що нас цікавлять.



Поняття статистики об'єднує цілий комплекс дисциплін, в якому можна виділити основні напрямки: методи збору даних; методи вимірювання; методи обробки та аналізу даних.

Слайд
№ 3

Статистичне дослідження
складається з трьох етапів.

1 етап

Збір і групування первинних статистичних даних, отриманих шляхом спостережень або вимірювань (результати соціологічного опитування, спостереження за природними явищами, проведення експериментів тощо).

Отримуємо статистичну вибірку.

2 етап

Аналіз отриманих статистичних даних (обчислення характеристик вибірок).

Отримуємо числові характеристики.

3 етап

Отримання висновків за даними спостережень (встановлення закономірностей, властивих генеральній сукупності).

Робимо висновки.



Сьогодні ми розглянемо перший і, частково, другий етапи.

Слайд
№ 4

На першому етапі визначають множину об'єктів для збору даних.

Генеральна сукупність складається з усіх об'єктів, які підлягають вивченню. Склад генеральної сукупності залежить від цілей дослідження. Іноді генеральна сукупність – це все населення певного регіону (наприклад, коли вивчається ставлення потенційних виборців до кандидата), але найчастіше задається кілька критеріїв відбору.

Наприклад, це можуть бути учні 10 класів, які навчаються за інформаційно-технологічним профілем.

Вибірка або вибіркова сукупність – множина об'єктів, випадковим чином вибраних з генеральної сукупності для участі в дослідженні.



Як ви думаєте, чому досліджують не всю генеральну сукупність, а вибірку?

Слайд
№ 5

Розглянемо ситуацію:

Є генеральна сукупність – **100 000** мешканців міста. Ми робимо вибірку – **10** мешканців, виходимо на вулицю та опитуємо **10** перехожих біля нашого помешкання або офісу, чи подобається їм така погода, як зараз.

Наприклад, 8 з 10 опитаних сказали, що їм подобається така погода.

Слайд
№ 6

Вибірка повинна бути **репрезентативною**, тобто її об'єкти повинні достатньо добре відображати властивості генеральної сукупності.

У наведеному вище прикладі ми не продумували, кого, де та як відбираємо, а просто почали спілкуватися з перехожими.

Озирнімося довкола. Можливо, ми перебуваємо поблизу університету в навчальний день?

Тоді серед перехожих здебільшого молоді люди, які загалом легше переносять високу температуру, ніж старші, й через це відсоток задоволених погодою може виявитися штучно завищеним.

Або, можливо, ми потрапили туди, де серед перехожих більше літніх людей, яким, імовірно, важко переносити задуху спекотних літніх днів? Тоді відсоток задоволених погодою може виявитися заниженим порівняно з усіма мешканцями населеного пункту.



Репрезентативність — обов'язкова вимога до статистичної вибірки.

Слайд
№ 7

На другому етапі статистичного дослідження **обчислюють статистичні характеристики вибірки**.

Вибірка містить набір значень певного параметра вибраних об'єктів, наприклад, набір оцінок учнів з математики. Ці значення називаються **варіантами**.

Для вивчення закономірностей варіанти піддають обробці:

ранжуванню (розташування в порядку зростання або зменшення значень) і **групуванню** (об'єднання в групи однакових чи близьких значень).

Впорядкований набір значень варіанти називається **варіаційним рядом**, або **рядом розподілу**.



У результаті ранжування або групування отримуємо статистичний ряд розподілу.

Слайд
№ 8

Значення (а,б,е,м,о,р,с,т) називаються **варіантами**.

Розташувавши варіанти у порядку зростання, отримали **варіаційний ряд**:

а б е м о р с т



Ось приклади варіант та варіаційного ряду.

Частота — це число, яке показує, скільки разів зустрічається кожна варіанта.

Зауважте, що кожен частоту треба ділити на суму частот, а не на кількість варіант!

Ось приклад дискретного варіаційного ряду.

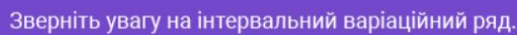
Слайд № 12

В інтервальному варіаційному ряді вказують не окремі значення варіанти, а **інтервали**, яким ці значення можуть належати. Частотами тоді будуть кількості значень, які потрапили в певний інтервал.

