МІНІСТЕРСТВО ОСВІТИ I НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ   
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

ФАКУЛЬТЕТ БІОМЕДИЧНОЇ ІНЖЕНЕРІЇ

КАФЕДРА БІОМЕДИЧНОЇ КІБЕРНЕТИКИ

**Розрахунково - графічна робота**

з дисципліни «Обробка біомедичних даних»

Варіант №9

**Виконав:**

студент гр. БС-83

Кучинський В.В,

Зараховано від \_\_\_.\_\_\_.\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис викладача)

Київ-2020

Зміст

[1. Комп'ютерний практикум №1 3](#_Toc41139898)

[2. Комп'ютерний практикум №2 "Візуалізація даних за допомогою бібліотеки Matplotlib" 6](#_Toc41139899)

[3. Комп'ютерний практикум №3 "Візуалізація даних за допомогою бібліотеки Seaborn" 14](#_Toc41139900)

[5. Комп'ютерний практикум №5 "Критерій Пірсона для порівняння декількох груп за розподіленням ознаки" 22](#_Toc41139901)

[6. Комп'ютерний практикум №6 "Парний t-критерій Стьюдента" 27](#_Toc41139902)

[7. Комп'ютерний практикум №7 "t-критерій Стьюдента для незалежних вибірок" 31](#_Toc41139903)

[8. Комп'ютерний практикум №8 "Дисперсійний аналіз" 35](#_Toc41139904)

# 1. Комп'ютерний практикум №1

В даному комп'ютерному практикумі ми знайомились із бібліотекою pandas та базовою роботою з даними в середовищі Jupiter Notebook.

pandas - програмна бібліотека мови Python для обробки і аналізу даних. Робота

pandas з даними будується поверх бібліотеки NumPy, що є інструментом

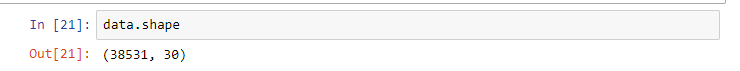
нижчого рівня. Надає спеціальні структури даних і операції для маніпулювання

числовими таблицями і тимчасовими рядами.

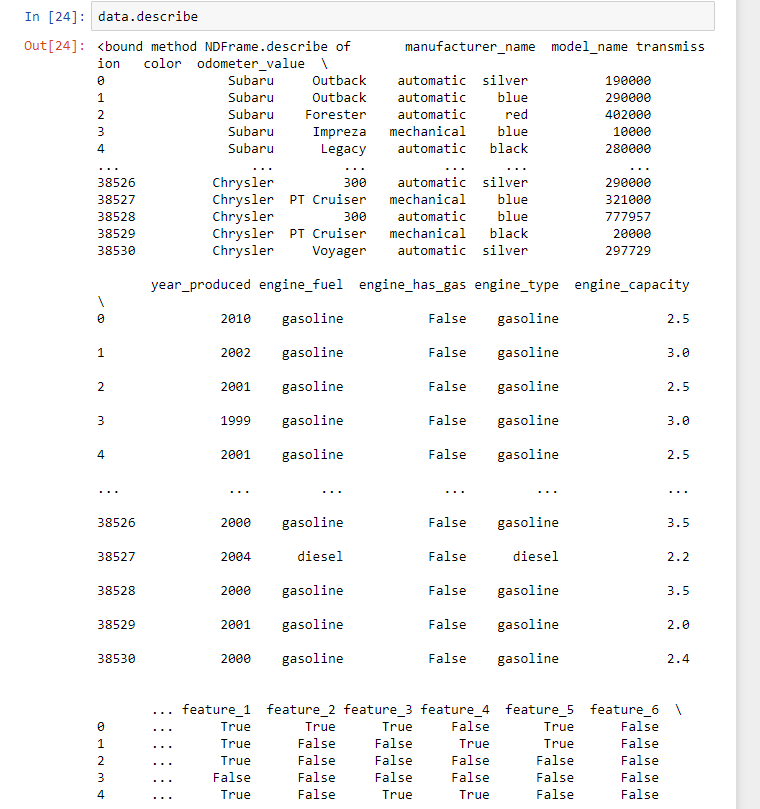
Підключення бібліотеки та відкривання файлу типу csv в ній



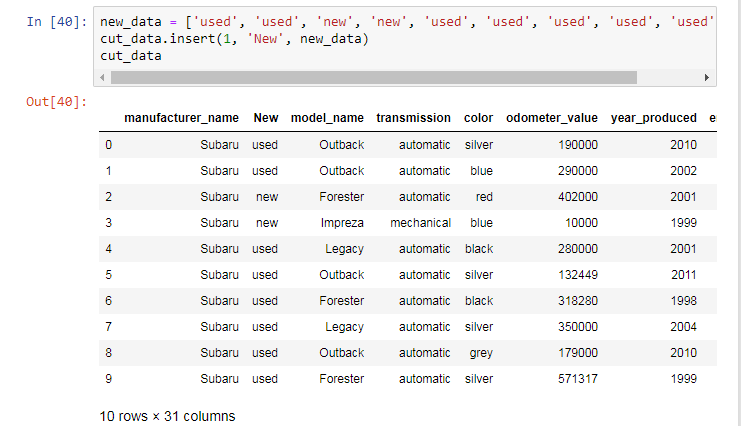
Запит на відображення кількості стовпців та рядків



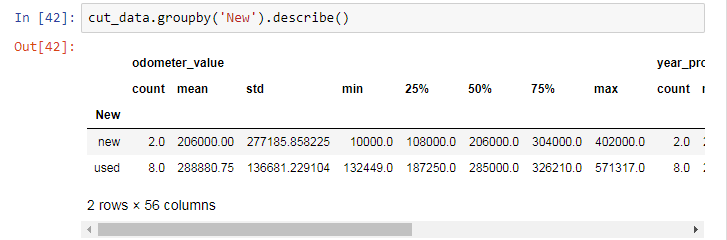
Застосування функції describe для опису різних типів даних



Додавання нового стовпчика з даними



Статистика по всім змінним, розділивши масив даних.



