

**МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ
НАЦІОНАЛЬНА МЕТАЛУРГІЙНА АКАДЕМІЯ УКРАЇНИ**

**Методичні вказівки
до виконання самостійних робіт з дисципліни
“ІТ-технології та програмування”
Розділ ”Комп'ютерні технології обробки даних в Excel”
для студентів напряму підготовки металургія 6.050401.**

Затверджено на засіданні кафедри
Протокол №4 від 12 лютого 2018

Методичні вказівки до вивчення розділу "Комп'ютерні технології обробки даних в Excel" з дисципліни "ІТ-технології та програмування" для студентів напряму підготовки Металургія 6.050401. / укладачі Соболенко А.В., Павленко Г.А.- Дніпро:НМетАУ, 2018.- 38с.

ВСТУП

Використання комп'ютерної техніки і застосування економіко-математичних методів в аналізі соціально-економічних явищ стало великим поштовхом у розвитку статистичної науки. Стандартні статистичні методи обробки даних увійшли до складу таких електронних таблиць як *LOTUS*, *EXCEL* та ін.; у математичні пакети загального призначення *MATHCAD*, *MATHLAB* та ін.; спеціалізовані пакети *STATISTICA*, *STATGRAPHICA*, *ОЛІМП* та ін.

В даний час у ділових колах найбільше поширення одержав табличний процесор *Microsoft EXCEL*. Електронні таблиці *EXCEL* мають визначений набір інструментів, цілком достатніх для проведення досить повного і якісного статистичного аналізу даних.

Метою даного лабораторного практикуму є вивчення можливостей і придбання практичних навичок роботи з використанням табличного процесора *EXCEL* для виконання статистичного аналізу даних.

Студентам пропонується виконати 7 лабораторних робіт, що охоплюють широкий спектр інструментів надбудови «Пакет аналізу», а також використання великої кількості статистичних функцій і графічних можливостей *Microsoft EXCEL* для виконання розрахунків і подання результатів при проведенні статистичної обробки даних.

Кожна лабораторна робота містить перелік завдань, а також коротку інструкцію про порядок їхнього виконання і прикладами по статистичній обробці результатів спостережень.

Передбачається, що до моменту виконання лабораторних робіт, студентом вже освоєні основні поняття *Microsoft EXCEL*, його меню і кнопки команд, а також робота з функціями, створення, редагування і копіювання формул.

Загальні відомості про надбудову «Пакет аналізу» і статистичні функції Microsoft EXCEL

1. 1 Установка надбудови «Пакет аналізу»

При створенні нової чи відкритті існуючої книги Microsoft Excel з'явиться вікно активного робочого листа. Відкрийте меню *Сервис*, виберіть команду *Анализ данных* і, клацнувши мишею по даній команді, ви попадаєте у вікно надбудови *Анализ данных ...*

Якщо в меню *Сервис* відсутня команда *Анализ данных* необхідно в тім же меню виконати команду *Надстройки...*(рис.1).

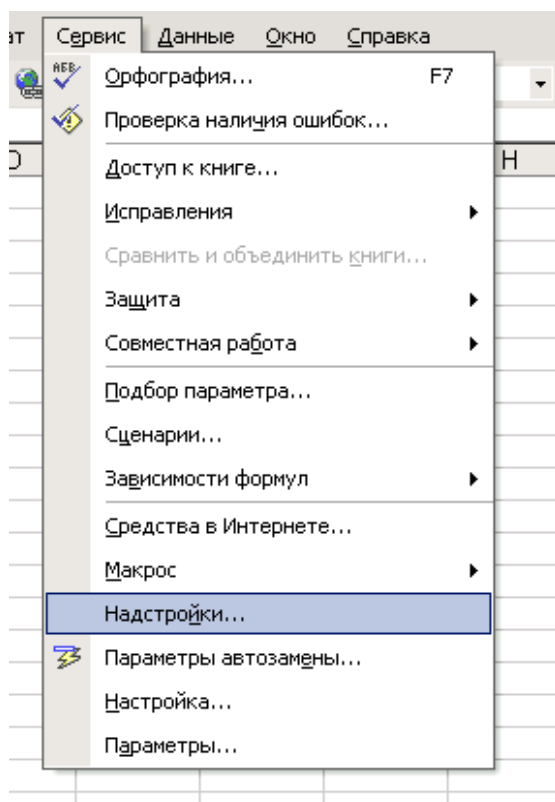


Рис.1

Розкриється вікно зі списком доступних надбудов (рис.2).

Потрібно поставити «галочку» біля рядка *Пакет анализа* і клацнути по кнопці *ОК*. Після цього в меню *Сервис* з'явиться команда *Анализ данных ...*(рис.3).

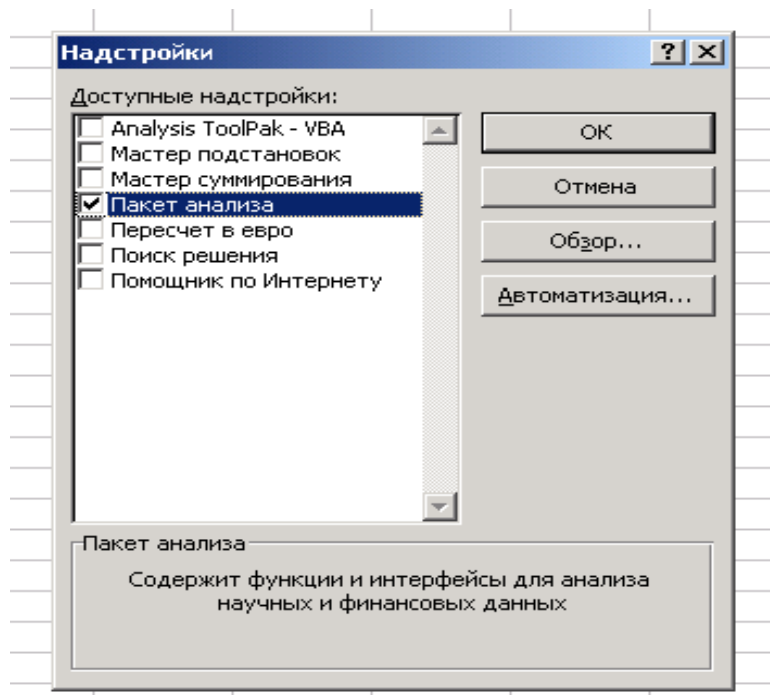


Рис. 2

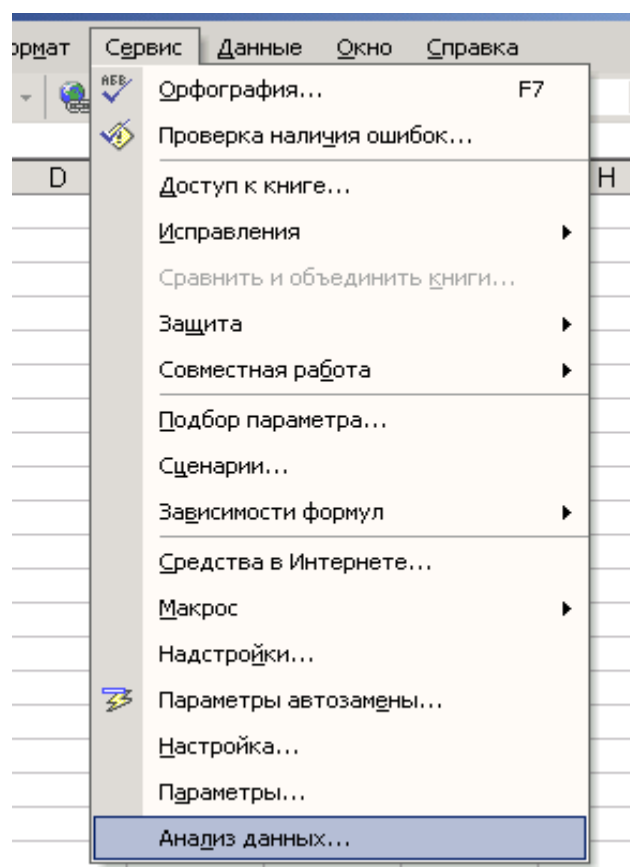


Рис. 3

1.2 Технологія роботи в режимі «Аналіз даних»

При виконанні команди меню *Сервис* пункт *Анализ данных...* з'являється вікно з однойменною назвою (рис.4).

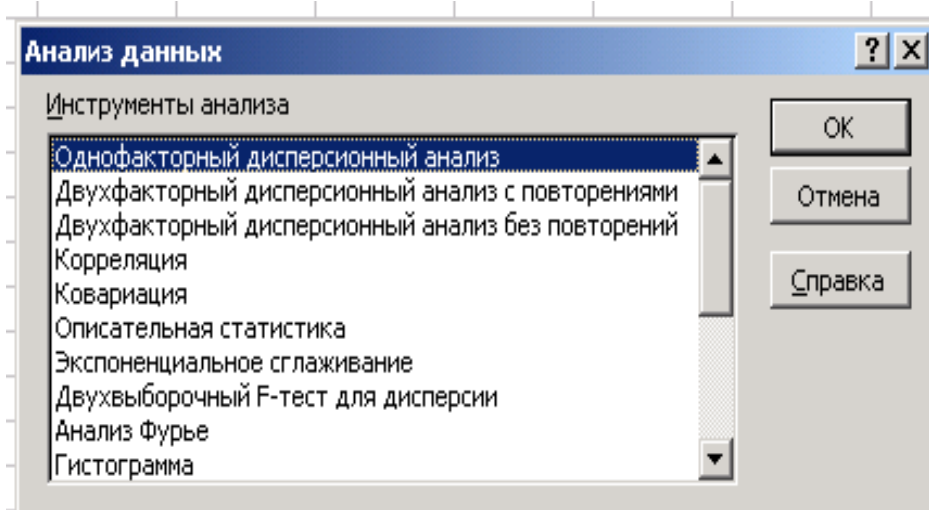


Рис. 4

Головним елементом вікна є область *Инструменты анализа*. У цій області представлений список реалізованих у Microsoft Excel методів статистичної обробки даних:

- «Однофакторний дисперсійний аналіз»;
- «Двофакторний дисперсійний аналіз»;
- «Кореляція»;
- «Ковариация»;
- «Регресія»;
- «Описова статистика»;
- «Гістограма»;
- «Вибірка»;
- «Генерація випадкових чисел»;
- «Ковзне середнє»;
- «Експонентне згладжування» та ін.

Кожний з названих методів реалізований у вигляді окремого режиму роботи. Для активації обраного режиму необхідно виділити відповідний метод і клацнути по кнопці *ОК*.

Діалогове вікно кожного режиму містить у собі елементи керування, за допомогою яких задаються наступні параметри:

1. Поле *Входной интервал* – уводиться посилання на клітинки, що містять аналізовані дані.
2. Прапорець *Метки* – встановлюється в активний стан, якщо рядок чи стовпець вхідного діапазону містить заголовки.
3. Перемикач *Группировка* – встановлюється в положення *По строкам* \ *По столбцам* у залежності від розташування аналізованих даних.
4. Перемикач *Выходной интервал* – активізується поле, у яке необхідно ввести посилання на ліву верхню клітинку вихідного діапазону.

Ви можете отримати справку *EXCEL* в будь який час, коли вона вам знадобиться при роботі з програмою. Для цього потрібно скористатися клавішею F1 (з'являється вікно спраки Excel Help. Ви також можете отримати контекстно-залежну справку Excel, якщо скористуетесь комбінацією клавіш Shift+F1.

