
ЛАБОРАТОРНЕ ЗАНЯТТЯ 3

Основні поняття математичної статистики. Первинне опрацювання статистичних даних

Мета: Навчитись визначати дискретний варіаційний ряд розподілу за означеннями та будувати графічне представлення дискретного ряду у *MS Excel*.

Згадаємо приклад з Лекції та проведемо розрахунки, використовуючи *MS Excel*.

Приклад. Проводяться спостереження над значеннями грошових виграшів у миттєвій лотереї. У результаті отримані наступні значення (у тис. грн.):

0, 1, 0, 0, 5, 0, 10, 0, 1, 0, 0, 1, 5, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 5, 0, 5, 0, 0, 1, 1, 1, 5,
10, 0, 1, 1, 0, 5, 0, 0, 0, 0, 1, 0, 1, 0, 5, 0, 0, 0, 0, 1, 0.

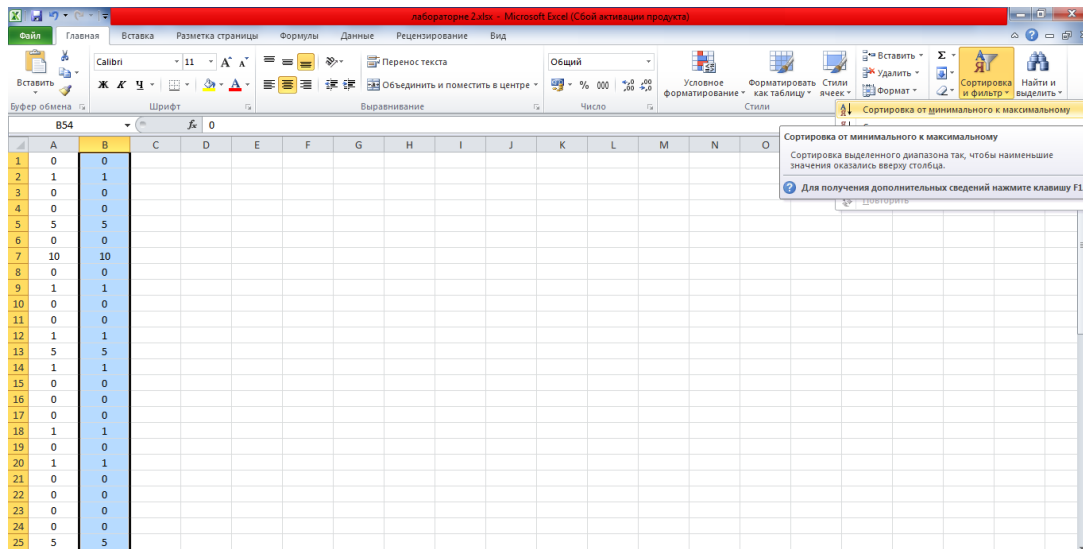
Проведемо первинне опрацювання статистичних даних. Для цього необхідно спочатку побудувати дискретний варіаційний ряд. Пригадаємо означення.

Означення. Дискретним варіаційним рядом розподілу (або дискретним статистичним розподілом, або розподілом частот) називається ранжована сукупність варіант x_i з відповідними частотами або відносними частотами.

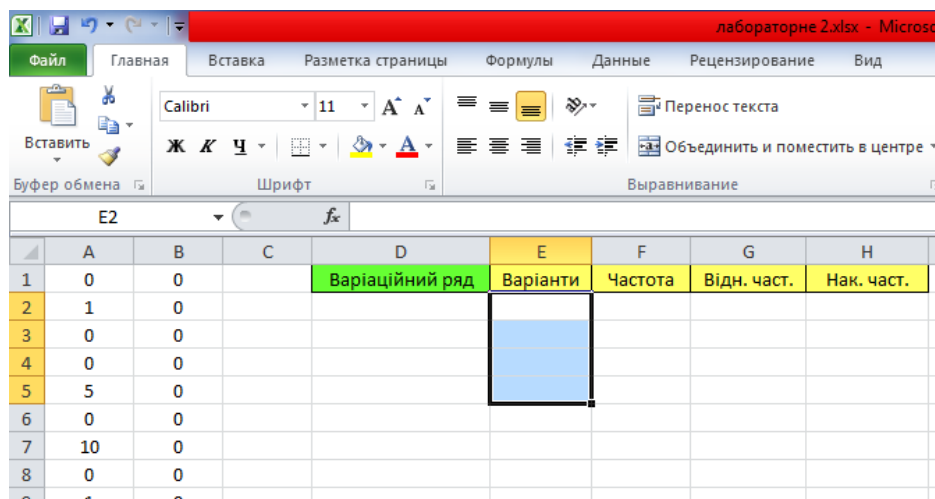
У табличній формі він має такий вигляд:

x_i	x_1	x_2	x_3	...	x_k
n_i	n_1	n_2	n_3	...	n_k
p_i^*	p_1^*	p_2^*	p_3^*	...	p_k^*

При опрацюванні статистичних даних в *MS Excel* зручніше заносити дані у стовпці. Запишемо дані вибірки у стовпець A1:A54 та використовуючи «Сортировка и фильтр» відсортуємо елементи від мінімального до максимального у стовпці B1:B54.



Побудуємо дискретний варіаційний ряд для заданої вибірки.



У стовпець E2:E5 внесемо значення варіант x_i : 0, 1, 5, 10.

	A	B	C	D	E	F	G	H
1	0	0		Варіацийний ряд	Варіанти	Частота	Відн. част.	Нак. част.
2	1	0			0			
3	0	0			1			
4	0	0			5			
5	5	0			10			
6	0	0						
7	10	0						

Заповнимо другий стовпець – підрахувати можна як вручну, так і використовуючи вбудовані функції. Можна просто виділити комірки, наприклад, з варіантою 5, і у нижньому правому куточку побачити кількість.