# 2. Комп'ютерний практикум №2 "Візуалізація даних за допомогою бібліотеки Matplotlib"

Бібліотека matplotlib - це бібіліотека двовимірної графіки для мови

програмування python за допомогою якої можна створювати високоякісні

малюнки різних форматів. Matplotlib є модуль-пакет для python.

Бібліотека matplotlib підтримує два методи побудови графіків:

- об’єктно-орієнтований метод, який полягає у створенні графіків як об'єктів та

роботу з ними як з об'єктами

-функціональний метод, який полягає у простій побудові графіків та їх

елементарному перетворенні.

Matplotlib є гнучким, легко конфігурованим пакетом, який разом з NumPy,

SciPy і IPython надає можливості, подібні MATLAB. В даний час пакет працює

з декількома графічними бібліотеками, включаючи wxWindows і PyGTK.

Пакет підтримує багато видів графіків і діаграм:

Графіки (line plot)

Діаграми розкиду (scatter plot)

Стовпчасті діаграми (bar chart) і гістограми (histogram)

Кругові діаграми (pie chart)

Стовбур-лист діаграми (stem plot)

Контурні графіки (contour plot)

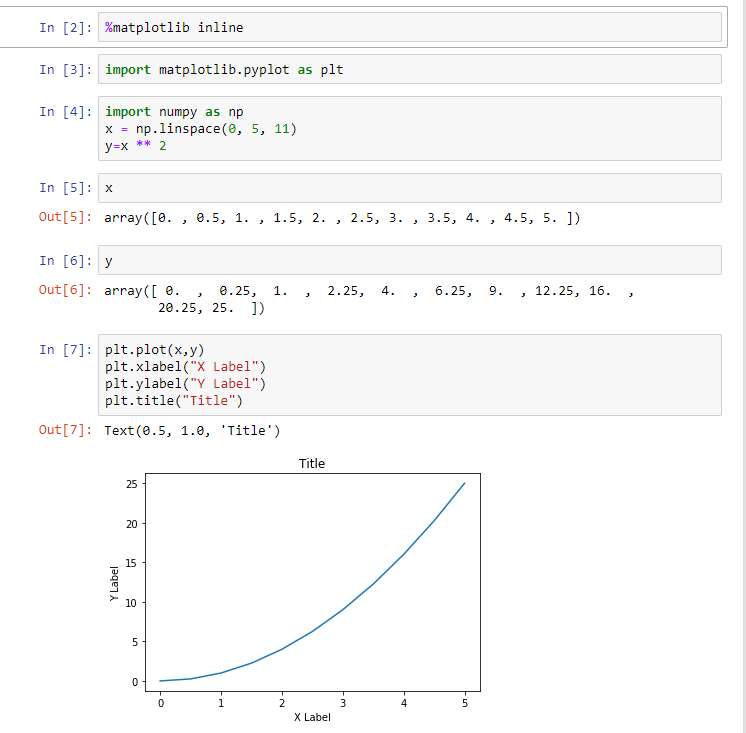
Поля градієнтів (quiver)

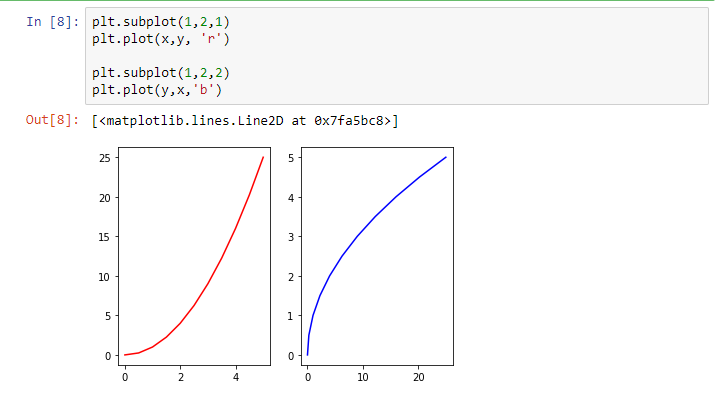
Спектральні діаграми (spectrogram)

Користувач може вказати осі координат, грати, додати написи і пояснення,

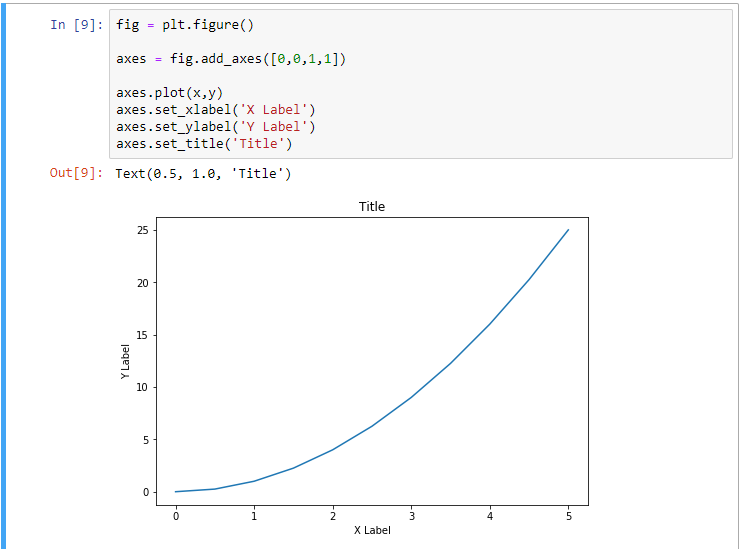
використовувати логарифмічну шкалу або полярні координати.