2. Види помилок при створенні формул

Формула, записана в клітинці, може містити одну чи кілька функцій, зв'язаних між собою арифметичними операціями, або вкладених друг у друга. Якщо при завданні формули були допущені помилки, результатом її обчислення буде значення помилки. У залежності від виду помилки в клітинці записуються різні значення. Нижче приведений список значень помилок з поясненням причин їхнього виникнення (табл. 1). Крім того, треба слідкувати, щоб усі скобки були парні, були введені усі необхідні аргументи та не було зайвих аргументів. У функцію можна вложити не більш ніж сім функцій.

Таблиця 1

Опис помилки	Причини виникнення помилки	Заходи для усунення помилки
####	Числове значення, що вводиться, не уміщається в клітинку	Треба збільшити ширину стовпця
#ИМЯ?	<p>1. Помилка в написанні імені функції</p> <p>2. У формулі введений текст не ув'язнений у подвійні лапки</p> <p>3. У посиланні на діапазон клітинок пропущена двокрапка</p>	<p>Виправити ім'я функції</p> <p>Установити подвійні лапки</p> <p>Виправте формулу, установивши двокрапку</p>
#ДЕЛ/0!	У дільнику використовується посилання на клітинку, що містить нульове або порожнє значення	Змінити посилання, або введіть ненульове значення в клітинку
#ЗНАЧ!	Замість числового чи логічного значення введений текст	Перевірте правильність завдання типів аргументів
#Н/Д	Невизначені дані, або посилання на порожню клітинку	Задайте правильно аргумент функції
#ЧИСЛО!	У функції з числовим аргументом використовується неприйнятний аргумент	Перевірте правильність використання у функції аргументів
#ССЫЛКА!	Клітинки, на які посилаються формули, були вилучені	Перевірте і змініте формулу
#ПУСТО!	Використано оператора, що задає перетинання діапазонів, які не мають загальних ячеек	Задайте правильно розмірність пересічних діапазонів

Лабораторна робота №1

Варіаційний ряд, статистичний ряд та його графічне зображення

Значення випадкової величини, які ми спостерігаємо будемо називати вибірковими і обозначати x_1, x_2, \dots, x_n . Впорядкований за збільшенням ряд вибіркових значень називається варіаційним рядом. При великій кількості спостережень, варіаційний ряд групується.

Якщо спостерігаєма величина X дискретна, то x_i - можливі значення дискретної випадкової величини, а m_i – кількість спостережень значення x_i . У результаті угруповання отримаємо варіаційний ряд:

x_i	x_1	x_2	$x_3 \dots$	x_n
m_i	m_1	m_2	$m_3 \dots$	m_n

При цьому значення $x_i \dots x_n$ розташовані за збільшенням.

Статистичний ряд подається у вигляді таблиці угрупованого варіаційного ряду з додатковою строкою, в яку розміщують частоти дискретної випадкової величини:

x_i	x_1	x_2	$x_3 \dots$	x_n
m_i	m_1	m_2	$m_3 \dots$	m_n
P_i^*	P_1^*	P_2^*	$P_3^* \dots$	P_n^*

Частоти P_i^* обчислюються за формулою:
$$P_i^* = \frac{m_i}{n},$$

де n об'єм вибірки $n = \sum m_i$

Якщо випадкова величина X , яку ми спостерігаємо безперервна, то діапазон спостерігаємих значень поділяють на інтервали і знаходять кількість значень належних до кожного інтервалу.

I_i	$[x_1; x_2)$	$[x_2; x_3)$...	$[x_i; x_{i+1})$...	$[x_k; x_{k+1})$
m_i	m_1	m_2	...	m_i	...	m_k

Довжину інтервалу Δx і кількість інтервалів k знаходять за допомогою формули Стерджеса:

$$k = [1 + 3,31 \lg n], \quad \Delta x = \frac{x_{\max} - x_{\min}}{k},$$

де x_{\max} , x_{\min} відповідно максимальне і мінімальне значення вибіркової величини. Ліва границя першого інтервалу $x_1 = x_{\min} - \Delta x/2$. Далі визначимо $x_{i+1} = x_i + \Delta x$.

Для дискретної випадкової величини статистичний ряд графічно оформлюється у вигляді полігона. Полігон дискретної випадкової величини це ламана лінія, яка з'єднує точки, абсциси котрих мають значення статистичного ряду x_i , а ординати дорівнюють або пропорційні частоті p_i^* .

Полігон безперервної випадкової величини це ламана лінія, яка з'єднує точки, абсциси яких – середини i -го інтервалу, а ординати дорівнюють або пропорційні частоті p_i^* .

Гістограму безперервної випадкової величини одержимо, якщо на кожному інтервалі зведемо прямокутник, висота якого дорівнює або пропорційна значенню відносної щільності експериментальних точок в цьому інтервалі $\frac{p_i^*}{\Delta x}$.

Завдання 1. За заданими дискретними вибілковими даними збудувати статистичний ряд, оформити його графічно у вигляді полігону.

Приклад 1.1. Результати вибіркового контролю якості продукції за 24 робочих дня місяця становлять:

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	1	0	1	0	7	0	1	1	0	0	0	6

i	13	14	15	16	17	18	19	20	21	22	23	24
x_i	1	1	0	2	0	0	1	1	1	0	2	1

Вигляд таблиці Excel для вирішення цієї задачі на рис. 5, необхідні формули наведені нижче.

Увага. У наведених формулах адреси треба скоригувати відповідно до своїх даних.

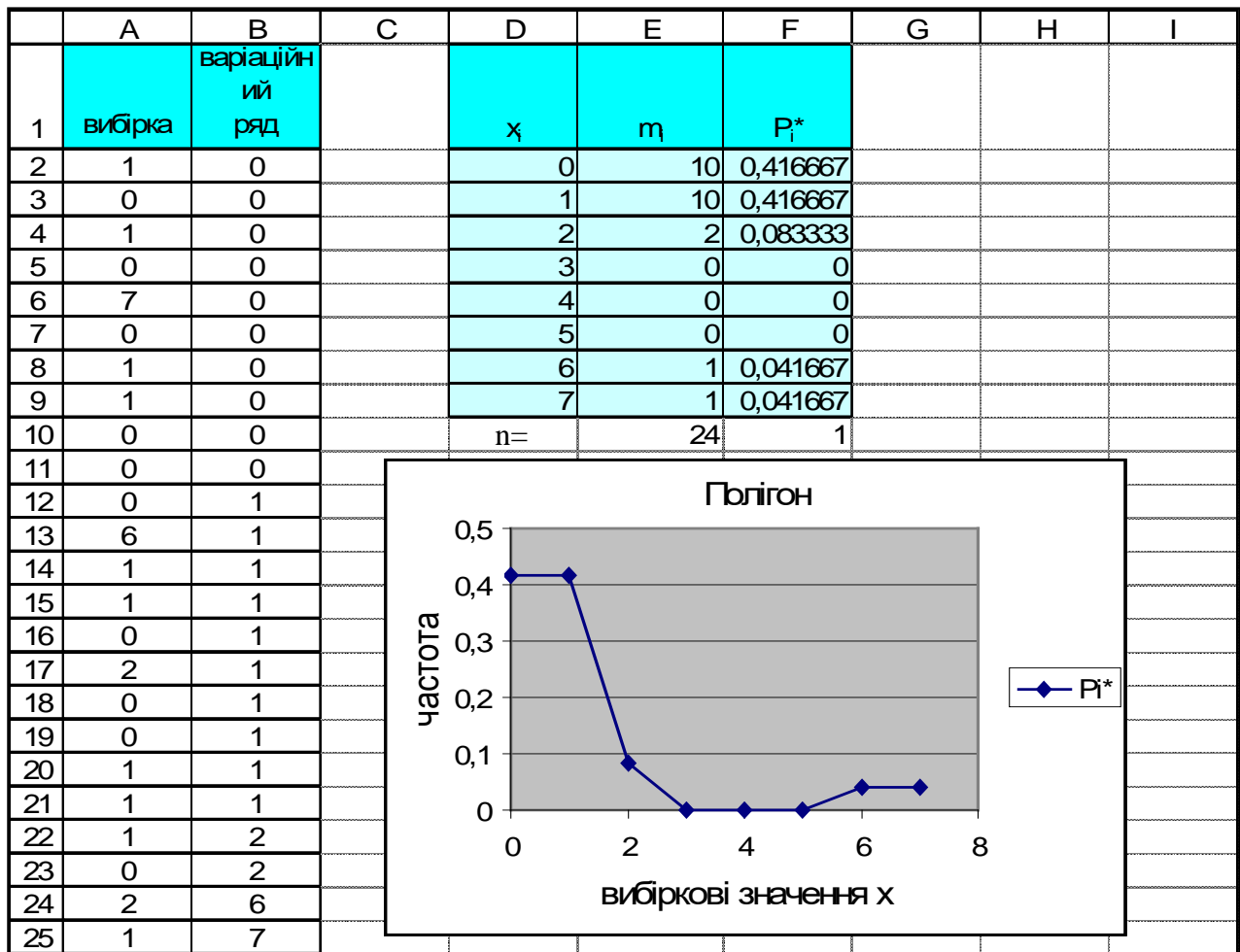


Рис. 5

В стовбець А заносимо вхідні дані. Для цього в клітинку А1 запишемо "вibірка" (лапки в клітинку не пишуться). В діапазон А2:А25 внесемо вибіркові (вхідні) значення. В клітинку В1 запишемо "варіаційний ряд", а в діапазон В2:В25 скопіюємо дані з діапазону А2:А25 і відсортуємо їх за збільшенням за допомогою команди:

Данные → Сортировка.

В стовбцях D, E, F побудуємо варіаційний ряд. В клітинку D1 запишемо " x_i ", в клітинку E1 – " m_i ", в клітинку F1 – " P_i^* ". В діапазон D2:D9 запишемо значення дискретного ряду від 0 (мінімум) до 7 (максимум), а в діапазоні E2:E9 розміщуємо формулу, за допомогою якої підрахуємо кількість повторювань значень вибіркової величини: "`=СЧЁТЕСЛИ(А$2:А$25;СЦЕПИТЬ(="";ТЕКСТ(D2;"Основной"))`)".

В клітинку D10 запишемо “n=”, в клітинку E10 - формулу “=СУММ(E2:E9)”, в клітинку F2 – “=E2/\$E\$10” і скопіюємо цю формулу на діапазон F2:F9, в клітинку F10 - формулу “=СУММ(F2:F9)”.

Графічне зображення дискретного варіаційного ряду (полігон) получимо виконавши команди:

Вставка → Диаграмма → График → Точечная.

Дані для полігону знаходяться у несуміжних діапазонах D1:D9 та F1:F9. (Несуміжні діапазони задаються при нажатій клавіші Ctrl.)

Завдання 2. За заданими вибірковими безперервними даними збудувати статистичний ряд, оформити його графічно.

Приклад 1.2. Проводились виміри діаметру деталі. Получені такі результати: 6,75; 6,77; 6,77; 6,73; 6,76 ;6,69; 6,81 ;6,80; 6,72; 6,75; 6,78; 6,82; 6,74; 6,76; 6,89; 6,70; 6,71; 6,76; 6,75; 6,78.

Вигляд таблиці Excel для вирішення цієї задачі на рис. 6, необхідні формули наведені нижче.

Увага. У наведених формулах адреси треба скоригувати відповідно до своїх даних.