44	1	1
45	0	1
46	1	5
47	0	5
48	5	5
49	0	5
50	0	5
51	0	5
52	0	5
53	1	10
54	0	10



Також можна використати функцію «ЧАСТОТА» (розберемо особливості її застосування у наступній лабораторній роботі).

У комірці F6 за допомогою функції СУММ знайдемо об'єм вибірки n .

Варіацийний ряд	Варіанти	Частота	Відн. част.	Нак. част.
	0	31		
	1	14		
	5	7		
	10	2		
	=			

Мастер функций - шаг 1 из 2

Поиск функции:
 Введите краткое описание действия, которое нужно выполнить, и нажмите кнопку "Найти"
 Категория: 10 недавно использовавшихся

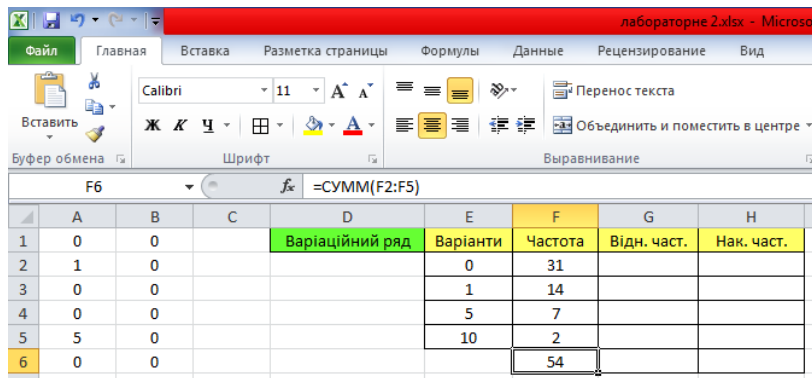
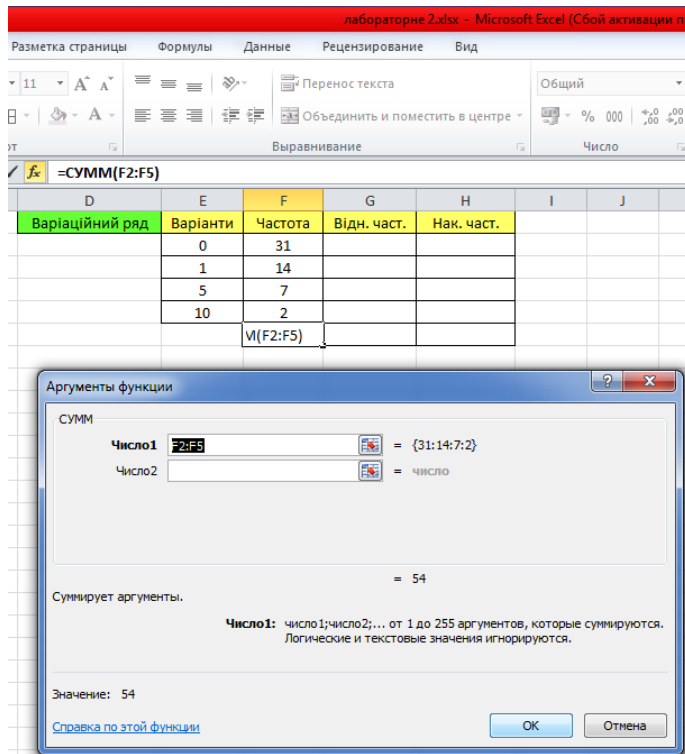
Выберите функцию:

- ЧИСЛКОМБ
- СУММ**
- КОРЕНЬ
- СУММПРОИЗВ
- ПЕРЕСТ
- ФАКТР
- КОРРЕЛ

СУММ(число1;число2;...)
 Суммирует аргументы.

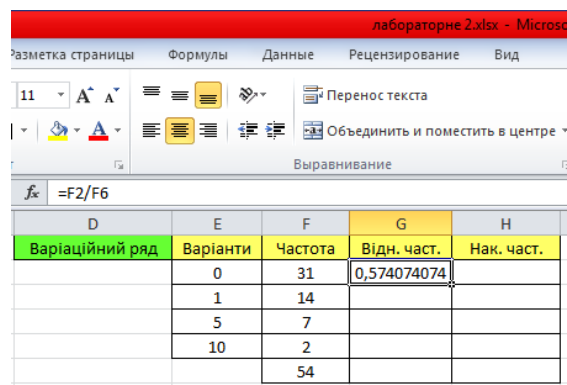
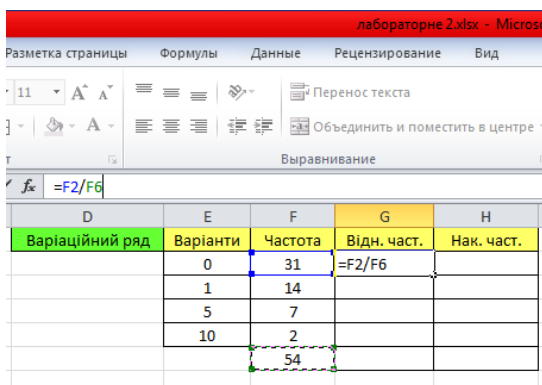
Справка по этой функции

OK Отмена



Заповнимо стовпець з відносними частотами – за означенням $p_i^* = \frac{n_i}{n}$, для

цього у комірці G2 задамо формулу =F2/F6.



Аналогічно обчислюємо всі відносні частоти.