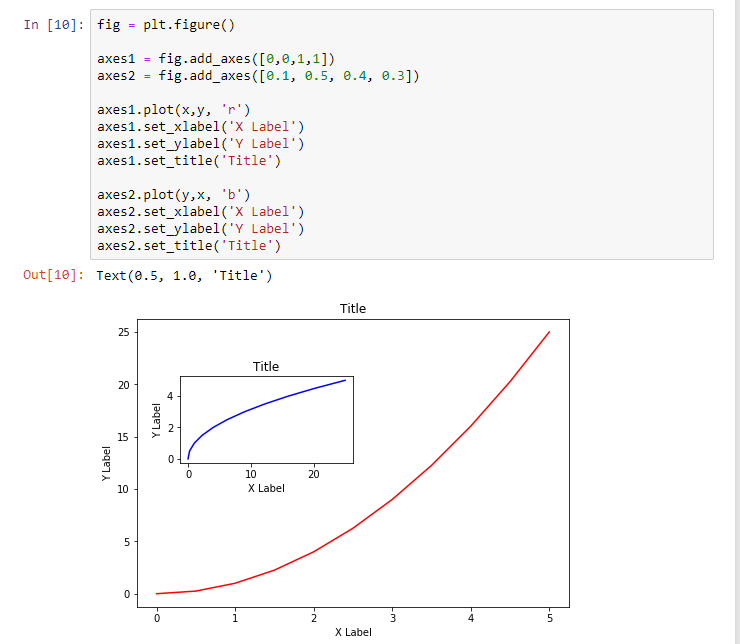
Створення графіка y(x) за допомогою функціонального методу

Побудова декількох графіків поруч:

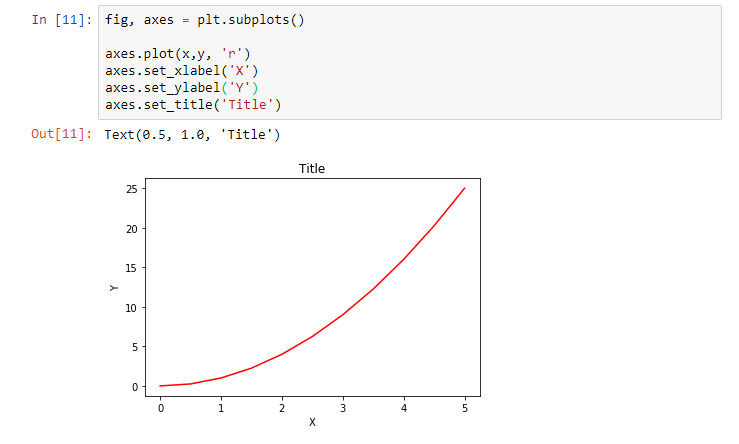


Побудова графіків за допомогою об’єктно-орієнтованого методу

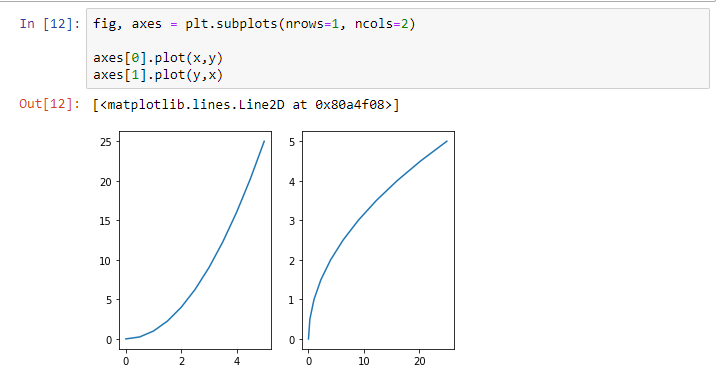


Одночасна побудова декількох графіків:

Приклад підписуваня графіків:

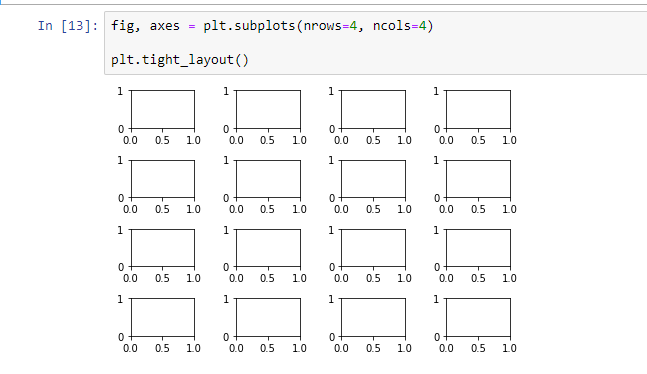


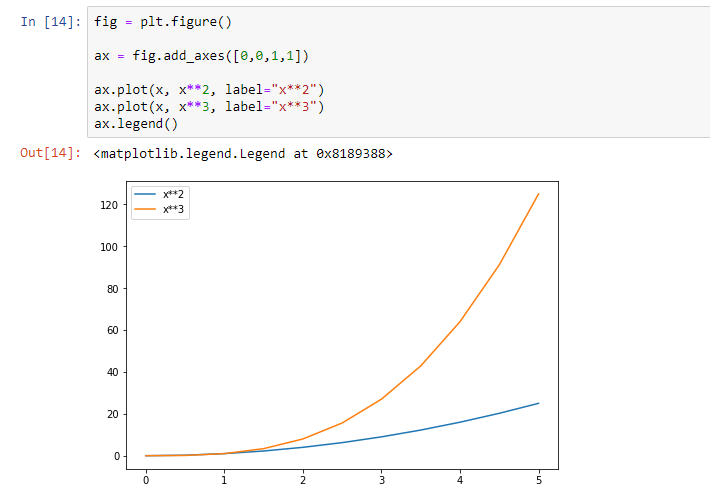
Вказування кількості рядків і стовпців під час створення об’єкта subplots ():



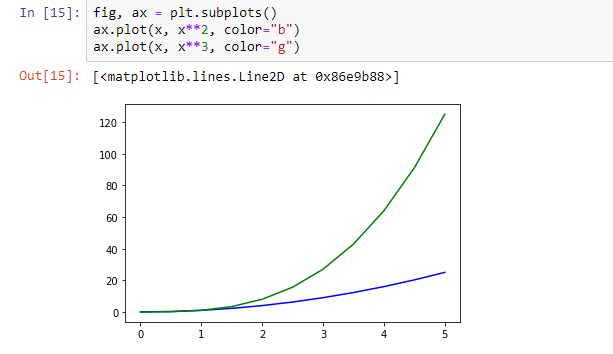
Вирівнювання декількох графіків при накладанні за допомогою

plt.tight\_layout():

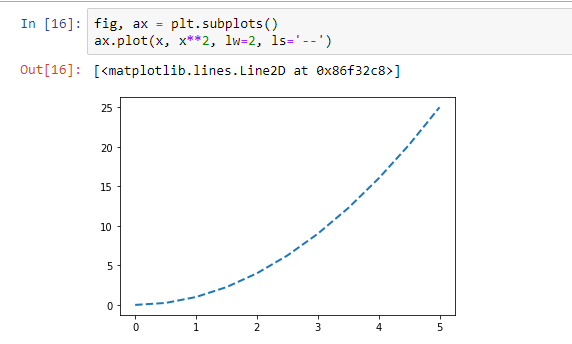


Додавання легенди до графіку: 

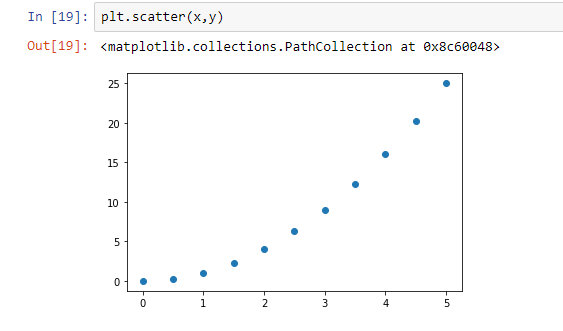
Зміна кольору графіку:



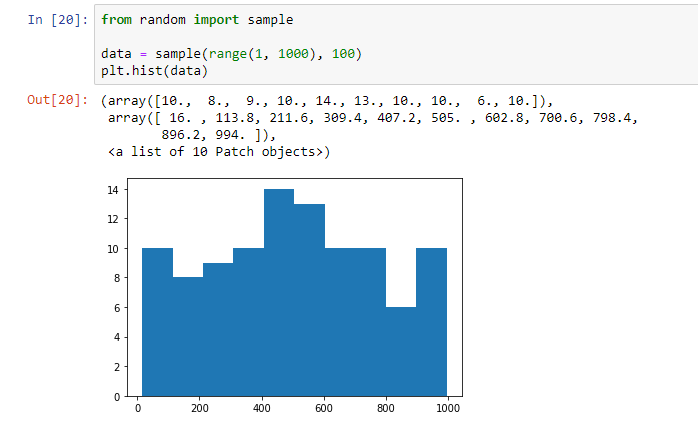
Зміна типу лінії графіку:



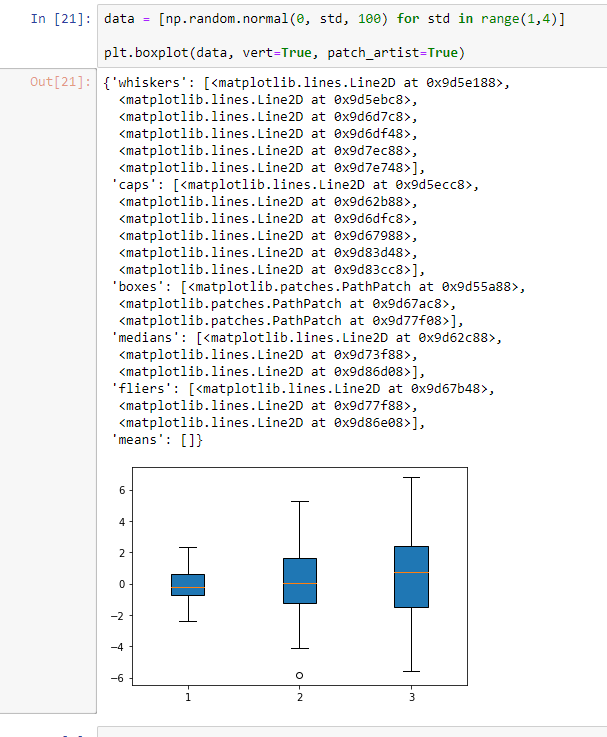
Точкова діаграма:



Гістограма:



Вox-plot діаграма:



Висновки: В даному комп'ютерному практикумі була проведена робота з

бібліотекою matplotlib. Було зформовано та подальше оброблено деяку

кількість різних графіків за допомогою функціонального та об’єктно-

орієнтованого методу.

