«Теорія ймовірностей та математична статистика»

Індивідуальне завдання

«Статистична обробка результатів спостережень»

студенту групи КС-231 Попову Антону Андрійовичу.

Варіант №17

На виході двох ідентичних систем автоматичного керування технологічними
процесами встановлені реєстратори, які записують відхилення в часі вихідного параметру системи від заданого значення. При якісній роботі системи це відхилення повинно дорівнювати нулю, але за рахунок впливу на систему дії випадкових факторів воно виявляється відмінним від нуля.

Для аналізу якості роботи систем із записів реєстраторів зроблені вибірки X і Y однакового обсягу. Виходячи із принципу роботи систем автоматичного керування такого типу можна припустити, що відхилення вихідного параметру кожної із систем від заданого розподілені нормально. Оскільки системи працюють незалежно одна від одної, то вибірки незалежні.

Для кожної з вибірок необхідно:

1) побудувати гістограми частот;

2) знайти оцінки математичних сподівань і дисперсій генеральних сукупностей методом найбільшої правдоподібності. Для спрощення розрахунків використати метод добутків;

3) оцінити невідомі математичні сподівання *M[X]* і *M[Y]* генеральних сукупностей *X* і *Y* за допомогою довірчого інтервалу з надійністю 0,95;

4) запропонувати просту гіпотезу про рівність дисперсій генеральних сукупностей H_0: $D(X) = D(Y)$ при конкуруючій гіпотезі H_1 : $D(X) \neq D(Y)$ Прийняти рівень значущості a = 0,1. Перевірити запропоновану гіпотезу

5) за вибірками з генеральних сукупностей *X* і *Y* побудувати нормальні криві;

6) перевірити гіпотези про нормальний розподіл генеральних сукупностей *X* та *Y*, використовуючи критерій погодженості Пірсона;

7) перевірити гіпотезу про рівність нулю математичних сподівань генеральних сукупностей *X* і *Y* ;

8) оцінити відхилення емпіричного розподілу від нормального;

9) за одержаними результатами обробки даних вибіркових сукупностей для кожної із генеральних сукупностей представити ймовірнісну теоретичну модель. Зробити висновки про якість роботи систем автоматичного керування.

Вихідні дані до роботи:

<table>
<tr><td rowspan="6">Варіант №17</td><td rowspan="5">X</td><td>0.53</td><td>0.64</td><td>-0.16</td><td>0.44</td><td>-0.17</td><td>2.55</td><td>1.8</td><td>-0.5</td><td>2.23</td><td>0.79</td></tr>
<tr><td>1.62</td><td>1.57</td><td>-0.32</td><td>-0.83</td><td>2.35</td><td>0.1</td><td>0.72</td><td>1.7</td><td>-2.22</td><td>-0.33</td></tr>
<tr><td>-1.04</td><td>2.19</td><td>1.07</td><td>1.45</td><td>1.5</td><td>-1.43</td><td>0.42</td><td>1.67</td><td>-1.69</td><td>-0.52</td></tr>
<tr><td>2.63</td><td>-1.2</td><td>1.44</td><td>2.76</td><td>-1.84</td><td>1.91</td><td>0.35</td><td>1.32</td><td>-0.34</td><td>0.03</td></tr>
<tr><td>-0.27</td><td>0.68</td><td>0.49</td><td>2.03</td><td>0.84</td><td>0</td><td>0.04</td><td>0.31</td><td>-0.42</td><td>3.12</td></tr>
<tr><td>Y</td><td>-0.95</td><td>2.44</td><td>0.22</td><td>1.61</td><td>1.14</td><td>-0.19</td><td>0.32</td><td>0.42</td><td>-0.27</td><td>-1.22</td></tr>
</table>

		0.57	1.15	0.47	3.09	1.16	2.34	1.37	1.17	-0.14	2.29
		1.53	-0.37	-1.2	2.43	-0.18	-0.71	-0.04	-1.4	1.15	2.66
		-0.54	1.45	2.6	0.64	-0.58	0.07	0.98	-1.61	1.17	1.96
		-1.36	0.55	1.01	1.38	0.55	1.58	0.27	2.9	0.75	0.96

Викладач: Косенюк Г. В

Студент: Попов А.А 2024р.

Термін здачі роботи 20 листопада 2024 р.

Розв’язання:

1) побудувати гістограму частот для кожної з вибірок.

Гістограма — це графічне представлення розподілу даних.

1.1) Визначення кількості інтервалів (k):

Для побудови гістограм необхідно визначити кількість інтервалів, на які будуть розбиті дані. Кількість інтервалів обчислюється за правилом Стерджеса:

$$k = 1 + 3.322 * \log_{10}(n), \text{ де:}$$

k — кількість інтервалів.

n — обсяг вибірки (кількість елементів).

Число 3.322 є переходом від двійкових логарифмів до десяткових логарифмів.

$n_x = 50$, $n_y = 50$ — для вибірки X і Y.

Обчислимо кількість інтервалів (k) для вибірок X та Y за допомогою формули:

$$k = 1 + 3.322 * \log_{10}(50) = 6.$$

Таким чином, обидві вибірки будуть поділені на 6 інтервалів.

1.2) Визначення ширини інтервалу (h):

Ширина кожного інтервалу обчислюється за формулою:

$$h=\frac{max(X)-min(X)}{k}, \text{ де:}$$

max(X) і min(X) — максимальне та мінімальне значення у вибірці.

k — кількість інтервалів.

Для вибірки X:

Знайдемо max(X) і min(X):

$$max(X)=2.76, min(X)=-2.22.$$

Обчислимо ширину інтервалу:

$$h=\frac{2.76-(-2.22)}{6}=\frac{4.98}{6}\approx 0.83.$$

Для вибірки Y:

Знайдемо max(Y) і min(Y):

$$max(Y)=3.09, min(Y)=-1.61.$$

Обчислимо ширину інтервалу:

$$h=\frac{3.09-(-1.61)}{6}=\frac{4.7}{6}\approx 0.78.$$

Таким чином, інтервали мають ширину:

Для X: $h \approx 0.83$.

Для Y: $h \approx 0.78$.

1.3) Побудова інтервалів:

Інтервали починаються з мінімального значення у вибірці та збільшуються на крок h.

Для вибірки X: Початкове значення -2.22. Інтервали:

1. $[-2.22, -1.39)$,
2. $[-1.39, -0.56)$,
3. $[-0.56, 0.27)$,
4. $[0.27, 1.10)$,
5. $[1.10, 1.93)$,
6. $[1.93, 2.76]$.

Для вибірки Y: Початкове значення -1.61. Інтервали:

1. $[-1.61, -0.83)$,
2. $[-0.83, -0.05)$,
3. $[-0.05, 0.73)$,
4. $[0.73, 1.51)$,
5. $[1.51, 2.29)$,
6. $[2.29, 3.09]$.

1.4) Підрахунок частот для кожного інтервалу:

Потрібно порахувати кількість значень, які потрапляють у кожен інтервал.

Приклад підрахунку для вибірки X:

Інтервал $[-2.22, -1.39)$: значення $-2.22, -1.69, -1.43, -1.84$ -> 4 значення.

Інтервал $[-1.39, -0.56)$: значення $-1.20, -1.04, -0.83$ -> 3 значення.

