

Робота з курсу «Прикладні Статистичні Методи Обчислень»

Мета роботи

Потрібно дати відповіді на поставлені запитання:

- Чи є різниця між успіхами учнів у першому та другому семестрах?
- Чи є різниця між групами учнів?
- Чи існує залежність між групою учнів та відвідуванням додаткових занять?
- Як впливає проведення додаткових занять на канікулах та поглиблених занять після уроків на результати учнів?

Вихідні дані

Маємо Excel файл з таким наповненням

	A	B	C	D	E	F	G
1	ID	Group	Term	Add	Ext	Result	
2	A-1	A	1	FALSE	TRUE	8,59	
3	A-2	A	1	TRUE	TRUE	11,44	
4	A-3	A	1	FALSE	FALSE	8,4	
5	A-4	A	1	TRUE	FALSE	10,03	
6	A-5	A	1	FALSE	FALSE	10,9	
7	A-6	A	1	FALSE	FALSE	8,73	
8	A-7	A	1	TRUE	FALSE	7,25	
9	A-8	A	1	TRUE	FALSE	8,97	
10	A-9	A	1	TRUE	FALSE	7,55	

Та додатком з описом змінних у таблиці

#	Змінна	Опис
1	ID	Персональний ID учня для визначення у базі даних.
2	Group	Змінна, що визначає клас, у якому навчається учень.
3	Term	Визначення семестру, у якому зібрана інформація.
4	Add	Наявність чи відсутність додаткових занять на канікулах (осінь/весна) у поточному семестрі.
5	Ext	Наявність чи відсутність поглиблених занять після уроків у поточному семестрі.
6	Result	Середнє арифметичне оцінок усіх семестрових оцінок з усіх дисциплін.

Усі тестування проводяться для рівня $\alpha = 5\%$.