Розподіл середньодобових температур у липні		
Температура °C	Інтервал температури °C	Кількість днів
18	23	8
19	28	18
20	32	5
20		
21		
21		
21		
22		
24		
24		
24		
24		
24		
24		
24		

Розподіл середньодобових температур у липні

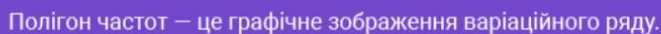
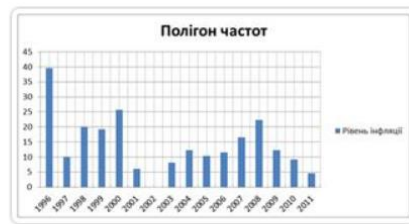
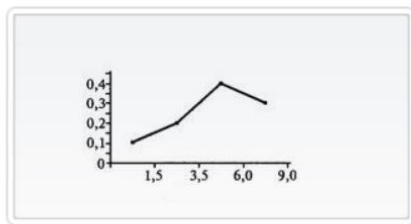
Температура (°C)	Кількість днів
23	8
28	18
32	5



**Слайд
№ 13**

Для графічного відображення дискретного та інтервального варіаційних рядів будують **полігони частот**.

Полігони частот можуть виглядати як ламана або як гістограма.



Вправа № 1



Задача.

Група з 15 учнів написала контрольну роботу з математики. Оцінки, одержані учнями за виконання контрольної роботи, виявилися такими:
12, 11, 7, 8, 9, 9, 7, 8, 10, 5, 6, 5, 6, 7, 3

Потрібно побудувати полігон частот, що відображає розподіл кількості учнів за рівнями успішності:

- ♦ **1 – 3** низький рівень
- ♦ **3 – 6** середній рівень
- ♦ **6 – 9** достатній рівень
- ♦ **9 – 12** високий рівень



Я покажу, як розв'язувати цю задачу в табличному процесорі крок за кроком.
Виконуйте кроки за мною!

Вправа № 1
Крок № 1



Крок 1

- ♦ Закачайте з електронного курсу файл-заготовку **урок6.xlsx** та відкрийте аркуш **оцінки**.
- ♦ Відсортуйте оцінки за зростанням.
- ♦ Введіть назви рівнів у клітинки **B2:B5**.
- ♦ Введіть верхні межі рівнів у клітинки **C2:C5**.

	A	B	C	D	E
	Оцінки		Інтервали	Частота	Відносна частота
1					
2	3	низький	3		
3	5	середній	6		
4	5	достатній	9		
5	6	високий	12		
6	6				
7	7				
8	7				
9	7				
10	8				
11	8				
12	9				
13	9				
14	10				
15	11				
16	12				



Щоб відсортувати діапазон чисел, скористайтеся меню **Дані > Сортувати**.

Вправа № 1
Крок № 2



Крок 2. Визначте частоту оцінок у кожному з інтервалів, скориставшись статистичною функцією **FREQUENCY (ЧАСТОТА)**.

- ♦ виділіть клітинки **D2:D5** (масив частот);
- ♦ відкрийте функцію **FREQUENCY**, натиснувши кнопку **fx**;
- ♦ в полі **Масив даних** укажіть діапазон **A2:A16**, в полі **Масив інтервалів** — діапазон **C2:C5**;
- ♦ функція **FREQUENCY** повертає не окреме значення, а масив значень; тому для отримання результату натисніть клавіші **Ctrl + Shift + Enter** (якщо не виділити діапазон для результатів або якщо натиснути лише клавішу **Enter**, то замість масиву значень буде отримано тільки 1 число).

	A	B	C	D	E	F
	Оцінки		Інтервали	Частота	Відносна частота	
1						
2	3	низький	3	C2:C5		
3	5	середній	6			
4	5	достатній	9			
5	6	високий	12			
6	6					
7	7					
8	7					
9	7					
10	8					
11	8					
12	9					
13	9					
14	10					
15	11					
16	12					

Аргументи функції

FREQUENCY

Масив_даних: A2:A16 = {3;5;5;6;7;7;8;8;9;9;10;11;12}

Масив_інтервалів: C2:C5 = {3;6;9;12}

Обчислює розподіл значень по інтервалах і повертає вертикальний масив значень, який містить на один елемент більше ніж масив_інтервалів.

