#### Комп'ютерний практикум №4

# СТАТИСТИЧНИЙ РОЗПОДІЛ ТА ЙОГО ЧИСЛОВІ ХАРАКТЕРИСТИКИ

Мета: Навчитись визначати числові характеристики вибірки.

Змістовність роботи: Статистичний розподіл: таблична та графічна форма запису. Числові характеристики вибірки: середнє значення, дисперсія, середнє квадратичне відхилення, коефіцієнт асиметрії та ексцес. Використання функцій та пакету «Аналіз даних» MS Excel для розрахунку числових характеристик вибірки та побудови графіків.

#### 4.1. Теоретичні відомості

Упорядкований перелік варіант і відповідних їм частот називається статистичним розподілом вибірки або дискретним варіаційним рядом. Якщо число різних значень у вибірці є досить великим, то розраховувати частоту кожного з них немає сенсу. У даному випадку складають інтервальний варіаційний ряд. Весь проміжок вибірки  $[x_{min}, x_{max}]$  (від максимального до мінімального) розбивають на часткові інтервали, тобто проводять групування.

Число інтервалів може бути розраховане за формулою Стерджерса:

$$k \approx 1 + \log_2 n, \tag{4.1}$$

де  $log_2 n = 3.322 lg n$ , значення k підбирається цілим.

Довжина інтервалу знаходиться за формулою (4.2):

$$h = \frac{x_{max} - x_{min}}{k}.$$
(4.2)

За початок першого часткового інтервалу, як правило, вибирається точка  $x_0 = x_{min} - \frac{h}{2}$ .

У перший рядок таблиці інтервального ряду вписують часткові інтервали  $[x_0, x_1], (x_1, x_2], ...(x_{k-1}, x_k]$ , що мають однакову довжину h, при цьому весь інтервал  $[x_0, x_k]$  повинен повністю покривати всі значення обраної ознаки, тобто  $x_0 \le x_{min}, x_{max} \le x_k$ . У другому рядку записується кількість попадань (частота)  $m_i$  (i=1,2...k) в кожен інтервал. Таким чином, статистичний розподіл має вигляд таблиці 4.1.

Таблиця 4.1 Інтервальний варіаційний ряд

$(x_{i-1}; x_i]$	$[x_0; x_1]$	$(x_1; x_2]$	•••	$(x_{k-1}; x_k]$	Σ
$m_i$	$m_1$	$m_2$	•••	$m_k$	$\sum_{i=1}^{k} m_i = n$
$p_i^* = \frac{m_i}{n}$	$p_1^*$	$p_2^*$	•••	$p_k^*$	$\sum_{i=1}^{k} p_i^* = 1$

При вивченні варіаційних рядів використовують також поняття накопиченої частоти  $m_i^{\mu\alpha\kappa}$ . Накопичена частота показує, скільки спостерігається варіантів із значенням ознаки, меншої за *x*. Відношення накопиченої частоти до загальної кількості спостережень називають накопиченою відносною частотою  $p_i^{\mu\alpha\kappa}$ . Накопичені частоти для кожного інтервалу знаходять послідовним сумуванням частот всіх попередніх інтервалів, включаючи даний.

Статистичний ряд може бути представлений у вигляді полігону, гістограми і кумуляти.

Полігоном частот або відносних частот називають ламану лінію, яка з'єднує точки дискретного ряду ( $x_i$ ,  $m_i$ ) або ( $x_i$ ,  $p^*_i$ ), відповідно (рис.4.1.а). Для неперервного ряду полігон будуються для середніх значень інтервалу  $x_{cep}$ .

Гістограмою відносних частот називаються ступінчаста фігура (рис.4.1.б), що складається з прямокутників, основою яких є часткові інтервали (довжини  $h_i = x_i - x_{i-1}$ ), а висоти дорівнюють відношенню  $f_i^* = \frac{p_i^*}{h}$  (або інколи  $f_i^* = p_i^*$ ).

**Кумулятою** називається крива накопичених частот (рис.4.1.в), яка має вигляд ламаної лінії, що з'єднує точки ( $x_i$ ,  $m_i^{\text{нак}}$ ) або ( $x_i$ ,  $p_i^{\text{нак}}$ ).