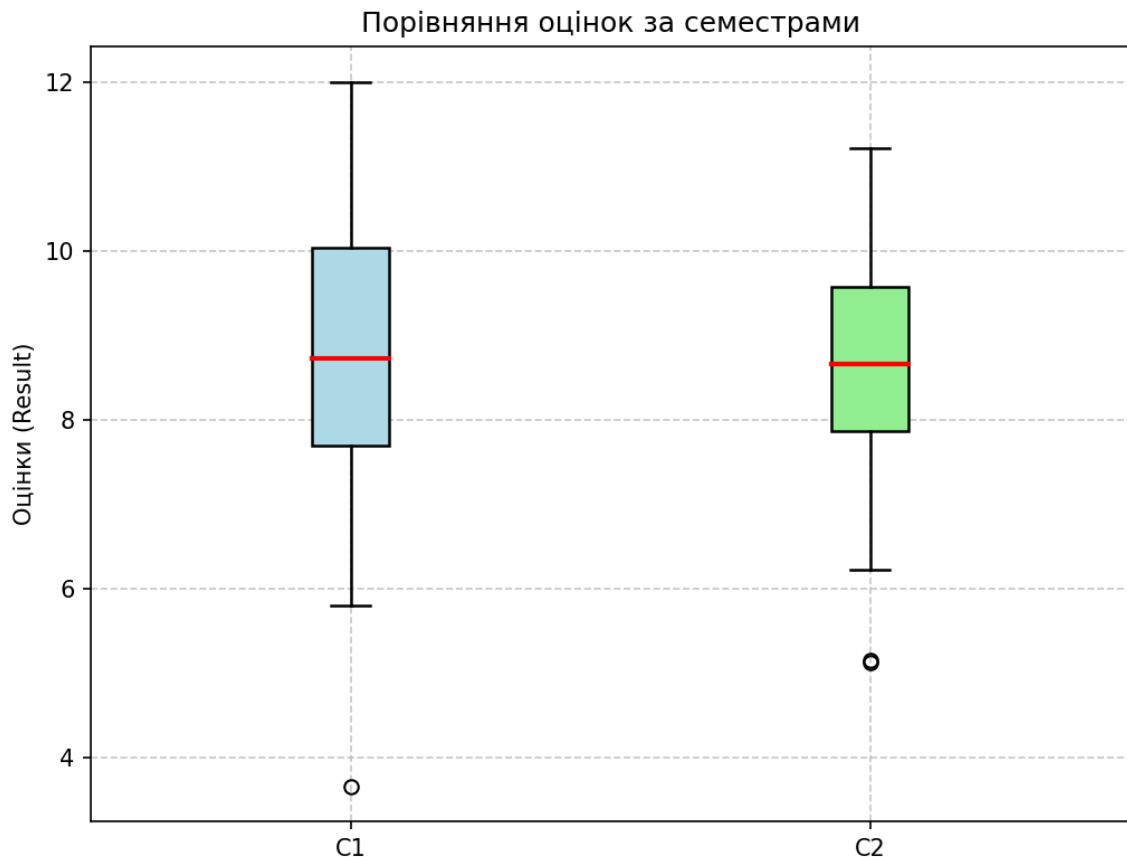
Хід виконання

1) Чи є різниця між успіхами учнів у першому та другому семестрах?

А) Формулювання нульової та альтернативної гіпотези:

Нульова гіпотеза (H_0): Середній бал учнів у першому семестрі не відрізняється від середнього балу у другому семестрі.

Альтернативна гіпотеза (H_1): Середній бал учнів у першому семестрі відрізняється від середнього балу у другому семестрі.



Статистики для 1-го семестру:

