

Обчислення статистичних характеристик

Виконайте

Практичні завдання.

Вправа 1. Порівняння двох вибірок

Вправа 2. Ранжування вибірки.

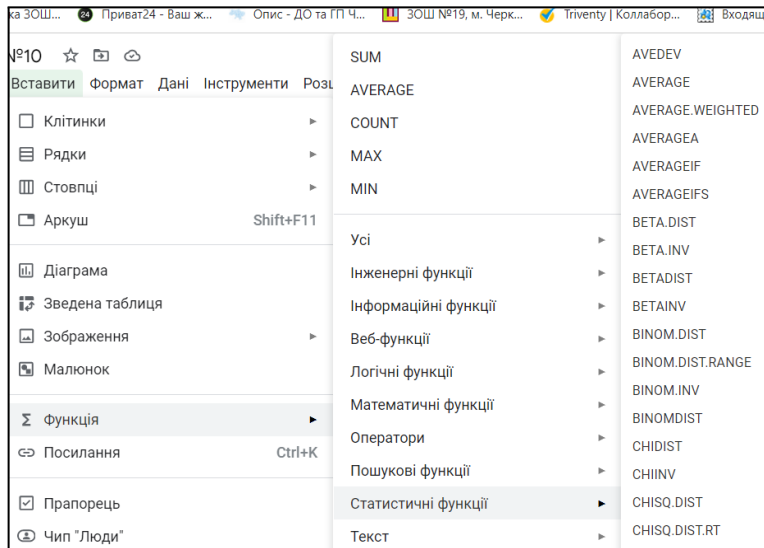
Вправа 3.

Вправа 4.

Табличний процесор Microsoft Excel / Google-таблиці оснащений засобами статистичної обробки даних, хоча й суттєво поступається спеціалізованим статистичним програмним пакетам. У табличному процесорі ці засоби представлені функціями, зібраними в категорії Статистичні, та майже всі вони призначені для обчислення узагальнюючих статистичних характеристик вибірки.

Статистичні функції, що розглядаються нижче (крім функції РАНГ), можуть мати до 30 аргументів, які повинні бути числами, масивами або посиланнями на діапазони клітинок, що містять числа. Якщо до діапазону-аргументу функції входять клітинки, які містять текст чи логічні значення або є порожніми, то вони ігноруються; але клітинки, що містять нульові значення, враховуються. Коли потрібно обчислити певну статистичну характеристику вибірки, діапазон, що містить елементи вибірки, слід зробити аргументом функції.

I) В Google-таблиці статистичні функції: Вставити -> Функції-> Статистичні -> обрати потрібну функцію



II) На панелі інструментів натиснути знак Σ -> Статистичні -> обрати потрібну функцію

III) Вручну ввести потрібну функцію, для цього поставити в потрібній клітинці знак =, далі ввести назву функції, а в дужках потрібні аргументи, наприклад

=AVERAGE(B3:B9)

Ознайомимось із кількома найважливішими функціями категорії Статистичні:

№	Функції кирилиці	Функції латинськими літерами	Призначення функції
1	=СРЗНАЧ(...)	=AVERAGE(...)	обчислює середнє значення;
2	=МАКС(..)	=MAX(..)	обчислює найбільше значення;
3	=МИН(..)	=MIN(..)	обчислює найменше значення;
4	=МЕДИАНА(..)	=MEDIAN(..)	повертає медіану (значення, що знаходиться посередині впорядкованої вибірки);
5	=МОДА(..)	=MODE(..)	повертає моду (значення, що зустрічається найчастіше);

6	=ДИСП(..)	=VAR(..)	обчислює дисперсію (найбільш споживана міра розсіювання, тобто відхилення від середнього);
7	=СТАНДОТКЛОН(..)	=STDEV(..)	обчислює стандартне відхилення;
8	=СКОС(..)	=SKEW(..)	повертає асиметрію розподілу;
9	=РАНГ(число; посилання; порядок)	=RANK(число; посилання; порядок)	повертає ранг числа у списку чисел, тобто його номер у впорядкованій послідовності чисел із вказаного діапазону.
10	=МАКС(..)-МИН(..)	=MAX(..)-MIN(..)	Розмах: різниця між найбільшим і найменшим значенням