Рис. 4.1 Графічний вигляд статистичного ряду: полігон відносних частот (а); гістограма відносних частот (б); кумулята відносних частот (в)

#### Числові та описові характеристики вибірки.

Для вибірки можна визначити ряд числових характеристик, аналогічних тим, що використовуються в теорії ймовірності та визначаються для випадкових величин. У таблиці 4.2 представлені найменування та позначення, що використовуються в статистиці для оцінки вибірки, і аналогічно – у теорії ймовірності. А також відповідні формули для розрахунку згрупованого статистичного ряду та відповідна функція MS Ехсеl. Оскільки в даній роботі розглянуто інтервальний ряд розподілу, то для розрахунку числових характеристика  $x_i = \tilde{x}_i$ , де  $\tilde{x}_i = \frac{x_{i-1} + x_i}{2}$  - середина інтервалу,  $m_i$  – відповідна частота, i = 1, 2, ..., k.  $\mu_3$  та  $\mu_4$  - центральні моменти третього та четвертого порядків (див.КП.3).

Статистичне	Теорія	Формула	Функція
позначення	ймовірності		MS Excel
Вибіркове	Математичне	<b>k</b> *	СРЗНАЧ()
середнє	сподівання	$x_e = \sum x_i p_i$	
$\overline{\mathrm{x}_{\scriptscriptstyle\mathrm{B}}}$ , $\overline{\mathrm{x}}$ , $\mathrm{M}^*[\mathrm{X}]$ , $m_{\chi}^*$	M(x)	<i>i</i> =1	
Вибіркова	Дисперсія	k - 2 *	ДИСП()
дисперсія	D(x)	$D_e = \sum (x_i - x_e)^2 \cdot p_i$	
$D_{\scriptscriptstyle  m B}$		<i>i</i> =1	
Вибіркове	Середньо-	$\sigma_{\rm p} = \sqrt{D_{\rm p}}.$	СТАНДОТКЛОН()
середньо-	квадратичне	D V D	
квадратичне	відхилення		
відхилення			
Мода <i>М</i>	Мода <i>М<sub>о</sub></i>	$M_{o}^{*} = x_{i-1} + h_{i} \cdot \frac{p_{i}^{*} - p_{i-1}^{*}}{(p_{i}^{*} - p_{i-1}^{*}) + (p_{i}^{*} - p_{i+1}^{*})},$	МОДА()
Медіана М <sub>е</sub> *	Медіана М <sub>е</sub>	$M_{e}^{*} = x_{i-1} + \frac{h_{i}}{p_{i}^{*}} \cdot \left(0, 5 - \sum_{j=1}^{i-1} \frac{h_{j}}{p_{j}^{*}}\right)$	МЕДИАНА()
Асиметрія, <i>а</i> <sub>S</sub>	Асиметрія, <i>А</i> <sub>S</sub>	$a_{s} = \frac{\mu_{3}}{\sigma_{e}^{3}}$	CKOC()
Ексцес, <i>еs</i>	Ексцес, <i>E</i> <sub>S</sub>	$e_{s} = \frac{\mu_{4}}{\sigma_{e}^{4}} - 3$	ЭКСЦЕСС()

Таблиця 4.2 Оцінка вибірки для статистики та теорії ймовірності

### 4.2. Завдання до виконання роботи

Перед початком виконання роботи встановити «Пакет анализа данных». Для цього виконати наступні дії: «Файл» → «Параметры» → «Надстройки» → «Пакет анализа данных» і натиснути кнопку «Перейти» (рис.4.2). Обрати «Пакет анализа» і натиснути кнопку «ОК».

Надстройки	? 🗙
Доступные надстройки:	
Инструменты для евро	ОК
Пакет анализа — VBA	Отмена
- nonek pemennin	06 <u>3</u> 0p
	<u>А</u> втоматизация
-	
Пакет анализа	·
Содержит инструменты для ан финансовых данных	ализа научных и

Рис. 4.2 Вікно налаштування «Пакета анализа»

Створити файл MS Excel, який буде мати наступну назву: КП4\_Прізвище студента\_№варіанту.

### Завдання 4.1.

Згенерувати вхідні данні в залежності від номеру варіанту (табл.4.3).

	1	2	3	4	5	6	7	8	9	10	11	12
n	100	110	120	130	100	110	120	130	100	110	120	100
$X_{_{\!$	20	20	20	0,64	0,64	0,64	40	40	40	0,93	60	45
$\sigma_{\scriptscriptstyle  m B}$	3	4	2	0,1	0.34	0,57	7	5	3	0,1	35	5
	13	14	15	16	17	18	17	19	20	21	22	23
n	120	130	100	110	120	130	100	110	120	130	100	120
$X_{_{\!$	0,93	0,93	80	80	80	0,85	0,85	0,85	50	50	0,85	60
$\sigma_{\scriptscriptstyle \mathrm{B}}$	0,05	0,24	5	6	9	0,5	0,2	0,7	3	7	0,2	24

Таблиця 4.3 Вхідні дані до завдання 4.1.

n – число випадкових чисел;

 $\overline{x_{e}}$  – середнє значення;

б<sub>в</sub> – середнє відхилення.