Математичне сподівання (середнє оцінок): 8.7242

Дисперсія: 2.6783

Стандартне відхилення: 1.6365

Квантиль рівня (25%, 50%, 75%): [7.69 8.73 10.03]

Статистики для 2-го семестру:

Математичне сподівання (середнє оцінок): 8.6397

Дисперсія: 1.8148

Стандартне відхилення: 1.3471

Квантиль рівня (25%, 50%, 75%): [7.86 8.66 9.57]

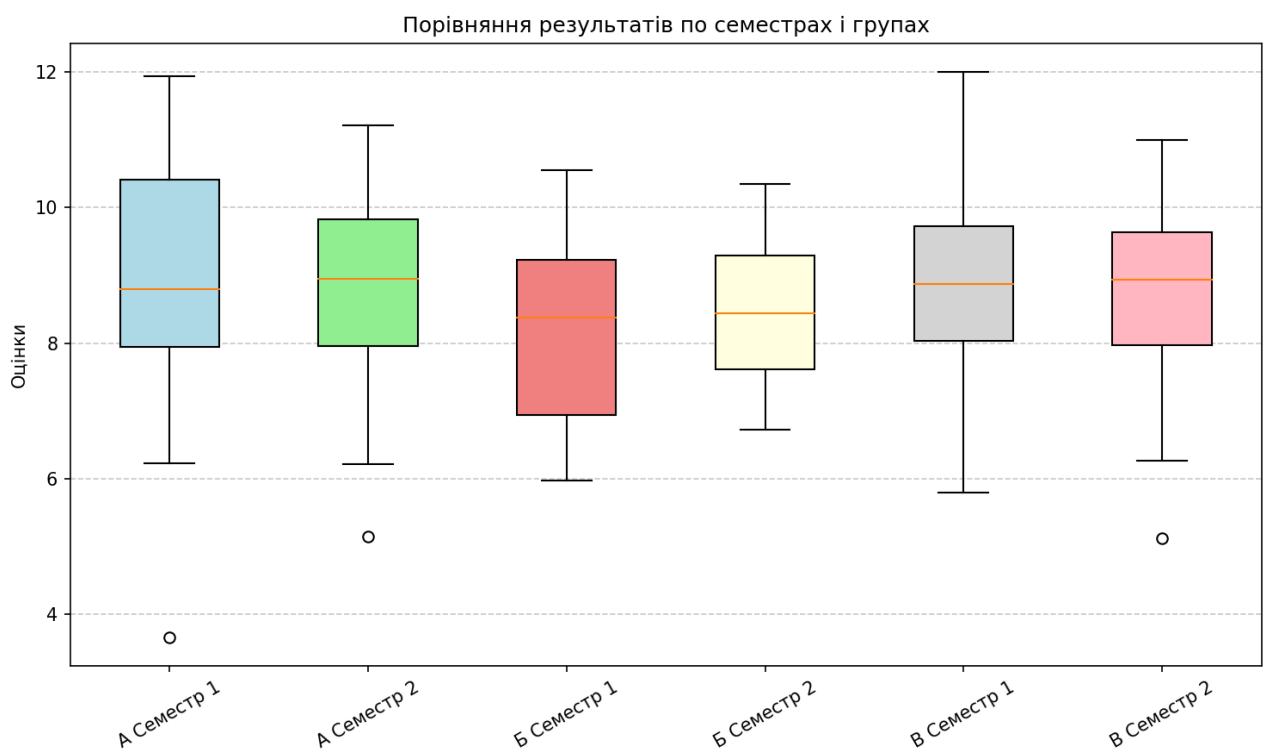
Парний t-тест: t-статистика = 1.109, p-value = 0.2710

