

ІНФОРМАТИКА

Основи статистичного аналізу даних. Ряди даних

10

За навчальною програмою 2018 року



Урок 8



1. Як обчислити середнє арифметичне кількох чисел? Для чого його використовують? Наведіть приклади.

2. Що таке ряди даних у табличному процесорі? Наведіть приклади.

3. З якою метою проводять опитування? У яких опитуваннях ви брали участь? Про які опитування ви чули або читали?





Вам уже відомо з курсу алгебри 9-го класу, що в багатьох:

дослідженнях для аналізу даних

установлення певних закономірностей

формулювання висновку

надання рекомендацій

прогнозування тощо

потрібно використати багато даних. Ви знаєте, що методи отримання, опрацювання й аналізу даних, які характеризують масові явища, вивчає наука **статистика.**





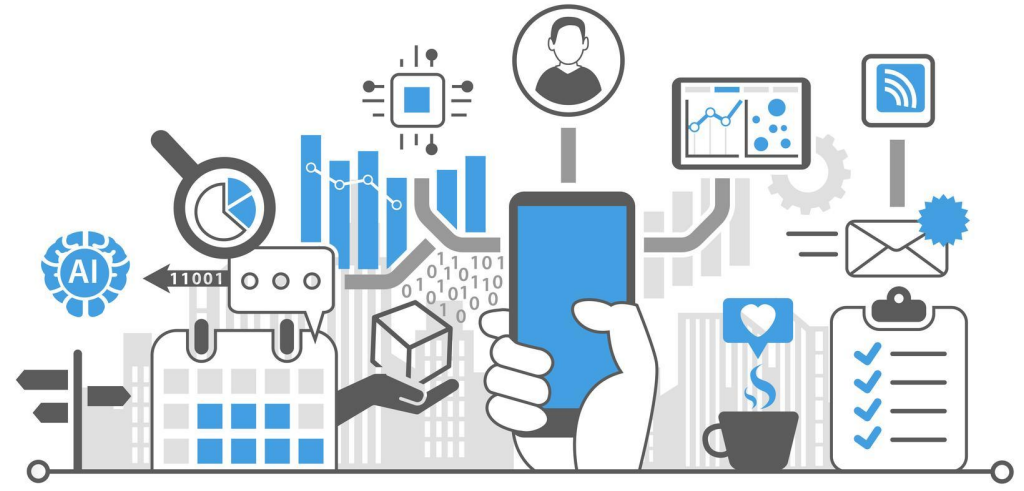
Так, наприклад,

для аналізу тенденцій змінення маси учнів 10-х класів України за останні роки

ризикі серцевих захворювань людей певного віку на планеті

популярності продуктів харчування серед населення певного регіону

Потрібно проаналізувати сотні тисяч або навіть мільйони даних.





Зрозуміло, що,

*провести
зважування*

*вивчити історії
хвороб*

*провести
анкетування*

*Сотень тисяч або навіть мільйонів людей практично неможливо. Тому для аналізу створюють певну **вибірку об'єктів дослідження**, тобто з усієї множини об'єктів дослідження відбирають певну кількість і на ній проводять дослідження. Що більше така вибірка, то точніше буде проведено аналіз і зроблено відповідні висновки. Тобто вибірка повинна бути **масовою**.*



Але не тільки кількість даних у вибірці визначає рівень точності аналізу і висновків. Так,

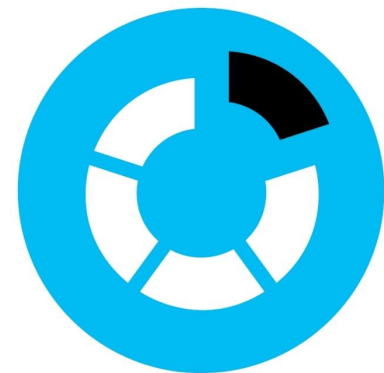
у першому і другому з наведених вище прикладів доцільно вибирати