Для парних варіантів – розподіл нормальний, для непарних – Пуассона.

# Спільні значення для всіх варіантів:

«Число переменных» - 1;

«Случайное рассеивание» - пусте поле;

«Параметры вывода» - «Выходной интервал» - виділити область для запису даних.

# Порядок виконання Завдання 4.1.

1. Перейменувати «Лист 1» в «Завдання 4.1».

2. Згенерувати вхідні дані. Для цього виконати наступні дії: натиснути

«Данные» → «Анализ нанных» → «Генерация случайных чисел» (рис.4.3).

Генерация случайных чисел	? <b>×</b>
Число переменных:	ОК
Цисло случайных чисел: 100	Отмена
<u>Р</u> аспределение: Нормальное	<u>С</u> правка
Параметры	
Ср <u>е</u> днее = 70	
Стандартное отклонение = 8	
Случ <u>а</u> йное рассеивание:	
Параметры вывода	
Выходной интервал: SA:SA	
Новый рабочий <u>л</u> ист:	
🔘 Новая рабочая <u>к</u> нига	

Рис. 4.3 Вікно «Генерация случайных чисел»

- 3. Заповнити дані відповідно до номеру варіанта.
- 4. Задати вхідні та натиснути кнопку «ОК».

### Завдання 4.2.

Сформувати з отриманих даних таблицю даних та скласти варіаційний ряд. Приклад виконання в додатку 4.1.

# Порядок до виконання Завдання 4.2.

1. Перейменувати «Лист 2» в «Завдання 4.2».

2. Сформувати таблицю «Вхідні дані». У даному прикладі 10 стовпців на 10 рядків.

3. Знайти мінімальне  $x_{min}$  та максимальне  $x_{max}$  значення вибірки за допомогою вбудованих функцій «МИН» та «МАКС», відповідно (рис.4.4). Для визначення  $x_{max}$  комірці D14 присвоїти значення «МАКС(D4:M13)», а для визначення  $x_{min}$  комірці D15 присвоїти значення «МИН(D4:M13)». (D4:M13) – масив варіаційного ряду.

Аргументы функции МАКС Число1 раниз Карана (188):104;91;97;77;103;86;79;86;100:82 Число2 Карана инслор	Аргументы функции  МИН  Число1  Ф.:М13  Ф:: В8:104;91;97;77;103;86;79;86;100:82  Число2  Ф:: В8:104;91;97;77;103;86;79;86;100:82
= 105	= 68
Возвращает наибольшее значение из списка аргументов. Логические и текстовые значения	Возвращает наименьшее значение из списка аргументов. Логические и текстовые значения
игнорируются.	игнорируются.
Число1: число1;число2; от 1 до 255 чисел, пустых ячеек, логических или	Число1: число1;число2; от 1 до 255 чисел, пустых ячеек, логических или
текстовых значений, среди которых ищется наибольшее значение.	текстовых значений, среди которых ищется наименьшее значение.
Значение: 105	Значение: 68
Справка по этой функции ОК Отмена	Справка по этой функции ОК Отмена
a)	ნ)

Рис. 4.4 Вікно функцій: «МАКС» (а) та «МИН» (б)

4. Розрахувати кількість часткових інтервалів за формулою Стерджеса (4.1), яка в MS Excel задається наступним чином: «=ОКРУГЛВВЕРХ((1+3,32\*LOG10(СЧЁТ(D4:M13)));0)», де функція «СЧЁТ(D4:M13)» визначає об'єм вибірки, а «0» – вказує на число знаків після коми.

5. Довжину інтервалу знаходимо за формулою (4.2), яка в MS Excel задається наступним чином: «=ОКРУГЛВВЕРХ(((D14-D15)/G14);0)».

6. Знайдемо початок інтервалу  $x_0$  за формулою (4.3), яка в MS Excel задається наступним чином: «=D15-(G15/2)». Якщо кількість інтервалів в п.2 була цілою, то  $x_0 = x_{min}$ .

7. Скласти таблицю 4.4.варіаційного ряду та відносних частот.

Таблиця 4.4 Інтервальний варіаційний ряд

k	Інтервал, ( <i>x</i> <sub>(<i>i</i>-1)</sub> ; <i>x</i> <sub><i>i</i></sub> ]	Середина інтервалу, $\tilde{x}_i$	Частота, <i>m<sub>i</sub></i>	Відносна частота, <i>p<sub>i</sub>*</i>	$f^*$	F*

8. Заповнити значення *k* від 1 до 10.