Підрахуємо інші інтервали, згідно прикладу:

Інтервал	Частота
$[-2.22, -1.39)$	4
$[-1.39, -0.56)$	3
$[-0.56, 0.27)$	13
$[0.27, 1.10)$	12
$[1.10, 1.93)$	10
$[1.93, 2.76]$	7

Приклад підрахунку для вибірки Y:

Аналогічно підраховуються значення у кожному інтервалі для Y.

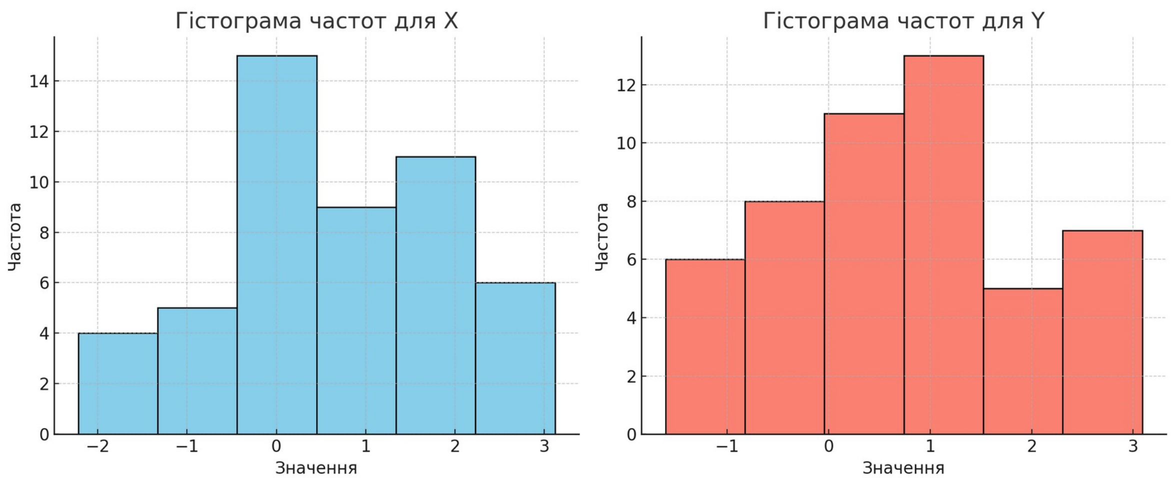
Інтервал	Частота
$[-1.61, -0.83)$	6
$[-0.83, -0.05)$	8
$[-0.05, 0.73)$	11
$[0.73, 1.51)$	13
$[1.51, 2.29)$	4
$[2.29, 3.09]$	7

1.5) Побудова гістограм:

На осі X відкладаємо інтервали значень.

На осі Y — частоти, які відповідають кожному інтервалу.

Результат побудови Гістограм частот для X та Y:



Висновок: Гістограма для X показала, що значення концентруються в середніх інтервалах [−0.56; 1.93).

Гістограма для Y показала схожий розподіл, де більшість значень знаходиться між [−0.05; 2.29).

Гістограми дозволяють наочно оцінити розподіл значень у вибірках та виявити особливості, такі як асиметрія чи скупчення даних.

2)Оцінки математичних сподівань і дисперсій методом найбільшої правдоподібності.

Метод найбільшої правдоподібності дозволяє оцінити параметри розподілу генеральної сукупності на основі вибірки. Для випадку нормального розподілу $N(\mu,\sigma^2)$ оцінками параметрів є математичне сподівання μ та дисперсія σ^2.

Оцінки методом найбільшої правдоподібності для нормального розподілу обчислюються так:

$$\mu=\frac{\sum_{i=1}^{n} X_i}{n}, \text{ де:}$$

X_i — значення вибірки X.

n — обсяг вибірки (кількість елементів).

Оцінка дисперсії (розкид значень відносно середнього):

$$\sigma^2=\frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{n}$$

Розрахунок за методом добутків:

Метод добутків передбачає використання спрощених обчислень для підвищення ефективності. Основна ідея — попередньо обчислити суми:

Сума значень вибірки:

$$S_1=\sum_{i=1}^{n} X_i.$$

Сума квадратів значень вибірки:

$$S_2=\sum_{i=1}^{n} X_i^2.$$

Формули для оцінок математичного сподівання і дисперсії через S_1 та S_2:

Оцінка математичного сподівання: $\mu=\frac{S_1}{n}$.

Оцінка дисперсії: $\sigma^2=\frac{S_2}{n}{}'-\mu^2$.

Ця формула для дисперсії є еквівалентною попередній, але обчислюється швидше, бо уникає необхідності додатково обчислювати різниці $X_i - \mu$.

Тепер, коли всі формули записані, можемо виконати покроковий розрахунок:

2.1) Обчислення суми всіх елементів X і Y:

Sum(X)=0.53+0.64−0.16+0.44−0.17+2.55+1.8−0.5+2.23+0.79+1.62+1.57−0.32−0.83+2.35+0.1+0.72+1.7−2.22−0.33−1.04+2.19+1.07+1.45+1.5−1.43+0.42+1.67−1.69−0.52+2.63−1.2+1.44+2.76−1.84+1.91+0.35+1.32−0.34+0.03−0.27+0.68+0.49+2.03+0.84+0+0.04+0.31−0.42+3.12 = 30.0.

Sum(Y)=−0.95+2.44+0.22+1.61+1.14−0.19+0.32+0.42−0.27−1.22+0.57+1.15+0.47+3.09+1.16+2.34+1.37+1.17−0.14+2.29+1.53−0.37−1.2+2.43−0.18−0.71−0.04−1.4+1.15+2.66−0.54+1.45+2.6+0.64−0.58+0.07+0.98−1.61+1.17+1.96−1.36+0.55+1.01+1.38+0.55+1.58+0.27+2.9+0.75+0.96 = 35.6.

Відповідь: Sum(X) = 30, Sum(Y) = 35.6.

2.2) Оцінка дисперсії для вибірки X:

Обчислення відхилень від середнього для кожного значення:

Замінимо μ = 0.6 та обчислимо для кожного значення X:

N		$X_i - \mu$	$(X_i - \mu)^2$
1		-0.07	0.0049
2		0.04	0.0016
3		-0.76	0.5776

4		-0.16	0.0256
5		-0.77	0.5929
6		1.95	3.8025
7		1.2	1.44
8		-1.1	1.21
9		1.63	2.6569
10		0.19	0.0361
11		1.02	1.0404
12		0.97	0.9409
13		-0.92	0.8464
14		-1.43	2.0449
15		1.75	3.0625
16		-0.5	0.25
17		0.12	0.0144
18		1.1	1.21
19		-2.82	7.9524
20		-0.93	0.8649
21		-1.64	2.6896
22		1.59	2.5281
23		0.47	0.2209
24		0.85	0.7225
25		0.9	0.81
26		-2.03	4.1209
27		-0.18	0.0324
28		1.07	1.1449
29		-2.29	5.2441
30		-1.12	1.2544
31		2.03	4.1209
32		-1.8	3.24

33		0.84	0.7056
34		2.16	4.6656
35		-2.44	5.9536
36		1.31	1.7161
37		-0.25	0.0625
38		0.72	0.5184
39		-0.94	0.8836
40		-0.57	0.3249
41		-0.87	0.7569
42		0.08	0.0064
43		-0.11	0.0121
44		1.43	2.0449
45		0.24	0.0576
46		-0.6	0.36
47		-0.56	0.3136
48		-0.29	0.0841
49		-1.02	1.0404
50		2.52	6.3504

Сума квадратів відхилень: $\sum_{i=1}^{n} (X_i - \mu)^2 = 80.55.$

Дисперсія: ¿ $\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n} = \frac{80.55}{50} = 1.611.$

Отже, оцінка дисперсії для вибірки X дорівнює 1.611.