• **РАНГ**(число; посилання; порядок)/ **=RANK**(число; посилання; порядок) — повертає ранг числа у списку чисел, тобто його номер у впорядкованій послідовності чисел із вказаного діапазону. Функція має такі аргументи: **число** — це число, для якого визначається ранг; **посилання** — це масив або посилання на список чисел; **порядок** — аргумент, який визначає спосіб упорядкування. Якщо цей аргумент відсутній або дорівнює нулю, то найбільше число має ранг 1. Якщо цей аргумент дорівнює будь-якому ненульовому числу, то ранг 1 має найменше число.

ПРИМІТКА. Функція **РАНГ** призначає повторюваним числам однаковий ранг. Проте наявність повторюваних чисел впливає на ранги наступних чисел. Наприклад, якщо у списку цілих чисел, відсортованих за зростанням, двічі трапляється число 10 із рангом 5, число 11 матиме ранг 7 і жодне з чисел не матиме рангу 6.

Практичні завдання.

Вправа 1. Порівняння двох вибірок

Туристична фірма, що реалізує путівки у семи містах країни, провела рекламну кампанію. У таблиці (рис. 9.1) наведено дані про кількість путівок, проданих протягом тижня до проведення рекламної кампанії та протягом тижня після неї. Потрібно, обчисливши

характеристики центру та варіації розподілу, порівняти дві вибірки та зробити висновки щодо впливу рекламної кампанії на обсяги продажу путівок.

- Відкрийте лист **Вправа1**

	А	В	С
1		Кількість проданих путівок	
2		З рекламою	Без реклами
3		162	135
4		156	126
5		144	115
6		137	140
7		125	121
8		145	112
9		151	130
10			
11	Середнє		
12	Медіана		
13	Розмах		
14	Станд. відхилення		
15	Асиметрія		

Рис. 9.1. Дані про діяльність турагенства.

• Такий показник, як мода, обчислювати недоцільно, оскільки відомості про те, яка кількість проданих путівок траплялася найчастіше, навряд чи можна якось використати. Тож обчисліть середнє значення для обох вибірок.

а) **СЕРЕДНЄ**: Скориставшись кнопкою Вставка функції, введіть у клітинку **B11** формулу **=AVERAGE(B3:B9)**. У цій клітинці буде обчислено середнє значення для стовпця **З рекламою**; воно дорівнює 145,71.

Скопіюйте формулу з клітинки **B11** у клітинку **C11**. У ній буде обчислено середнє значення для стовпця **Без реклами**.

Аналогічним чином у діапазоні **B12:C15** обчисліть для обох вибірок медіану, розмах, стандартне відхилення та асиметрію.

б) **МЕДІАНА**: **=MEDIAN(B3:B9)**

в) **РОЗМАХ**: Для обчислення розмаху скористайтеся функціями **MIN** та **MAX**, оскільки розмах — це різниця найбільшого та найменшого елемента вибірки **=MAX(B3:B9)-MIN(B3:B9)**.

г) Стандартне відхилення: **=STDEV(B3:B9)**

д) **АСИМЕТРІЯ**: **=SKEW(B3:B9)**

Ви маєте отримати такі результати, як на рис. 9.2.

	A	B	C
1		Кількість проданих путівок	
2		З рекламою	Без реклами
3		162	135
4		156	126
5		144	115
6		137	140
7		125	121
8		145	112
9		151	130
10			
11	Середнє	145,7142857	125,5714286
12	Медіана	145	126
13	Розмах	37	25
14	Станд. Відхилення	12,29788987	10,27711281
15	Асиметрія	-0,501821668	0,032744545
16			

дані числа не можна вводити вручну, їх ви отримаєте, якщо використаєте відповідні формули

Зробіть висновки.