З отриманих даних робимо висновок, що нульову гіпотезу потрібно прийняти, особливої різниці між результатами учнів у першому та другому семестрах не спостерігаються, на графіку присутні викиди значень, що вказує на те, що в декількох учнів з обох груп були значно гірші результати ніж у інших.

Б) Формулювання нульової та альтернативної гіпотези:

Н₀: Середній бал у 1-му семестрі не відрізняється від балу у 2-му семестрі для групи А, Б, В

Н₁: Середній бал у 1-му семестрі відрізняється від балу у 2-му семестрі для групи А, Б, В



Група А: t-статистика = 0.409, p-value = 0.6856

Група Б: t-статистика = -1.375, p-value = 0.1851

Група В: t-статистика = 1.542, p-value = 0.1355

Для всіх трьох груп (А, Б, В) можемо прийняти нульову гіпотезу.

Висновок:

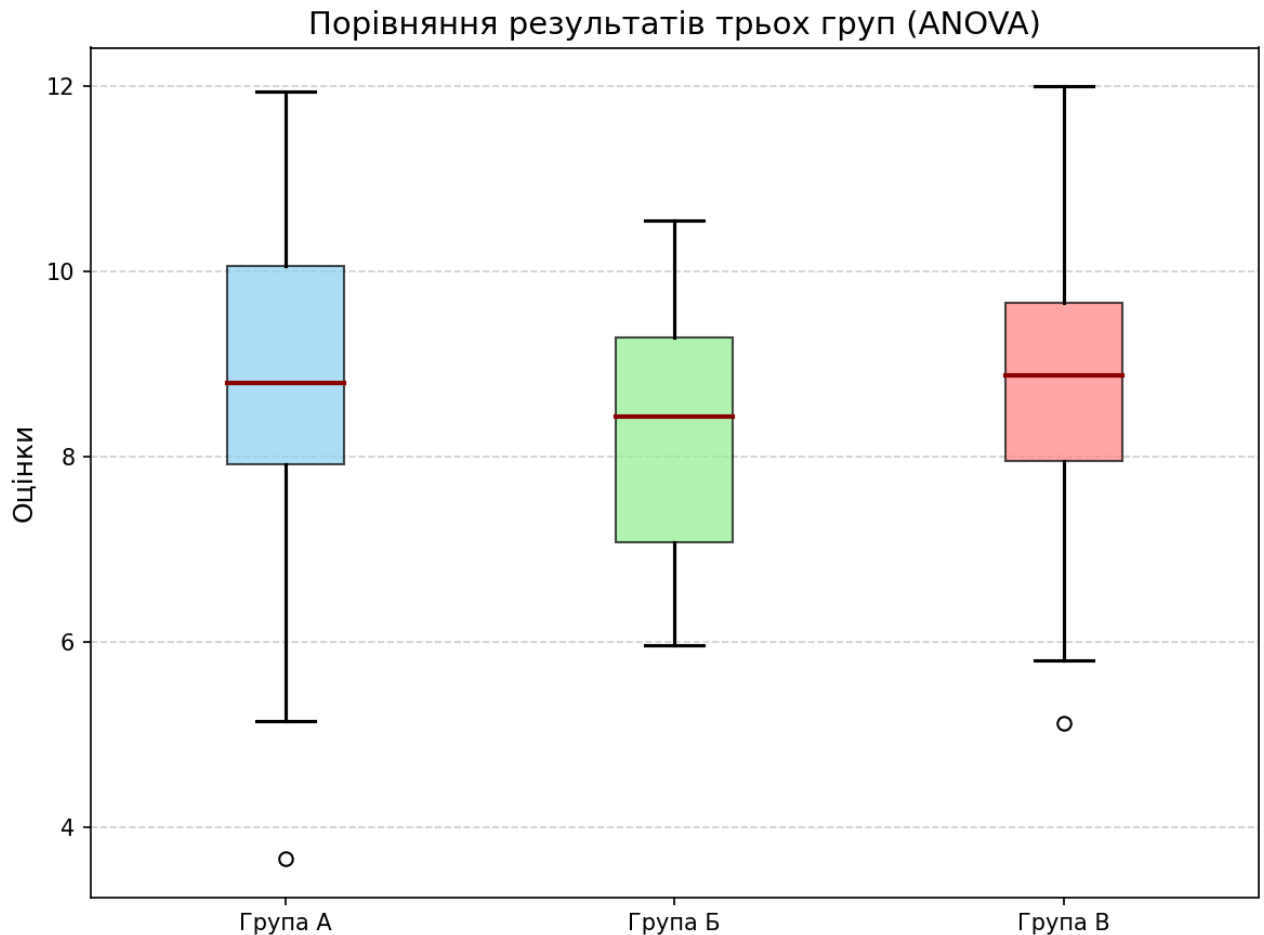
На основі аналізу оцінок учнів у першому та другому семестрах за допомогою парного t-тесту не виявлено статистично значущих змін в тому числі і між жодними з трьох груп.

2) Чи є різниця між групами учнів?

А) Формулювання нульової та альтернативної гіпотези:

Нульова гіпотеза (H_0): Середні оцінки учнів не відрізняються між групами А, Б і В

Альтернативна гіпотеза (H_1): Хоча б одна з груп має середню оцінку, що відрізняється від інших.



ANOVA F-статистика: 1.669

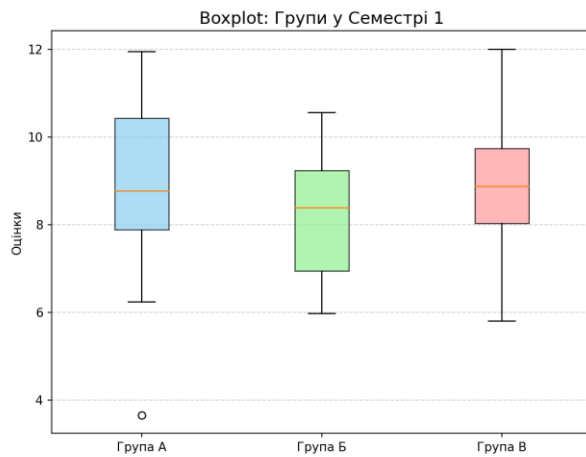
p-value: 0.1920

На основі результатів дисперсійного аналізу можна зробити висновок, що статистично значущої різниці між середніми оцінками учнів у групах А, Б, В не виявлено. За результатами ANOVA-аналізу середні оцінки у групах не відрізняються з точки зору статистики.