у третьому доцільно вибирати

людей різних регіонів і різної статі

людей різного віку

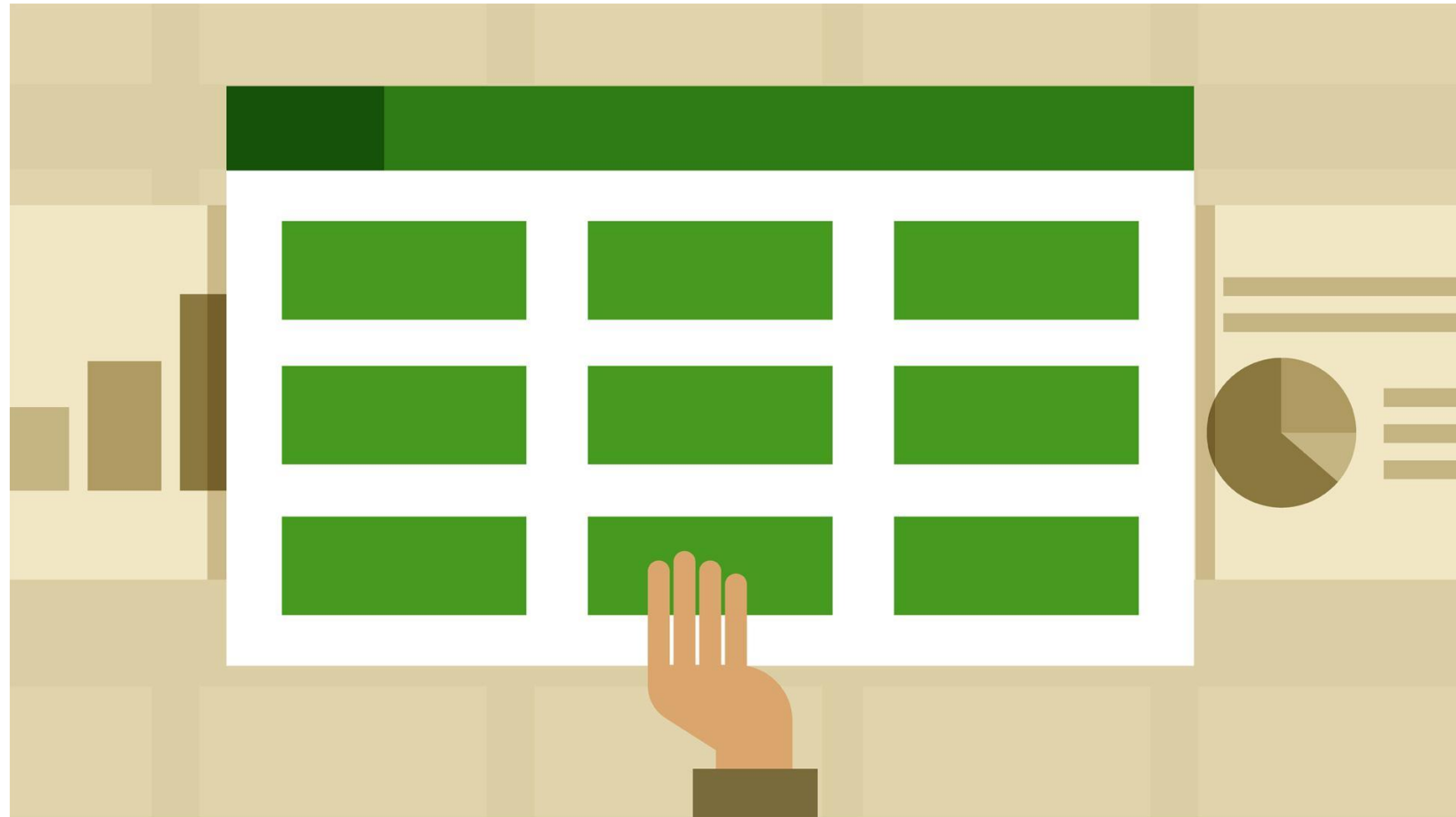
*Кажуть, що вибірка даних має бути **репрезентативною** (франц. *Representatif* — показовий, характерний, типовий).*





Дані, отримані з дослідженої вибірки, найчастіше заносять у таблицю.

*Така форма подання даних з вибірки зручна для їх аналізу та прогнозів. Дані з кожного рядка і стовпця такої таблиці утворюють **ряди даних**.*





Наведемо кілька прикладів вибірок і рядів даних.