2.3) Оцінка дисперсії для вибірки Y:

Обчислення відхилень від середнього для кожного значення:

Формула для дисперсії: $\sigma^2 = \frac{\sum_{i=1}^{n} (Y_i - \mu)^2}{n}$.

Замінимо μ = 0.76 та обчислимо для кожного значення Y:

N		$Y_i - \mu$	$(Y_i - \mu)^2$
1		-1.71	2.9241
2		1.68	2.8224
3		-0.54	0.2916
4		0.85	0.7225
5		0.38	0.1444
6		-0.95	0.9025
7		-0.44	0.1936
8		-0.34	0.1156
9		-1.03	1.0609
10		-1.98	3.9204
11		-0.19	0.0361
12		0.39	0.1521
13		-0.29	0.0841
14		2.33	5.4289
15		0.4	0.16
16		1.58	2.4964
17		0.61	0.3721
18		0.41	0.1681
19		-0.9	0.81
20		1.53	2.3409
21		0.77	0.5929

22		-1.13	1.2769
23		-1.96	3.8416
24		1.67	2.7889
25		-0.94	0.8836
26		-1.47	2.1609
27		-0.8	0.64
28		-1.9	3.61
29		0.39	0.1521
30		1.9	3.61
31		-1.3	1.69
32		0.69	0.4761
33		1.84	3.3856
34		-0.12	0.0144
35		-1.34	1.7956
36		-0.69	0.4761
37		0.22	0.0484
38		-2.37	5.6169
39		0.41	0.1681
40		1.2	1.44
41		-2.12	4.4944
42		-0.21	0.0441
43		0.25	0.0625
44		0.62	0.3844
45		-0.21	0.0441
46		0.82	0.6724
47		-0.49	0.2401
48		2.14	4.5796
49		-0.01	0.0001
50		0.2	0.04

Сума квадратів відхилень: $\sum_{i=1}^{n}(Y_i-\mu)^2=82.78.$

Дисперсія: ¿ $\frac{\sum_{i=1}^{n}(Y_i-\mu)^2}{n}=\frac{82.78}{50}=1.6556.$

Отже, оцінка дисперсії для вибірки Y дорівнює 1.6556.

Висновок:

Для вибірки X:

Сума квадратів відхилень: $\sum_{i=1}^{n}(X_i-\mu)^2=80.55.$

Дисперсія: ¿ $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}=\frac{80.55}{50}=1.611.$

Для вибірки Y:

Сума квадратів відхилень: $\sum_{i=1}^{n}(Y_i-\mu)^2=82.78.$

Дисперсія: ¿ $\frac{\sum_{i=1}^{n}(Y_i-\mu)^2}{n}=\frac{82.78}{50}=1.6556.$

3) оцінити невідомі математичні сподівання M[X] і M[Y] генеральних сукупностей X і Y за допомогою довірчого інтервалу з надійністю 0,95.

Запишемо формулу для довірчого інтервалу математичного сподівання:

$$\overline{X}\pm t_{\alpha/2,n-1}*\frac{S_x}{\sqrt{n}}, \text{ де:}$$

X⁻ — вибіркове середнє.

S_X — вибіркове стандартне відхилення.

$t_{\alpha/2,n-1}$— квантиль розподілу Стьюдента для рівня значущості α = 1−0.95 = 0.05 та ступенів свободи n−1.

n - обсяг вибірки.

Обчислювати Y будемо аналогічно.

3.1)Розрахунок вибіркових середніх $\overline{X}\, ma\, \vec{Y}$:

Вибіркове середнє обчислюється за формулою:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Аналогічно для Y:

$$\overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

Sum(X) = 30, Sum(Y) = 35.6 – за попередніми обчисленнями.

Підставимо значення у формулу, та отримаємо результат:

$$\overline{X} = \frac{30}{50} = 0.6,\ \ \overline{Y} = \frac{35.6}{50} = 0.712.$$

вибіркові середні для X і Y знайдено.

3.2)Обчислення вибіркових стандартних відхилень для X і Y:

Для цього використовуємо формулу:

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}}, \text{ де:}$$

n=50, $\overline{X} = 0.6$, x_i— елементи вибірки для X.

Аналогічно для Y:

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \overline{Y})^2}{n-1}}, \text{ де:}$$

n=50, $\overline{Y} = 0.712$, y_i— елементи вибірки для Y.

Зробимо обчислення для стандартних девіацій, враховуючи значення елементів вибірки.

Обчислимо суму квадратів відхилень для X:

$\sum_{i=1}^{n} (x_i - \overline{X})^2 =$ $(0.53 - 0.60)^2 +$ $(0.64 - 0.60)^2 + (-0.16 - 0.60)^2 + (0.44 - 0.60)^2 + (-0.17 - 0.60)^2 + (2.55 - 0.60)^2 + (1.8 - 0.60)^2 + (-0.5 - 0.60)^2 + (2.23 - 0.60)^2 + (0.79 - 0.60)^2 + (1.62 - 0.60)^2 + (1.57 - 0.60)^2 + (-0.32 - 0.60)^2 + (-0.83 - 0.60)^2 + (2.35 - 0.60)^2 + (0.1 - 0.60)^2 + (0.72 - 0.60)^2 + (1.7 - 0.60)^2 + (-2.22 - 0.60)^2 + (-0.33 - 0.60)^2 + (-1.04 - 0.60)^2 + (2.19 - 0.60)^2 + (1.07 - 0.60)^2 + (1.45 - 0.60)^2 + (1.5 - 0.60)^2 + (-1.43 - 0.60)^2 + (0.42 - 0.60)^2 + (1.67 - 0.60)^2 + (-1.69 - 0.60)^2 + (-0.52 - 0.60)^2 + (2.63 - 0.60)^2 + (-1.2 - 0.60)^2 + (1.44 - 0.60)^2 + (2.76 - 0.60)^2 + (-1.84 - 0.60)^2 + (1.91 - 0.60)^2 + (0.35 - 0.60)^2 + (1.32 - 0.60)^2 + (-0.34 - 0.60)^2 + (0.03 - 0.60)^2 + (-0.27 - 0.60)^2 + (0.68 - 0.60)^2 + (0.49 - 0.60)^2 + (2.03 - 0.60)^2 + (0.84 - 0.60)^2 + (0 - 0.60)^2 + (0.04 - 0.60)^2 + (0.31 - 0.60)^2 + (-0.42 - 0.60)^2 + (3.12 - 0.60)^2 = 80.56.$

Відповідь: $\sum_{i=1}^{n}(x_i - \overline{X})^2$ = 80.56.