Б) Формулювання нульової та альтернативної гіпотези:

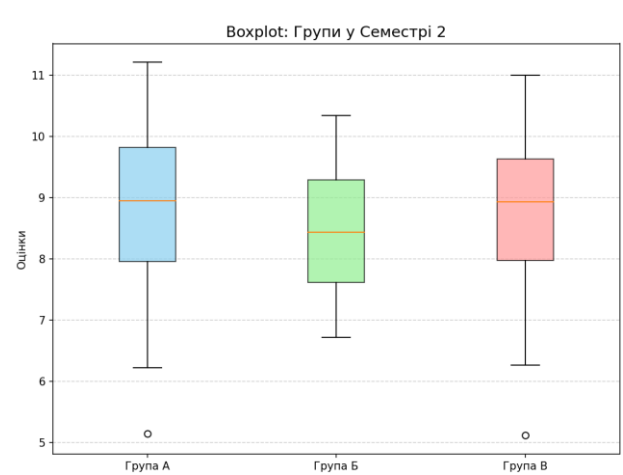
H_0 : Середні оцінки груп А, Б, В у першому та другому семестрі не відрізняються.

H_1 : Хоча б одна група має середній бал, що відрізняється.



Семестр 1:

F-статистика: 1.164, p-value: 0.3180



Семестр 2:

F-статистика: 0.519, p-value: 0.5974

Висновок:

Для обох семестрів не виявлено статистично значущої різниці між групами учнів. Це означає, що на основі даних не можна стверджувати, що середні значення оцінок між групами суттєво відрізняються. В обох випадках ми приймаємо нульову гіпотезу.

При цьому візуалізація показала невеликі зміщення в медіанах та деякі відмінності в розподілах, проте без різких розривів.

3) Чи існує залежність між групою учнів та відвідуванням додаткових занять?

А) Формулювання нульової та альтернативної гіпотези:

Нульова гіпотеза (H_0): Відвідуваність занять на канікулах не залежить від групи.

Альтернативна гіпотеза (H_1): Відвідуваність занять на канікулах залежить від групи.

Нище зображено таблицю з підрахованою кількістю учнів, що відвідували заняття влітку і що не відвідували їх, для кожної групи.

Add Group	False	True
A	28	27
Б	15	25
В	12	40

Критичне значення для $\chi^2 = 5.9914$

$\chi^2 = 8.8423$

Ступені свободи: 2

p-value = 0.0120

Результати тесту показали, що існує статистично значимий зв'язок між групою учня та фактом відвідування занять на канікулах. Тобто, відвідування додаткових занять на канікулах залежить від того, до якої групи належить учень. З таблиці видно, що від групи А до В, частка учнів, які відвідують додаткові заняття на канікулах зростає. Це може свідчити про різний рівень засвоєння матеріалу групами, що збільшує попит груп Б та В у додаткових заняттях влітку.

Б) Формулювання нульової та альтернативної гіпотези:

Нульова гіпотеза (H_0): Відвідуваність додаткових занять не залежить від групи.

Альтернативна гіпотеза (H_1): Відвідуваність додаткових занять залежить від групи.

Нище зображено таблицю з підрахованою кількістю учнів, що відвідували заняття влітку і що не відвідували їх, для кожної групи.

Ext	False	True
Group		
A	27	28
Б	23	17
В	35	17

Критичне значення $\chi^2 = 5.9914$

$\chi^2 = 3.6394$

Ступені свободи: 2

p-value = 0.1620

За результатами тесту, ми не можемо відхилити нульову гіпотезу. Між групами учнів та фактом відвідування додаткових занять після уроків не виявлено статистично значущого зв'язку. Хоча по таблиці видно невелику тенденцію до зменшення частки учнів що відвідують додаткові заняття від групи А до В.