У таблиці подано результати виступів команди учнівства України на міжнародних олімпіадах з інформатики з 2005 по 2017 рік.

Тут вибіркою є вказані в таблиці роки, а рядами даних — загальна кількість медалей у ці роки, а також кількість золотих, срібних і бронзових медалей у вказані роки.

Рік	Кількість медалей	Золоті	Срібні	Бронзові
2005	4	2	1	1
2006	4	1	2	1
2007	4	1	2	1
2008	3	0	1	2
2009	4	1	1	2
2010	3	0	1	2
2011	3	0	1	2
2012	4	1	1	2
2013	4	0	1	3
2014	3	0	1	2
2015	4	0	3	1
2017	3	1	2	0

Вибірка

Ряди даних



Зауважимо, що Україна бере участь у таких олімпіадах починаючи з 1992 року. Команда кожної країни складається із чотирьох учнів. У 2016 році команда України не брала участі у такій олімпіаді.

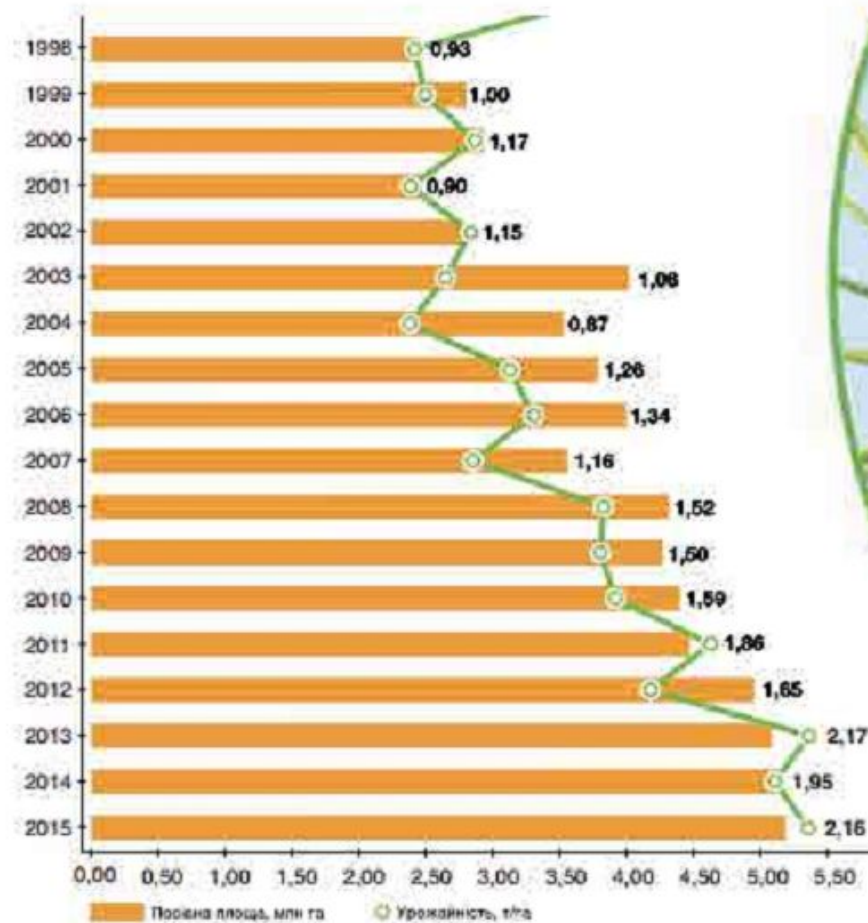
За цими рядами даних або за деякими з них можна побудувати графіки або діаграми і візуалізувати їх, використовуючи, наприклад, **табличний процесор.**





На малюнку подано діаграму динаміки посівних площ соняшнику в Україні в 1998–2015 роках і графік врожайності соняшнику в Україні за ці роки. Тут вибіркою є 1998–2015 роки, а рядами даних — посівні площі та врожайність соняшнику в ці роки.

Динаміка посівних площ та врожайності соняшнику в Україні





За цими діаграмами і графіками ряди даних або частини рядів даних можна, за потреби, занести в таблицю.