9. Заповнити значення інтервалів ( $x_{i-1}$ ; $x_i$ ). Для цього виконати наступні дії: комірці D18 присвоїти значення J14 ( $x_0$ ); комірці E18 присвоїти значення D18+G15 (h), скопіювати його та поширити на наступні комірки (E19:E27); D19 присвоїти значення E18, скопіювати та поширити на наступні комірки (D20:B27).

10. Знайти значення середини інтервалу. Комірці F18 присвоїти значення «=(E18+D18)/2», скопіювати та поширити на наступні комірки (F19: F27).

11. Визначити частоту. Для цього виділити відповідний масив розміщення частот (G18:G27) та за допомогою «Майстра функцій» задати функцію «ЧАСТОТА()» (рис.4.5). У полі «Массив\_данных» задати дані вибірки (D4:M13), в у полі «Массив\_интервалов» - масив визначених інтервалів (D18:E27). Натиснути разом клавіші CTRL+SHIFT+ENTER.

Аргументы функции			? 💌
ЧАСТОТА			
Массив_данных	D4:M13	- (88	3;104;91;97;77;103;86;79;86;100:
Массив_интервалов	D18:E27	<b>(66</b>	;70:70;74:74;78:78;82:82;86:86;
Вычисляет распределе: содержащий на один э. Масси	ние значений по интервалам и лемент больше, чем массив инт в_интервалов массив интервал группируются зн	= (0: возвращает ервалов. юв или ссыл ачения из г	чжээж лотожлалый массия, вертикальный массия, ака на интервалы, в которых массива данных.
Значение: О			
Справка по этой функц	ии		ОК Отмена

Рис. 4.5 Вікно функції «ЧАСТОТА»

12. Визначити відносну частоту. Для цього комірці H18 присвоїти значення «=G18/\$J\$15», де G18 – частота, \$J\$15 – об'єм вибірки.

13. Визначити частоту накопичення для побудови емпіричного графіку. Для цього комірці J18 присвоїти значення відносної частоти H18. Комірці J19 присвоїти значення «=J18+ H19». Скопіювати J19 та вставити до кінця ряду (J18: J27).

14. Зробити перевірку. Знайти суму частот та відносних частот.

### Завдання 4.3.

Побудувати полігон відносних частот, гістограму та кумуляту. Приклад виконання завдання показаний у додатку 4.1.

### Порядок виконання завдання 4.3.

1. Побудувати полігон відносних частот. Для цього виконати наступну послідовність: «Вставка»→«Диаграммы» → «Точечная с прямыми отрезнами и маркерами» → «Выбрать данные» (рис.4.6.а). Натиснути кнопку «Добавить» і у вікні «Изменение ряда» в полі «Имя ряда» напистати «Полігон відносних частот», в полі «Значения Х» - вставити ряд із значеннями середини інтервалу (F18:F27), у полі «Значения Y» - вставити ряд із значеннями відносної частоти (H18:H27).



Рис. 4.6 Вікна вибору даних для побудови полігону

2. Побудувати гістограму. Для цього виконати наступну послідовність: «Вставка» — «Диаграммы» — «Гистограма» — «Выбрать данные» (рис.4.6.а). Натиснути кнопку «Добавить» і у вікні «Изменение ряда» в полі «Имя ряда» напистати «Гістограма», в полі «Значения - вставити ряд із значеннями  $f^*$ (I18:I27) (4.7.а). Натиснути кнопку «ОК» і у вікні «Выбору источника данных» натиснути кнопку «Изменить» та ввести значення ряду середини інтервалу (F18:F27) (4.7.6). Перейти в «Экспресс-макет» і обрати «Макет 8».



Рис. 4.7 Вікна вибору даних для побудови гістограми

3. Побудувати кумуляту. Для цього виконати наступну послідовність: «Вставка»—«Диаграммы» — «Точечная с прямыми отрезнами и маркерами» — «Выбрать данные» (рис.4.4.а). Натиснути кнопку «Добавить» і у вікні «Изменение ряда» в полі «Имя ряда» напистати «Кумулята», в полі «Значения Х» - вставити ряд із значеннями середини інтервалу (F18:F27), в полі «Значения Y» - вставити ряд із значеннями  $F^*$  (J18:J27).

### Завдання 4.4.

Знайти числові та описові характеристики вибірки. Приклад виконання завдання показаний у додатку 4.1.

### Порядок виконання Завдання 4.4.

 Знайти середнє вибірки. Присвоїти комірці О18 значення «=СРЗНАЧ(D4:M13)», де масив (D4:M13) – вхідні дані.