Висновок:

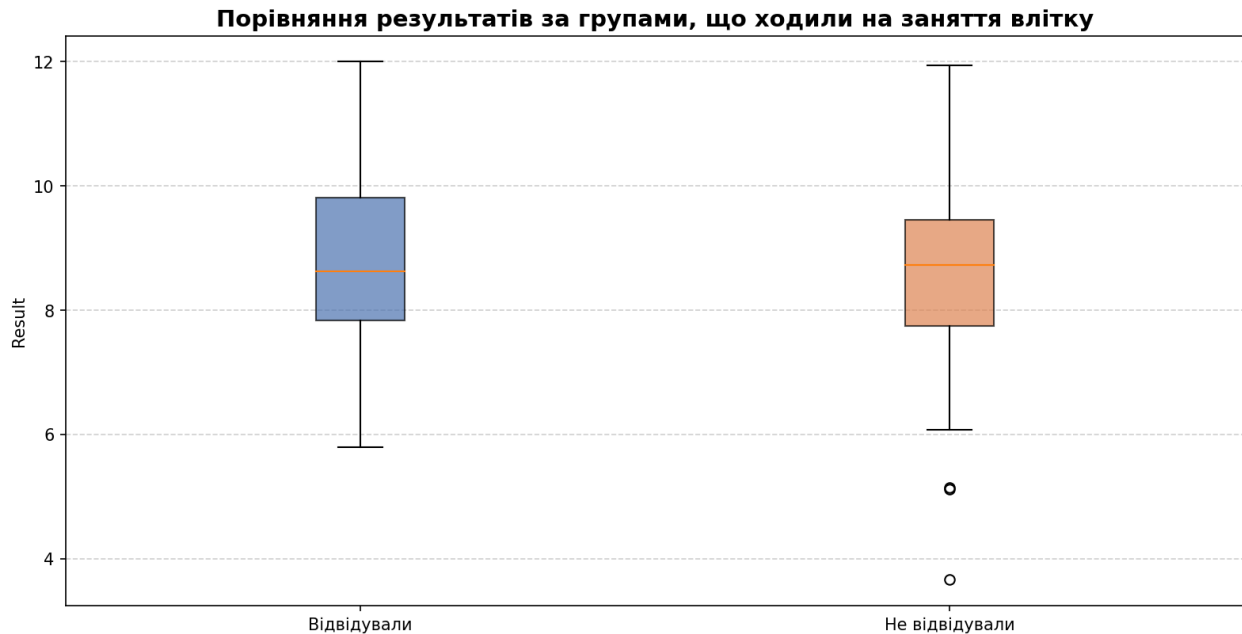
Дані свідчать, що учні різних груп по-різному відвідують заняття на канікулах, натомість відвідування занять після уроків є більш однорідним серед учнів різних груп – статистично значущої різниці не зафіксовано. Це може свідчити про більш цікаву форму викладання зі сторони викладачів, що ведуть літні заняття для груп Б та В.

4) Як впливає проведення додаткових занять на канікулах та поглиблених занять після уроків на результати учнів?

А) Формулювання нульової та альтернативної гіпотези:

Нульова гіпотеза (H_0): Середній бал учнів, які відвідували додаткові заняття на канікулах, не відрізняється від середнього балу учнів, які не відвідували такі заняття.

Альтернативна гіпотеза (H_1): Середній бал учнів відрізняється залежно від того, чи відвідували вони додаткові заняття на канікулах.