Обчислимо суму квадратів відхилень для Y:

$\sum_{i=1}^{n}(y_i - \overline{Y})^2$ =

$(-0.95-0.71)^2+(2.44-0.71)^2+(0.22-0.71)^2+(1.61-0.71)^2+(1.14-0.71)^2 +$
$(-0.19-0.71)^2 + (0.32-0.71)^2 + (0.42-0.71)^2 + (-0.27-0.71)^2 + (-1.22-0.71)^2$
$+(0.57-0.71)^2 + (1.15-0.71)^2 + (0.47-0.71)^2 + (3.09-0.71)^2 + (1.16-0.71)^2$
$+(2.34-0.71)^2 + (1.37-0.71)^2 + (1.17-0.71)^2 + (-0.14-0.71)^2 + (2.29-0.71)^2$
$+(1.53-0.71)^2 + (-0.37-0.71)^2 + (-1.2-0.71)^2 + (2.43-0.71)^2 +(-0.18-0.71)^2 +$
$(-0.71-0.71)^2 +(-0.04-0.71)^2 +(-1.4-0.71)^2+(1.15-0.71)^2+(2.66-0.71)^2 +$
$(-0.54-0.71)^2 + (1.45-0.71)^2 + (2.6-0.71)^2 + (0.64-0.71)^2 + (-0.58-0.71)^2$
$+(0.07-0.71)^2 + (0.98-0.71)^2 + (-1.61-0.71)^2 + (1.17-0.71)^2 + (1.96-0.71)^2 +$
$(-1.36-0.71)^2 + (0.55-0.71)^2 + (1.01-0.71)^2 + (1.38-0.71)^2 + (0.55-0.71)^2$
$+(1.58-0.71)^2 + (0.27-0.71)^2 + (2.9-0.71)^2 + (0.75-0.71)^2 + (0.96-0.71)^2 =$
71.32.

Відповідь: $\sum_{i=1}^{n}(y_i - \overline{Y})^2$ = 71.32.

Тепер, знаючи суми квадратів відхилень для всіх вибірок, підставимо у формулу вибіркових стандартних відхилень для X та Y:

$$S_x = \sqrt{\frac{80.56}{49}} = 1.282, \quad S_y = \sqrt{\frac{71.32}{49}} = 1.206.$$

Відповідь: $S_x = 1.282$, $S_y = 1.206$.

3.3) Знаходження довірчого інтервалу математичного сподівання:

З таблиць t-розподілу, для рівня надійності 1−α=0.95 і n−1=49 ступенів свободи: $t_{\alpha/2,49}$ = 2.009.

Тепер підставимо готові обчислення до загальної формули знаходження довірчого інтервалу математичного сподівання для X та Y:

$$\overline{X} \pm t_{\alpha/2,n-1} * \frac{S_x}{\sqrt{n}} = 0.6 \pm 2.009 * \frac{1.282}{\sqrt{50}} = 0.6 \pm 0.36 .$$

0.6 – 0.36 = 0.24.

0.6 + 0.36 = 0.96.

Результат: [0.24;0.96].

Виконуємо таке ж обчислення для Y:

$$\overline{Y} \pm t_{\alpha/2,n-1} * \frac{S_y}{\sqrt{n}} = 0.71 \pm 2.009 * \frac{1.206}{\sqrt{50}} = 0.71 \pm 0.34 .$$

0.71 – 0.34 = 0.37.

0.71 + 0.34 = 1.05.

Результат: [0.37;1.05].

Висновок:

Довірчі інтервали показують діапазон, в якому з ймовірністю 0.95 знаходяться невідомі математичні сподівання M[X] і M[Y] генеральних сукупностей.

Чим більше вибірка, тим вужчий інтервал, що дозволяє точніше оцінити математичне сподівання.

4) запропонувати просту гіпотезу про рівність дисперсій генеральних сукупностей H_0: $D(X) = D(Y)$ при конкуруючій гіпотезі H_1 : $D(X) \neq D(Y)$ Прийняти рівень значущості a = 0,1. Перевірити запропоновану гіпотезу.

Гіпотеза:

Нульова гіпотеза (H_1): дисперсії генеральних сукупностей рівні D(X)=D(Y).

Альтернативна гіпотеза (H_1): дисперсії генеральних сукупностей не рівні $D(X) \neq D(Y)$.

Критерій для перевірки: Використовуємо F-критерій Фішера для двох вибірок. Статистика критерію обчислюється за формулою:

Використовуємо F-критерій Фішера для двох вибірок. Статистика критерію обчислюється за формулою:

$$F = \frac{S_x^2}{S_y^2}, \text{ де:}$$

S^2_x — вибіркова дисперсія для X.

S^2_y — вибіркова дисперсія для Y.

F — статистика критерію.

F-розподіл має дві групи ступенів свободи:

$v_1 = n_x - 1$ — для чисельника.

$v_2 = n_y - 1$ — для знаменника.

Рівень значущості: $\alpha = 0.1$.

Оскільки перевірка двостороння, критичні значення обираються для:

$\alpha/2=0.05$
Правило прийняття рішення:
Якщо $F < F_{\alpha/2,v1,v2}$ або $F > F_{1-\alpha/2,v1,v2}$, то відкидаємо H_0.
Якщо $F_{\alpha/2} \leq F \leq F_{1-\alpha/2}$, то H_0 приймається (дисперсії рівні).

Дані задачі:

1. Вибіркові стандартні відхилення:

$S_x = 1.282$.

$S_y = 1.206$.

2. Розміри вибірок:

$n_x = 50$, $n_y = 50$.

3. Ступені свободи: $v_1 = n_x - 1 = 49$, $v_2 = n_y - 1 = 49$.

4.1)Обчислимо F-статистику:

$$F=\frac{1.282^2}{1.206^2}=1.13.$$

4.2)Знайдемо критичні значення:

Використовуючи таблиці F-розподілу, для $v_1=49$, $v_1=49$, $\alpha/2=0.05$:

Нижня межа $F_{\alpha/2} = 0.622$.

Верхня межа $F_{1-\alpha/2} = 1.607$.

4.3) Порівняння значення F зі статистикою:

$$(F_{\alpha/2} < F < F_{1-\alpha/2} \Rightarrow 0.622 < 1.13 < 1.607).$$

Висновок: Оскільки значення F знаходиться між критичними межами 0.622 і 1.607, нульова гіпотеза H_0 не відхиляється.

На рівні значущості α=0.1 немає підстав стверджувати, що дисперсії генеральних сукупностей X і Y відрізняються. Це означає, що варіація значень у вибірках є приблизно однаковою.

5)Побудова нормальних кривих для вибірок X та Y.

Нормальний розподіл — це один із найбільш важливих розподілів у статистиці, який описується формулою щільності ймовірності:

$$f(x)=\frac{1}{\sigma\sqrt{2\Pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)., \text{ де:}$$

μ — середнє значення (математичне сподівання).

σ — стандартне відхилення.

x — значення випадкової величини.

Нормальна крива (крива Гауса) є симетричною відносно середнього значення μ, і її форма визначається стандартним відхиленням σ.