Таблиця даних про врожайність соняшнику і посівну площу в Україні в 2006-2015 роках

<i>Рік</i>	<i>Урожайність, т/га</i>	<i>Посівна площа, млн га</i>
2006	1,34	4,2
2007	1,16	3,6
2008	1,52	4,4
2009	1,5	4,3
2010	1,59	4,5
2011	1,66	4,6
2012	1,65	5
2013	2,17	5,2
2014	1,95	5,3
2015	2,16	5,3



Як приклад розглянемо роботу синоптичної служби. Синоптична служба збирає і зберігає дані про температуру, опади, атмосферний тиск та інші показники вже понад 160 років.





Для прогнозування температури та ймовірності опадів у Львові в першій декаді червня наступного року для проведення екскурсії потрібно вибрати відповідні дані, наприклад за останні 10–15 років саме про Львів і саме про першу декаду червня, проаналізувати отримані два ряди даних:

про температуру



кількість опадів



Зробити відповідний прогноз погоди.



Розглянемо деякі статистичні характеристики ряду даних:

середнє арифметичне

AVERAGE

стандартне відхилення

STDEV.P

мода

MODE.SNGL (MODE.MULT)

медіана

MEDIAN



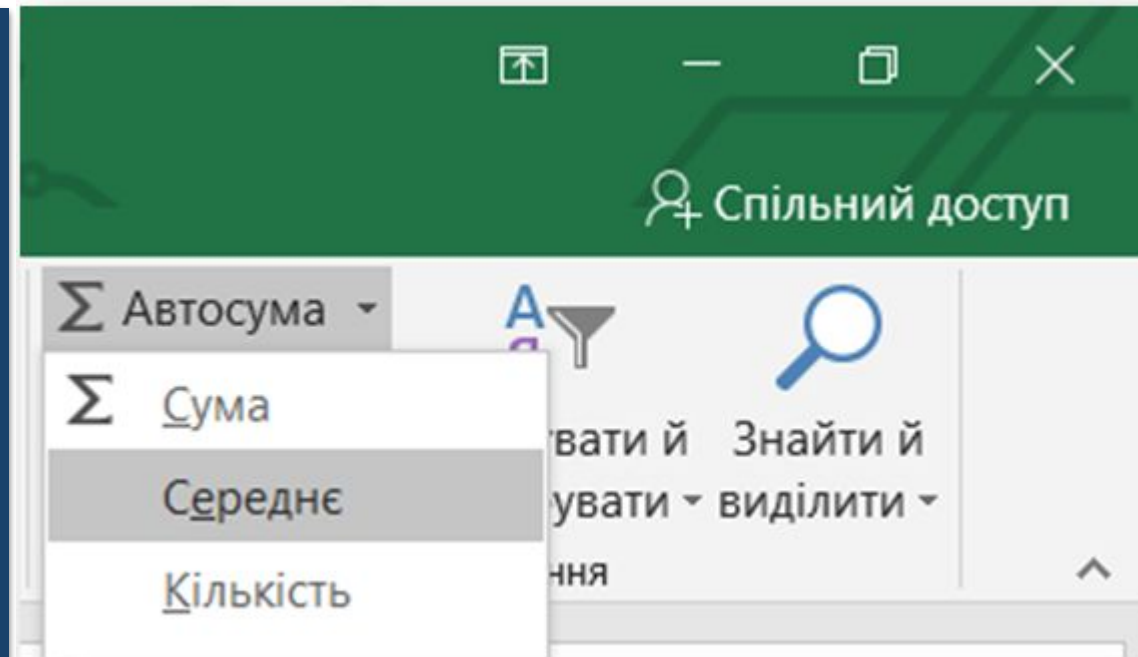
Так, можна знайти середнє арифметичне врожайності соняшнику в Україні за 2006–2015 роки, використовуючи, наприклад, табличний процесор.