t-статистика: 0.6114

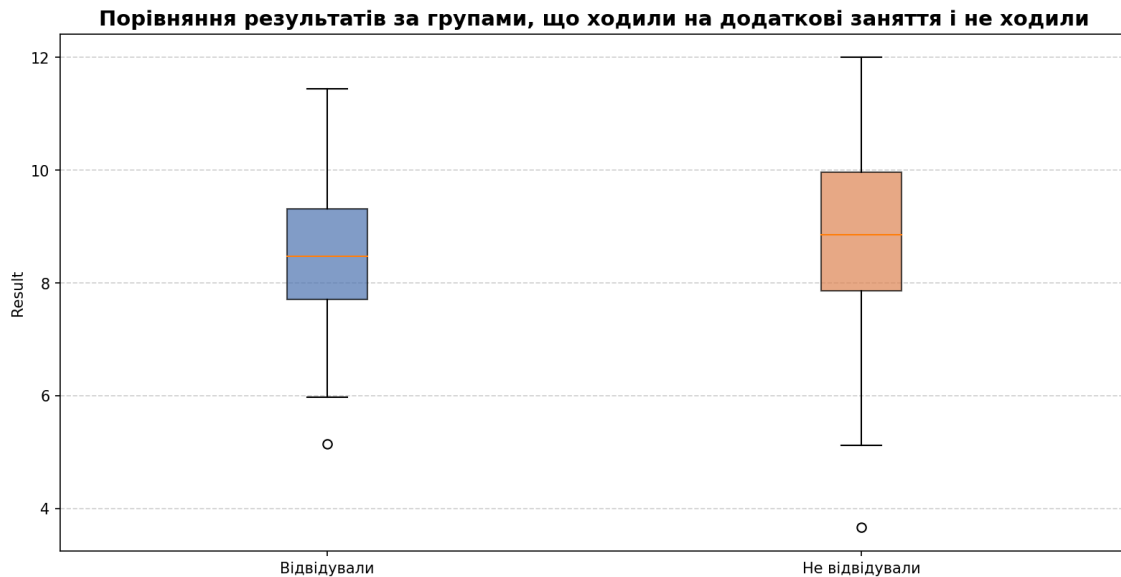
p-value: 0.5418

За результатами тесту ми приймаємо нульову гіпотезу. Це означає, що немає різниці між оцінками тих, хто ходив на заняття влітку, і тими, хто не відвідував їх. Проте на boxplot-ах видно викиди значень для учнів, що не відвідували заняття влітку, що вказує на наявність серед цієї групи учнів з дуже поганими результатами.

Б) Формулювання нульової та альтернативної гіпотези:

Нульова гіпотеза (H_0): Середній бал учнів, які відвідували поглиблені заняття після уроків, не відрізняється від середнього балу тих, хто їх не відвідував.

Альтернативна гіпотеза (H_1): Середній бал залежить від того, чи відвідував учень поглиблені заняття після уроків.