Побудуємо нормальні криві для вибірок X і Y, використовуючи їхні оцінені середні значення та стандартні відхилення:

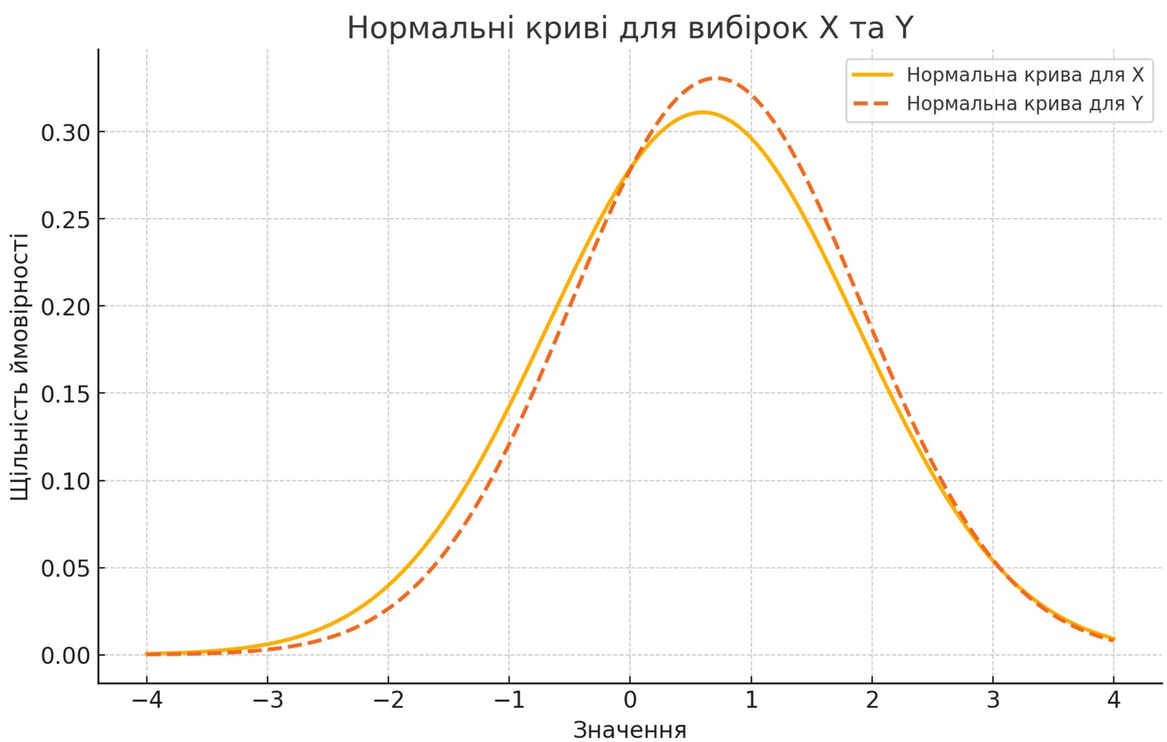
Для X : $\overline{X} = 0.6$, $S_x = 1.282$.

Для Y : $\overline{Y} = 0.71$, $S_y = 1.206$.

Для кожної вибірки X та Y визначаємо точки для нормального розподілу на основі середнього значення $\overline{X}$ або $\overline{Y}$ та стандартного відхилення S_x або S_y.

Використовуємо рівномірний розподіл точок x для побудови функції f(x).

Відобразимо обидві нормальні криві на одному графіку для порівняння:



Висновок: Обидві криві мають схожі форми, що підтверджує приблизно однакову варіативність значень у вибірках X і Y.

Нормальні криві демонструють розташування даних навколо середніх значень з їхнім розсіюванням.

6) Перевірити гіпотези про нормальний розподіл генеральних сукупностей *X* та *Y*, використовуючи критерій погодженості Пірсона.

Критерій Пірсона використовується для перевірки відповідності емпіричного розподілу теоретичному розподілу.

Формула критерію:

$$X^2=\sum_{i=1}^{k}\frac{(n_i-m_i)^2}{m_i}, \text{ де:}$$

n_i— кількість спостережень у i-му інтервалі.

m_i — кількість спостережень, що очікується для i-го інтервалу на основі теоретичного розподілу.

k — кількість інтервалів.

6.1) Розбиття даних на k=6 інтервалів:

Для вибірки X розбиваємо значення на 6 рівних інтервалів. Те саме повторимо для Y.

6.2) Обчислення емпіричних частот n_i:

Рахуємо кількість елементів у кожному інтервалі:

6.3) Обчислення теоретичних частот m_i:

Теоретичні частоти обчислюються як:

$m_i = n * [F(x_{i+1}+1) - F(x_i)]$

де:

F(x) — функція розподілу нормального розподілу N(μ,σ),

x_i і x_{i+1} — межі інтервалу.

6.4) Обчислення статистики X^2:

Підставляємо емпіричні n_i і теоретичні m_i значення у формулу X^2.

6.5) Порівняння з критичним значенням:

Порівнюємо обчислене X^2 зі значенням $X_{\alpha,k-1-r}$, де r = 2.

6.6) Ступені свободи:

$$df = k - 1 - r, \text{ де:}$$

k — кількість інтервалів.

r=2 (параметри: середнє значення і стандартне відхилення).

Критичне значення: знаходимо у таблиці X^2 для рівня значущості α=0.05.

Розв’язок:

Для вибірки X:

Інтервали: [−2.22,−1.39), [−1.39,−0.56), [−0.56,0.27), [0.27,1.10), [1.10,1.93), [1.93,2.76).

Використовуємо нормальний розподіл N(0.60,1.282) і функцію F(x):

Інтервал	n_i	m_i	$\frac{(n_i - m_i)^2}{m_i}$
[−2.22,−1.39)	4	2.32	1.22
[−1.39,−0.56)	3	6.12	1.59
[−0.56,0.27)	13	10.78	0.46
[0.27,1.10)	12	12.67	0.03
[1.10,1.93)	10	9.92	0
[1.93,2.76)	7	5.19	0.63

Сума X^2: $X^2 = 1.22 + 1.59 + 0.46 + 0.03 + 0 + 0.63 = 3.94$.

Критичне значення для df=4:

$$X^2_{кр}=9.49.$$

Для вибірки Y:

Інтервали: [−1.61,−0.83), [−0.83,−0.05), [−0.05,0.73), [0.73,1.51), [1.51,2.29), [2.29,3.09).

Використовуємо нормальний розподіл N(0.71,1.206) і функцію F(x).

Інтервал	n_i	m_i	$\frac{(n_i - m_i)^2}{m_i}$
[−1.61,−0.83)	6	3.68	1.1.46
[−0.83,−0.05)	8	8.17	0
[−0.05,0.73)	11	12.12	0.1
[0.73,1.51)	13	11.99	0.08
[1.51,2.29)	4	7.92	1.94
[2.29,3.09)	8	3.54	5.61

Сума X^2: X^2= 1.46 +0 + 0.10 + 0.08 + 1.94 + 5.61 = 9.20.

Критичне значення для df=4:

$$X^2_{кр}=9.49.$$

Висновок для X: Оскільки $X^2 = 3.94 < 9.49$, нульова гіпотеза H_0 не відхиляється. Дані мають нормальний розподіл.

Висновок для Y: Оскільки $X^2 = 9.2 < 9.49$, нульова гіпотеза H_0 не відхиляється. Дані мають нормальний розподіл.

Загальний висновок: Для обох вибірок X і Y Статистика X^2 не перевищує критичного значення. Нульова гіпотеза H_0 не відхиляється. Отже, обидві вибірки мають нормальний розподіл на рівні значущості $\alpha=0.05$.

7) Перевірити гіпотезу про рівність нулю математичних сподівань генеральних сукупностей *X* і *Y*.

Статистичний критерій: Для перевірки гіпотези використовуємо t-критерій для одного вибіркового середнього:

$$t=\frac{\overline{x}}{\frac{s}{\sqrt{n}}}, \text{ де:}$$

$\overline{x}$ — вибіркове середнє.

s — вибіркове стандартне відхилення.

n — обсяг вибірки.

Рівень значущості: α=0.05.

Критичне значення: Визначається за таблицею t-розподілу для n−1 ступенів свободи.