# 3. Комп'ютерний практикум №3 "Візуалізація даних за допомогою бібліотеки Seaborn"

Бібліотека візуалізації даних Seaborn також заснована на Matplotlib.

Вона поставляється з повнофункціональним інтерфейсом для малювання

привабливою і інформативною статистичної графіки. Його API для створення

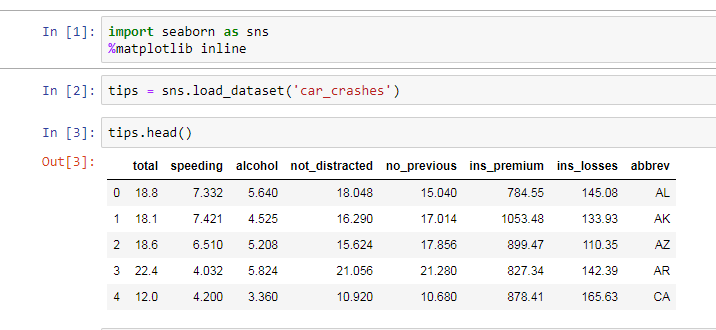
візуалізацій на основі KDE більш лаконічний, ніж у інших схожих бібліотек.

Seaborn старанно працює над тим, щоб візуалізація стала центральною

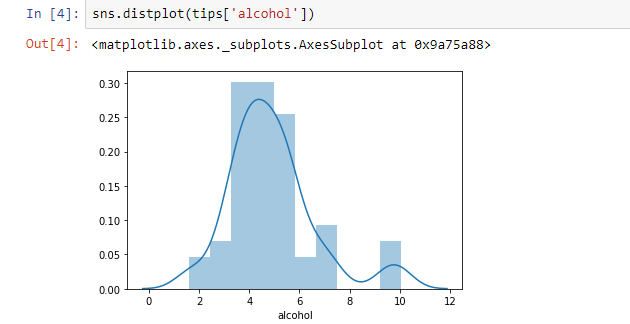
частиною розуміння і вивчення даних. Seaborn повністю інтегрований зі стеком

PyData, включаючи підтримку структур даних numpy і pandas.

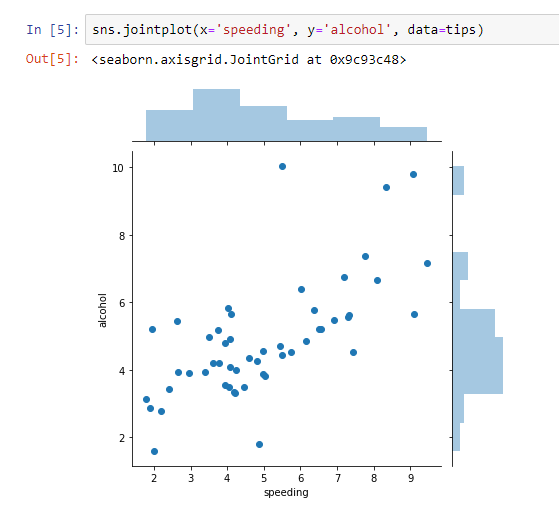
Відкривання файлу



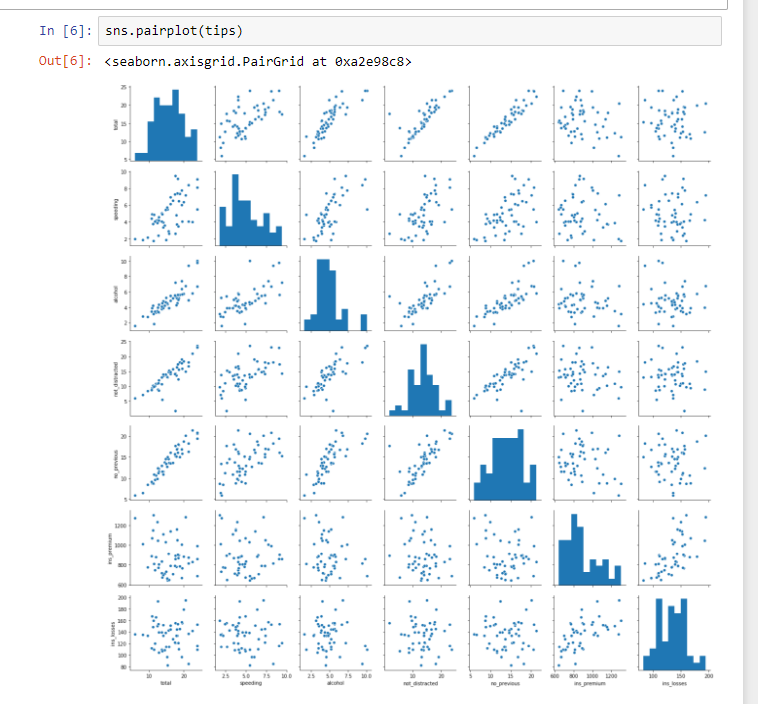
Розподіл одновимірного набору спостережень



Розподіл двох змінних та їх взаємозв’язок за допомогою діаграми розсіювання.