лабораторне 2.xlsx - Microsoft Excel

Разметка страницы Формулы Данные Рецензирование Вид

11 А А Перенос текста

Объединить и поместить в центре

Выравнивание

Вариационный ряд	Варианты	Частота	Відн. част.	Нак. част.
	0	31	0,574074074	
	1	14	0,259259259	
	5	7	0,12962963	
	10	2	0,037037037	
		54		

Та змінимо формат комірок відносних та накопичувальних частот.

лабораторне 2.xlsx - Microsoft Excel (Сбой активации продукта)

Разметка страницы Формулы Данные Рецензирование Вид

11 А А Перенос текста

Объединить и поместить в центре

Число

Выравнивание

fx =F2/F6

Вариационный ряд	Варианты	Частота	Відн. част.	Нак. част.
	0	31	0,574074074	
	1	14	0,259259259	
	5	7	0,12962963	
	10	2	0,037037037	
		54		

Врезатель
Копировать
Параметры вставки:
Специальная вставка...
Вставить...
Удалить...
Очистить содержимое
Фильтр
Сортировка
Вставить примечание
Формат ячеек...
Выбрать из раскрывающегося списка...
Присвоить имя...
Гиперссылка...

лабораторне 2.xlsx - Microsoft Excel (Сбой активации продукта)

Разметка страницы Формулы Данные Рецензирование Вид

11 А А Перенос текста

Объединить и поместить в центре

Число Условное форматирование

Выравнивание

fx =F2/F6

Вариационный ряд	Варианты	Частота	Відн. част.	Нак. част.
	0	31	0,574074074	
	1	14	0,259259259	
	5	7	0,12962963	

Формат ячеек

Число Выравнивание Шрифт Граница Заливка Защита

Числовые форматы:

- Общий
- Числовой
- Денежный
- Финансовый
- Дата
- Время
- Процентный
- Дробный
- Экспоненциальный
- Текстовый
- Дополнительный (все форматы)

Образец: 0,574

Число десятичных знаков: 8

Разделитель групп разрядов ()

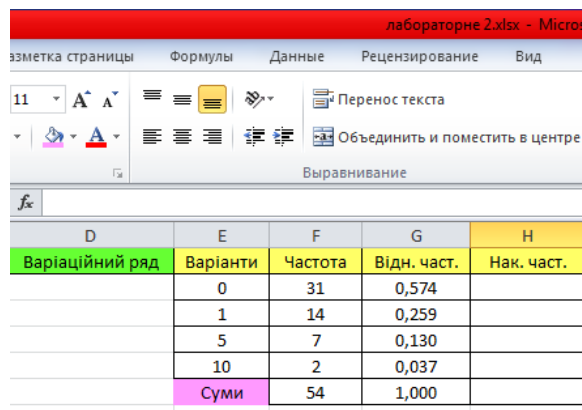
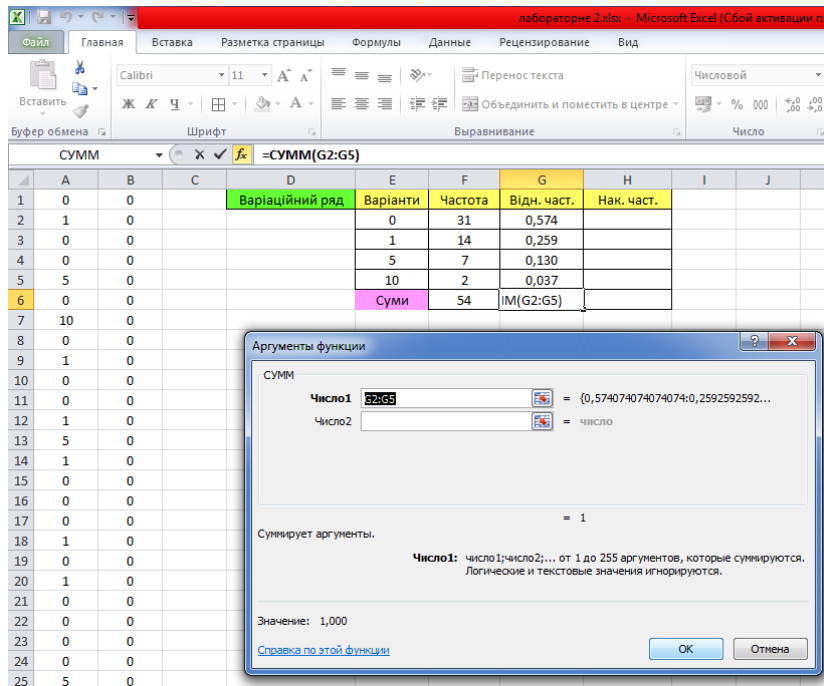
Отрицательные числа:

- 1234,210
- 1234,210
- 1234,210
- 1234,210

Числовой формат является наиболее общим способом представления чисел. Для вывода денежных значений используются также форматы "Денежный" и "Финансовый".

OK Отмена

І обчислимо суму відносних частот у комірці G6.



Заповнимо останній стовпець з накопичувальними частотами, який нам знадобиться для побудови емпіричної функції розподілу. Комірці H0 присвоїмо значення 0, наступні комірки будемо заповнювати за означенням емпіричної функції $F^*(x) = \frac{n_x}{n}$, де n_x – число вибірових значень величини X , менших за x , а n – об'єм вибірки.

лабораторне 2.xlsx [Последнее сохранение поль...

Разметка страницы Формулы Данные Рецензирование Вид

11 A A⁺ Перенос текста

 Объединить и поместить в центре

Выравнивание

f_x =H2+G2

Вариацийний ряд	Варіанти	Частота	Відн. част.	Нак. част.
	0	31	0,574	0,000
	1	14	0,259	=H2+G2
	5	7	0,130	
	10	2	0,037	
	Суми	54	1,000	

лабораторне 2.xlsx [Последнее сохранение поль...