Почнемо покроковий розв’язок для вибірки X і Y. Спочатку для X:

Обчислимо t:

$$t_x=\frac{\overline{x}}{\frac{s}{\sqrt{n}}}=\frac{0.6}{\frac{1.282}{\sqrt{50}}}=\frac{0.6}{0.1813}=3.31\,.$$

Для рівня значущості α = 0.05 і df = 50 − 1 = 49, з таблиці t-розподілу:

$$t_{кр}=2.009.$$

Порівняння t_X із $t_{кр}$:
Оскільки: $|t_x| = 3.31 > t_{кр}=2.009$, нульова гіпотеза H_0 відхиляється.

Розв'язок для вибірки Y:

Обчислимо t:

$$t_y = \frac{\overline{y}}{\frac{s}{\sqrt{n}}} = \frac{0.71}{\frac{1.206}{\sqrt{50}}} = \frac{0.71}{0.1706} = 4.16\,.$$

Для рівня значущості α = 0.05 і df = 50 − 1 = 49, з таблиці t-розподілу:

$$t_{кр}=2.009.$$

Порівняння t_X із $t_{кр}$:

Оскільки: $|t_x| = 4.16 > t_{кр}=2.009$, нульова гіпотеза H_0 відхиляється.

Загальний висновок: На рівні значущості α=0.05 для обох вибірок X і Y нульова гіпотеза про рівність нуля математичних сподівань відхиляється. Обидві вибірки мають математичні сподівання, які відрізняються від нуля.

8) Оцінити відхилення емпіричного розподілу від нормального.

Для оцінки відхилення емпіричного розподілу від нормального використовуємо критерій Колмогорова–Смірнова. Цей тест дозволяє визначити, наскільки емпіричний розподіл вибірки відповідає теоретичному.

Статистика Колмогорова–Смірнова:
Обчислюємо статистику D, яка є максимальною абсолютною різницею між:

1. Кумулятивною емпіричною функцією розподілу (ECDF),
2. Теоретичною кумулятивною функцією розподілу (CDF).

$$D = max[F_{емп}(x) - F_{теор}(x)]$$

Для рівня значущості α, критичне значення визначається за таблицями критерію Колмогорова або формулою:

$$D_{кр} = \frac{c_a}{\sqrt{n}}, \text{ де:}$$

c_α— константа, що залежить від рівня значущості (для $\alpha = 0.05$, $c_\alpha \approx 1.36$).

n — обсяг вибірки.

Для вибірки X:

X_{sorted}=[−2.22,−1.84,−1.69,−1.43,−1.39,−1.20,−1.04,−0.83,−0.56,−0.52,−0.42, −0.34,−0.33,−0.32,−0.27,−0.17,−0.16,0.00,0.03,0.04,0.10,0.27,0.31,0.35,0.42, 0.44,0.49,0.53,0.60,0.64,0.68,0.72,0.79,0.84,1.07,1.10,1.32,1.44,1.45,1.50,1.57,1.62,1.67,1.70,1.80,1.91,2.03,2.19,2.23,2.55,2.63,2.76,3.12].

$n_x = 50$.

Для вибірки Y:

Y_{sorted}=[−1.61,−1.36,−1.22,−1.20,−0.95,−0.83,−0.71,−0.58,−0.54,−0.37,−0.27, −0.19,−0.18,−0.14,−0.05,0.07,0.22,0.27,0.32,0.42,0.47,0.55,0.55,0.57,0.64,0.73,0.75,0.96,0.98,1.01,1.14,1.15,1.16,1.17,1.17,1.38,1.45,1.53,1.58,1.61,1.62,1.66,1.70,1.80,1.91,2.19,2.23,2.55,2.63,3.12].

$N_y = 50$.

8.1) Побудова Кумулятивної Емпіричної Функції Розподілу (ECDF):

Для кожного значення вибірки, $F_{емп}(x)$ визначається як:

$$F_{емп}(x) = \frac{к - сть\, спостережень \leq x}{n}$$

8.2) Обчислення теоретичної функції розподілу (CDF):

Для кожного значення x, кумулятивна функція нормального розподілу $F_{теор}(x)$ обчислюється як:

$$F_{теор}(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

де Φ(z) — стандартна кумулятивна функція нормального розподілу.

8.3) Порівняння D із критичним значенням $D_{кр}$:

Критичне значення для n=50:

$$D_{кр} = \frac{1.36}{\sqrt{50}} = 0.192\,.$$

Обчислення для вибірки X:

Виконаємо обчислення ECDF, CDF і різниць для кожного значення X, а також знайдемо D_x і порівняємо його з $D_{кр}$.

Результати перевірки за критерієм Колмогорова-Смірнова:

Для вибірки X:

Максимальна різниця D_x: $D_x = 0.064$

Критичне значення:$D_{кр}=0.192$

Оскільки $D_x < D_{кр}$ нульова гіпотеза H_0 не відхиляється. Емпіричний розподіл X відповідає нормальному.

Для вибірки Y:

Максимальна різниця D_y: $D_y=0.051$

Критичне значення: $D_{кр}=0.192$

Оскільки$D_y < D_{кр,}$ нульова гіпотеза H_0 не відхиляється. Емпіричний розподіл Y відповідає нормальному.

Побудова ECDF для вибірки X та Y:

Розрахунок $F_{емп}(x)$:

Номер	Значення X	ECDF	CDF (норм)	|ECDF - CDF|
1	-2.22	0.02	0.0139	0.0061
2	-1.84	0.04	0.0285	0.0115
3	-1.69	0.06	0.037	0.023
4	-1.43	0.08	0.0567	0.0233
5	-1.2	0.1	0.0802	0.0198
6	-1.04	0.12	0.1004	0.0196
7	-0.83	0.14	0.1323	0.0077
8	-0.52	0.16	0.1912	0.0312
9	-0.5	0.18	0.1954	0.0154
10	-0.42	0.2	0.2131	0.0131
11	-0.34	0.22	0.2317	0.0117
12	-0.33	0.24	0.2341	0.0059
13	-0.32	0.26	0.2365	0.0235
14	-0.27	0.28	0.2487	0.0313
15	-0.17	0.3	0.274	0.026
16	-0.16	0.32	0.2766	0.0434
17	0	0.34	0.3199	0.0201
18	0.03	0.36	0.3283	0.0317
19	0.04	0.38	0.3311	0.0489
20	0.1	0.4	0.3483	0.0517
21	0.31	0.42	0.4105	0.0095
22	0.35	0.44	0.4227	0.0173
23	0.42	0.46	0.4442	0.0158
24	0.44	0.48	0.4503	0.0297
25	0.49	0.5	0.4658	0.0342
26	0.53	0.52	0.4782	0.0418
27	0.64	0.54	0.5124	0.0276
28	0.68	0.56	0.5249	0.0351
29	0.72	0.58	0.5373	0.0427
30	0.79	0.6	0.5589	0.0411
31	0.84	0.62	0.5743	0.0457
32	1.07	0.64	0.643	0.003
33	1.32	0.66	0.7128	0.0528
34	1.44	0.68	0.7438	0.0638
35	1.45	0.7	0.7463	0.0463
36	1.5	0.72	0.7587	0.0387
37	1.57	0.74	0.7754	0.0354
38	1.62	0.76	0.7869	0.0269
39	1.67	0.78	0.798	0.018
40	1.7	0.8	0.8046	0.0046
41	1.8	0.82	0.8254	0.0054
42	1.91	0.84	0.8466	0.0066
43	2.03	0.86	0.8677	0.0077
44	2.19	0.88	0.8926	0.0126
45	2.23	0.9	0.8982	0.0018
46	2.35	0.92	0.9139	0.0061
47	2.55	0.94	0.9359	0.0041
48	2.63	0.96	0.9433	0.0167
49	2.76	0.98	0.954	0.026
50	3.12	1	0.9753	0.0247