В стовбець А заносимо вхідні дані. Для цього в клітинку A1 запишемо ”вбірка” (лапки в клітинку не пишуться). В діапазон A2:A21 внесемо вибіркові значення. В клітинку B1 запишемо ”варіаційний ряд”, а в діапазон B2:B21 скопіюємо дані з діапазону A2:A21 і відсортуємо їх за збільшенням за допомогою команди: *Данные → Сортировка.*

В клітини C1, C3, C5 запишемо відповідно ”n”, ”k”, ”Δx”, а в клітини C2, C4, C6 формули для розрахунку цих величин – “=СЧЁТ(A2:A21)”, “=ОКРУГЛВВЕРХ(1+3,32*LOG10(20);0)”, “=(B21-B2)/ОКРУГЛВНИЗ(1+3,32*LOG10(20);0)”.

В клітини E1, F1 запишемо відповідно заголовки ” x_i ”, ” x_{i+1} ”, в клітинку E2-формулу розрахунку для лівої границі першого інтервалу ” =B2- $\Delta x/2$ ”, в клітинку E3 - формулу розрахунку для лівої границі інших інтервалів “=E2+ Δx ” і скопіюємо її в діапазон E4:E7, в клітинку F2 - формулу розрахунку для правої границі інтервалів ” =E2+ Δx ” і скопіюємо її в діапазон F3:F7

Частоту влучання у кожний інтервал визначимо як різницю між кількістю вибіркових даних менших за праву границю інтервалу та даних менших за ліву

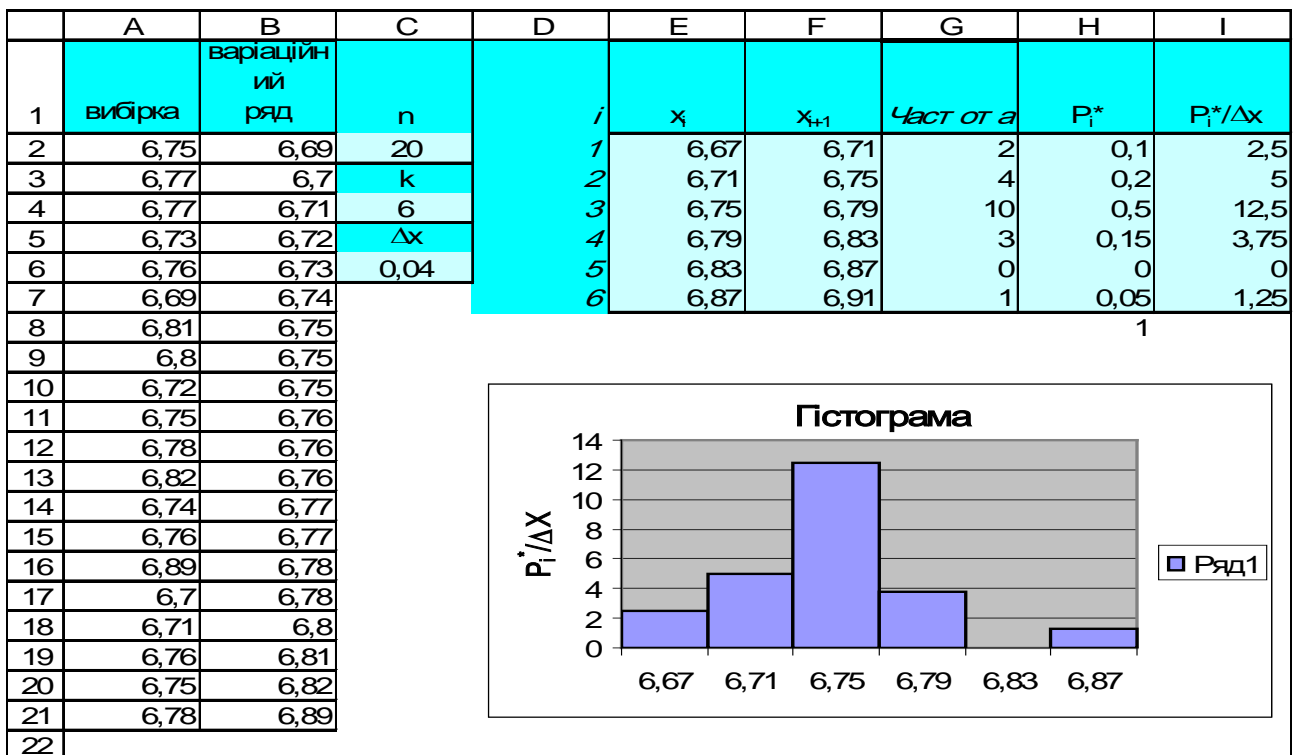


Рис. 6

границю. То в діапазон клітин G2:G7 запишемо формулу ”=СЧЁТЕСЛИ (A\$2:A\$21; СЦЕПИТЬ("<";ТЕКСТ(F2; "Основной"))) – СЧЁТЕСЛИ (A\$2:A\$21; СЦЕПИТЬ("<";ТЕКСТ(E2;"Основной")))”.

В клітини H1 та I1 запишемо заголовок ” P_i^* ”, ” $P_i/\Delta x$ ”, в діапазон клітин H2:H7 - формулу розрахунку для частоти ” =G2/\$C\$2 ”, в діапазон клітин I2:I7 – формулу ” =H2/\$C\$6 ”.

Графічне зображення дискретного варіаційного ряду (гістограму) получимо виконавши команди:

Вставка → Диаграмма → гистограмма.

Дані для гістограми знаходяться у діапазоні клітин I1:I9 (вкладка *Диапазон данных*), підписи осі x - у діапазоні клітин E2:E7 (вкладка *Ряд*).

Зберегти результати роботи у файлі.

Варіанти завдання №1

1. 8, 3, 9, 6, 7, 4, 9, 9, 9, 0, 6, 5, 3, 3, 2, 4, 2, 3, 0, 9, 8, 4, 7, 1, 5, 6, 3, 9, 1, 9, 9, 7, 5, 9, 5, 1, 7, 1, 1, 6;
2. 3, 4, 0, 3, 7, 0, 8, 8, 1, 3, 6, 7, 1, 1, 5, 1, 4, 5, 0, 6, 7, 0, 2, 5, 8, 8, 3, 3, 6, 9, 8, 3, 8, 1, 1, 1, 4, 9, 2, 4;
3. 6, 0, 1, 0, 2, 3, 3, 3, 9, 3, 9, 2, 5, 9, 2, 2, 7, 4, 2, 1, 2, 6, 5, 2, 8, 0, 7, 6, 6, 4, 3, 7, 9, 5, 4, 6, 6, 4, 9, 5;
4. 7, 2, 5, 0, 6, 0, 9, 7, 1, 3, 5, 1, 8, 5, 2, 3, 1, 4, 6, 0, 1, 4, 3, 2, 8, 8, 4, 0, 0, 2, 8, 9, 5, 4, 1, 8, 1, 4, 6, 5;
5. 5, 5, 6, 2, 0, 3, 6, 7, 3, 5, 4, 4, 5, 6, 1, 3, 1, 1, 4, 3, 2, 6, 0, 1, 1, 7, 1, 8, 7, 7, 1, 5, 8, 7, 6, 8, 9, 9, 3, 7;
6. 5, 3, 4, 8, 8, 4, 2, 6, 5, 2, 9, 5, 7, 8, 8, 2, 3, 7, 9, 0, 0, 7, 1, 9, 2, 3, 3, 8, 6, 7, 9, 5, 6, 7, 2, 6, 0, 6, 3, 3;
7. 5, 6, 6, 1, 9, 5, 4, 9, 4, 7, 8, 4, 8, 1, 0, 0, 2, 0, 5, 1, 1, 6, 4, 8, 8, 0, 4, 0, 6, 3, 4, 1, 8, 0, 8, 8, 3, 9, 7, 7;
8. 7, 8, 3, 0, 0, 4, 2, 6, 4, 9, 8, 4, 5, 7, 3, 1, 4, 4, 5, 3, 4, 1, 0, 5, 0, 6, 2, 8, 0, 2, 2, 6, 9, 4, 8, 8, 1, 1, 7, 9;
9. 4, 9, 3, 5, 8, 7, 1, 6, 0, 7, 5, 6, 5, 8, 7, 3, 0, 1, 0, 9, 7, 8, 2, 3, 2, 7, 3, 3, 5, 6, 2, 2, 5, 5, 0, 1, 0, 7, 5, 2;
10. 7, 2, 9, 2, 9, 1, 1, 2, 3, 3, 2, 8, 5, 2, 4, 7, 0, 3, 7, 0, 7, 0, 7, 8, 2, 2, 2, 2, 2, 2, 0, 5, 0, 2, 4, 9, 8, 0, 7, 3;
11. 2, 8, 3, 4, 5, 0, 0, 4, 4, 9, 6, 8, 2, 6, 9, 5, 8, 3, 4, 1, 8, 1, 2, 0, 4, 8, 7, 4, 3, 2, 9, 0, 6, 9, 2, 1, 1, 7, 8, 6;
12. 5, 7, 7, 2, 5, 0, 0, 5, 0, 5, 3, 7, 2, 5, 3, 1, 9, 4, 3, 8, 3, 2, 6, 8, 1, 3, 0, 4, 3, 2, 9, 9, 1, 1, 1, 1, 7, 1, 2, 1;
13. 6, 3, 4, 4, 0, 0, 8, 0, 5, 3, 2, 2, 8, 0, 0, 2, 9, 2, 9, 3, 5, 1, 6, 1, 2, 6, 4, 5, 7, 5, 0, 4, 4, 3, 3, 5, 5, 2, 9, 1;
14. 1, 2, 0, 8, 5, 7, 2, 4, 1, 6, 9, 3, 5, 0, 2, 8, 9, 2, 5, 2, 4, 9, 1, 4, 7, 0, 4, 3, 7, 6, 5, 4, 5, 3, 3, 3, 0, 0, 9;
15. 6, 0, 8, 6, 3, 3, 2, 4, 29, 7, 5, 1, 7, 1, 7, 4, 6, 3, 8, 2, 9, 8, 4, 2, 4, 2, 0, 0, 5, 1, 9, 3, 2, 8, 5, 3, 3, 9, 7.

