

Excel

Аналіз вибірових даних

Мета лабораторного заняття: Ознайомити студентів з методами розрахунку математичного сподівання, дисперсії і будови частотних гістограм.

Необхідне обладнання та матеріали:

ПЕОМ із встановленою ОС Microsoft Windows, програмне забезпечення MS Excel, Інтернет браузер, методичні вказівки

Питання для актуалізації знань і перевірки готовності студентів до виконання конкретної лабораторної роботи.

1. Структура, призначення і особливості програмного додатку Excel/
2. Визначення, значення і формула математичного сподівання.
3. Визначення, значення і формула дисперсії.
4. Визначення гістограм, види гістограм.

Короткі теоретичні відомості.

Microsoft Excel (повна назва Microsoft Office Excel) — табличний процесор, програма для роботи з електронними таблицями. Програма входить до складу офісного пакета Microsoft Office. Лист Excel являє собою готову таблицю, тому Excel часто використовують для створення документів, що мають табличне представлення, крім того, створюють різні види графіків і діаграм, які беруть дані для побудови з комірок таблиць. Excel містить багато математичних і статистичних функцій, завдяки чому його використовують для елементарних розрахунків. Excel може працювати як база даних, хоча, звичайно, до повноцінної бази даних йому далеко.

Microsoft Excel (повна назва Microsoft Office Excel) — табличний процесор, програма для роботи з електронними таблицями. Програма входить до складу офісного пакета Microsoft Office (рис. 43).

Типові області застосування Excel:

- завдяки тому, що лист Excel являє собою готову таблицю, Excel часто використовують для створення документів без усіляких розрахунків, що просто мають табличне представлення (наприклад, прайсь листи в магазинах, розклади);
- у Excel легко можна створювати різні види графіків і діаграм, які беруть дані для побудови з комірок таблиць (графік зниження ваги тіла за вказаний період від початку занять спортом);
- його можуть використовувати звичайні користувачі для елементарних розрахунків (скільки витратив за цей місяць, що/кому/коли дав/взяв);
- Excel містить багато математичних і статистичних функцій, завдяки чому його можуть використовувати школярі і студенти для розрахунків курсових, лабораторних робіт;
- Excel інтенсивне використовується в бухгалтерії — у багатьох фірмах це

основний інструмент для оформлення документів, розрахунків і створення діаграм. Природно, він має в собі відповідні функції;

- Excel може навіть працювати як база даних. Хоча, звичайно, до повноцінної бази даних йому далеко.