Разметка страницы Формулы Данные Рецензирование Вид

11 A A⁺ Перенос текста

 Объединить и поместить в центре

Выравнивание

f_x =H3+G3

Вариацийний ряд	Варіанти	Частота	Відн. част.	Нак. част.
	0	31	0,574	0,000
	1	14	0,259	0,574
	5	7	0,130	=H3+G3
	10	2	0,037	
	Суми	54	1,000	

І таким чином заповнюємо увесь стовпчик.

лабораторне 2.xlsx [Последнее сохранение поль...

Разметка страницы Формулы Данные Рецензирование Вид

11 A A⁺ Перенос текста

 Объединить и поместить в центре

Выравнивание

f_x

Вариацийний ряд	Варіанти	Частота	Відн. част.	Нак. част.
	0	31	0,574	0,000
	1	14	0,259	0,574
	5	7	0,130	0,833
	10	2	0,037	0,963
	Суми	54	1,000	1,000

Побудуємо полігон відносних частот. Пригадаємо означення. Для графічного представлення дискретного варіаційного ряду розподілу будують *полігон (або багатокутник) частот* та *полігон (або багатокутник) відносних частот*.

Означення. Ламана лінія, відрізки якої послідовно з'єднують точки з координатами (x_i, n_i) , $i = \overline{1, k}$, називається *полігоном частот*, а ламана лінія, відрізки якої послідовно з'єднують точки з координатами (x_i, p_i^*) , $i = \overline{1, k}$, називається *полігоном відносних частот*.

Для побудови полігона відносних частот використаємо Вставка→Гистограмма→Точечная→Точечная с прямыми отрезками и маркерами.

	A	B	C	D	E	F	G	H	I
1	0	0		Варіаційний ряд	Варіанти	Частота	Відн. част.	Нак. част.	
2	1	0			0	31	0,574	0,000	
3	0	0			1	14	0,259	0,574	
4	0	0			5	7	0,130	0,833	
5	5	0			10	2	0,037	0,963	
6	0	0			Суми	54	1,000	1,000	
7	10	0							

	A	B	C	D	E	F	G	H	I
1	0	0		Варіаційний ряд	Варіанти	Частота	Відн. част.	Нак. част.	
2	1	0			0	31	0,574	0,000	
3	0	0			1	14	0,259	0,574	
4	0	0			5	7	0,130	0,833	
5	5	0			10	2	0,037	0,963	
6	0	0			Суми	54	1,000	1,000	
7	10	0							
8	0	0							
9	1	0							
10	0	0							
11	0	0							
12	1	0							
13	5	0							
14	1	0							
15	0	0							
16	0	0							
17	0	0							
18	1	0							
19	0	0							
20	1	0							
21	0	0							
22	0	0							
23	0	0							
24	0	0							

Обрати Выбрать данные→Добавить. У полі Имя ряда вносимо «Полігон відносних частот», у полі Значения X стовпець варіант E2:E5, у полі Значения Y стовпець відносних частот G2:G5.

Выбор источника данных

Диапазон данных для диаграммы: []

Элементы легенды (ряды):

Подписи горизонтальной оси (категории):

Скрытые и пустые ячейки:

OK Отмена

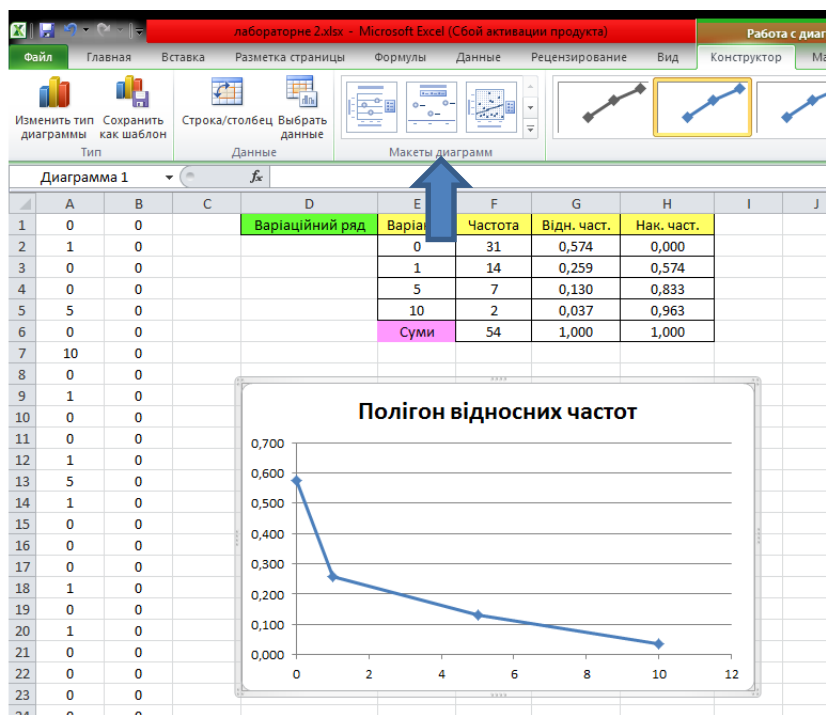
Изменение ряда

Имя ряда: [Полигон відносних частот] Выберите диапазон

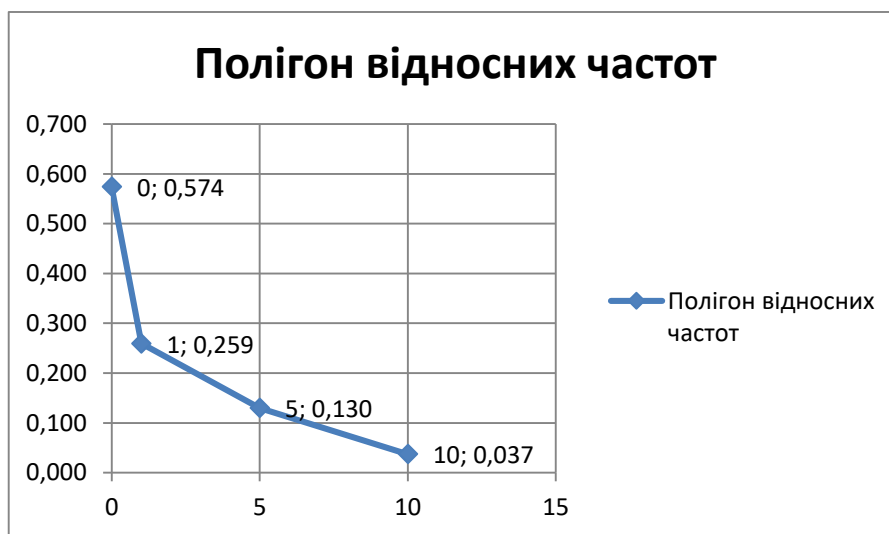
Значения X: [=Дискр!\$E\$2:\$E\$5] = 0; 1; 5; 10

Значения Y: [=Дискр!\$G\$2:\$G\$5] = 0,574; 0,259; ...

OK Отмена



В Макеты диаграмм обираємо більш зручний формат, наприклад, Макет 7.



Задачі для лабораторного роботи

За результатами спостережень, наданих у другому листі файлу (за варіантами) побудувати дискретний варіаційний ряд розподілу, полігон частот (так само, як у першому листі).

Розберемо на прикладі (з Лекції 8) побудову інтервального варіаційного ряду у *MS Excel*.

Приклад. При вимірювання зросту 105 студентів (хлопців) отримана вибірка результатів:

155, 170, 185, 180, 188, 152, 173, 178, 178, 168, 185,
173, 170, 183, 175, 173, 170, 183, 175, 180, 175, 193,
178, 183, 180, 197, 178, 181, 187, 168, 174, 179, 184,
183, 178, 180, 178, 163, 166, 178, 175, 182, 190, 167,
170, 178, 183, 170, 178, 181, 173, 168, 185, 175, 170,
155, 169, 186, 179, 189, 155, 174, 179, 179, 169, 186,
174, 171, 184, 175, 193, 178, 184, 180, 196, 175, 181,
188, 168, 179, 178, 183, 184, 178, 181, 177, 163, 166,
178, 175, 183, 190, 167, 170, 178, 183, 170, 178, 182,
173, 168, 186, 176, 171, 188.

Внесемо дані вибірки у стовпець A1:A105 (файл *лабораторне 3.xlsx*) і в цьому ж стовпці використовуючи «Сортировка и фильтр» відсортуємо елементи від мінімального до максимального.

Складемо інтервальний варіаційний ряд. Знайдемо мінімальне та максимальне значення випадкової величини: $x_{\min} = 152$, $x_{\max} = 197$. Тобто розмах – інтервал варіювання дорівнює $R = x_{\max} - x_{\min} = 197 - 152 = 45$. Нехай за умовами задане число частинних інтервалів, яке дорівнює 10.

	A	B	C	D
1	152		Мінімальне значення	152
2	155		Максимальне значення	197
3	155		Розмах варіації	45
4	155		Число частинних інтервалів	10
5	163		Довжина інтервалу	4,5
6	166			

Для отримання більш зручних значень границь інтервалів, нижню границю першого інтервалу, тобто x_1 , можна зсунути відносно x_{\min} у бік менших значень, а верхню границю останнього інтервалу відповідно у бік більших значень, але цей зсув не повинен перевищувати $\frac{h}{2}$. Зсунувши нижню границю з 152 до 150, а верхню з 197 до 200, отримаємо довжину частинного інтервалу, яка дорівнює $\frac{200-150}{10} = 5$. Внесемо ці дані – зсув границь першого та останнього інтервалів та довжину інтервалу 5, у другий стовпець таблиці.

	A	B	C	D	E
1	152		Мінімальне значення	152	-2
2	155		Максимальне значення	197	3
3	155		Розмах варіації	45	
4	155		Число частинних інтервалів	10	
5	163		Довжина інтервалу	4,5	5

Зауважимо, що в нашому прикладі нам було вказано кількість інтервалів, але у загальному випадку кількість інтервалів потрібно визначати досліднику. Найчастіше використовується формула Стерджесса:

$$k = 1 + 3,322 \cdot \lg n,$$

де n – обсяг вибірки, k – кількість частинних інтервалів.

Обчислимо за цією формулою оптимальну кількість частинних інтервалів (як приклад, в самій лабораторній будемо використовувати задану в умові кількість 10). В комірці D8 введемо відповідну формулу, а у комірці E8 – значення округлене.

D8		fx		=1+3,322*LOG(105;10)	
	A	B	C	D	E
1	152		Мінімальне значення	152	-2
2	155		Максимальне значення	197	3
3	155		Розмах варіації	45	
4	155		Число частинних інтервалів	10	
5	163		Довжина інтервалу	4,5	5
6	166				
7	167				
8	167		Число частинних інтервалів (опт.)	7,714391	7
9	168				

В наступній лабораторній ми вже будемо самі визначати число інтервалів як раз за цією формулою.

Складемо сам інтервальний варіаційний ряд (запишемо його у зручному для подальших обчислень в *MS Excel* вигляді). Як і у попередній лабораторній роботі, розташовувати за стовпцями.