Варіанти завдання №2

1. 2.4; 2.6; 2.7; 2.5; 2.5; 2.8; 2.8; 2.8; 2.5; 2.6; 2.6; 2.7; 2.4; 2.4; 2.5; 2.1; 2.6; 2.5; 2.3; 2.2.
2. 2.4; 2.7; 2.5; 2.4; 2.6; 2.4; 2.4; 2.4; 2.6; 2.6; 2.5; 2.5; 2.3; 2.5; 2.6; 2.7; 2.3; 2.5; 2.8; 2.6.
3. 2.5; 2.6; 2.4; 2.7; 2.6; 2.5; 2.9; 2.4; 2.7; 2.5; 2.7; 2.6; 2.6; 2.8; 2.6; 2.7; 2.4; 2.6; 2.6; 2.6.
4. 2.4; 2.8; 2.5; 2.4; 2.4; 2.4; 2.6 2.5; 2.2; 2.6; 2.1; 2.6; 2.4; 2.3; 2.5; 2.6; 2.7; 2.5; 2.6; 2.4.
5. 2.7; 2.3; 2.5; 2.4; 2.4; 2.4; 2.2; 2.4; 2.6; 2.5; 2.5; 2.7; 2.3; 2.7; 2.5; 2.4; 2.5; 2.4; 2.4; 2.3.
6. 2.7; 2.6; 2.4; 2.5; 2.5; 2.4; 2.4; 2.5; 2.5; 2.5; 2.6; 2.5; 2.5; 2.5; 2.7; 2.4; 2.8; 2.5; 2.4; 2.3.
7. 2.2; 2.5; 2.6; 2.5; 2.4; 2.4; 2.4; 2.2; 2.5; 2.5; 2.5; 2.7; 2.2; 2.4; 2.5; 2.2; 2.5; 2.4; 2.4; 2.5.
8. 2.2; 2.3; 2.4; 2.4; 2.4; 2.5; 2.4; 2.5; 2.6; 2.5; 2.4; 2.6; 2.6; 2.6; 2.4; 2.5; 2.5; 2.4; 2.5; 2.5.
9. 2.6; 2.5; 2.6; 2.6; 2.5; 2.5; 2.3; 2.5; 2.5; 2.6; 2.5; 2.5; 2.1; 2.4; 2.4; 2.5; 2.4; 2.6; 2.5; 2.4.
10. 2.4; 2.4; 2.5; 2.3; 2.2; 2.2; 2.2; 2.4; 2.4; 2.6; 2.7; 2.3; 2.5; 2.5; 2.6; 2.3; 2.5; 2.6; 2.4; 2.6.
11. 2.4; 2.4; 2.5; 2.4; 2.7; 2.7; 2.2; 2.2; 2.6; 2.4; 2.6; 2.7; 2.6; 2.8; 2.3; 2.6; 2.5; 2.3; 2.5; 2.4.
12. 2.7; 2.5; 2.4; 2.5; 2.5; 2.4; 2.6; 2.4; 2.6; 2.5; 2.5; 2.4; 2.7; 2.4; 2.4; 2.5; 2.5; 2.4; 2.6; 2.3.
13. 2.8; 2.4; 2.5; 2.4; 2.4; 2.4; 2.4; 2.5; 2.5; 2.2; 2.5; 2.4; 2.4; 2.5; 2.4; 2.5; 2.2; 2.5; 2.5; 2.4.
14. 2.6; 2.6; 2.7; 2.2; 2.6; 2.5; 2.5; 2.6; 2.5; 2.4; 2.6; 2.6; 2.5; 2.6; 2.5; 2.5; 2.3; 2.8; 2.6; 2.6.
15. 2.7; 2.5; 2.7; 2.1; 2.5; 2.6; 2.5; 2.4; 2.5; 2.7; 2.4; 2.5; 2.6; 2.6; 2.7; 2.5; 2.7; 2.6; 2.5; 2.3.

Емпірична функція розподілу

Емпіричною функцією розподілу називають функцію $F^*(x)$, яка дорівнює для кожного значення x частоті події $X < x$ у даній вибірці:

$$F^*(x) = P^*(X < x) = \frac{m^*}{n},$$

де m^* - кількість вибірових значень, менших за x у даній вибірці.

При великих n функція розподілу вибірки $F^*(x)$ наближається до функції розподілу генеральної сукупності $F(x)$.

Таким чином, емпіричну функцію розподілу вибірки можна вважати статистичним аналогом функції розподілу генеральної сукупності.

Приклад 2.1. Знайти емпіричну функцію розподілу вибірки, яка є дискретним варіаційним рядом для умов прикладу 1.1.

$$F^*(0) = P^*(X < 0) = \frac{0}{24}, F^*(1) = P^*(X < 1) = \frac{10}{24}, F^*(2) = P^*(X < 2) = \frac{20}{24},$$

$$F^*(4) = P^*(X < 4) = \frac{22}{24}, F^*(5) = P^*(X < 5) = \frac{22}{24}, F^*(6) = P^*(X < 6) = \frac{22}{24},$$

$$F^*(7) = P^*(X < 7) = \frac{23}{24}, F^*(8) = P^*(X < 8) = \frac{24}{24}.$$

Вигляд таблиці Excel для вирішення цієї задачі на рис. 7, необхідні формули наведені нижче.

Загрузимо файл, який ми отримали в результаті виконання лабораторної роботи №1.

В клітину G1 запишемо " F_i^* ", а в діапазон G2:G10 запишемо формулу для розрахунку значення функції розподілу $F^*(x_i)$ "`=СЧЁТЕСЛИ(A$2:A$25; СЦЕПИТЬ("<"; ТЕКСТ(D2;0))) /E$10`".

Графік функції розподілу $F^*(x)$ получимо клацнувши на панелі інструментів кнопку *майстер діаграм*, указав тип *гістограма* і адресу за якою знаходяться дані(G2:G10).

Для інтервального варіаційного ряду статистичну функцію розподілу $F^*(x)$ безперервної випадкової величини X будують за значеннями, які функція набуває на кінцях i -го інтервалу.

$$F^*(x_1) = 0;$$

$$F^*(x_2) = P_1^*;$$

$$F^*(x_3) = P_1^* + P_2^*;$$

$$F^*(x_4) = P_1^* + P_2^* + P_3^*; \dots;$$

$$F^*(x_i) = \sum_{k=1}^{i-1} P_k^* .$$

Приклад 2.2. Знайти емпіричну функцію розподілу вибірки, яка є інтервальним варіаційним рядом для умов прикладу 1.2.

$$F^*(6,67) = 0; \quad F^*(6,71) = 0,1; \quad F^*(6,75) = 0,1+0,3=0,3;$$

$$F^*(6,79) = 0,3+0,5=0,8; \quad F^*(6,88) = 0,8+0,15=0,95;$$

$$F^*(6,87) = 0,95+0=0,95; \quad F^*(6,91) = 0,95+0,05=1.$$

Вигляд таблиці Excel для вирішення цієї задачі на рис. 8, необхідні формули наведені нижче.

Завантажимо файл, який ми отримали в результаті виконання лабораторної роботи №1.

В клітину I1 запишемо " F_i^* ", в клітину J2 значення функції розподілу $F^*(x_i)$, що дорівнює нулю, а в діапазон J3:J10 запишемо формулу для розрахунку значення функції розподілу $F^*(x_i)$ " $=I2+H2$ ".

Графік функції розподілу $F^*(x)$ получимо клацнувши на панелі інструментів кнопку *майстер діаграм*, указав тип *точечная* і адресу за якою знаходяться дані (це несуміжні діапазони E1:E8; I1:I8).

Завдання.. Получити емпіричну функцію розподілу вибірки для дискретного та інтервального рядів

Увага. У наведених в прикладах формулах адреси треба скоригувати відповідно до своїх даних.

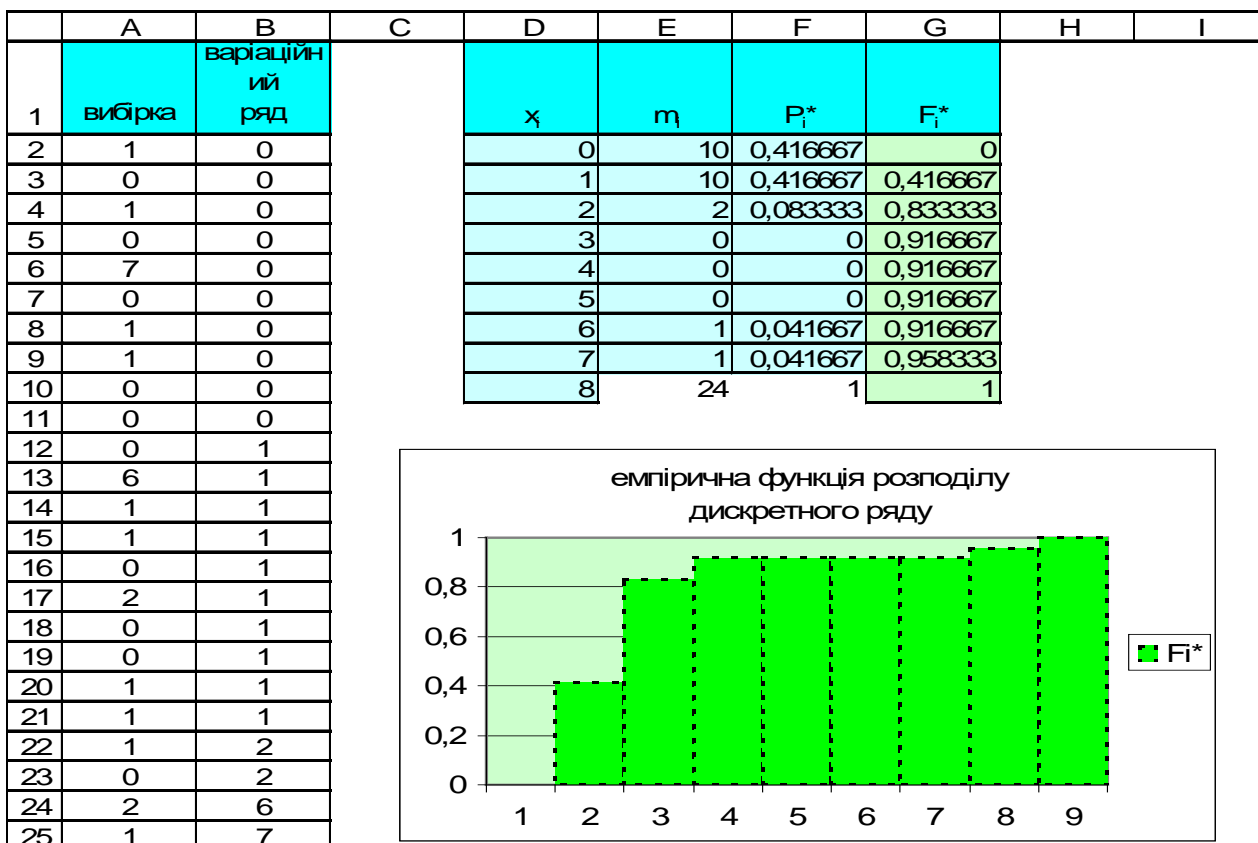


Рис. 7

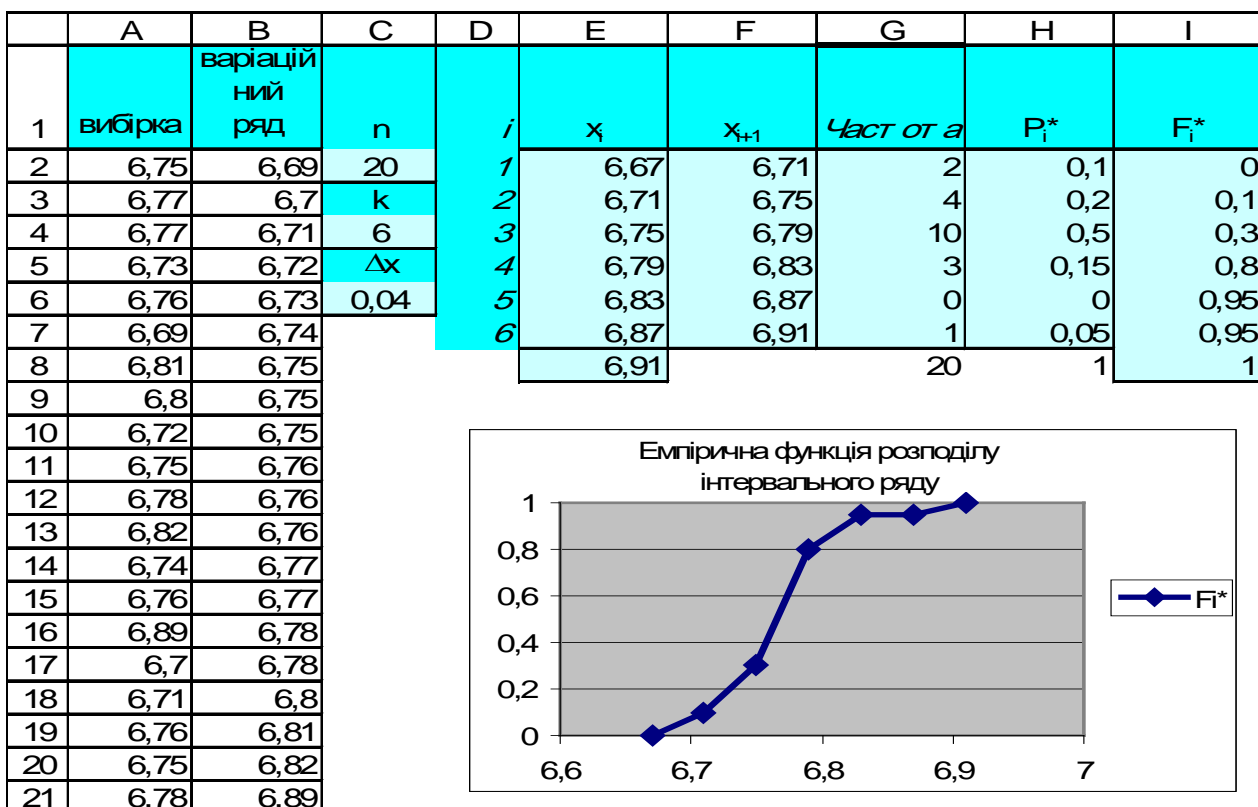


Рис. 8

Лабораторна робота №3

Точкові оцінки числових характеристик та параметрів розподілу випадкової величини

Вибіркові характеристики будемо називати оцінками відповідних характеристик генеральної сукупності. Точковою називають оцінку яка відповідає одному числу.