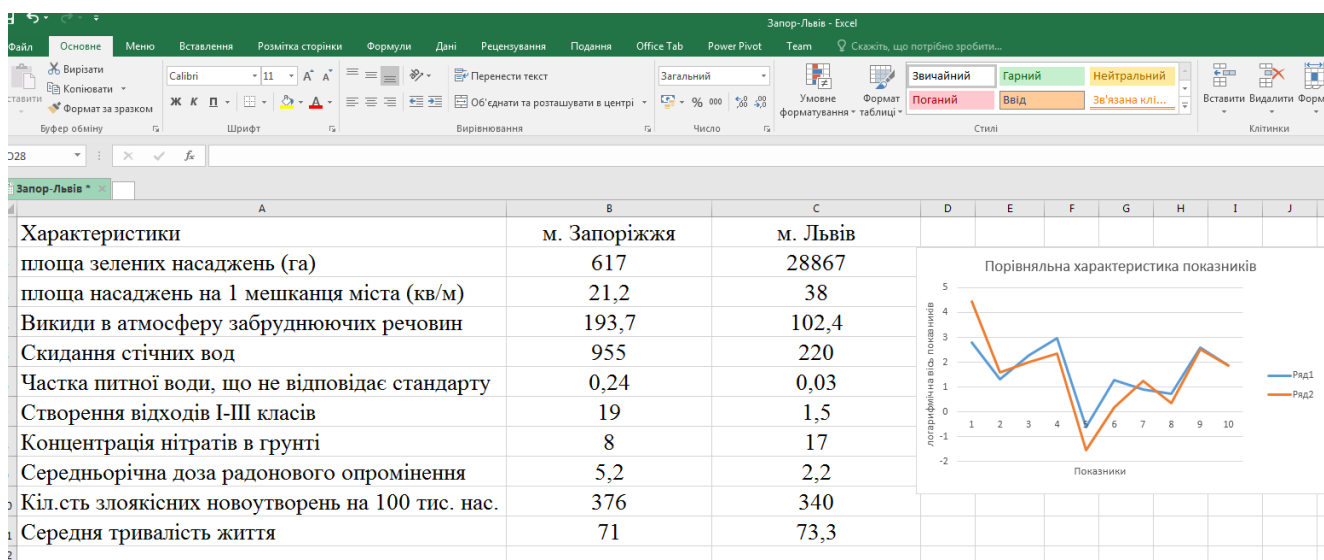


Рис. 43 – Приклад аркуша Excel

Excel — програмований табличний калькулятор. Всі розрахунки в Excel виконують формули. Excel вважає формулою все, що починається із знаку "=". Якщо в комірці написати просто «1 +1», то Excel не буде обчислювати цей вираз. Однак, якщо написати «=1 +1» і натиснути клавішу Enter, в комірці з'явиться результат обчислення виразу — число 2. Після натискання клавіші Enter формула не пропадає, її можна побачити в панелі інструментів «Рядок формул».

У формулі можна використовувати різні типи операторів (арифметичні і таке інше), текст, посилання на комірку або діапазон комірок, круглі дужки, іменовані діапазони. Природно, в формулах дотримується пріоритет виконання операцій (множення виконується раніше додавання і таке інше). Для зміни порядку виконання операцій використовуються круглі дужки.

При написанні складних формул, особливо використовують вкладені функції, використання майстра функцій — найкраще рішення. Він дуже полегшує і прискорює введення формул, і робить багато речей за нас: автоматично вставляє знак «дорівнює», ім'я функції, круглі дужки, розставляє крапки з комою. Дозволяє переглядати значення посилань і результати проміжних обчислень, так наприклад:

- у вікні майстра функцій, виділяючи посилання і натискаючи «F4» один, або кілька разів, можна поміняти тип посилання (зробити абсолютної або змішаної);
- якщо у вікні майстра функцій натиснути «F3», відкриється вікно для вставки іменованих посилань (якщо вони існують);
- якщо якусь функцію у формулі потрібно замінити на іншу, виділяємо цю

функцію в рядку формул і вставляємо потрібну функцію. Виділена функція автоматично вставляється;

- клавіша «Tab» служить для перемикання на наступний аргумент (текстове поле), а поєднання «Shift + Tab» – на попереднє;
- якщо виділити комірку, що містить формули з функціями і натиснути кнопку F_x , то відкриється вікно майстра функцій для однієї з функцій у формулі;
- для того, щоб написати формулу типу «= СУММ (J1: J3) / СУММ (K1: K3)», відкрийте майстер функцій, виберете функцію «СУМ», виділіть потрібний діапазон, потім клацнете мишкою в рядок формул і в ній вручну наберіть « / », потім натисніть на кнопку для вставки функції, і вставте другу функцію « СУМ »;
- якщо в текстовому полі потрібно ввести ТІЛЬКИ текст, не обов'язково вручну ставити подвійні лапки. Можна написати текст без лапок і натиснути Tab, або просто клацнути в інше поле. Excel проставить лапки автоматом (працює не у всіх текстових полях);
- у тих полях, де потрібно ввести логічне значення «БРЕХНЯ» або «ІСТИНА», досить ввести «0» для «БРЕХНЯ», і будь-який не нульове значення для «ІСТИНА» (прийнято використовувати 1).

Математичне сподівання це одна з основних числових характеристик кожної випадкової величини. В теорії ймовірностей, математичне сподівання випадкової величини є середнім значенням при довгостроковому повторенні одного і того ж експерименту. Закон великих чисел стверджує, що середнє арифметичне всіх значень майже певно збігається до математичного сподівання, із тим як кількість повторів даного експерименту прямує до нескінченності. Математичне сподівання також іноді називають сподіванням, середнім, середнім значенням. У більш практичному розумінні, математичне сподівання дискретної випадкової величини є середнім зваженим по імовірності для всіх можливих значень. Іншими словами, кожне можливе значення випадкової величини фактично є помножене на його імовірність виникнення, і отриманий добуток складається у загальну суму, яка утворює математичне сподівання. Математичного сподівання визначають за формулою:

$$M = \sum_1^n x_i p_i.$$

Дисперсія це міра розсіяння значень випадкової величини відносно середнього значення розподілу. Більші значення дисперсії свідчать про більші відхилення значень випадкової величини від центру розподілу.