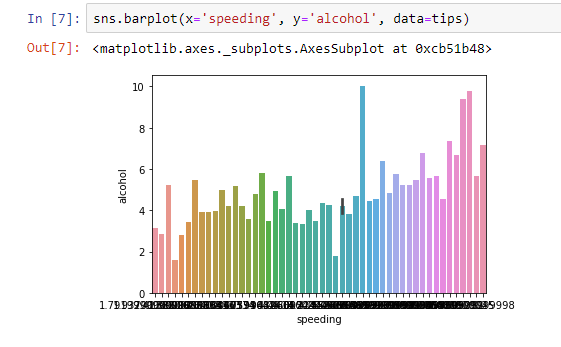


Попарні взаємозв’язки між змінними в повному наборі даних.



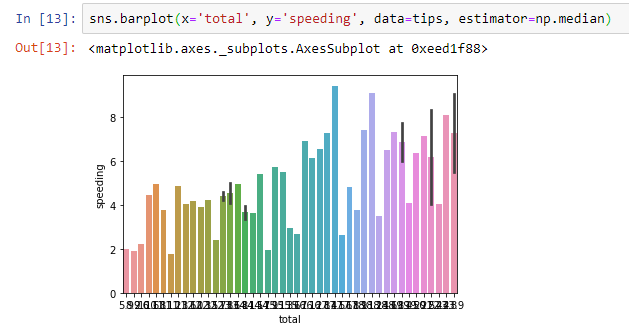
Отримання сукупних даних по кількісним даним в залежності від

категоріальної змінної.

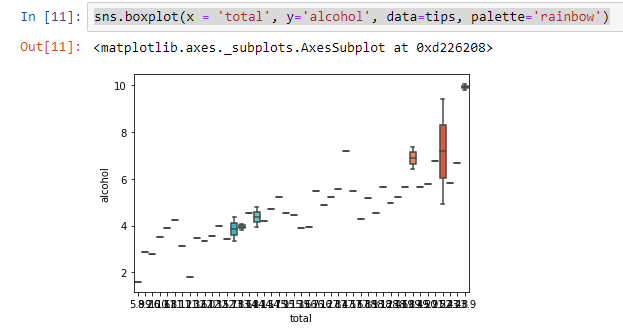


Отримання сукупних даних по кількісним даним в залежності від

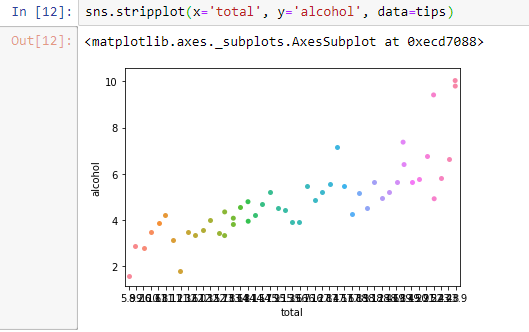
категоріальної змінної зі зміною варіанту відображення кількосної змінної .



Візуалізація розподілу категоріальних даних за певною кількісною змінною.



Діаграма розсіювання, де одна за змінних категоріальна.



Висновки: В цьому комп'ютерному практикумі ми ознайомилися з бібліотекою

seaborn та роботою з нею. За допомогою цієї бібліотеки можна будувати більш

складні і наглядні графіки за допомогою більш зрозумілої візуалізації.

**4. Комп'ютерний практикум №4 "Перевірка даних на нормальність за допомогою критерію Шапіра-Уілка"**

Критерій Шапіро-Уілкі використовується для перевірки гіпотези:

«випадкова величина X розподілена нормально» і є одним найбільш

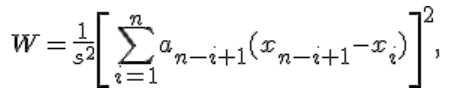
ефективних критеріїв перевірки нормальності. Критерії, перевіряючі

нормальність вибірки, є окремим випадком критеріїв згоди. Якщо вибірка

нормальна, можна далі застосовувати потужні параметричні критерії,

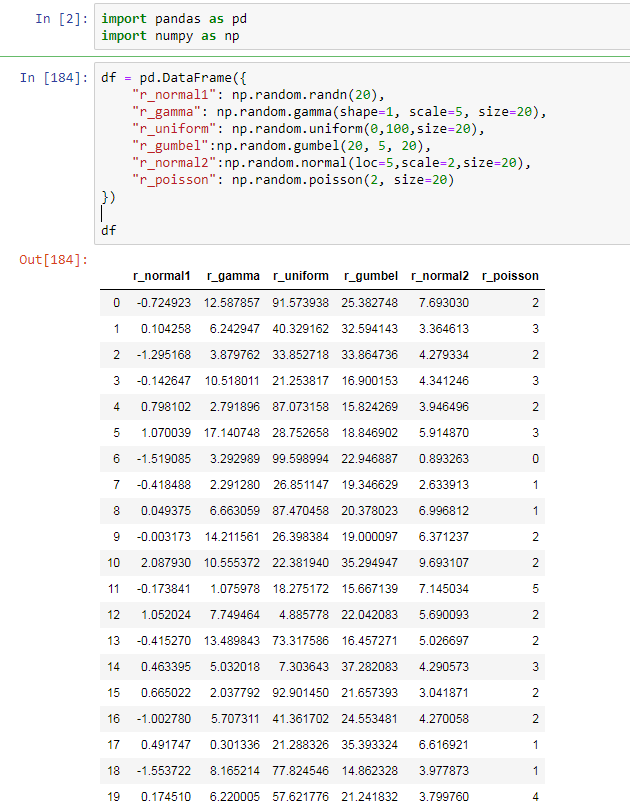
наприклад, критерій Фішера.