Загальними характеристиками випадкової величини є статистичні моменти. Найважливіші з них перший початковий (математичне очікування m_x) та другий центральний (дисперсія D_x) моменти.

Коли вибірка значень випадкової величини X представлена дискретним варіаційним рядом

x_i	x_1	x_2	...	x_i	...	x_k
m_i	m_1	m_2	...	m_i	...	m_k

або інтервальним варіаційним рядом,

I_i	$[x_1; x_2)$	$[x_2; x_3)$...	$[x_i; x_{i+1})$...	$[x_k; x_{k+1})$
m_i	m_1	m_2	...	m_i	...	m_k

то в першу чергу визначається об'єм вибірки

$$n = \sum_{i=1}^k m_i .$$

Незміщені оцінки математичного очікування та дисперсії випадкової величини можна визначити за формулами:

$$\tilde{m}_x = \sum_{i=1}^n \frac{x_i}{n} ,$$
$$\tilde{D}_x = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} ,$$

для інтервального варіаційного ряду:

$$\tilde{m}_x = \sum_{i=1}^k \frac{m_i \bar{x}_i}{n} ,$$
$$\tilde{D}_x = \sum_{i=1}^k \frac{m_i (x_i - \bar{x})^2}{n-1} ,$$

де $\bar{x}_i = \frac{x_{i+1} + x_i}{2} = x_i + \frac{\Delta x}{2}$ - середина i -го інтервалу.

Приклад 3.1. Знайти точкові оцінки числових характеристик випадкової величини, яка є дискретним рядом для умов прикладу 1.1.

У разі коли ряд не сгрупований, найпростіший спосіб отримати

характеристики усі разом застосувати команду

Сервис → Анализ данных → Описательная статистика

Указати: вхідний інтервал A3:A22, вихідний інтервал C2.

Кожну з цих характеристик можна одержати окремо, застосувавши відповідні функції, приведені на рис. 9.

Для інтервального варіаційного ряду середнє та дисперсію можна отримати за формулами: ” =СУММПРОИЗВ((B2:B7+C2:C7)/2;D2:D7)/D8”, ”=СУММПРОИЗВ(D2:D7;((B2:B7+C2:C7)/2-G1)^2)” (Рис. 11).

	A	B	C	D	E	F
1	Точенні оцінки					
2	вибрка		вибрка			
3	6,75		Среднее	6,762		=СРЗНАЧ(A3:A22)
4	6,77		Стандартная ошибка	0,010223811216398		
5	6,77		Медиана	6,76		=МЕДИАНА(A3:A22)
6	6,73		Мода	6,75		=МОДА(A3:A22)
7	6,76		Стандартное отклонение	0,0457222737379815		=СТАНДОТКЛОН(A3:A22)
8	6,69		Дисперсия выборки	0,0020905263157709		=ДИСП(A3:A22)
9	6,81		Эксцесс	2,09378674237219		=ЭКЦЕСС(A3:A22)
10	6,8		Асимметричность	0,976163767456007		
11	6,72		Интервал	0,19999999999999999		=F14-F13
12	6,75		Минимум	6,69		=МИН(A3:A22)
13	6,78		Максимум	6,89		=МАКС(A3:A22)
14	6,82		Сумма	135,24		=СУММ(A3:A22)
15	6,74		Счет	20		=СЧЕТ(A3:A22)
16	6,76					
17	6,89					
18	6,7					
19	6,71					
20	6,76					
21	6,75					
22	6,78					

Рис. 9

	A	B	C	D	E	F	G
1	/	χ	χ _{n1}	η		середнє	=СУММПРОИЗВ((B2:B7+C2:C7)/2;D2:D7)/D8
2	1	6,67	6,71	3		дисперсія	=СУММПРОИЗВ(D2:D7;((B2:B7+C2:C7)/2-G1)^2)
3	2	6,71	6,75	6			
4	3	6,75	6,79	7			
5	4	6,79	6,83	3			
6	5	6,83	6,87	0			
7	6	6,87	6,91	1			
8				=СУММ(D2:D7)			

Рис. 10

**Вилучення результатів спостережень,
що різко вирізняються**

Маємо вибірку x_1, x_2, \dots, x_n . Якщо якесь вибіркове значення визиває підозру, чи не є воно грубою помилкою спостережень, то відповідь на це питання дає статистичний критерій v , величину якого знаходять з виразу

$$v_{cn} = \frac{|x_{\max(\min)} - \tilde{m}_x|}{\tilde{\sigma}_x},$$

де $\tilde{m}_x, \tilde{\sigma}_x$ - незміщені оцінки математичного очікування і середньоквадратичного відхилення.

Задаючи рівень значності α та знаючи об'єм вибірки n , з таблиці 2 знаходимо $v_{кр}(n, \alpha)$. Якщо нерівність $v_{cn} \leq v_{кр}(n, \alpha)$ порушена, то підозріле значення x_{\max} або x_{\min} вилучається з вибірки. Оцінки \tilde{m}_x та $\tilde{\sigma}_x$ перераховують за спостереженнями, які залишилися.

Завдання. Перевірити при рівні значності $\alpha=0,05$, чи не є якесь максимальне (мінімальне) значення даної вибірки грубою помилкою спостережень.

Приклад. Маємо вибірку 12 спостережень: 3, 7, 9, 4, 11, 9, 6, 10, 5, 17, 7,

	A	B	C	D	E	F	G
1	вбірка з генеральної сукупності	варіаційний ряд		математичне очікування	середньоквадратичне відхилення	спостережуваний критерій	критерій
2	3	3	min	8	3,717	2,421	2,387
3	7	4		перераховано			
4	9	5		7,182	2,523		
5	4	6					
6	11	7					
7	9	7					
8	6	8					
9	10	9		H_0 : значення	17	не є різко вирізняючимся	
10	5	9					
11	17	10					
12	7	11		Висновок: гіпотеза не справедлива			
13	8	17	max				

8. Проводимо дослідження згідно з загальною схемою перевірки статистичних гіпотез.

1. Гіпотеза H_0 : значення $x_{max} = 17$ не є різко вирізняючимся.
2. Використовуємо статистичний критерій v .
3. Знайдемо v_{cn} та зробимо висновки.

Вигляд таблиці Excel для вирішення цієї задачі та необхідні формули наведені нижче (рис. 11).

До діапазону клітин A2:A13 вводимо дану вибірку. Копіюємо дані до діапазону B2:B13 та робимо сортування за зростанням.

До клітини D2 запишемо формулу “=СРЗНАЧ(B2:B13)”, до клітини E2 — “=СТАНДОТКЛОН(B2:B13)”, до клітини F2 — “=ABS(B13-D2)/E2”.

До клітини G2 заносимо значення $v_{кр}(12;0,05)$, яке дорівнює 2,387, до клітини D9 - “ H_0 : значення “, до клітини E9 — “=B13”, до клітини F9 - “не є різко вирізняючимся”.

До клітини D12 заносимо “ Висновок: “, до клітини E12 заносимо формулу “=ЕСЛИ(F2<=G2;"гіпотеза справедлива";"гіпотеза не справедлива")”.

До клітини D3 заносимо формулу “=ЕСЛИ(F2<=G2;" ";"перераховано)”, до клітини D4 — “=ЕСЛИ(F2<=G2;" ";"СРЗНАЧ(B2:B12))”, до клітини E4 — “=ЕСЛИ(F2<=G2;" ";"СТАНДОТКЛОН(B2:B12))”.

Увага. У всіх вище наведених формулах підозріле значення знаходиться у клітині B13. Адресу B13 та B12 треба скоригувати відповідно до своїх даних. А також необхідно скоригувати відповідно до своїх даних значення $v_{кр}$.

Варіанти лабораторного завдання 4.

Для приведених нижче вибірок при рівні значності $\alpha=0,05$ зробити:

1. Вилучення результатів спостережень, які різко вирізняються.
2. Перевірка випадковості та незалежності результатів спостережень

1) $n=25$

26, 10, 16, 7, 17, 19, 13, 21, 29, 4, 1, 28, 37, 45, 15, 17, 14, 9, 10, 7, 11, 24

2) $n=18$

28, 33, 29, 16, 44, 21, 16, 17, 19, 1, 22, 28, 22, 14, 7, 13, 21, 34

3) $n=21$

21, 19, 11, 9, 11, 13, 16, 15, 14, 14, 8, 2, 5, 7, 10, 6, 19, 27, 17, 15, 19

4) $n=17$

37, 48, 19, 27, 18, 5, 11, 42, 64, 32, 36, 21, 20, 39, 15, 16, 24

5) $n=26$

14, 9, 10, 7, 11, 24, 26, 10, 16, 7, 17, 19, 13, 21, 29, 4, 1, 28, 37, 45, 15, 17, 21

6) $n=19$

28, 34, 33, 29, 16, 44, 21, 22, 28, 22, 14, 7, 13, 21, 15, 16, 17, 19, 1

7) $n=21$

27, 11, 9, 11, 21, 19, 13, 16, 15, 14, 14, 8, 2, 5, 7, 10, 6, 19, 17, 19, 15

8) $n=17$

27, 18, 5, 11, 42, 24, 64, 32, 37, 48, 19, 36, 21, 20, 15, 39, 16

9) $n=27$

4, 1, 28, 37, 45, 15, 17, 14, 9, 10, 26, 10, 16, 7, 17, 24, 19, 13, 21, 29, 7, 11, 15, 18

10) $n=20$

28, 33, 29, 16, 44, 21, 16, 17, 19, 1, 22, 28, 22, 14, 7, 13, 21, 15, 34, 31

11) $n=21$

13, 16, 15, 14, 21, 19, 11, 9, 17, 15, 19, 11, 14, 8, 2, 5, 7, 10, 6, 19, 27

12) $n=17$

32, 36, 37, 48, 19, 27, 18, 5, 21, 20, 39, 15, 16, 24, 11, 42, 64

13) $n=26$

28, 37, 45, 15, 17, 26, 10, 16, 7, 29, 4, 1, 14, 9, 10, 7, 11, 24, 17, 19, 13, 21, 14