Дисперсія випадкової величини – це один з параметрів розподілу ймовірностей – це середньоквадратичне відхилення від середнього значення. Інакше кажучи, це математичне сподівання квадрату відхилення цієї змінної від її очікуваного значення (її математичного сподівання). Отже дисперсія є

вимірюванням величини розпорошеності значень цієї змінної, беручи до уваги всі її значення і їхні ймовірності. Формула розрахунку дисперсії:

$$D = \sum_1^n x_i^2 p_i - (M)^2.$$

У простому розумінні, дисперсія дозволяє виміряти наскільки далеко випадкові значення розподілені від їх середнього значення. Дисперсія відіграє важливу роль в статистиці.

Гістограма – спосіб графічного представлення табличних даних, приблизне представлення розподілу числових даних. Являє собою діаграму, що складається з прямокутників без розривів між ними. Кількісні співвідношення деякого показника представлені у вигляді прямокутників, площі яких пропорційні. Найчастіше для зручності сприйняття ширину прямокутників беруть однакою, при цьому їх висота визначає співвідношення відображуваного параметра.

Хід виконання лабораторної роботи (методика експериментальних досліджень).

1. Згенерувати вибірку одним із способів які показано нижчі.
2. Записати частотний ряд.
3. Створити формулу в Excel математичного сподівання та зробити розрахунок.
4. Створити в Excel формулу дисперсії та зробити розрахунок.
5. За отриманими даними побудувати частотну діаграму.
6. Використовуючи описову статистику зробити порівняльний аналіз отриманих результатів.

Щоб виконати завдання а саме, скласти варіаційний ряд та побудувати полігон розподілу у MS Excel, де складання варіаційних рядів та побудова графіків розподілу ґрунтується на застосуванні лише трьох функцій: МІН(), МАКС(), які визначають найменше та найбільше значення вибірових даних, та функції ЧАСТОТА(), яка визначає частоту зустрічальності кожної з варіант, треба виконати наступні дії:

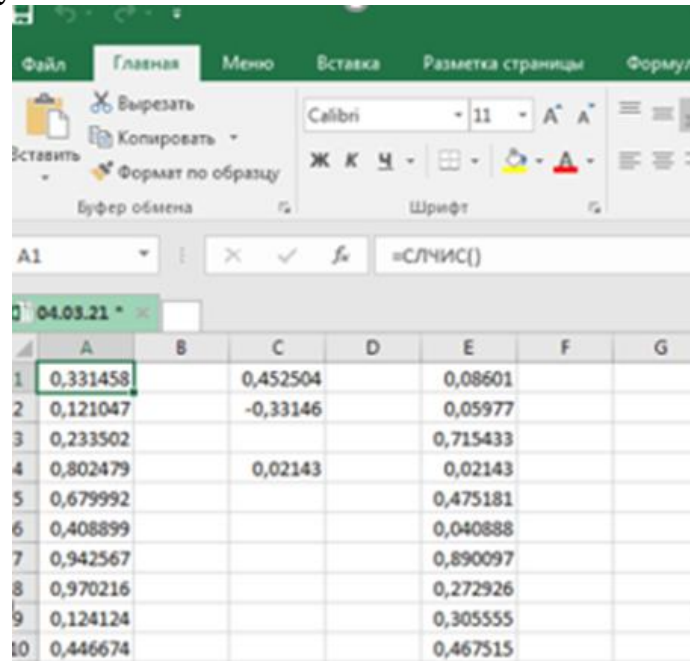
- 1) скласти варіаційний ряд;
- 2) сортувати дані за збільшенням;
- 3) визначити частоти зустрічальності певної величини/ознаки (побудова варіаційного ряду частот);
- 4) побудувати та налаштувати графічне зображення.

Приклади виконання завдання в Excel.

1. Згенерувати вибірку одним із способів:

- 1) На новий лист Excel, як показано на рисунку 10, виділити комірку A1 і в рядку для формул ввести «СЛЧИС» подвійним натисканням на праву кнопку мишки і закрити дужку натиснувши одночасно кнопку «Shift» і «9». В комірки

з'явиться число в зеленій рамці, підвести курсор до нижнього правого кута рамки, курсор набирає вигляду чорного хрестика і не відпускаючи затиснувши правою кнопкою мишки протягнути на ту кількість рядків на яку треба отримати вибірку.