На малюнку наведено приклад обчислення середньої врожайності соняшнику за 2006–2015 роки і формулу для її обчислення
=AVERAGE (C3:C12)

	A	B	C	D	E
1					
2		Рік	Урожайність, т/га		
3		2006	1,34		
4		2007	1,16		
5		2008	1,52		
6		2009	1,5		
7		2010	1,59		
8		2011	1,66		
9		2012	1,65		
10		2013	2,17		
11		2014	1,95		
12		2015	2,16		
13		Середнє:	1,67		



Для обчислення середнього арифметичного в табличному процесорі можна використати відому вам функцію **AVERAGE** (англ. *average* — середній). Нагадаємо, що аргументами цієї функції може бути:



діапазон клітинок

список клітинок

їх комбінації

наприклад

AVERAGE (B2:D5; F4; E7)



Обчислене в наведеному прикладі середнє арифметичне визначає, яка б була врожайність кожного року (1,67 т/га), якщо вона щороку була б однаковою.

Аналогічно середнє арифметичне будь-якого ряду даних визначає, які б були значення у цьому ряді, якщо б вони всі були однакові.



Деякі статистичні характеристики ряду даних



Зазначимо, що не для всіх рядів даних середнє арифметичне є показовою характеристикою самого цього ряду. Наприклад, для ряду даних:

2,5; 2,8; 2,3; 2,55; 2,47

У якому дані незначно відрізняються одне від одного, середнє арифметичне дорівнює:

2,524

Що незначно відрізняється від усіх членів цього ряду, а значить, достатньо показово характеризує весь цей ряд даних.



А для ряду:

$4,7; 6,2; 5,1; 12,4; 14,1$

У якому дані значно відрізняються одне від одного, середнє арифметичне дорівнює:

$8,5$

Що значно відрізняється від усіх членів цього ряду, а значить, недостатньо показово характеризує весь цей ряд даних.



Для визначення, наскільки показово середнє арифметичне ряду даних характеризує весь ряд даних,

можна використати таку характеристику ряду даних, як **стандартне відхилення**. Стандартне відхилення характеризує, наскільки широко розташовані значення ряду даних відносно їх середнього арифметичного.





Стандартне відхилення обчислюється за формулою:

$$S = \sqrt{\frac{(x_1 - x_0)^2 + (x_2 - x_0)^2 + (x_3 - x_0)^2 + \dots + (x_n - x_0)^2}{n}}$$

Де,

x_1, x_2, \dots, x_n

члени ряду даних

x_0

**середнє арифметичне цього ряду
даних**



Для першого з вищенаведених двох прикладів рядів даних стандартне відхилення дорівнює:

$$s = \sqrt{\frac{(2,5 - 2,524)^2 + (2,8 - 2,524)^2 + (2,3 - 2,524)^2 + (2,55 - 2,524)^2 + (2,47 - 2,524)^2}{5}} \approx 0,16$$

А для другого:

$$s = \sqrt{\frac{(4,7 - 8,5)^2 + (6,2 - 8,5)^2 + (5,1 - 8,5)^2 + (12,4 - 8,5)^2 + (14,1 - 8,5)^2}{5}} \approx 3,95$$



Очевидно,

що середнє арифметичне першого ряду даних менше відрізняється від усіх членів ряду даних, а значить,

більш показово характеризує весь цей ряд даних.

А середнє арифметичне другого ряду даних більше відрізняється від усіх членів ряду даних, а значить,

менш показово характеризує весь цей ряд даних.



Автоматизувати обчислення стандартного відхилення в табличному процесорі можна, використавши функцію:

STDEV.P (англ.
standard deviation —
стандартне
відхилення)

(Для версії нижче
2010 — *STDEV.P*).

	A	B	C	D	E
1					
2		x1	2,5	4,7	
3		x2	2,8	6,2	
4		x3	2,3	5,1	
5		x4	2,55	12,4	
6		x5	2,47	14,1	
7		Середнє	2,524	8,5	
8		Стандартне відхилення	0,1615673	3,946137	



Ще однією характеристикою ряду даних є **мода**.

Мода — це значення в ряді даних, яке повторюється найчастіше. Таке значення є показовим, наприклад,

*під час дослідження цін на
ринку*

*під час дослідження попиту
взуття, одягу*

*ціна, яка
трапляється
найчастіше*



*розміри, які
купують
найбільше*



Деякі статистичні характеристики ряду даних

Розділ 2
§ 2.2



У розглянутому вище прикладі, «Результати виступу команди учнівства України на міжнародних олімпіадах з інформатики за 2005–2017 роки».