Критерій Шапіро-Уілкі заснований на оптимальній лінійної несмещённой оцінці дисперсії до її звичайної оцінці методом максимальної правдоподібності. Статистика критерію має вигляд:

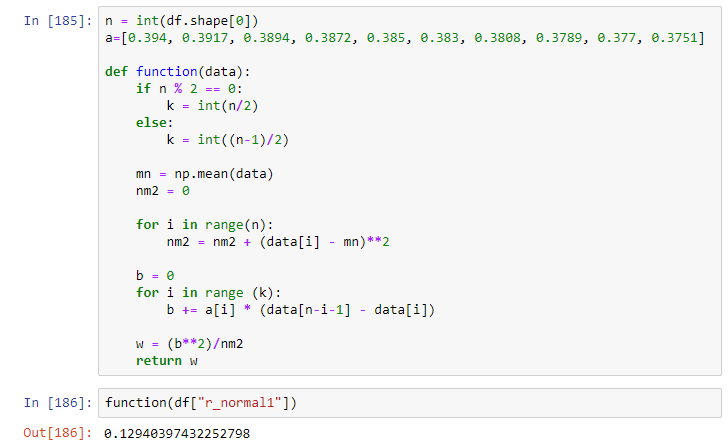
s^2=\sum_{i=1}^n (x_i -\overline{x})^2, \overline{x}=\frac{1}{n}\sum_{i=1}^n x_i.

Чисельник є квадратом оцінки середньоквадратичного відхилення Ллойда.

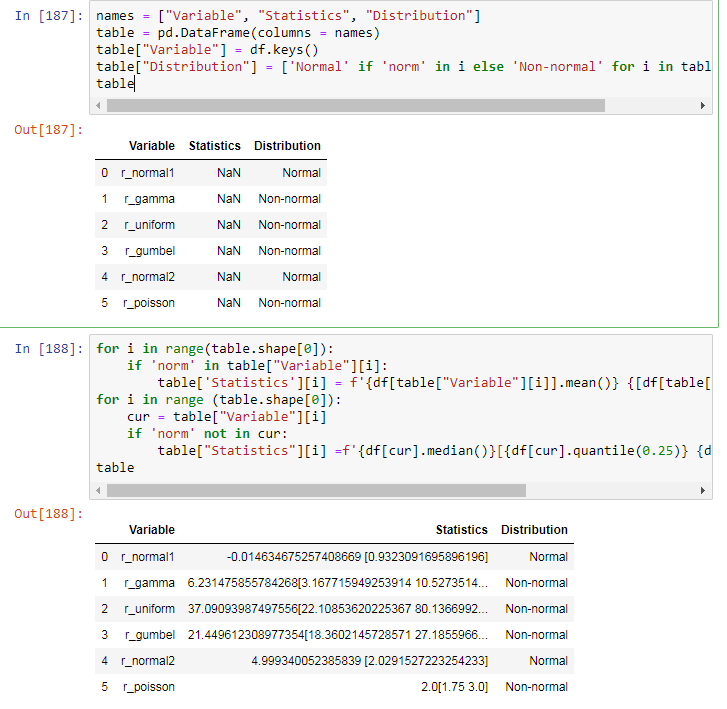
Коефіцієнти беруться з таблиць

Генерація даних за різними видами розподілу

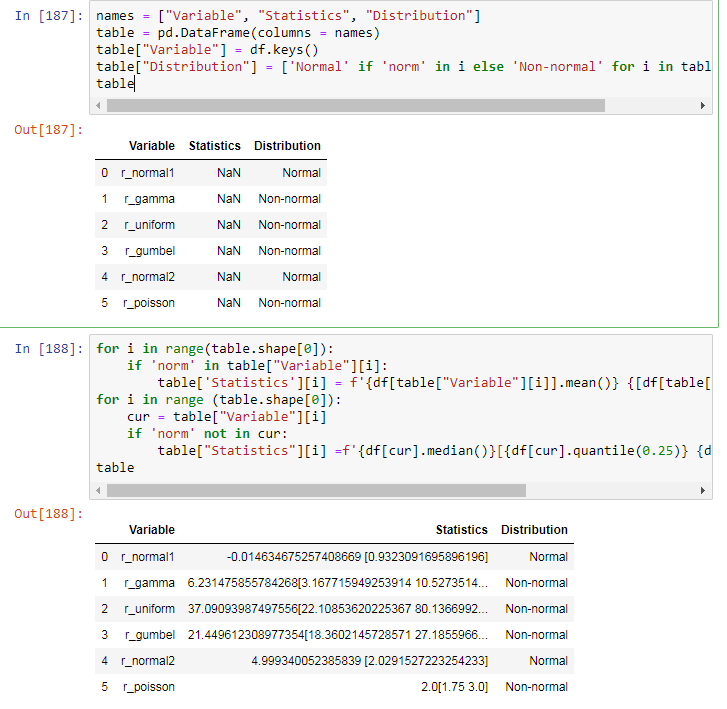
Створення функції, що реалізує критерій Шапіро-Уілка



