**Практична робота № 2**

**ВИКОРИСТАННЯ СТАТИСТИЧНИХ ФУНКЦІЙ**

**ПРОГРАМИ MATLAB**

**Мета роботи:**

* освоєння принципів роботи зі статистичними функціями програми **MatLab**;
* вивчення принципів роботи електронних таблиць **MatLab** з експериментальними даними;
* освоєння методики прийняття рішень на основі отриманих статистичних показників.

**Короткі теоретичні відомості**

Програма **MatLab** має багатий набор статистичних функцій, які використовуються для швидкої і елегантною оцінки різного роду імовірнісних та статистичних параметрів. Для їх застосування треба розбиратися в математичній статистиці.

Пакет аналізу є надбудовою **MatLab**. Нижче указані основні види (показники) аналізу з короткою характеристикою, підтримувані пакетом **MatLab**.

• Розрахунок набору статистичних показників (мода, медіана, дисперсія і т.д.) одновимірного набору даних.

• t-тест (t-тест Стьюдента) з однаковими і різними дисперсіями для перевірки гіпотези про рівність (розходженні) середніх двох вибірок.

• Однофакторний і двофакторний дисперсійний аналіз двох або більше вибірок, що належать одній і тій же генеральної сукупності.

• F-тест для порівняння дисперсій двох генеральних сукупностей.

• Кореляція для кількісної оцінки взаємозв'язку двох наборів даних, представлених у безрозмірному вигляді.

• Коваріація для обчислення середнього твору відхилень точок даних від відносних середніх.

• Експоненційне згладжування для передбачення значення на основі прогнозу для попереднього періоду, скоригованого з урахуванням похибок у цьому прогнозі.

• Аналіз Фур'є для вирішення завдань в лінійних системах та аналізу періодичних даних з використанням методу швидкого перетворення Фур'є.

• Гістограма для обчислення вибіркових та інтегральних частот попадання даних у вказані інтервали значень, при цьому генеруються числа влучень для заданого діапазону комірок.

• Ковзаюче середнє для розрахунку значень у прогнозованому періоді на основі середнього значення змінної для вказаного числа попередніх періодів.

• Генерація випадкових чисел, витягнутих з одного або декількох розподілів, для моделювання об'єктів, що мають випадкову природу.

• Регресія для аналізу впливу на окрему залежну змінну значень декількох незалежних змінних.

Найпростіший аналіз даних, що містяться в деякому масиві, полягає в пошуку його елементів з максимальним і мінімальним значеннями. В системі MATLAB визначені наступні швидкі функції для знаходження мінімальних і максимальних елементів масиву:

***mах (А)*** - повертає найбільший елемент, якщо А - вектор; або повертає вектор-рядок, що містить максимальні елементи кожного стовпця, якщо А - матриця, в багатовимірних масивах працює з першою не одиничною розмірності;

mах (А.В) - повертає масив того ж розміру, що А і В, кожен елемент якого є максимальний з відповідних елементів цих масивів;

max (A. [], dim) - повертає найбільші елементи по стовпцях або по рядках матриці в залежності від значення скаляра dim. Наприклад, тах (А, [], 1) повертає максимальні елементи кожного стовпця матриці А;

[C.I] = max (A) - крім максимальних значень повертає вектор індексів I цих елементів.

Для швидкого знаходження елемента масиву з мінімальним значенням служить наступна функція:

***min (A)*** - повертає мінімальний елемент, якщо А - вектор; або повертає вектор-рядок, що містить мінімальні елементи кожного стовпця, якщо А - матриця;

min (A.B) - повертає масив того ж розміру, що А і В, кожен елемент якого є мінімальний з відповідних елементів цих масивів;

min (A, [], dim) - повертає найменший елемент по стовпцях або по рядках матриці в залежності від значення скаляра dim. Наприклад, тах (А, [], 1) повертає мінімальні елементи кожного стовпця матриці А;

[C, I] = min (A) - крім мінімальних значень повертає вектор індексів цих елементів.

Елементарна статистична обробка даних в масиві зазвичай зводиться до знаходження їх середнього значення, медіани (серединного значення) і стандартного відхилення. Для цього в системі MATLAB визначені наступні функції:

***mean (А)*** - повертає арифметичне середнє значення елементів масиву, якщо А - вектор; або повертає вектор-рядок, що містить середні значення елементів кожного стовпця, якщо А - матриця. Арифметичне середнє значення є сума елементів масиву, поділена на їх число;

mean (A.dim) - повертає середнє значення елементів по стовпцях або по рядках матриці в залежності від значення скаляра dim (dim = l за стовпцями і dim = 2 по рядках відповідно).

median (A) - повертає медіану, якщо А - вектор; або вектор-рядок медіан для кожного стовпця, якщо А - матриця;

median (A.dim) - повертає значення медіан для стовпців або рядків матриці в залежності від значення скаляра dim.

***std (X)*** - повертає стандартне відхилення елементів масиву, що обчислюється за формулою якщо X - вектор. Якщо X - матриця, то std (X) повертає вектор-рядок, що містить стандартне відхилення елементів кожного стовпця (зверніть увагу, що воно відрізняється від середньоквадратичного відхилення);

std (X.flag) - повертає те ж значення, що і std (X), якщо flag = 0; якщо flag = l, функція std (X.l) повертає середньоквадратичне відхилення (квадратний корінь з несмещенной дисперсії), що обчислюється за формулою

**std (X.flag.dim)** - повертає стандартне або середньоквадратичне відхилення по рядах (dim = 2) або за стовпцями (dim = 1) матриці X в залежності від значення змінної dim.