	A	B	C	D	E	F	G
1	0,331458		0,452504		0,08601		
2	0,121047		-0,33146		0,05977		
3	0,233502				0,715433		
4	0,802479		0,02143		0,02143		
5	0,679992				0,475181		
6	0,408899				0,040888		
7	0,942567				0,890097		
8	0,970216				0,272926		
9	0,124124				0,305555		
10	0,446674				0,467515		

Рис. 44 – Приклад генерування вибірки у Excel

- 2) Закріпити генеровані цифри: виділити весь діапазон, копіювати і вставити через спеціальну вставку "Вставка значень" бажано на інший лист.
- 3) Округлити дані вибірки використовуючи `=ОКРУГЛ(A1;2)`, як показано на рисунку 45, далі копіюємо і вставляємо через спеціальну вставку для значень.

	A	B	C	D	E	F
1	0,008503	0,010000				
2	0,02143	0,020000				
3	0,040888	0,040000				
4	0,05977	0,060000				
5	0,064496	0,060000				
6	0,08601	0,090000				
7	0,272926	0,270000				
8	0,305555	0,310000				
9	0,368218	0,370000				
10	0,448468	0,450000				

Рис. 45 – Приклад генерування вибірки у Excel (округлення)

- 4) В стовпчик А вставте значення вашої вибірки, а у стовпчик В число зустрічаємості даного числа (використовуючи функцію «Счетесли», або вручну записати формулу «=счетесли(\$A\$1:\$A\$37;A1)».
- 5) В стовпчик С у першу комірку вставити знак «=» активізувати комірку В1, вставити знак поділу «/» загальну кількість значень (n), або в стрічки формул ввести формулу =C2/\$A\$2.
- 6) В стовпчике D вставити «= A1 * C1», далі комірку D1 активізувати та отримавши тонкий чорний хрестик потягнути на всі строчки стовпчика.
- 7) Розрахувати суму значень стовпчика D1 відомим вам способом.

Увага! Не забувайте робити перевірку:

Сума ймовірності повинна дорівнювати **1**

Далі працюємо з даними. Перший спосіб генерую дані від 0 до 1, другий спосіб генерує значення від мінімального вказаного числа до максимального вказаного числа.

8) На новий лист Excel виділити комірку A1 і в рядку для формул ввести «=СЛУЧМЕЖДУ(11;53)» подвійним натисканням на праву кнопку мишки і закрити дужку натиснувши одночасно кнопку «Shift» і «9». В комірки з'явиться число в зеленій рамці, як показано на рисунку 46, підвести курсор до нижнього правого кута рамки, курсор набирає вигляду чорного хрестика і не відпускаючи затиснувши правою кнопкою мишки протягнути на ту кількість рядків на яку треба отримати вибірку.