	A	B	C	D	E	F	G	H
1	152		Мінімальне значення	152	-2		Інтервали	
2	155		Максимальне значення	197	3		150	155
3	155		Розмах варіації	45			155	160
4	155		Число частинних інтервалів	10			160	165
5	163		Довжина інтервалу	4,5	5		165	170
6	166						170	175
7	167						175	180
8	167		Число частинних інтервалів (опт.)	7,714391	7		180	185
9	168						185	190
10	168						190	195
11	168						195	200

Спочатку заповнимо комірки, відповідні лівій та правій межам інтервалів: починаємо зі 150 у G2, в якій прописано формулу D1+E1, тобто відступили від мінімального значення на 2 одиниці для зручності. В комірці H2 – правій межі першого інтервалу, запишемо формулу G2+E\$5 (знак долара пишемо для того, щоб потім протягнути формулу на комірки донизу). Заповнюємо стовпці таблиці до верхньої межі останнього інтервалу.

	C	D	E	F	G	H
Мінімальне значення		152	-2		Інтервали	
Максимальне значення		197	3		150	155
Розмах варіації		45			155	160
Число частинних інтервалів		10			160	165
Довжина інтервалу		4,5	5		165	170
					170	175
					175	180
Число частинних інтервалів (опт.)		7,714391	7		180	185
					185	190
					190	195
					195	200

Після чого переходимо до заповнення наступних табличних даних – середин інтервалів, частоти та відносної частоти.

	A	B	C	D	E	F	G	H	I	J	K
1	152		Мінімальне значення	152	-2		Інтервали	y_i	Част.	Відн. част.	
2	155		Максимальне значення	197	3		150	155	152,5	4	0,038
3	155		Розмах варіації	45			155	160	157,5	0	0,000
4	155		Число частинних інтервалів	10			160	165	162,5	1	0,010
5	163		Довжина інтервалу	4,5	5		165	170	167,5	19	0,181
6	166						170	175	172,5	18	0,171
7	167						175	180	177,5	27	0,257
8	167		Число частинних інтервалів (опт.)	7,714391	7		180	185	182,5	22	0,210
9	168						185	190	187,5	10	0,095
10	168						190	195	192,5	2	0,019
11	168						195	200	197,5	2	0,019
12	168								Сума	105	1
13	168										

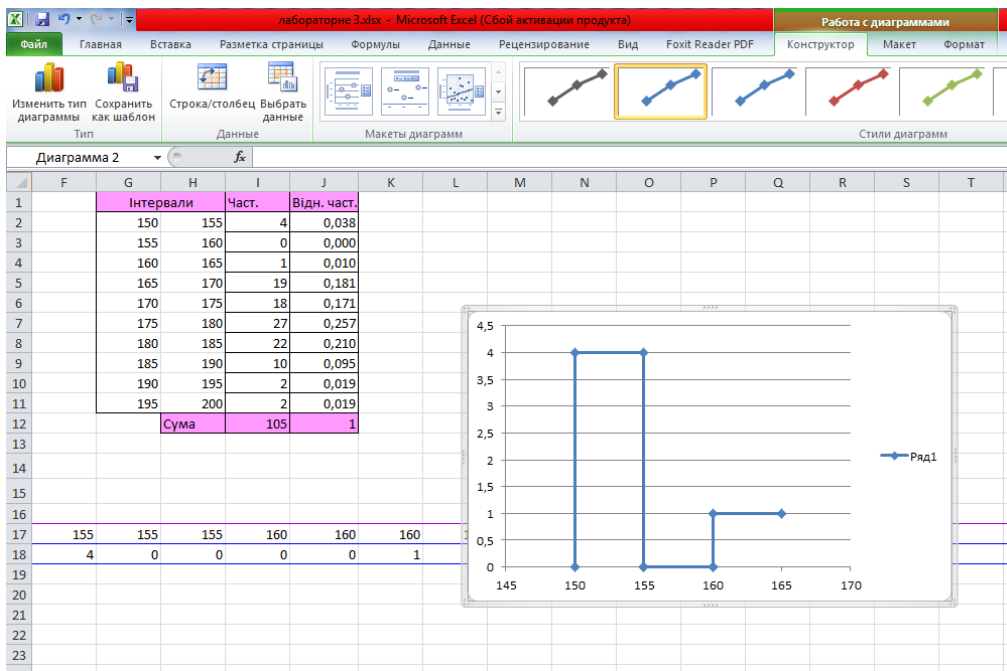
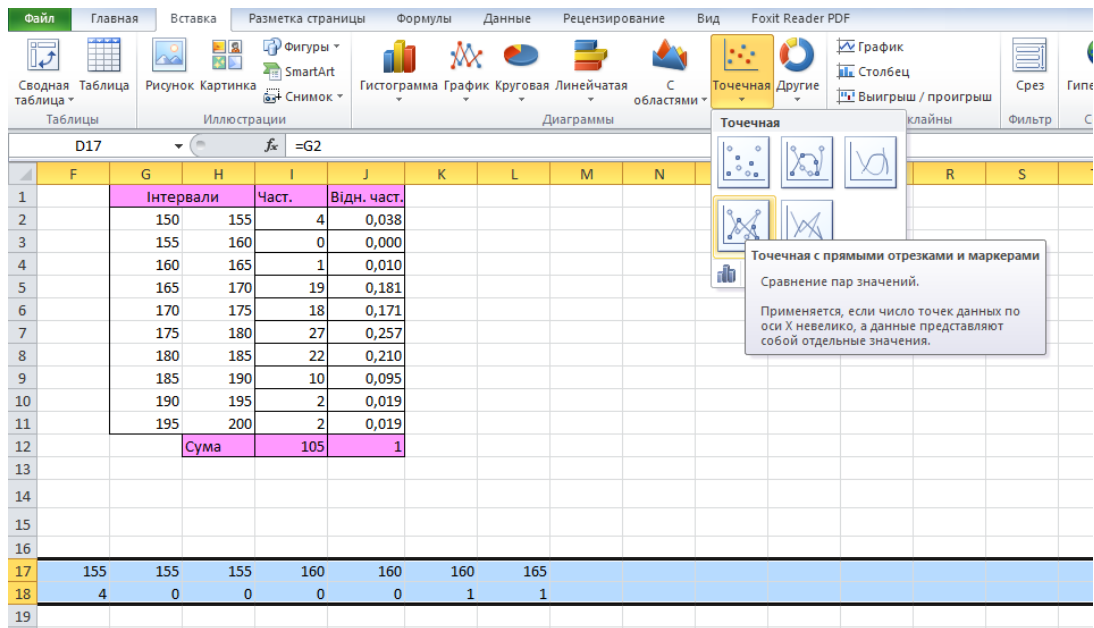
Як і у попередній лабораторній, перевіряємо суми частот та відносних частот.