*Statistics Toolbox* пропонує широкий спектр інструментів для статистичних обчислень. Основні можливості включають: регресійний аналіз і діагностика з вибором змінної, нелінійне моделювання, моделювання ймовірностей і оцінка параметрів, аналіз чутливості з використанням генератора випадкових чисел, управління статистичними процесами і планування експерименту. Пакет включає 20 різних розподілів ймовірностей, включаючи T, F і Хі-квадрат.Закони розподілу випадкових величин

***cdf (X, V)*** служить для розрахунку значень функції розподілу ймовірностей закону Стьюдента для значень випадкової величини Х і ступеня свободи V. Розмірність векторів або матриць X, V повинна бути однаковою. Розмірність скалярного параметра збільшується до розмірності іншого вхідного аргументу. Значення числа ступенів свободи V має бути позитивним цілим числом.

***fcdf (X, V1, V2)*** служить для розрахунку значення функції розподілу ймовірностей закону Фішера для параметрів розподілу V1, V2 і значення випадкової величини X. Розмірність векторів або матриць X, V1 і V2 повинна бути однаковою. Розмірність скалярного параметра збільшується до розміру інших вхідних аргументів. Параметри V1 і V2 повинні бути позитивними цілими числами.

**Виконання роботи**

* 1. Ввести в вигляді вектора в **MatLab** два набори вихідних даних з таблиці 2.1 відповідно до варіанта завдання, зазначеному у таблиці 2.2.

Таблиця 2.1

|  |
| --- |
| **Номер наборів вихідних даних** |
| **№ 1** | **№ 2** | **№ 3** | **№ 4** | **№ 5** | **№ 6** | **№ 7** | **№ 8** |
| 96 | 276 | 366 | 322 | 239 | 216 | 256 | 229 |
| 181 | 265 | 315 | 177 | 325 | 153 | 123 | 506 |
| 352 | 419 | 439 | 309 | 224 | 312 | 312 | 271 |
| 296 | 378 | 548 | 138 | 441 | 283 | 328 | 362 |
| 322 | 353 | 353 | 282 | 219 | 185 | 185 | 130 |
| 130 | 315 | 315 | 289 | 352 | 419 | 419 | 285 |
| 491 | 456 | 115 | 273 | 321 | 418 | 418 | 325 |
| 121 | 366 | 466 | 257 | 617 | 296 | 296 | 61 |
| 258 | 314 | 314 | 124 | 183 | 332 | 332 | 95 |
| 423 | 279 | 279 | 267 | 294 | 317 | 317 | 257 |
| 236 | 388 | 388 | 256 | 298 | 281 | 281 | 160 |
| 287 | 398 | 390 | 184 | 361 | 372 | 372 | 269 |
| 276 | 393 | 393 | 187 | 350 | 241 | 241 | 281 |
| 535 | 172 | 448 | 170 | 49 | 501 | 518 | 271 |
| 288 | 340 | 340 | 63 | 359 | 286 | 286 | 329 |
| 337 | 383 | 383 | 176 | 235 | 311 | 311 | 266 |
| 394 | 256 | 356 | 481 | 338 | 406 | 649 | 319 |
| 287 | 392 | 392 | 231 | 299 | 535 | 305 | 99 |
| 235 | 529 | 553 | 361 | 298 | 290 | 290 | 198 |
| 229 | 339 | 339 | 284 | 298 | 216 | 216 | 189 |

Таблиця 2.2

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| варіант | Номера наборів | варіант | Номера наборів | варіант | Номера наборів | варіант | Номера наборів | варіант | Номера наборів |
| 1 | 1,2 | 6 | 1,7 | 11 | 2,6 | 16 | 3,6 | 21 | 4,7 |
| 2 | 1,3 | 7 | 1,8 | 12 | 2,7 | 17 | 3,7 | 22 | 4,8 |
| 3 | 1,4 | 8 | 2,3 | 13 | 2,8 | 18 | 3,8 | 23 | 5,6 |
| 4 | 1,5 | 9 | 2,4 | 14 | 3,4 | 19 | 4,5 | 24 | 5,7 |
| 5 | 1,6 | 10 | 2,5 | 15 | 3,5 | 20 | 4,6 | 25 | 5,8 |

2. Використовуючи статистичні функції програми розрахувати: середнє значення, дисперсію і стандартне відхилення для кожного набору.

3. Визначити відносне відхилення мінімального і максимального значень вихідних даних в кожному наборі за формулою:

, (2.1)

де Хmin/max - мінімальне або максимальне значення в оброблюваному наборі,

S - експериментальне значення середньоквадратичного відхилення.

4. Визначити можливість виключення розглянутого показання шляхом порівняння отриманої величини зі значеннями t-критерію одностороннього розподілу Стьюдента. Аналізований показник може бути виключений з подальшої обробки, якщо ймовірність помилкової оцінки менше 0,025. Для решти даних провести розрахунки за п. 3.

5. Визначити значимість відмінностей розбіжностей середніх значень двох решти наборів даних. Для чого:

* оцінити можливу дисперсію узагальненого набору даних за формулою:

; (2.2)

* розрахуватикоефициент Стьюдента:

, (2.3)

* порівняти отриману величину з табличній при обраному рівні значущості і числі ступенів свободи ().

Якщо отримані значення перевершують табличні значення t-критерію двостороннього розподілу Стьюдента (ймовірність помилки не більше 0,025), то розбіжності середніх величин двох розподілів можна визнати значущими.

7. Порівняти дисперсії двох наборів і перевірити значимість їх розбіжностей по F-розподілу.