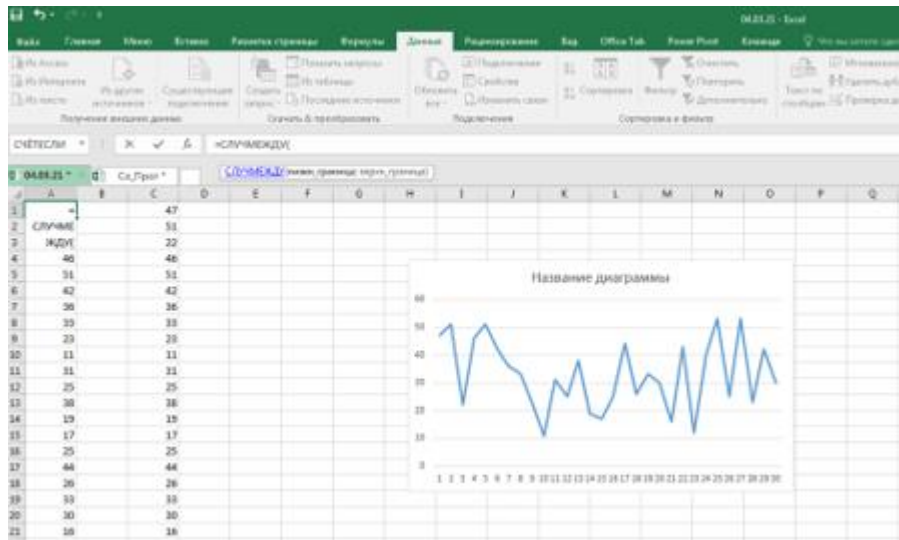


Рис. 46 – Приклад генерування вибірки у Excel різного діапазону

- 9) Далі повторюємо пункти 4 і 5 та отримуємо вибірні значення та частотний ряд.
- 10) Видалити значення які повторюються, використовуючи вкладку «Дані» та «Видалити дублі», як показано на рисунку 47.
2. Записати частотний ряд.
 - 1) Вибірку записати у ранжирований ряд (використовуючи вкладку «Головна»→ «Сортування і фільтр»).
 - 2) Записати частотний ряд, тобто яка кількість раз зустрічається величина, (використовуючи функцію «Счетесли», або вручну записати формулу =счетесли(\$A\$1:\$A\$37;A1). Визначити частоту зустрічаємості.

Файл Главная Меню Вставка Разметка страницы Формулы

Получение внешних данных: Из Access, Из Интернета, Из текста, Из других источников, Существующие подключения, Создать запрос

Скачать & преобразовать: Показать запросы, Из таблицы, Последние источники

Формула: =СЧЁТЕСЛИ(\$C\$1:\$C\$30;C1)

	A	B	C	D	E	F	G
1	0,008503	0,010000	0,01	1			
2	0,02143	0,020000	0,02	1			
3	0,040888	0,040000	0,04	1			
4	0,05977	0,060000	0,06	2			
5	0,064496	0,060000	0,06	2			
6	0,08601	0,090000	0,09	1			
7	0,272926	0,270000	0,27	1			
8	0,305555	0,310000	0,31	1			
9	0,368218	0,370000	0,37	1			
10	0,448468	0,450000	0,45	1			
11	0,467515	0,470000	0,47	1			
12	0,475181	0,480000	0,48	1			
13	0,533643	0,530000	0,53	2			
14	0,534165	0,530000	0,53	2			
15	0,537984	0,540000	0,54	1			
16	0,583941	0,580000	0,58	1			
17	0,611125	0,610000	0,61	1			
18	0,631552	0,630000	0,63	1			
19	0,685791	0,690000	0,69	1			
20	0,715433	0,720000	0,72	1			
21	0,750172	0,750000	0,75	1			
22	0,758852	0,760000	0,76	1			
23	0,76772	0,770000	0,77	1			
24	0,845018	0,850000	0,85	1			
25	0,870589	0,870000	0,87	1			
26	0,890097	0,890000	0,89	1			
27	0,920817	0,920000	0,92	1			
28	0,936532	0,940000	0,94	1			
29	0,973184	0,970000	0,97	1			
30	0,99196	0,990000	0,99	1			

Рис. 47 – Приклад генерування вибірки у Excel (повторюваність)

- 3) Видалити значення які повторюються, використовуючи вкладку «Дані», як показано на рисунку 48.