14) $n=19$

19, 1, 22, 28, 33, 29, 16, 44, 21, 15, 34, 21, 16, 17, 28, 22, 14, 7, 13

15) $n=22$

16, 15, 14, 21, 5, 7, 10, 6, 19, 27, 17, 15, 19, 19, 11, 9, 11, 13, 14, 8, 2, 22

Табл. 2

n	α			n	α		
	0,001	0,01	0,05		0,001	0,01	0,05
3	1,414	1,414	1,412	28	3,528	3,124	2,764
4	1,731	1,723	1,689	29	3,548	3,140	2,778
5	1,990	1,955	1,869	30	3,567	3,156	2,792
6	2,203	2,130	1,996	31	3,586	3,171	2,805
7	2,377	2,265	2,093	32	3,603	3,185	2,818
8	2,521	2,374	2,172	33	3,620	3,199	2,830
9	2,643	2,464	2,238	34	3,636	3,212	2,842
10	2,747	2,540	2,294	35	3,652	3,224	2,853
11	2,837	2,606	2,343	36	3,667	3,236	2,864
12	2,915	2,663	2,387	37	3,681	3,248	2,874
13	2,984	2,713	2,426	38	3,695	3,259	2,885
14	3,046	2,759	2,461	39	3,708	3,270	2,894
15	3,102	2,800	2,494	40	3,720	3,281	2,904
16	3,152	2,837	2,523	41	3,733	3,291	2,913
17	3,198	2,871	2,551	42	3,745	3,301	2,922
18	3,240	2,903	2,557	43	3,756	3,310	2,931
19	3,278	2,932	2,601	44	3,767	3,320	2,940
20	3,314	2,959	2,623	45	3,778	3,329	2,948
21	3,347	2,984	2,644	46	3,788	3,338	2,956
22	3,378	3,008	6,664	47	3,798	3,346	2,964
23	3,407	3,030	2,683	48	3,808	3,354	2,972
24	3,434	3,051	2,707	49	3,818	3,363	2,980
25	3,459	3,071	2,718	50	3,827	3,370	2,987
26	3,483	3,089	2,734	51	3,836	3,378	2,994
27	3,506	3,107	2,749	52	3,845	3,386	3,001

Перевірка результатів на випадковість і незалежність

Застосування більшості статистичних методів засновано на результатах рядів незалежних спостережень, які утворюють так звану вибірку із даної генеральної сукупності.

Тому необхідно пересвідчитись, що результати спостережень дійсно утворюють випадкову вибірку та стохастично незалежні.

Для вирішення цього питання маємо критерій серій, критерій "східних" та "низхідних" серій та інші критерії.

Критерій серій

Маємо вибірку x_1, x_2, \dots, x_n . Знайдемо вибіркоче значення медіани. Воно дорівнює або середньому члену варіаційного ряду з номером $(n+1)/2$, якщо об'єм вибірки n непарний, або середньому суми членів варіаційного ряду з номерами $(n/2)$ та $(n/2+1)$, якщо об'єм вибірки n парний.

Далі утворюємо послідовність плюсів та мінусів: якщо в вибірці з генеральної сукупності $x_i > Me$ ставимо "плюс", якщо $x_i < Me$ - "мінус", якщо $x_i = Me$, то такі члени вибірки не беруться до уваги.

Утворена послідовність плюсів та мінусів характеризується числом серій v_{cn} та протяжністю найбільшої серії τ_{cn} (під серією розуміється послідовність плюсів або мінусів, які йдуть поспіль).

Гіпотеза про випадковість і незалежність вибірки приймається, коли виконуються нерівності:

$$v_{cn} > v_{кр};$$

$$\tau_{cn} < \tau_{кр}.$$

Критичні значення $v_{кр}$, $\tau_{кр}$ залежать від об'єму вибірки n та рівня значності α . Коли $\alpha=0,05$:

$$v_{кр} = \left\lceil \frac{1}{2}(n+1 - 1,96\sqrt{n-1}) \right\rceil;$$

$$\tau_{кр} = \lfloor 3,3(\lg n + 1) \rfloor.$$

Критерій "східних" та "низхідних" серій

Критерій "східних" та "низхідних" серій виявляє систематичну помилку (поступове зміщення середнього) в розподілі не тільки монотонного (як критерій серій), але й більш загального характеру, наприклад періодичного.

Тут також вивчається послідовність плюсів та мінусів, утворена за даними вибірки з генеральної сукупності. На i -му місці послідовності ставиться "+", якщо $x_{i+1} - x_i > 0$ та "-" якщо $x_{i+1} - x_i < 0$. Якщо два або декілька поряд значень рівні, то приймається до уваги тільки одне.

Таким чином, послідовність плюсів відповідає зростанню (східна серія), а послідовність мінусів – зменшенню (низхідна серія) результатів спостережень.

Утворена послідовність плюсів та мінусів характеризується числом серій v_{cn} та протяжністю найбільшої серії τ_{cn} (під серією розуміється послідовність плюсів або мінусів, які йдуть поспіль).

Гіпотеза про випадковість і незалежність вибірки приймається, коли виконуються нерівності:

$$v_{cn} > v_{кр};$$

$$\tau_{cn} < \tau_{кр}.$$

Критичні значення $v_{кр}$, $\tau_{кр}$ залежать від об'єму вибірки n та рівня значності α .

Коли $\alpha=0,05$

$$v_{кр} = \left\lceil \frac{1}{3}(2n - 1) - 1,96 \sqrt{\frac{16n - 29}{90}} \right\rceil,$$

а $\tau_{кр}$ знаходиться за нижче даною таблицею

n	$n \leq 26$	$26 < n \leq 153$	$153 < n \leq 1170$	$n > 1170$
$\tau_{кр}$	5	6	7	8

Завдання. Перевірити при рівні значності $\alpha=0,05$ випадковість і незалежність результатів спостережень.

Приклад. Маємо вибірку, об'єм якої $n=11$: 3, 7, 9, 4, 11, 9, 6, 10, 5, 7, 8.

Проводимо дослідження згідно загальної схеми перевірки статистичних гіпотез.

1. Гіпотеза H_0 : вибірка випадкова та незалежна.

2. Використовуємо статистичні критерії серій та критерій "східних" та "низхідних" серій

3. Знайдемо v_{cn} і τ_{cn} , підрахуємо $v_{кр}$, $\tau_{кр}$ та зробимо висновки.

Критерії серій. Варіаційний ряд : 3, 4, 5, 6, 7, 7, 8, 9, 9, 10, 11. Вибіркове значення медіани $Me = x\left(\frac{n+1}{2}\right) = x_6 = 7$.

Утворюємо послідовність плюсів та мінусів 3, 7, 9, 4, 11, 9, 6, 10, 5, 7, 8.

-	+	-	+	+	-	+	-	+
---	---	---	---	---	---	---	---	---

Знаходимо $v_{cn} = 8$; $\tau_{cn} = 2$,

$$v_{кр} = \lceil \frac{1}{2}(11+1 - 1,96\sqrt{11-1}) \rceil = \lceil 2,75 \rceil = 3; \quad \tau_{кр} = \lfloor 3,3(\lg 11+1) \rfloor = \lfloor 6,7 \rfloor = 6.$$

Висновок: гіпотеза справедлива.

Критерій "східних" та "низхідних" серій.

Утворюємо послідовність плюсів та мінусів 3, 7, 9, 4, 11, 9, 6, 10, 5, 7, 8.

+	+	-	+	-	-	+	-	+	+
---	---	---	---	---	---	---	---	---	---

Знаходимо $v_{cn} = 7$; $\tau_{cn} = 2$,

$$v_{кр} = \lceil \frac{1}{3}(2 \cdot 11 - 1) - 1,96\sqrt{\frac{16 \cdot 11 - 29}{90}} \rceil = \lceil 4,49 \rceil = 5,$$

$\tau_{кр} = 5$ (за таблицею).

Висновок: гіпотеза справедлива.

Вигляд таблиці Excel для вирішення цієї задачі та необхідні формули наведені нижче (рис. 12).

Критерій серій

До діапазону клітин A3:A13 вводимо дану вибірку. До клітини A14 заносимо "медіана", до клітини A15 — "=МЕДИАНА(A3:A13)".

Утворюємо послідовність плюсів та мінусів за визначеним правилом. До клітини B3 вводимо формулу "=ЕСЛИ(A3<A\$15;"-";ЕСЛИ(A3=A\$15;"-";"+"))". Копіюємо цю формулу до діапазону B3:B13.

Утворюємо послідовність для підрахунку серій в стовбці C. Нова серія починається тоді, коли змінюється "+" на "-" або навпаки, а також на пробіл (" "). Початок нової серії позначається одиницею, продовження - нулем. До клітини C3 вводимо одиницю, до клітини C4 формулу - "= ЕСЛИ (A4=A\$15 ;

0; ЕСЛИ (ИЛИ (И (A4<A\$15; A3>=A\$15); И(A4>A\$15; A3<=A\$15)));1;0)". Копіюємо цю формулу до діапазону C4:C13.

Утворюємо послідовність для підрахунку протяжності серій в стовбці D. Проти початку нової серії в стовбці C стоїть одиниця, тому до лічильника протяжності додається +1 доки ми не зустрінемо ознаку початку нової серії в стовпці C. До клітини D3 вводим формулу-" =ЕСЛИ(A3=A\$15; 0; ЕСЛИ(ИЛИ(И(A2<A\$15; A3<A\$15); И(A2>A\$15; A3>A\$15))); D2+1;1)". Копіюємо цю формулу до діапазону D4:D13.

Підраховуємо значення $v_{кр}$, до клітини A18 вводим формулу - "
=ОКРУГЛВВЕРХ(0,5*(СЧЁТ(A3:A13)+1-1,96*КОРЕНЬ(СЧЁТ(A3:A13)-1));0)
".

Підраховуємо значення $\tau_{кр}$, до клітини B18 вводим формулу -
"=ОКРУГЛВНИЗ(3,3*LOG10(СЧЁТ(A3:A13)+1);0)".

Підраховуємо значення $v_{сн}$, до клітини C18 вводим формулу -
"=СУММ(C3:C13)".

Підраховуємо значення $\tau_{сн}$, до клітини D18 вводим формулу -
"=МАКС(D3:D13)".

До клітини A19 вводим текст " Н0:вибірка випадкова та результати спостережень незалежні "

До клітини A21 вводим текст " Висновок: " До клітини B21 вводим формулу " =ЕСЛИ(И(C18>A18;D18<B18);"гіпотеза справедлива";"гіпотеза не справедлива")".

Критерій "східних" та "низхідних" серій

Утворюємо послідовність плюсів та мінусів. До клітини E3 вводим формулу " =ЕСЛИ(A3<A4;"+";ЕСЛИ(A3=A4;" ";"-"))". Копіюємо цю формулу до діапазону E3:E12.

Утворюємо послідовність для підрахунку серій. Початок нової серії позначається одиницею, продовження – нулем. До клітини F4 формулу -
"=ЕСЛИ(E3=E2;0;1)". Копіюємо цю формулу до діапазону F4:F12.

Утворюємо послідовність для підрахунку протяжності серій. До клітини G3 вводим одиницю, до клітини G4 вводим формулу
"=ЕСЛИ(E4=E3;G3+1;1). Копіюємо цю формулу до діапазону G4:G12.

Підраховуємо значення $v_{сп}$, до клітини E18 вводимо формулу “=СУММ(F3:F12)“.

Підраховуємо значення $\tau_{сп}$, до клітини F18 вводимо формулу “=МАКС(G3:G12)“.

Підраховуємо значення $v_{кр}$, до клітини G18 вводимо формулу “=ОКРУГЛВВЕРХ((1/3)*(2*СЧЁТ(A3:A13)-1)-1,96*КОРЕНЬ((16*СЧЁТ(A3:A13)-29)/90);0)“.

Підраховуємо значення $\tau_{кр}$, до клітини H18 вводимо формулу “=ЕСЛИ(G14>1170;8;ЕСЛИ(G14>153;7;ЕСЛИ(G14<26;5;6)))“.