*мода
кількостей
медалей*

*мода
кількостей
золотих
медалей*

*мода
кількостей
срібних
медалей*

*мода
кількостей
бронзових
медалей*

дорівнює 4

0

1

2

Тому що найчастіше в ці роки команда нашої країни завойовувала 4 медалі.



Якщо в ряді даних два або більше значень повторюються найбільшу кількість разів, то кожне з них вважається модою ряду даних.

Так, наприклад, у ряді даних

2, 3, 3, 2, 1

Модою є:

і число 2

число 3



У табличному процесорі є спеціальна функція для обчислення моди ряду даних, якщо вона одна — **MODE.SNGL** (англ. *mode single* — мода одинарна).

Аргументами цієї функції може бути:

діапазон клітинок

список клітинок

їх комбінації

Наприклад,

MODE.SNGL (B2:D5; F4; E7)

Деякі статистичні характеристики ряду даних

Розділ 2
§ 2.2



На малюнку наведено приклад обчислення моди для кількостей завойованих медалей і формула для її обчислення:

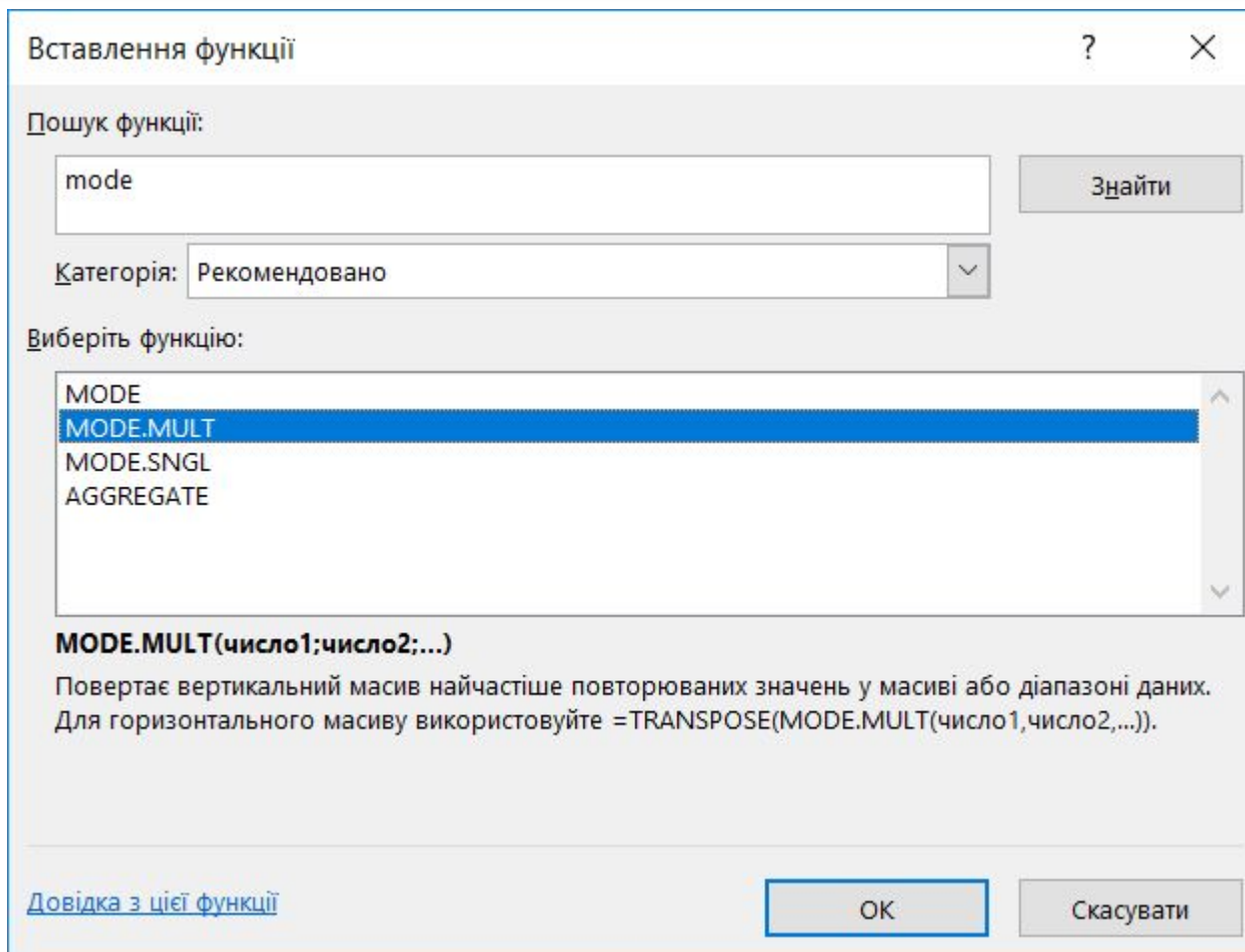
`=MODE.SNGL(C3:C14)`

	A	B	C	D	E	F
1						
2		Рік	Кількість медалей	Золоті	Срібні	Бронзові
3		2005 рік	4	2	1	1
4		2006 рік	4	1	2	1
5		2007 рік	4	1	2	1
6		2008 рік	3	0	1	2
7		2009 рік	4	1	1	2
8		2010 рік	3	0	1	2
9		2011 рік	3	0	1	2
10		2012 рік	4	1	1	2
11		2013 рік	4	0	1	3
12		2014 рік	3	0	1	2
13		2015 рік	4	0	3	1
14		2017 рік	3	1	2	0
15		Середнє	3,58	0,58	1,42	1,58
16		Мода	4	0	1	2



Якщо в ряді даних два або більше значень можуть повторюватися однаково найбільшу кількість разів,

то в Excel потрібно використати функцію **MODE.MULT (англ. *mode multiple* — мода багаторазова), результатом якої будуть усі значення з ряду, які повторюються найбільшу кількість разів.**





На малюнку наведено приклад обчислення моди, якщо їх у ряді даних кілька.

`{=MODE.MULT(D3:D10)}`

	A	B	C	D	E	F
1						
2						
3			x1	2		
4			x2	2		