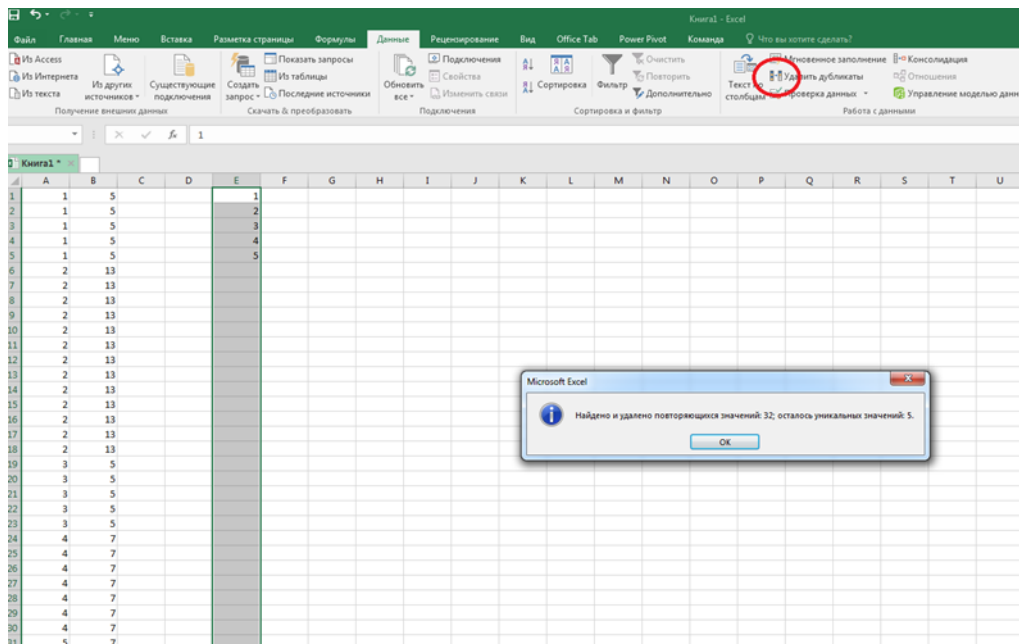


Рис.48 – Використовуючи вкладку «Дані» прибираємо повторення

3. Створити формулу в Excel математичного сподівання та зробити розрахунок.

Створення формул завжди починаються зі знаку дорівнює (=), за яким слідує константи, які є числами і операторами обчислень, такими як «плюс» (+), «мінус» (-), «зірочка» (*) множення або «коса риска» (/) ділення на початку.

Як приклад розглянемо формулу математичного сподівання:

$$M = \sum_1^n x_i p_i.$$

Виділіть на аркуші комірку, в яку необхідно ввести формулу.

Введіть = (знак рівності), а потім константи і оператори (не більше 8192 знаків), які потрібно використовувати при обчисленні.

4. Створити в Excel формулу дисперсії та зробити розрахунок.

Використовуючи формулу розрахунку дисперсії проводимо розрахунок, як показано на рисунку 49.

$$D = \sum_1^n x_i^2 p_i - (M)^2 .$$

Проводячи розрахунки не забувайте проводити перевірку.

Для використання вбудованих формул визначення математичного сподівання та дисперсії у програмному середовищі Excel необхідно перейти во вкладку «Данные», далі «Анализ данных», «Описательная статистика», «Ок».

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	n	Xi	Wi	Pi	M	M2	Xi2	Xi2*P			D	s	
2	30	11	1	0,033333	0,366667	0,134444	121	4,033333			1166,55	150,956667	
3		12	1	0,033333	0,4	0,16	144	4,8					
4		16	1	0,033333	0,533333	0,284444	256	8,533333					
5		17	1	0,033333	0,566667	0,321111	289	9,633333					
6		19	1	0,033333	0,633333	0,401111	361	12,033333				M=ΣXiPi	
7		22	1	0,033333	0,733333	0,537778	484	16,133333					
8		23	2	0,066667	1,533333	2,351111	529	35,266667				D=ΣX2P-M2	
9		25	3		0,1	2,5	6,25	62,5					
10		26	1	0,033333	0,866667	0,751111	676	22,533333					
11		30	2	0,066667	2	4	900	60					
12		31	1	0,033333	1,033333	1,067778	961	32,033333					
13		33	2	0,066667	2,2	4,84	1089	72,6					
14		36	1	0,033333	1,2	1,44	1296	43,2					
15		38	1	0,033333	1,266667	1,604444	1444	48,133333					
16		40	1	0,033333	1,333333	1,777778	1600	53,333333					
17		42	2	0,066667	2,8	7,84	1764	117,6					
18		43	1	0,033333	1,433333	2,054444	1849	61,633333					
19		44	1	0,033333	1,466667	2,151111	1936	64,533333					
20		46	1	0,033333	1,533333	2,351111	2116	70,533333					
21		47	1	0,033333	1,566667	2,454444	2209	73,633333					
22		51	2	0,066667	3,4	11,56	2601	173,4					
23		53	2	0,066667	3,533333	12,48444	2809	187,2667					
24			30	1	32,9	66,81667		1233,367					
25					1082,41								

Рис.49 – Приклад розрахунків математичного сподівання та дисперсії

5. За отриманими даними будуюмо гістограму даної вибірки (рис. 50) та частотну діаграму (рис.51).