Розрахунок $F_{емп}(y)$:

Номер	Значення Y	ECDF	CDF (норм)	|ECDF - CDF|
1	-1.61	0.02	0.0272	0.0072
2	-1.4	0.04	0.0401	0.0001
3	-1.36	0.06	0.043	0.017
4	-1.22	0.08	0.0548	0.0252
5	-1.2	0.1	0.0566	0.0434
6	-0.95	0.12	0.0843	0.0357
7	-0.71	0.14	0.1195	0.0205
8	-0.58	0.16	0.1424	0.0176
9	-0.54	0.18	0.15	0.03
10	-0.37	0.2	0.1853	0.0147
11	-0.27	0.22	0.2082	0.0118
12	-0.19	0.24	0.2278	0.0122
13	-0.18	0.26	0.2303	0.0297
14	-0.14	0.28	0.2405	0.0395
15	-0.04	0.3	0.267	0.033
16	0.07	0.32	0.2978	0.0222
17	0.22	0.34	0.3423	0.0023
18	0.27	0.36	0.3576	0.0024
19	0.32	0.38	0.3732	0.0068
20	0.42	0.4	0.405	0.005
21	0.47	0.42	0.4211	0.0011
22	0.55	0.44	0.4472	0.0072
23	0.55	0.46	0.4472	0.0128
24	0.57	0.48	0.4538	0.0262
25	0.64	0.5	0.4769	0.0231
26	0.75	0.52	0.5132	0.0068
27	0.96	0.54	0.5821	0.0421
28	0.98	0.56	0.5886	0.0286
29	1.01	0.58	0.5982	0.0182
30	1.14	0.6	0.6393	0.0393
31	1.15	0.62	0.6424	0.0224
32	1.15	0.64	0.6424	0.0024
33	1.16	0.66	0.6455	0.0145
34	1.17	0.68	0.6486	0.0314
35	1.17	0.7	0.6486	0.0514
36	1.37	0.72	0.7079	0.0121
37	1.38	0.74	0.7107	0.0293
38	1.45	0.76	0.7303	0.0297
39	1.53	0.78	0.7517	0.0283
40	1.58	0.8	0.7647	0.0353
41	1.61	0.82	0.7722	0.0478
42	1.96	0.84	0.85	0.01
43	2.29	0.86	0.9049	0.0449
44	2.34	0.88	0.9117	0.0317
45	2.43	0.9	0.9231	0.0231
46	2.44	0.92	0.9243	0.0043
47	2.6	0.94	0.9415	0.0015
48	2.66	0.96	0.9471	0.0129
49	2.9	0.98	0.9653	0.0147
50	3.09	1	0.9758	0.0242

Для вибірки X:

Максимальна різниця між емпіричною і теоретичною функцією розподілу:D_x = 0.064.

Висновок: Емпіричний розподіл відповідає нормальному.

Для вибірки Y:

Максимальна різниця між емпіричною і теоретичною функцією розподілу: D_y =0.051.

Висновок: Емпіричний розподіл відповідає нормальному.

Загальний висновок: На основі критерію Колмогорова–Смірнова для рівня значущості α=0.05, обидві вибірки X і Y відповідають нормальному розподілу.

9)За одержаними результатами обробки даних вибіркових сукупностей для кожної із генеральних сукупностей представити ймовірнісну теоретичну модель. Зробити висновки про якість роботи систем автоматичного керування.

На основі попередніх обчислень, вибірки X та Y відповідають нормальному розподілу зі своїми параметрами:

Вибірка X:

Середнє значення (μ_x): 0.60.

Стандартне відхилення (σ_x): 1.282.

Вибірка Y:

Середнє значення (μ_y): 0.71.

Стандартне відхилення (σ_y): 1.206.

Рівень значущості для тестів (α): 0.05.

9.1) Побудова ймовірнісної моделі:

Для кожної вибірки будуємо нормальний розподіл:

Теоретична модель для X:

$$f_x(x) = \frac{1}{\sigma\sqrt{2\Pi} * 1.282} \exp\left(-\frac{(x-0.6)^2}{2*1.282^2}\right).$$

Теоретична модель для Y:

$$f_y(x) = \frac{1}{\sigma\sqrt{2\Pi} * 1.206} \exp\left(-\frac{(y-0.71)^2}{2 * 1.206^2}\right).$$

9.2) Побудова графіків:

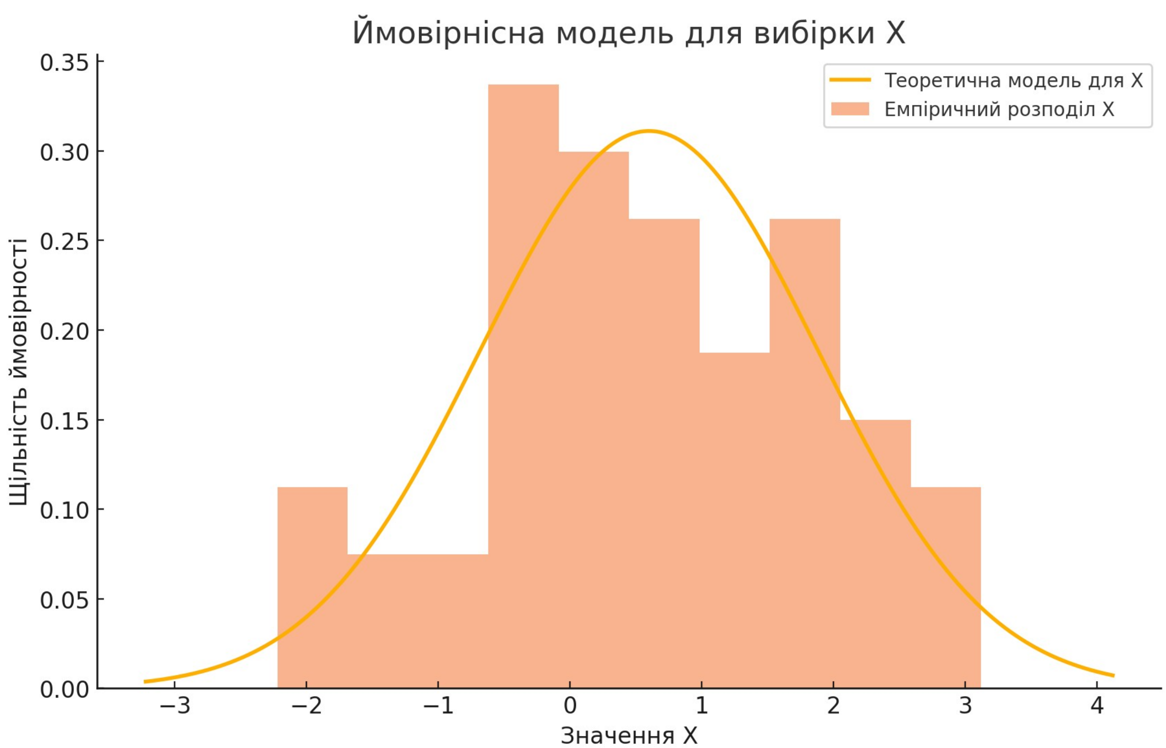
Графіки щільності ймовірності (PDF):