• Середнє значення та медіана вибірки **З рекламою** , що свідчить про після рекламної кампанії собою, стали менш однорідними та передбачуваними.
Інакше кажучи, незважаючи на загальний позитивний вплив рекламної кампанії, у деяких містах вона фактично провалилася.

Вправа 2. Ранжування вибірки.

У таблиці, наведеній на рис. 9.3, подано результати змагань групи студентів з бігу на 100 м. Потрібно ранжувати студентів за цими результатами, присвоївши номер 1 студенту з найкращим результатом, номер 2 — студенту з другим результатом і т. д. Також необхідно визначити, до яких результатів — високих чи низьких — тяжіє основна маса студентів.

- Відкрийте лист Вправа2

	А	В	С
1	Результати змагань з бігу		
2	Студент	Результат, с	Місце
3	Курило Тетяна	12,8	
4	Поданчук Дмитро	13,2	
5	Надеев Дмитро	13	
6	Харів Ольга	12,9	
7	Завгородній Дмитро	13,5	
8	Рачинський Руслан	13	
9			

Рис. 9.3. Таблиця з вихідними даними

- Для ранжування студентів скористайтесь функцією **РАНГ**, яку потрібно ввести у клітинку **С3** і скопіювати у діапазон **С4:С8**. Формула у клітинці **С3** має бути такою: **=RANK(B3;B\$3:B\$8;1)**. Клітинка **С3** містить число, яке ранжується, **B\$3:B\$8** — діапазон усіх результатів. Зверніть увагу, що номери стовпців у діапазоні результатів фіксовані, а посилання на клітинку, ранг якої обчислюється, відносне. Тому після копіювання цієї формули значення, ранг якого обчислюється, змінюватиметься, а весь діапазон значень залишатиметься тим самим. Місця студентів у змаганнях з бігу мають відобразитися у стовпці С.

- Щоб визначити, до яких значень — високих чи низьких — тяжіють результати основної маси студентів, варто обчислити асиметрію вибірки. Отже, уведіть у будь-яку клітинку поза межами таблиці формулу **=SKEW(B3:B8)**, і буде обчислено асиметрію розподілу результатів змагань з бігу. Вона дорівнює приблизно **—1,1**. Знайдіть найкращий результат (це студент з рангом 1), він дорівнює **12,8 с**. Зменште його до **10 с**, і асиметрія стане меншою, оскільки найкращий результат став ще рідкіснішим.

Вправа 3.

За таблицею з відомостями про кількість опадів, яку складали кожного місяця протягом трьох років (див. рис.), потрібно визначити та проаналізувати показники центру та варіації розподілу опадів за роками і місяцями.

Хід виконання

Відкрийте лист Вправа3.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Дані про кількість опадів																
2	Місяць	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Середня кількість опадів	Стандартне відхилення	Асиметрія	
3	Рік																
4	2009	37,2	11,4	16,5	19,5	11,7	129	57,1	43,8	85,7	86	12,5	21,2				
5	2010	34,5	51,3	20,5	26,9	45,5	71,5	153	96,6	74,8	14,5	21	22,3				
6	2011	8	1,2	3,8	11,9	66,3	60	50,6	14,5	79,9	74,9	56,6	9,4				
7	Середнє за місяць																
8	Ранг місяця за вологістю																
9	Ранг місяця за засушливістю																
10																	
11																	
12																	
13																	

• Для знаходження **середньої кількості опадів за кожний місяць** у клітинку **B6** уведіть формулу обчислення **середнього значення** за діапазоном **B3:B5**.

• Скопіюйте формулу у клітинки діапазону **C6:M6**.

• Для знаходження **рангів місяців** за вологістю у клітинку **B7** уведіть формулу для обчислення **рангу** клітинки **B6** у діапазоні **B6:M6** за спаданням кількості опадів. Скопіюйте формулу у клітинки діапазону **C7:M7**. **Зверніть увагу**, що посилання на діапазон має бути абсолютним, а порядок ранжування (за спаданням) визначає третій аргумент функції **RANK**.