 Знайти дисперсію вибірки. Присвоїти комірці О19 значення «=ДИСП(D4:M13)», де масив (D4:M13) – вхідні дані.

3. Знайти середнє відхилення вибірки. Присвоїти комірці О20 значення «=СТАНДОТКЛОНА(D4:M13)», де масив (D4:M13) – вхідні дані.

4. Знайти коефіцієнт асиметрії. Присвоїти комірці О21 значення «=СКОС(D4:M13)», де масив (D4:M13) – вхідні дані.

 Знайти ексцес. Присвоїти комірці О22 значення «=ЭКСЦЕСС(D4:M13)», де масив (D4:M13) – вхідні дані.

53

6. Знайти моду. Присвоїти комірці О23 значення «=МОДА(D4:M13)», де масив (D4:M13) – вхідні дані.

7. Знайти медіану. Присвоїти комірці О24 значення
 «=МЕДИАНА(D4:M13)», де масив (D4:M13) – вхідні дані.

#### Завдання 4.5.

Знайти числові характеристики та побудувати графіки статистичного розподілу за допомогою вбудованого аналізу даних. Приклад виконання завдання показаний у додатку 4.2.

#### Порядок виконання завдання 4.5.

1. Повернутися до даних, отриманих в «Завдання 4.1».

2. Знайти числові характеристики. Для цього виконати наступні дії: натиснути «Данные» → «Анализ нанных» → «Описательная статистика» (рис.4.8).

Описательная статистика		? 💌
Входные данные		
В <u>х</u> одной интервал:	\$A\$3:\$A\$100	
Группирование:	по стол <u>б</u> цам	Отмена
	○ по с <u>т</u> рокам	<u>С</u> правка
Метки в первой строке		
Параметры вывода		
	\$C\$1-\$E\$16	
о выходной интервал:		
Повыи рабочии <u>л</u> ист:		
🔘 Новая рабочая <u>к</u> нига		
Итоговая статистика		
Уровень надежности:	95 %	
🔲 К-ый <u>н</u> аименьший:	1	
К-ый наибольший:	1	

Рис. 4.8 Вікно «Описательная статистика»

У поле «Входной интервал» вставити ряд отриманих вхідних даних з завдання 4.1. Відмітити «Группирование» по стовпцях. У полі «Выходной интервал» виділити область виводу даних. Відмітити позначкою «Итоговая статистика». 3. Побудувати гістограму та кумуляту. Для цього виконати наступні дії: натиснути «Данные» → «Анализ нанных» → «Описательная статистика» (рис.4.9).

Гистограмма		? 💌
Входные данные В <u>х</u> одной интервал: И <u>н</u> тервал карманов: П Метки	SAS1:SAS100 📧	ОК Отмена <u>С</u> правка
Параметры вывода Выходной интервал: Новый рабочий дист: Новая рабочая <u>к</u> нига Парето (отсортированна <u>И</u> нтегральный процент Вывод <u>г</u> рафика	SDS88:SGS104 💽	

Рис .4.9 Вікно «Гистограмма»

У полі «Входной интервал» ввести вхідні дані. У полі «Интервал карманов» - ввсети дані з завдання 4.2 – середина інтервалу. Виділити поле для формування даних. Обрати вид гістограми.

4. Порівняти отримані дані з даними завдання 4.3 та 4.4.

# 4.3. Вимоги до оформлення звіту

1. Звіт має бути представлений в електронному вигляді.

2. Назва електронного файлу КП4\_Прізвище студента\_№варіанту.

3. Файл повинен містити наступні елементи:

Дано: вхідні дані – позначити жовтим;

Розв'язок: проміжні результати – позначити зеленим.

4. При заміні тільки вхідних даних (виділено жовтим) повинен змінюватись вихідний результат.

5. Оформлення завдань згідно Додатків 4.1-4.2.

6. Висновки по роботі.

Якщо звіт не відповідає вимогам, то КП не приймається!

# 4.4. Контрольні запитання

- 1. Що таке статистичний ряд? Які форми його запису?
- 2. Який вигляд має полігон частот?
- 3. Який вигляд має гістограма?
- 4. Який вигляд має кумулята?

5. Якими числовими характеристиками описується вибірка? Приведіть аналоги з теорії ймовірності.

6. Які функції MS Excel використовуються для створення варіаційного ряду?

- 7. Які два методи розрахунку в MS Excel числових характеристик ви знаєте?
- 8. Якими функціями MS Excel описують числові характеристики вибірки?
- 9. Назвіть переваги пакету «Аналіз даних».