t-статистика: -1.4032

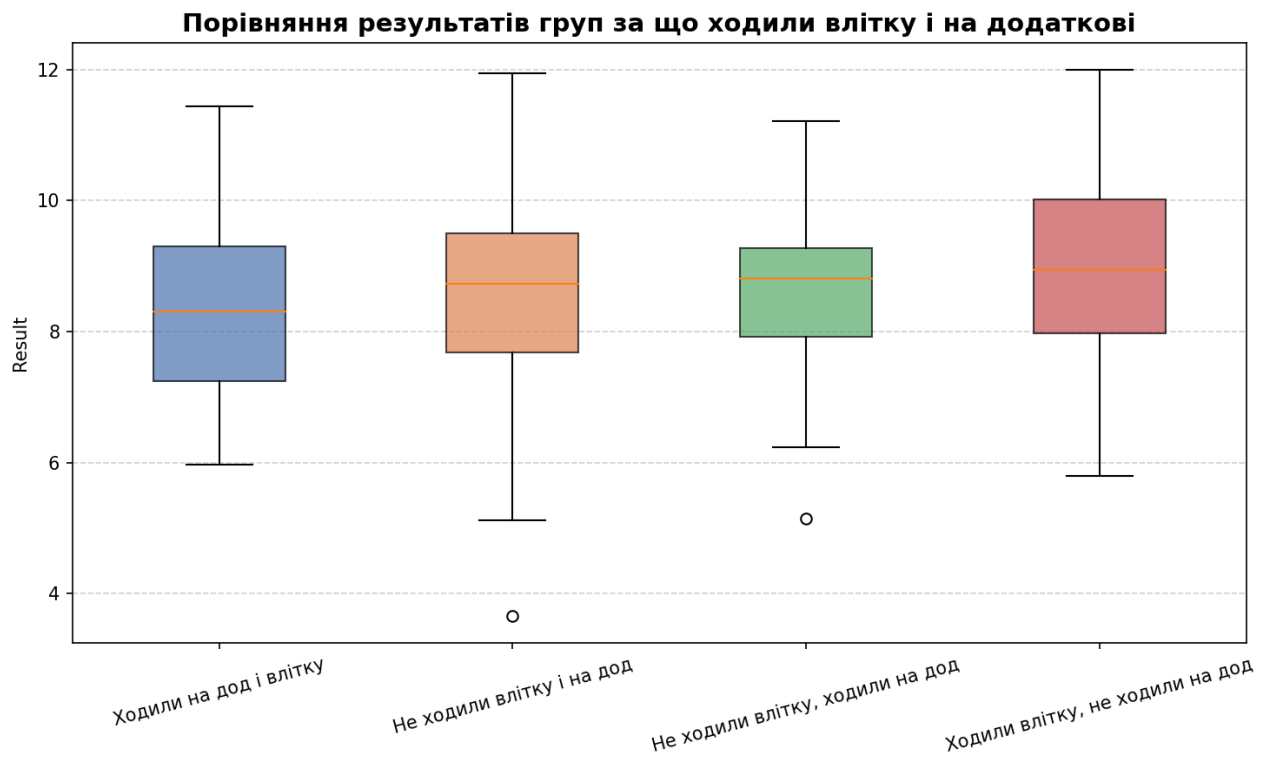
p-value: 0.1626

За результатами тесту ми не можемо прийняти альтернативну гіпотезу, різниці між середніми оцінками учнів, які відвідували та не відвідували додаткові заняття після уроків, не є статистично значущими. Тобто, відвідування таких занять не виявило достатнього впливу на успішність. Проте можна помітити з boxplot-ів, що учні, які не відвідували заняття мають більше відхилення від середньої оцінки, тоді як оцінки учнів, які ходили на заняття більш наближені до усередненого значення.

В) Формулювання нульової та альтернативної гіпотези:

Нульова гіпотеза (H_0): Середні оцінки учнів не відрізняються між чотирма комбінаціями груп (тобто фактори не мають впливу на результат).

Альтернативна гіпотеза (H_1): Принаймні одна з груп має середній бал, що значимо відрізняється від інших.



ANOVA F-статистика: 1.450

p-value: 0.2308

Не відхиляємо нульову гіпотезу. Між чотирма комбінаціями груп немає значущої різниці у середніх оцінках. Тобто жодна із стратегій відвідування занять не дала достатньо кращого чи гіршого результату в порівнянні з іншими.

Висновок: Отримані результати можуть свідчити про те, що самі по собі додаткові заняття не є головним фактором, що впливає на результати. Можливо, важливішу роль відіграє якість занять, мотивація учнів, індивідуальна підготовка, тощо.

Загальний висновок з дослідження:

У ході проведеного дослідження було проаналізовано успішність учнів у першому та другому семестрах, порівняння між трьома навчальними групами, а також вплив відвідування додаткових занять на результати навчання.

Встановлено, що статистично значущої різниці між оцінками учнів у першому та другому семестрах немає. Аналогічно, порівняння середніх балів між групами А, Б та В не виявило суттєвих відмінностей.

Дослідження залежності між групою учнів та відвідуванням додаткових занять показало, що відвідування літніх занять статистично залежить від групи, при цьому учні груп Б і В частіше користуються такою можливістю. Водночас відвідування поглиблених занять після уроків не має значущої кореляції з групою.

Щодо впливу додаткових занять на успішність, було встановлено, що відвідування як літніх, так і післяшкільних, поглиблених занять не призводить до статистично значущих змін у середніх оцінках учнів.

Отже, основні залежності, виявлені у дослідженні, свідчать, що рівень успішності учнів незначно відрізняється між семестрами та групами, а вплив додаткових занять на результати навчання є незначним. Це вказує на необхідність зосередитися не лише на кількості додаткових занять, а й на їхній якості, мотивації учнів та індивідуальному підході до навчання.