• Аналогічним чином у клітинках **B8:M8** обчисліть **ранг місяців за засушливістю**, тобто за зростанням кількості опадів.

• У клітинках **N3:P5** обчисліть **середньомісячну кількість опадів** для кожного з трьох (2002—2004) років, **стандартне відхилення** та **асиметрію вибірки** щомісячних опадів для кожного року (усі ці функції використовувалися у вправі 1 урок 56). Формули потрібно вводити тільки у клітинки **M3:P3**, а в діапазон **N4:P5** їх можна скопіювати.

Результат

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Дані про кількість опадів															
2	Місяць	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Середня кількість опадів	Стандартне відхилення	Асиметрія
3	Рік															
4	2009	37,2	11,4	16,5	19,5	11,7	129	57,1	43,8	85,7	86	12,5	21,2	44,30	38,04	1,192
5	2010	34,5	51,3	20,5	26,9	45,5	71,5	153	96,6	74,8	14,5	21	22,3	52,70	40,79	1,508
6	2011	8	1,2	3,8	11,9	66,3	60	50,6	14,5	79,9	74,9	56,6	9,4	36,43	30,68	0,164
7	Середнє за місяць	26,6	21,3	13,6	19,4	41,2	86,8	86,9	51,6	80,1	58,5	30,0	17,6			
8	Ранг місяця за вологістю	8	9	12	10	6	2	1	5	3	4	7	11			
9	Ранг місяця за засушливістю	5	4	1	3	7	11	12	8	10	9	6	2			

За результатами аналізу самостійно зробіть такі висновки та внесіть на цьому ж аркуші нижче:

- Які місяці є найбільш та найменш засушливими?
- Чи можна вважати розподіл кількості опадів за місяцями рівномірним? Який показник дозволяє дати відповідь на це питання?
- Які місяці — найбільш чи найменш засушливі — є рідкіснішими у цій вибірці?

Вправа 4.

За даними про щомісячні виторги двох магазинів протягом року потрібно визначити:

1. який магазин має більші середньомісячні виторги;
2. який магазин працює стабільніше;
3. які середньомісячні виторги — великі чи маленькі — є більш типовими для кожного з магазинів.

Крім того, потрібно ранжувати місяці за сумарним виторгом обох магазинів.

Хід виконання

Відкрийте лист **Вправа4**.

	A	B	C
1	Щомісячні виторги магазинів		
2	Місяць	Виторг магазину 1	Виторг магазину 2
3	січень	12,36	9
4	лютий	12	11
5	березень	11,23	25,3
6	квітень	16	6
7	травень	12,3	24,5
8	червень	18,5	16,6
9	липень	13	11,4
10	серпень	15	16,5
11	вересень	12,23	11,5
12	жовтень	15	12
13	листопад	16	13,5
14	грудень	16	12,22
15			

- Уведіть у клітинки **B15:B17** формули, що дозволять визначити **середньомісячні виторги, стабільність роботи та типовість великих чи маленьких виторгів для магазину 1**. Визначте **самостійно**, якими статистичними функціями для цього краще скористатися.

- Скопіюйте введені формули в клітинки **C15:C17**, щоб обчислити статистичні показники для другого магазину.

- **Зробіть висновки**, давши відповіді на три питання(на цьому ж аркуші, нижче таблиці), сформульовані в умові практичної роботи.

- У клітинках **D3:D14** обчисліть **сумарні виторги двох магазинів** за кожен місяць (уведіть формулу у клітинку D3 і скопіюйте її у клітинки діапазону D4:D14).

- Скориставшись **функцією РАНГ**, у діапазоні **E3:E14** обчисліть ранг місяців за сумарним виторгом (на першому місці має розташовуватися місяць з найбільшим виторгом). **Зверніть увагу** на те, що посилання на діапазон має бути **абсолютним**, а порядок ранжування (за спаданням) визначає третій аргумент функції ранг.