Масив_інтервалів: масив або посилання на інтервали, по яких необхідно розподілити значення з масиву_даних.

Значення: 1

Додати з цієї функції

OK Скасувати



Функція **FREQUENCY** належить до категорії **Статистичні**.

Результат
вправи № 1,
кроку № 3

Результат

	A	B	C	D
	Оцінки		Інтервали	Частота
1				
2	3	низький	3	1
3	5	середній	6	4
4	5	достатній	9	7
5	6	високий	12	3
6	6			
7	7			
8	7			
9	7			
10	8			



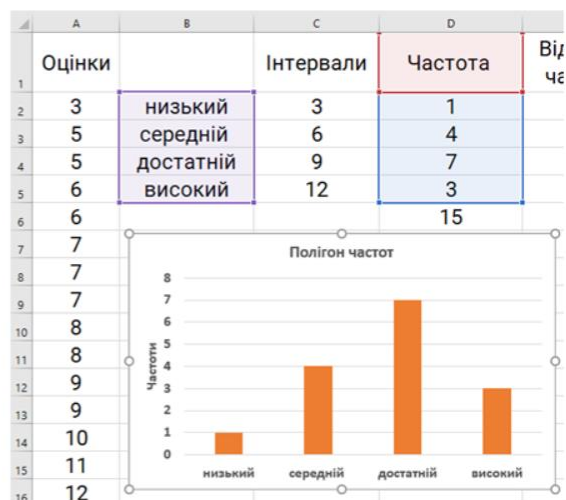
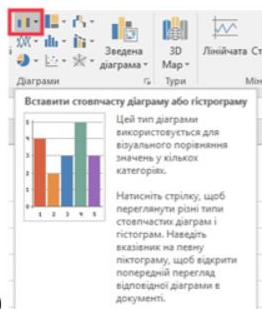
Для перевірки правильності обчислення частот можете порахувати їхню суму в клітинці **D6**. Має вийти 15.

Вправа № 1
Крок № 3



Крок 3. Побудуйте полігон частот оцінок з математики.

Для цього виділіть діапазон клітинок **B1:D5** і виберіть команди **Вставлення > Діаграми > Столпчаста діаграма**.



Не забудьте зберегти файл і відправити на перевірку вчителю.

Вправа № 2



Задача.

З партії деталей випадковим чином вибрали 25 виробів. Їхня маса (в г) виявилася такою, як вказано на аркуші **вироби** файлу **урок6.xlsx**.

Побудуйте інтервальний варіаційний ряд з 8 інтервалів, полігон частот та відносних частот.

- Розмах вибірки, заданої в діапазоні A2:A26, можна визначити за допомогою формули **=MAX(A2:A26)-MIN(A2:A26)**
- Величина кожного з 8 інтервалів ряду розподілу дорівнюватиме **R/8**, де **R** — розмах вибірки.

	A	B	C	D
	Вага			Відносні
1	деталей	Інтервали	Частоти	частоти
2	54	48	5	0,2
3	51	49	5	0,2
4	55	50	4	0,16
5	47,5	51	2	0,08
6	49,5	52	3	0,12
7	47	53	1	0,04
8	49,5	54	2	0,08
9	52	55	3	0,12
10	51,5	Разом:	25	
11	48			
12	49			
13	54,5			
14	54	Розмах:	8	
15	47	Довжина інтервалу:	1	
16	49			
17	48			
18	51,5			



Виконавши завдання, збережіть файл і відправте на перевірку вчителю.

Коментар до вправи № 2

Як обчислити відносні частоти?

- У клітинці **C10** обчисліть суму всіх частот за допомогою кнопки  або функції **SUM**.
- Введіть у клітинку **D2** формулу **=C2/C\$10** і скопіюйте її в діапазон **D3:D9**.
- Діапазону клітинок **D2:D9** надайте відсоткового формату, натиснувши кнопку  на вкладці **Основне**.

D2					=C2/\$C\$10				
	A	B	C	D					
1	Варг	Інтервали	Частоти	Відносні частоти					
2	47	48	5	=C2/\$C\$10					
3	47	49	5						
4	47,5	50	3						
5	48	51	3						
6	48	52	3						
7	48,5	53	1						
8	49	54	3						
9	49	55	2						
10	49		25						
11	49								
12	49,5								



Згадайте, що означає символ **\$** у формулі.