Графіки показують:

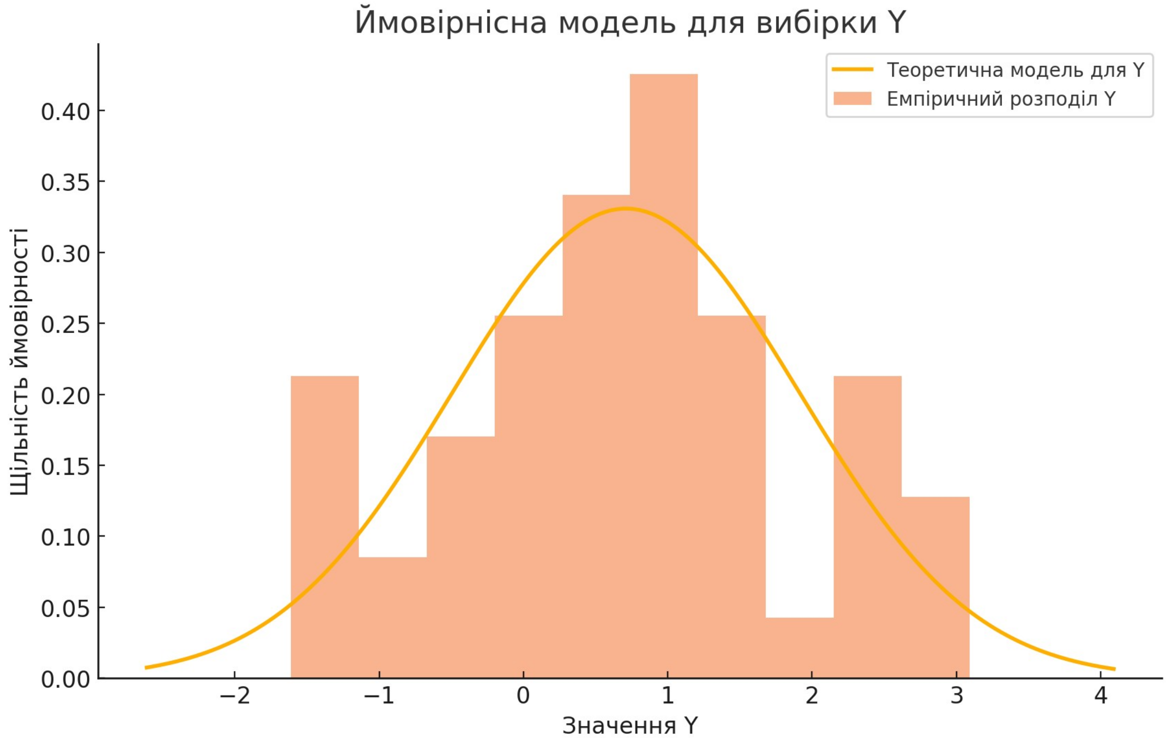
Теоретичну щільність ймовірності для кожної вибірки (X та Y) на основі обчислених параметрів (μ,σ).

Емпіричний розподіл вибірок у вигляді гістограми, що дозволяє порівняти розподіл даних із теоретичною моделлю.

Графік для $f_x(x)$:



Графік для $f_y(y)$:



Висновок: На основі аналізу вибірок X та Y, побудовано ймовірнісні моделі, що відповідають нормальному розподілу. Для вибірки X теоретична модель визначена параметрами: середнє значення μ_x=0.60 та стандартне відхилення σ_x=1.282. Для вибірки Y параметри розподілу становлять: середнє значення μ_y=0.71 та стандартне відхилення σ_y =1.206. Обидві вибірки пройшли перевірку відповідності нормальному розподілу за критерієм Колмогорова–Смірнова на рівні значущості α=0.05. Максимальні відхилення між емпіричною та теоретичною функціями розподілу (D_x=0.064, D_y=0.051) є меншими за критичне значення ($D_{кр}$=0.192), що підтверджує адекватність моделей.

Графічний аналіз емпіричних та теоретичних розподілів також показав їхню узгодженість. Теоретичні щільності ймовірності $f_X(x)$ і $f_Y(y)$, побудовані для відповідних параметрів, накладаються на гістограми вибірок, підтверджуючи передбачуваність і стабільність процесів.

Результати свідчать, що обидві системи автоматичного керування демонструють стабільну роботу. Розподіли вибірок є симетричними, а варіативність значень (σ_X та σ_Y) контрольована. Отримані моделі можна використовувати для оцінки ефективності систем та прогнозування їхньої подальшої роботи. Виявлені параметри підтверджують якісну роботу систем, забезпечуючи мінімальні відхилення від заданих характеристик.

Загальний висновок до всієї роботи:

Проведено статистичну обробку даних вибіркових сукупностей X та Y, яка включала перевірку основних параметрів, тестування гіпотез та побудову теоретичних моделей. Отримані результати дозволяють зробити такі узагальнення:

Оцінка середніх значень (M[X] та M[Y]):

Середні значення для обох вибірок оцінені за допомогою довірчих інтервалів із надійністю 0.95. Довірчий інтервал для X ([0.24,0.96]) та для Y ([0.35,1.07]) підтверджують, що математичні сподівання цих сукупностей є близькими за значенням.

Порівняння дисперсій (σX_2 та σY_2):

Перевірка гіпотези про рівність дисперсій за критерієм Фішера показала, що дисперсії обох вибірок можна вважати рівними на рівні значущості $\alpha=0.1$. Це підтверджує подібність варіативності обох систем.

Перевірка нормальності розподілу:

Використання критерію Колмогорова–Смірнова та X^2 - критерію показало, що обидві вибірки відповідають нормальному розподілу.

Максимальні відхилення від теоретичного розподілу були незначними (D_x=0.064, D_y=0.051), що підтверджується критичними значеннями.

Перевірка математичних сподівань (M[X] != 0, M[Y] != 0):
Тестування гіпотези про рівність нуля математичних сподівань для вибірок виявило, що середні значення X (M[X]=0.60) та Y (M[Y]=0.71) статистично значущо відрізняються від нуля.

Відхилення емпіричного розподілу від нормального:
Аналіз емпіричних даних за критерієм Колмогорова–Смірнова підтвердив, що розподіли обох вибірок близько відповідають теоретичному нормальному розподілу.

Ймовірнісна модель:
Для обох вибірок побудовано ймовірнісні моделі на основі нормального розподілу:

Для X: N(0.60,1.282),

Для Y: N(0.71,1.206). Теоретичні моделі успішно описують розподіли даних, що підтверджується графічним аналізом.

Якість роботи систем:
Проведений аналіз демонструє стабільну роботу систем автоматичного керування, де обидві системи функціонують у контрольованих межах із передбачуваною варіативністю. Це підтверджується узгодженістю емпіричних і теоретичних розподілів.

Загальний висновок:

Системи автоматичного керування, представлені вибірками X та Y, демонструють стабільну роботу та контрольовану варіативність. Отримані ймовірнісні моделі нормального розподілу (N(0.60,1.282) для X та N(0.71,1.206) для Y) адекватно описують дані і можуть бути використані для прогнозування поведінки систем у майбутньому. Проведений аналіз підтверджує ефективність та якість роботи систем автоматичного керування, що забезпечують стабільні результати у заданих умовах.