Для графічного представлення інтервального варіаційного ряду розподілу будують *гістограму частот* або *відносних частот*. Для її побудови у прямокутній системі координат на осі Ox відкладають відрізки частинних інтервалів варіювання та на цих відрізках як на основах будують прямокутники з висотами, які дорівнюють частотам або відносним частотам відповідних інтервалів.

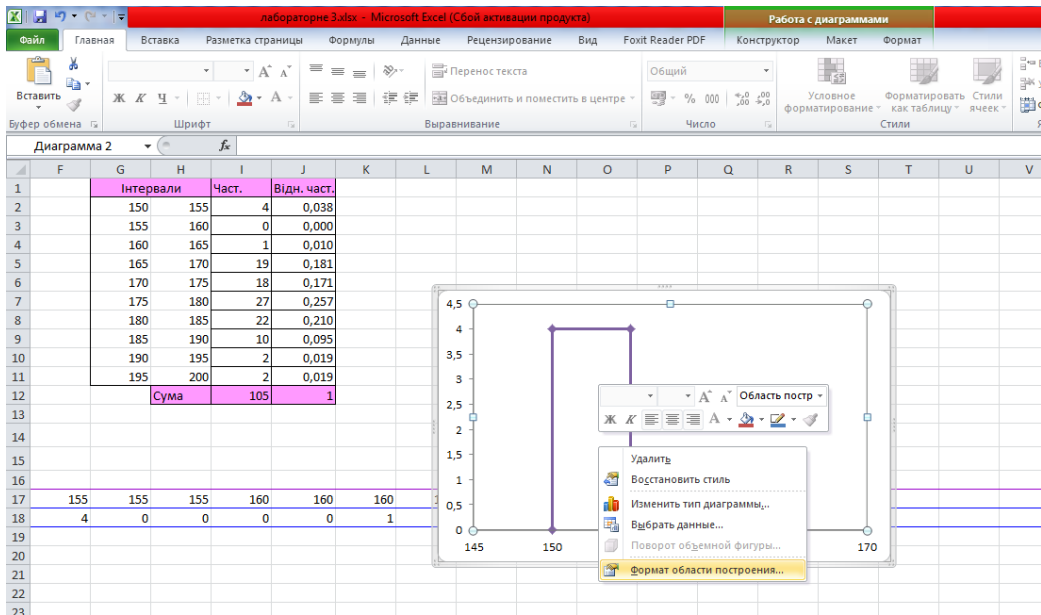
Означення. Ламана лінія, відрізки якої послідовно з'єднують точки з координатами (y_i, n_i) , $i = \overline{1, k}$, називається *полігоном частот*, а ламана лінія, відрізки якої послідовно з'єднують точки з координатами (y_i, p_i^*) , $i = \overline{1, k}$, називається *полігоном відносних частот*.

Для графічного представлення інтервального варіаційного ряду розподілу побудуємо *гістограму частот*. Для цього використаємо можливості *MS Excel*.

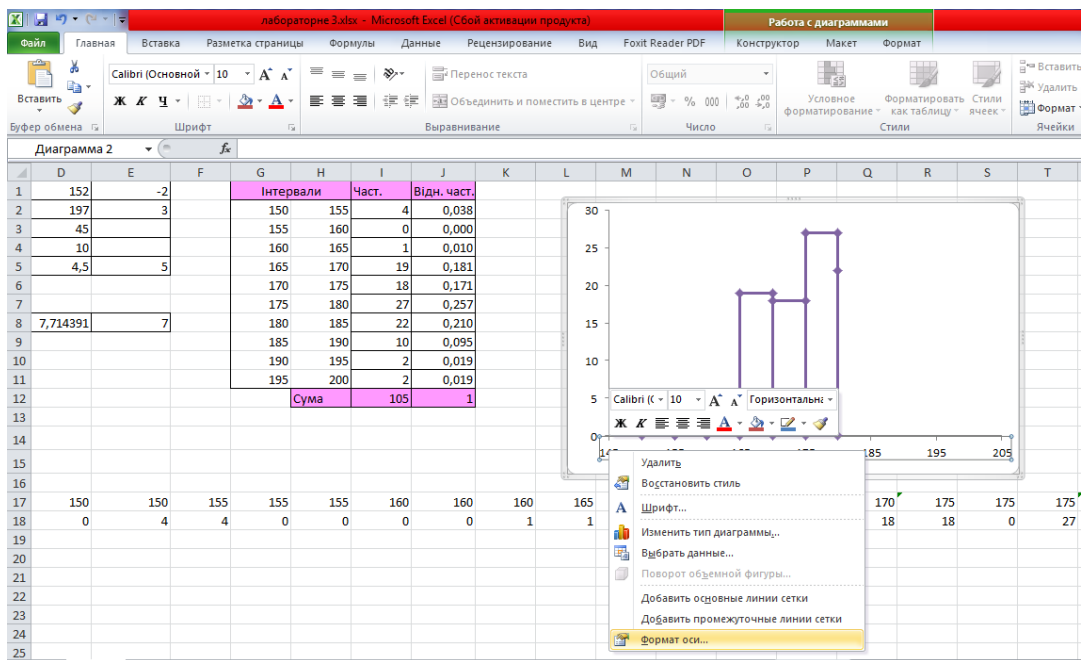
Для побудови гістограми зробимо допоміжну таблицю у комірках D17:AH17, D18:AH18 (це координати точок, за якими ми будемо пряму, беремо дані з таблиці – це значення меж інтервалів та відповідні частоти) та оберемо «Точечная с прямыми отрезками и маркерами».