Рис.50 – Приклад гістограми вибірки у програми Excel

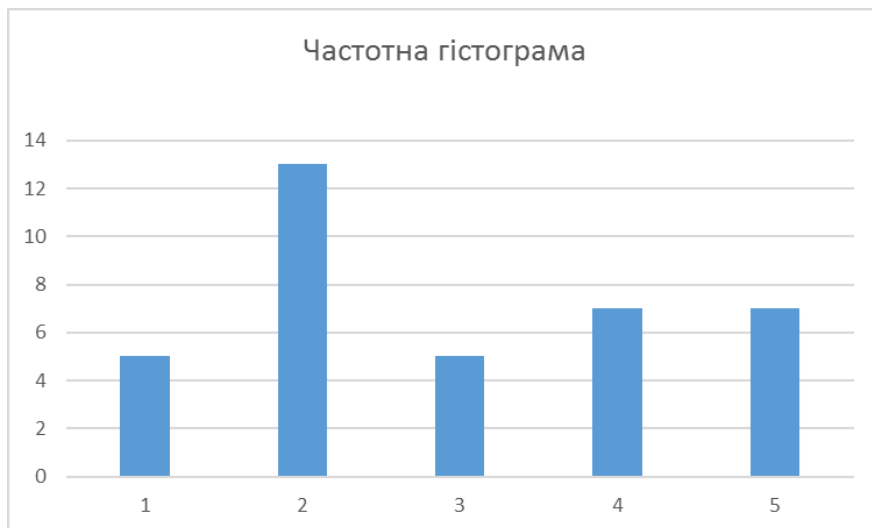


Рис.51 – Приклад частотної гістограми у програми Excel

6. Використовуючи описову статистику зробити порівняльних аналіз отриманих результатів.

Використовуючи формули розрахунку математичного сподівання та дисперсії робимо розрахунки, як показано на рисунку 52, праворуч показана загальна описова статистика, робимо висновки.

№	n	Xi	Wi	Рi	M	M2	Xi2	Xi2*P	D	s	Статистика	Значення
1	30	11	1	0,033333	0,366667	0,134444	121	4,033333	1166,55	150,956667	Середнє	32,9
2	12	1	0,033333	0,4	0,16	144	4,8				Стандартна помилка	2,281535075
3	16	1	0,033333	0,533333	0,284444	256	8,533333				Медіана	32
4	17	1	0,033333	0,566667	0,321111	289	9,633333				Мода	25
5	19	1	0,033333	0,633333	0,401111	361	12,03333				Стандартне відхилення	12,49648226
6	22	1	0,033333	0,733333	0,537778	484	16,13333				Дисперсія виборки	156,162069
7	23	2	0,066667	1,533333	2,351111	529	35,26667				Екцес	-1,069457082
8	25	3	0,1	2,5	6,25	625	62,5				Асиметричність	0,024870142
9	26	1	0,033333	0,866667	0,751111	676	22,53333				Інтервал	42
10	30	2	0,066667	2	4	900	60				Мінімум	11
11	31	1	0,033333	1,033333	1,067778	961	32,03333				Максимум	53
12	33	2	0,066667	2,2	4,84	1089	72,6				Сума	987
13	36	1	0,033333	1,2	1,44	1296	43,2				Счет	30
14	38	1	0,033333	1,266667	1,604444	1444	48,13333				Уровень надежности(95.0%)	4,666263165
15	40	1	0,033333	1,333333	1,777778	1600	53,33333					
16	42	2	0,066667	2,8	7,84	1764	117,6					
17	43	1	0,033333	1,433333	2,054444	1849	61,63333					
18	44	1	0,033333	1,466667	2,151111	1936	64,53333					
19	46	1	0,033333	1,533333	2,351111	2116	70,53333					
20	47	1	0,033333	1,566667	2,454444	2209	73,63333					
21	51	2	0,066667	3,4	11,56	2601	173,4					
22	53	2	0,066667	3,533333	12,48444	2809	187,2667					
23	30	1	0,033333	32,9	66,81667	1233,367						
24					1082,41							

Рис. 52 – Перевірка розрахунків у програми Excel

Вимоги до змісту та оформлення звіту за результатами виконаної роботи.

1. Титульний лист.
2. Виконати всі завдання за варіантами в СЕЗН Moodle
3. Оформити завдань з коментарями та описати отримані результати.
4. Висновки.