До клітини E19 вводимо текст “ H_0 :вибірка випадкова та результати спостережень незалежні “

До клітини E21 вводимо текст “ Висновок: “ До клітини F21 вводимо

	A	B	C	D	E	F	G	H
1		критерій серій			критерій "східних" та "низхідних" серій			
2	вибірка з генеральної сукупності							
3	3	-	1	1	+	1	1	
4	7		0	0	+	0	2	
5	9	+	1	1	-	1	1	
6	4	-	1	1	+	1	1	
7	11	+	1	1	-	1	1	
8	9	+	0	2	-	0	2	
9	6	-	1	1	+	1	1	
10	10	+	1	1	-	1	1	
11	5	-	1	1	+	1	1	
12	7		0	0	+	0	2	
13	8	+	1	1				
14	медіана				n = 11			
15	7							
16								
17	$v_{кр}$	$\tau_{кр}$	кількість серій $v_{сп}$	максимальна серія $\tau_{сп}$	кількість серій $v_{сп}$	максимальна серія $\tau_{сп}$	$v_{кр}$	$\tau_{кр}$
18	3	3	8	2	7	2	5	5
19	H_0 :вибірка випадкова та результати спостережень незалежні				H_0 :вибірка випадкова			
20								
21	Висновок: гіпотеза справедлива				Висновок: гіпотеза справедлива			

Рис. 12

формулу “ =ЕСЛИ(И(E18>G18;F18<H18);"гіпотеза справедлива";"гіпотеза не справедлива")“.

Увага. У вище наведених формулах адреси треба скоригувати відповідно до своїх даних.

Перевірка вибірки на однорідність за критерієм χ^2 .

Маємо дві вибірки об'ємами n_1 та n_2 . Кожна з них поділена на k інтервалів. Кількість спостерігаємих на кожному інтервалі значень $m_1^1, m_2^1, \dots, m_k^1$ та $m_1^2, m_2^2, \dots, m_k^2$ відповідно для першої та другої вибірок.

Нульова гіпотеза H_0 : вибірки 1 та 2 однорідні приймається, якщо виконується нерівність:

$$\chi_{набл}^2 \leq \chi_{кр}^2(f, \alpha)$$

де $\chi_{набл}^2 = n_1 \cdot n_2 \cdot \sum_{i=1}^k \frac{1}{m_i^1 + m_i^2} \left(\frac{m_i^1}{n_1} - \frac{m_i^2}{n_2} \right)^2$ - значення критерію; яке ми спостерігаємо;

$\chi_{кр}^2(f, \alpha)$ - критичне значення критерію для рівня значності α та степенів вільності $f = k - 1$.

Завдання. Перевірити при рівні значності $\alpha=0,05$ вибірки об'ємами n_1 та n_2 на однорідність за критерієм χ^2 .

Приклад. Маємо першу вибірку, об'єм якої $n_1=11$: 3, 7, 9, 4, 11, 9, 6, 10, 5, 7, 8; та другу вибірку, об'єм якої $n_2=15$: 5, 6, 10, 9, 6, 4, 7, 10, 5, 8, 6, 3, 9, 6, 11.

Вигляд таблиці Excel для вирішення цієї задачі та необхідні формули наведені на рис. 13.

До клітини B1 вводимо кількість інтервалів 4.

До діапазону клітин A3:A13 вводимо вибірку №1, до діапазону клітин B3:B17 - вибірку №2.

Спочатку знайдемо границі (ліву та праву) інтервалів, для чого необхідно знати найменше та найбільше значення, за якими ми можемо знайти довжину інтервалу. До клітини E11 вводимо формулу “=МАКС(A3:A13;B3:B17)”, до E12 – “=МИН(A3:A13;B3:B17), а до E13 – “=(E11-E12)/B1”. Номери інтервалів введемо до діапазону клітин C4:C7. Ліву границю порахуємо за формулою “=E\$12+(C4-1)*E\$13”, яку вводимо до

діапазону D4:D7. Праву границю - за формулою ” = $\$E\$12+C4*\$E\13 ”, яку вводимо до діапазону E4:E7.

Скориставшись командою меню *Сервис-Анализ данных...-Гистограмма* знайдемо кількість спостережуваних на кожному інтервалі значень $m_1^1, m_2^1, \dots, m_k^1$ та $m_1^2, m_2^2, \dots, m_k^2$ відповідно для першої та другої вибірок. Для першої вибірки вказуємо вхідний інтервал A3:A13, інтервал карманів E4:E8, вихідний інтервал F3. Для другої вибірки вказуємо вхідний інтервал B3:B17, інтервал карманів E4:E7, вихідний інтервал H3. Таким чином, маємо в G4:G7, I4:I7 необхідні нам частоти.

Кількість значень для першої та другої вибірок (n_1 та n_2) порахуємо скориставшись формулою ”=СЧЁТЗ(A3:A17)” та ”=СЧЁТЗ(B3:B17)”, які внесемо відповідно до клітин G11 та G12.

Для підрахунку $\chi_{набл}^2$ визначимо в діапазоні J4:J7 значення виразу $\frac{1}{m_i^1 + m_i^2} \left(\frac{m_i^1}{n_1} - \frac{m_i^2}{n_2} \right)^2$, для чого введемо в цей діапазон формулу

	A	B	C	D	E	F	G	H	I	J
1	$k=$	4								
2	вибірка з генеральної сукупності	вибірка з генеральної сукупності	№ інтервалу (i)	ліва границя i-го інтервалу	права границя i-го інтервалу	сукупність№1		сукупність№2		допоміжний
3	3	5				<i>Карман част от</i>		<i>Карман част от a</i>		
4	7	6	1	3	5	5	3	5	4	5E-06
5	9	10	2	5	7	7	3	7	5	0,0005
6	4	9	3	7	9	9	3	9	3	0,0009
7	11	6	4	9	11	11	2	11	2	0,0006
8	9	4				Еще	0	Еще	0	
9	6	7								
10	10	10								
11	5	5		$\max=$ 11			$n_1=$ 11			
12	7	8		$\min=$ 3			$n_2=$ 15			
13	8	6		$\Delta=$ 2						
14		3								
15		9								
16		6								
17		10		$\chi_{набл}^2 =$	0,32	$\chi_{кр}^2 (f, \alpha) =$	7,81			
18										
19										
20				H ₀ : вибірки 1 та 2 однорідні						
21				Висновок: гіпотеза справедлива						

"=1/(G4+I4)*(G4/H\$11-I4/H\$12)^2", нарешті у клітині E17 запишемо формулу "=СУММ(J4:J7)*H11*H12".

Значення $\chi_{кр}^2(f, \alpha)$ одержимо в клітині H17 за допомогою формули "=ХИ2ОБР(0,05;B1-1)".

Залишилось перевірити виконання нерівності та зробити висновок.

До клітини D20 вводимо "H0:вибірki 1 та 2 однорідні", до клітини D21 – "Висновок:", до клітини E21 – формулу "=ЕСЛИ(E17<=H17;"гіпотеза справедлива";"гіпотеза не справедлива)".

Увага. У вище наведених формулах адреси треба скоригувати відповідно до своїх даних.

Лабораторна робота №7

Перевірка закону розподілу генеральної сукупності за даними вибірки.

Критерій конкордації χ^2 (критерій Пірсона).

Співставлення гіпотези про розподіл генеральної сукупності за вибірковими даними здійснюється за допомогою статистичних критеріїв, які називаються критеріями конкордації.

Нульова гіпотеза H_0 формулюється так: вибірка вилучена з генеральної сукупності, яка має закон розподілу $F(x)$.

Розходження між гіпотетичним і емпіричним законами розподілу визначається величиною

$$\chi^2 = \sum_{i=1}^k \frac{(m_i - np_i)^2}{np_i}, \quad (1)$$

де n - об'єм вибірки, m_i – абсолютна частота значення x_i (для дискретних величин) або частота, відповідна i -му інтервалу (для безперервних величин), k - кількість різних значень випадкової величини X (для дискретних величин) або інтервалів (для безперервних величин), p_i – теоретичні ймовірності влучення випадкової величини X в i -й інтервал. У цієї величини закон розподілу приблизно збігається з розподілом χ^2 зі ступінем вільності $f = k - l$ -

1, де l - кількість параметрів гіпотетичного розподілу, які оцінюються за вибіркою.

Критичне значення критерію визначається в залежності від рівня значності α та ступіня вільності f , таким чином $\chi_{кр}^2 = \chi^2(f, \alpha)$.

Перевірка гіпотез про закон розподілу за критерієм конкордації χ^2 здійснюється у такий послідовності:

1. Утворюється статистичний ряд розподілу.
2. Статистичний ряд розподілу зображується графічно (гістограма або полігон).
3. Формулюється нульова гіпотеза H_0 про закон розподілу $F(x)$ випадкової величини, яку ми спостерігаємо.
4. Визначаються оцінки для кожного із параметрів гіпотетичного розподілу.
5. Визначається значення критерію $\chi_{сн}^2$ за формулою (1). (Інтервали з $m_i < 5$ об'єднуються)
6. Визначаються значення рівня значності α та ступіня вільності f .
7. Визначається критичне значення $\chi_{кр}^2 = \chi^2(f, \alpha)$.
8. Порівнюються $\chi_{сн}^2$ з $\chi_{кр}^2$, якщо $\chi_{сн}^2 > \chi_{кр}^2$, гіпотеза H_0 відхиляється, якщо $\chi_{сн}^2 \leq \chi_{кр}^2$, гіпотеза H_0 приймається.

Застосування критерію χ^2 для різних розподілів відрізняється тільки способом розрахунку величини p_i (теоретичної імовірності).

Для нормального розподілу
$$p(x_i < X < x_{i+1}) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_x} \int_{x_i}^{x_{i+1}} e^{-\frac{(x-\tilde{m}_x)^2}{2\tilde{\sigma}_x^2}} dx,$$

або

$$p(x_i < X < x_{i+1}) = \Phi\left(\frac{x_{i+1} - \tilde{m}_x}{\tilde{\sigma}_x}\right) - \Phi\left(\frac{x_i - \tilde{m}_x}{\tilde{\sigma}_x}\right),$$

де $\Phi\left(\frac{x - \tilde{m}_x}{\tilde{\sigma}_x}\right) = \int_0^{\frac{x - \tilde{m}_x}{\tilde{\sigma}_x}} e^{-\frac{(x - \tilde{m}_x)^2}{2\tilde{\sigma}_x^2}} dx$ - функція Лапласа, x_i, x_{i+1} - начало та кінець i -го інтервалу.

Для розподілу Пуассона $p_i = p(x_i) = \frac{\tilde{a}^{x_i}}{x_i!} e^{-\tilde{a}}$.

Для показового розподілу $p_i = e^{-\tilde{\lambda}x_i} - e^{-\tilde{\lambda}x_{i+1}}; (i = 1, 2, \dots, k)$.

Для рівномірного розподілу $p_1 = \frac{|x_1 - \tilde{a}|}{b - a};$

$p_i = \frac{|x_i - x_{i+1}|}{b - a} (i = 2, \dots, k - 1);$

$p_k = \frac{|\tilde{b} - x_{k-1}|}{b - a}.$

Завдання. Сформулювати гіпотезу про закон розподілу генеральної сукупності за даними вибірки. Перевірити гіпотезу при рівні значності $\alpha=0,01$ за критерієм χ^2 .

Приклад. Були виміряні 100 деталей. Відхилення від заданого розміру наведені в табл. 2.

Вигляд таблиці Excel для вирішення цієї задачі та необхідні формули наведені на рис. 14.