5			x3	2		
6			x4	3		
7			x5	3		
8			x6	4		
9			x7	4		
10			x8	4		
11			Мода	2		
12				4		
13				#N/A		
14				#N/A		



Оскільки в загальному випадку невідомо, скільки мод буде в ряді даних, потрібно:

- 1. Виділити для мод певний діапазон клітинок. У прикладі було виділено діапазон клітинок **D11:D14**.**
- 2. Увести формулу в першу клітинку виділеного діапазону.**
- 3. Натиснути **Ctrl+Shift+Enter**.**

Оскільки було виділено діапазон з чотирьох клітинок, а мод у наведеному ряді даних виявилось лише дві, то в двох інших клітинках виділеного діапазону відобразилося **#N/A.**



Розглянемо ще одну характеристику ряду даних — **медіану**.

Медіаною

впорядкованого ряду даних називається значення, яке поділяє ряд даних на дві рівні частини, тобто зліва і справа від цього значення знаходиться однакова кількість членів упорядкованого ряду даних.





Якщо у впорядкованому ряді даних:

непарна кількість членів

парна кількість членів

***то медіана такого ряду
даних дорівнює значенню
його середнього члена***

***то його медіана
обчислюється як середнє
арифметичне значень двох
середніх членів***



Наприклад, для ряду даних,

2; 3; 5; 6; 7

***медіана
дорівнює***

5

2; 3; 5; 6; 7; 9

***медіана
дорівнює***

$(5 + 6) : 2 = 5,5$

2; 2; 4; 4; 4; 5; 6

***медіана
дорівнює***

4

Деякі статистичні характеристики ряду даних



Медіана використовується, наприклад, для визначення місця побудови шкіл, дитячих садочків, магазинів, підприємств побуту тощо.

Потрібно визначити ряд відстаней, які слід подолати мешканцям певної місцевості до цього закладу, і побудувати його в точці, яка визначається медіаною цього ряду.





У табличному процесорі є спеціальна функція для обчислення медіани ряду даних — **MEDIAN** (англ. *median* — середній).