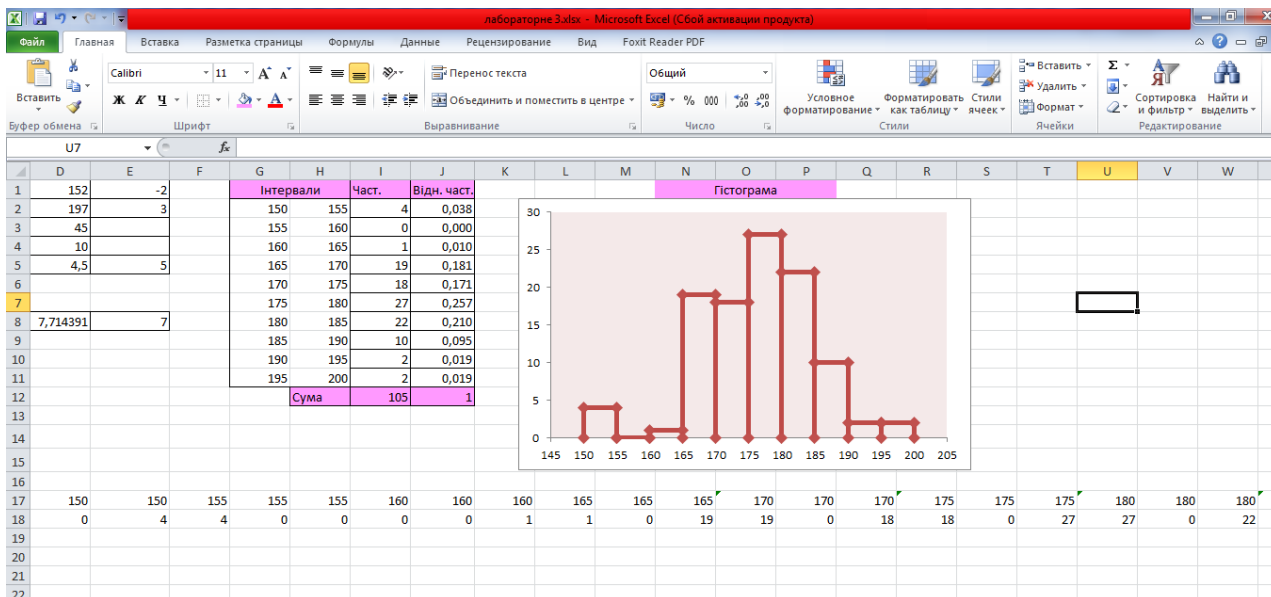
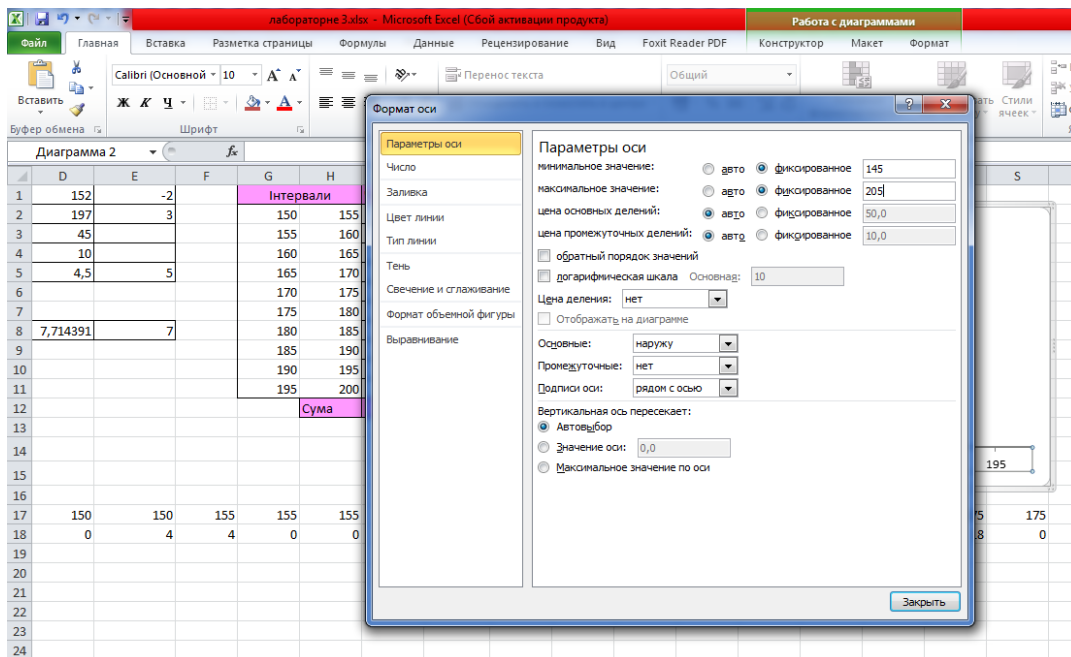


Оберемо зручний формат побудови гістограми.



Оберемо формат осі Ox .





Для побудови полігону частот на гистограмі натиснемо праву кнопку миші та оберемо «Выбрать данные» та «Добавить».