1. Утворимо інтервальний варіаційний ряд, тому що відхилення від заданого розміру є безперервною випадковою величиною.

Заголовки стовпців занесемо у перший рядок. У стовбець А внесемо дані вибірки. Скопіюємо ці дані у стовбець В та відсортуємо за зростанням.

Далі у клітини С2, С4, С6 відповідно запишемо формули $"=(B101-B2)/(1+3,3*LOG10(100))"$, $"=ОКРУГЛВВЕРХ((B101-B2)/C2;0)+1"$, $"СЧЁТ(A2:A101)"$ для визначення довжини інтервалу, кількості інтервалів та об'єму вибірки.

У діапазоні клітин D2:E10 визначаємо ліві та праві границі інтервалів:
D2 - $"=B2-C2/2"$, D2: D10 - $"=ЕСЛИ(D2+С$2<В$101;D2+С$2;"інтервалів")"$, E2:E10 - $"=D6+С$2"$.

У діапазоні клітин F2:F10 визначаємо кількість значень, що потрапили в інтервал $"=СЧЁТЕСЛИ(A$2:A$101;СЦЕПИТЬ("<";ТЕКСТ(E2; "Основной"))"$ - $СЧЁТЕСЛИ(A$2:A$101;СЦЕПИТЬ("<";ТЕКСТ(D2; "Основной")))"$.

2. Получимо графічне зображення наших даних – гістограму (статистичний аналог щільності розподілу генеральної сукупності). Зробимо це за допомогою команди:

Сервис→Анализ данных→Гистограмма...

Указати: вхідний інтервал A2:A101, інтервал карманів D2:D10, поставити флажок *вывод графика*, вихідний інтервал C14.

3. Зважаючи на гістограму висуваємо гіпотезу H_0 : відхилення від заданого розміру підпорядковані нормальному закону розподілу.

Занесемо в клітини D34 "H0:вибірка вилучена з генеральної сукупності, яка має закон розподілу нормальний", E37 – "кількість параметрів", G37- "2", F38 – "середнє", F39 - " σ_x ".

4. Визначимо оцінки параметрів нормального закону розподілу. Занесемо в клітини C8, C10 формули для розрахунку середнього та стандартного (квадратичного) відхилення відповідно
"=СУММПРОИЗВ((D2:D10+E2:E10)/2; F2:F10)/\$C\$6",
"=КОРЕНЬ(СУММПРОИЗВ(F2:F10;((D2:D10+E2:E10)/2-\$C\$8)^2)/(\$C\$6-1)",
а в клітинах G38, G39 - "=C8", "=C10".

5. Визначимо значення критерію χ^2_{cn} за формулою (1).

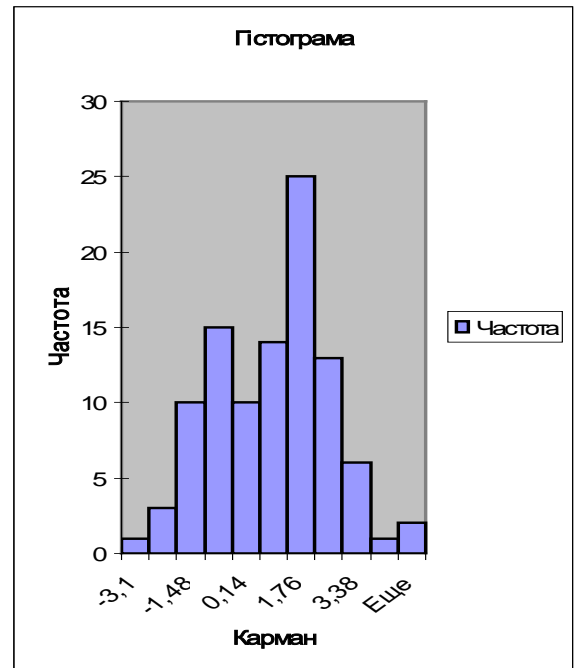
У діапазоні клітин G2:G10 визначаємо імовірність "=НОРМРАСП(E2; \$C\$8; \$C\$10;ИСТИНА) - НОРМРАСП(D2;\$C\$8;\$C\$10;ИСТИНА)", у H2:H10 – добуток імовірності та об'єму вибірки, у діапазоні клітин I2:I10 - "(F2-H2)^2/H2". В клітині I11 запишемо " =СУММ(I2:I10)", це і буде спостерігаємо значення критерію χ^2_{cn} .

6. Значення критерію $\chi^2_{кр}$ визначається в клітині I12 за формулою
"=ХИ2ОБР(0,01;\$E\$11-\$G\$37-1)".

7. Залишається порівняти критичне і спостерігаємо значення та зробити висновок. В клітинах D40, E40 відповідно запишемо "Висновок:",
"=ЕСЛИ(I11<I12;"гіпотеза справедлива";"гіпотеза не справедлива)".

Якщо висновок вас не влаштує, спробуйте об'єднати інтервали з $m_i < 5$, бо саме ці інтервали вносять найбільший вклад до спостерігаємого значення критерію.

вибірка з генеральної сукупності	варіаційний ряд	довжина інтервалу Δx	x_i	x_{i+1}	m_i	p_i	np_i	$\frac{(m_i - np_i)^2}{np_i}$	
-1,8	-3,1	1,07	-3,63	-2,57	3	0,03	2,972	0,000	
0,3	-3	кількість інтервалів в	-2,57	-1,50	11	0,09	8,550	0,702	
0,1	-2,8	9	-1,50	-0,44	15	0,17	17,040	0,244	
1,3	-2,5	об'єм вибірки n	-0,44	0,63	22	0,24	23,531	0,100	
2,1	-1,9	100	0,63	1,70	27	0,23	22,518	0,892	
-2,5	-1,9	середнє	1,70	2,76	13	0,15	14,933	0,250	
-0,9	-1,9	0,502368	2,76	3,83	6	0,07	6,861	0,108	
1,4	-1,9	σ_x	3,83	4,89	2	0,02	2,183	0,015	
3,1	-1,9	1,732365	4,89	5,96	1	0,00	0,481	0,560	
2,1	-1,8	інтервалів			9	100	0,99	$\chi^2_{оп} =$	2,872
2,1	-1,8						$\chi^2_{кр} =$	16,812	
-0,9	-1,8								
1,1	-1,8	Карман	Частота						
1,2	-1,8		-3,1	1					
0,1	-0,9		-2,29	3					
0,7	-0,9		-1,48	10					
1,3	-0,9		-0,67	15					
2,2	-0,9		0,14	10					
-1,8	-0,9		0,95	14					
1,4	-0,9		1,76	25					
1,1	-0,9		2,57	13					
3,1	-0,8		3,38	6					
-0,9	-0,8		4,19	1					
4,1	-0,8	Еще		2					
-1,8	-0,8								
0,5	-0,8								
0,1	-0,8								
2,1	-0,8								
-0,9	-0,7								
0,3	0,1								
2,1	0,1								
1,3	0,1								
1,2	0,1								
-0,8	0,1								
0,3	0,1								
1,2	0,1								
-0,8	0,1								
0,1	0,1								
-1,9	0,1								
-3,1	0,2								
-0,7	0,2								
2,2	0,2								
0,1	0,2								
1,2	0,2								
0,4	0,3								

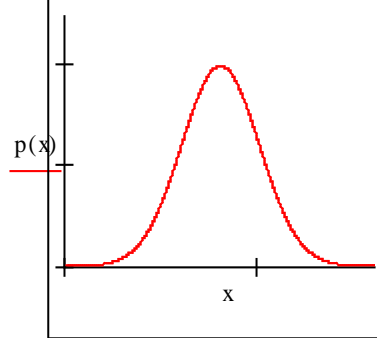
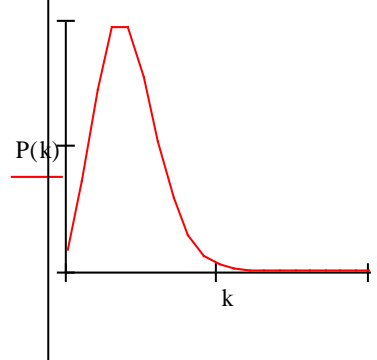
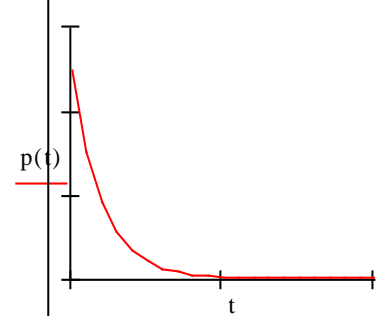
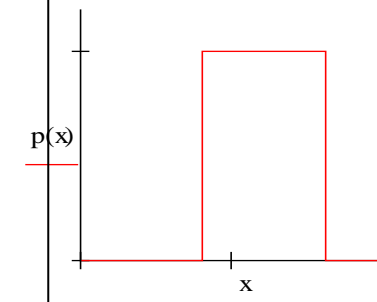


H_0 : вибірка вилучена з генеральної сукупності, яка має закон розподілу нормальний

кількість параметрів 2
середнє 0,502368
 σ_x 1,732365

Висновок: гіпотеза справедлива

Рис. 14

<p>Нормальний розподіл</p> 	$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}.$ <p>Для нормального розподілу оцінюються за вибіркою два параметри a та σ:</p> $\tilde{a} = \tilde{m}_x = \frac{1}{n} \sum_{i=1}^k m_i \bar{x}_i,$ $\tilde{\sigma}_x = \sqrt{\frac{1}{n} \sum_{i=1}^k m_i (\bar{x}_i - \tilde{m}_x)^2}.$
<p>Розподіл Пуассона</p> 	$P(k) = \frac{a^k}{k!} e^{-a} \quad (k = 0, 1, 2, \dots).$ <p>Для розподілу Пуассона оцінюється за вибіркою один параметр a:</p> $\tilde{a} = \tilde{m}_x = \frac{1}{n} \sum_{i=1}^k m_i \bar{x}_i.$ <p>Характерною ознакою розподілу Пуассона є рівність m_x та D_x.</p>
<p>Показовий розподіл</p> 	$p(t) = \begin{cases} \lambda e^{-\lambda t} & \text{при } t \geq 0 \\ 0 & \text{при } t < 0 \end{cases}.$ <p>Для показового розподілу оцінюється за вибіркою один параметр λ.</p> $\tilde{\lambda} = \frac{1}{\frac{1}{n} \sum_{i=1}^k m_i \bar{x}_i}.$
<p>Рівномірний розподіл</p> 	$p(x) = \begin{cases} C & \text{при } x \in (a, b) \\ 0 & \text{при } x \notin (a, b) \end{cases}.$ <p>Для рівномірного розподілу оцінюються за вибіркою два параметри a та b:</p> $\tilde{a} = \tilde{m}_x - \sqrt{3} \tilde{\sigma}_x,$ $\tilde{b} = \tilde{m}_x + \sqrt{3} \tilde{\sigma}_x.$

Зміст

ВСТУП.....	3
Лабораторна робота №1. Варіаційний ряд, статистичний ряд та його графічне зображення.....	9
Лабораторна робота №2. Емпірична функція розподілу.....	16
Лабораторна робота №3. Точечні оцінки числових характеристик та параметрів розподілу випадкової величини.....	19
Лабораторна робота №4. Вилучення результатів спостережень, які різко вирізняються.....	21
Лабораторна робота №5. Перевірка результатів на випадковість і незалежність.....	25
Лабораторна робота №6. Перевірка вибірки на однорідність за критерієм χ^2	30
Лабораторна робота №7 Перевірка закону розподілу генеральної сукупності за даними вибірки. Критерій конкордації χ^2 (критерій Пірсона).....	32