Аргументами цієї функції може бути:

діапазон клітинок

список клітинок

їх комбінації

Наприклад,

MEDIAN(B2:D5; F4; E7)



На малюнку наведено приклад обчислення медіани ряду даних урожайності соняшнику з використанням табличного процесора за формулою:

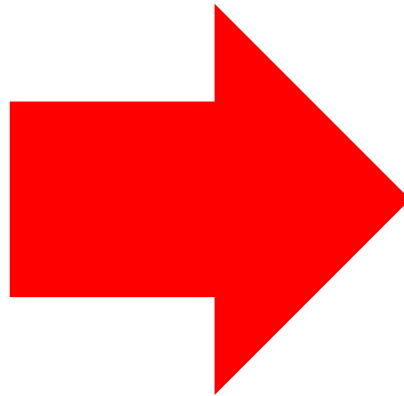
`=MEDIAN(C3:C12)`

	A	B	C	D	E
1					
2		Рік	Урожайність в т/га		
3		2006	1,34		
4		2007	1,16		
5		2008	1,52		
6		2009	1,5		
7		2010	1,59		
8		2011	1,66		
9		2012	1,65		
10		2013	2,17		
11		2014	1,95		
12		2015	2,16		
13		Середнє	1,67		
14		Медіана	1,62		



Звертаємо вашу увагу, що в електронній таблиці для знаходження медіани ряд даних не обов'язково має бути впорядкований. Табличний процесор:

***спочатку
впорядковує ряд
даних***



***а потім визначає
його медіану***



Зазначимо, що коли члени ряду даних незначно відрізняються одне від одного, то і середнє арифметичне, і медіана більш показово характеризують весь цей ряд. А якщо члени ряду даних значно відрізняються одне від одного, то медіана більш показово характеризує весь цей ряд даних, ніж середнє арифметичне.

x1	2,5	4,7
x2	2,8	6,2
x3	2,3	5,1
x4	2,55	12,4
x5	2,47	14,1
Середнє	2,524	8,5
Стандартне відхилення	0,161567	3,946137
Медіана	2,5	9,3



- 1. Що вивчає наука статистика? Наведіть приклади.**
- 2. Яким вимогам має відповідати вибірка об'єктів дослідження? Поясніть кожну з них.**
- 3. Що таке ряд даних? Наведіть приклади.**
- 4. Що характеризує середнє арифметичне ряду даних?**
- 5. У яких випадках середнє арифметичне більш показово характеризує весь ряд даних, а у яких — менш показово?**





6. **Що таке стандартне відхилення ряду даних? Що воно характеризує?**
7. **Що таке мода ряду даних? Що вона характеризує?**
8. **Що таке медіана ряду даних? Що вона характеризує?**
9. **У впорядкованому ряді чисел найбільше число збільшили на 5. Як зміняться значення середнього арифметичного, моди і медіани цього ряду? Відповідь поясніть.**



Розділ 2
§ 2.2

Домашнє завдання



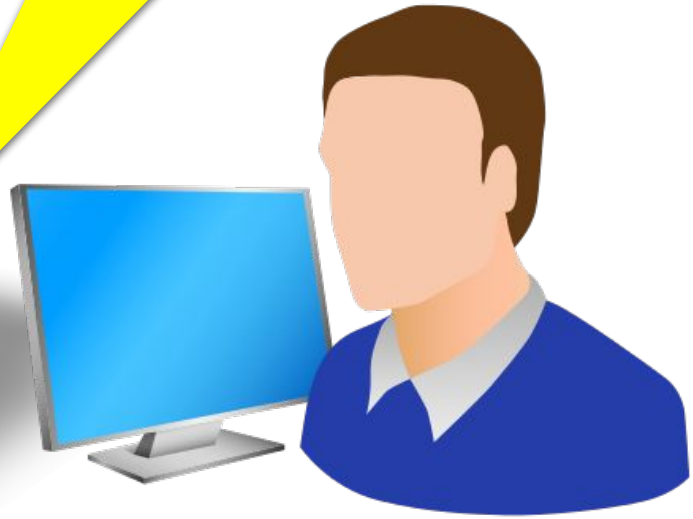
Проаналізувати
§ 2.2, ст. 42-48

Працюємо за комп'ютером

Розділ 2
§ 2.2



**Сторінка
47-48**



ІНФОРМАТИКА

Дякую за увагу!

10

За навчальною програмою 2018 року



Урок 8