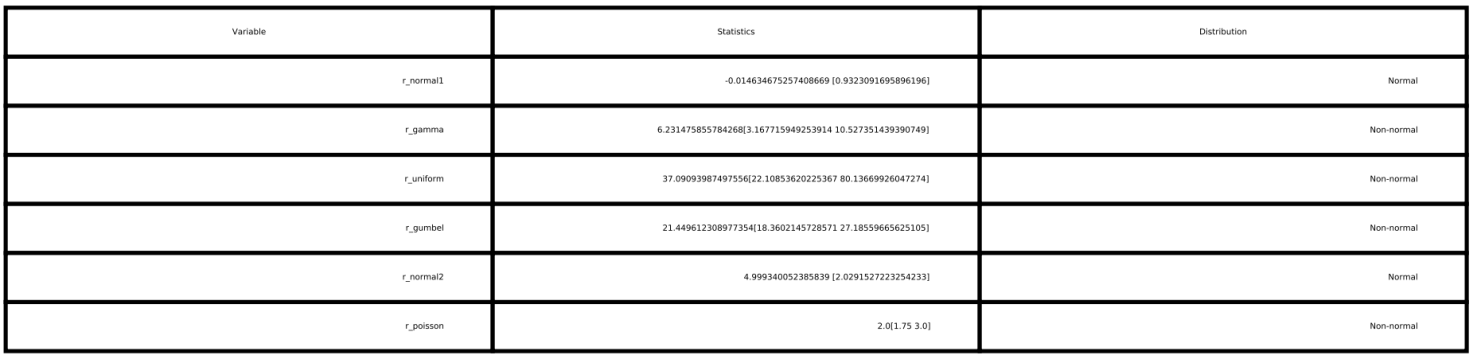
Створення таблиці та заповнення її результатом перевірки функцією згенерованих даних на нормальність.



Знаходження відхилень та медіани



Збереження даних у PDF - форматі

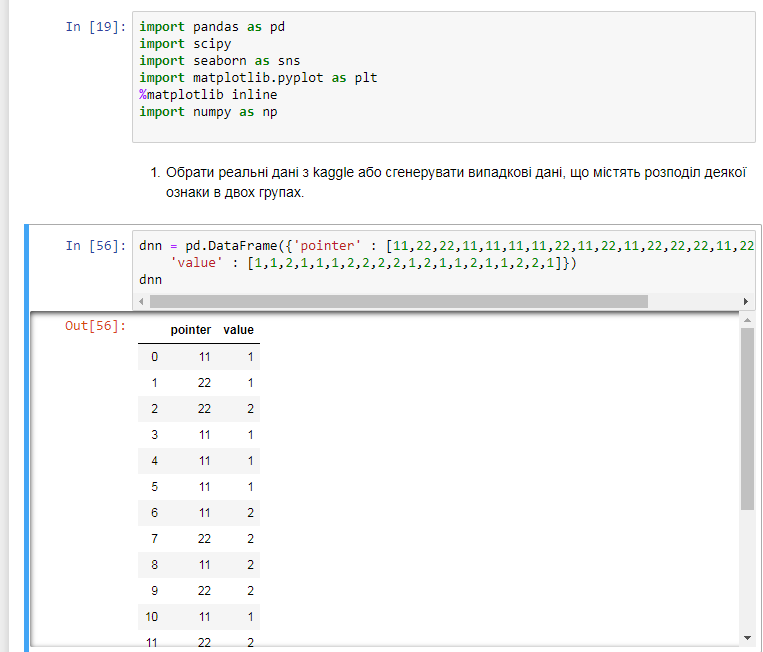


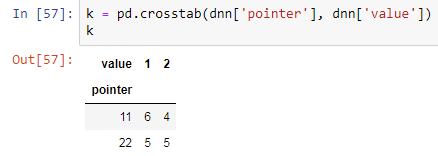
Висновки: Створена функція, що реалізує критерій Шапіро - Уілка та застосована до кожного набору даних. Функція знайшла всі нормальні розподіли, тому її можна вважати вірною.

# 5. Комп'ютерний практикум №5 "Критерій Пірсона для порівняння декількох груп за розподіленням ознаки"

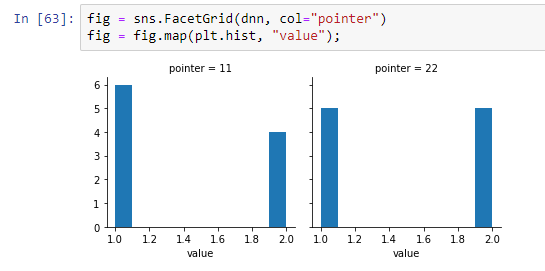
Таблиця спряженості, або таблиця контингентні, факторна таблиця в статистиці - засіб представлення спільного розподілу двох змінних, призначене для дослідження зв'язку між ними. Таблиця спряженості є найбільш універсальним засобом вивчення статистичних зв'язків, так як в ній можуть бути представлені змінні з будь-яким рівнем вимірювання. Таблиці спряженості часто використовуються для перевірки гіпотези про наявність зв'язку між двома ознаками з використанням точного тесту Фішера або критерію згоди Пірсона.

Критерій згоди Пірсона або критерій згоди (Хі-квадрат) - це непараметричний метод, який дозволяє оцінити значимість відмінностей між фактичним (виявленим в результаті дослідження) кількістю результатів або якісних характеристик вибірки, що потрапляють в кожну категорію, і теоретичним кількістю, яке можна очікувати в досліджуваних групах при справедливості нульової гіпотези. Висловлюючись простіше, метод дозволяє оцінити статистичну значущість відмінностей двох або декількох відносних показників (частот, часткою). Найбільш часто вживається критерій для перевірки гіпотези про приналежність спостережуваної вибірки  об'ємом n деякого теоретичного закону розподілу 

Генеруємо випадкові дані, що містять розподіл деякої ознаки в двох групах. 

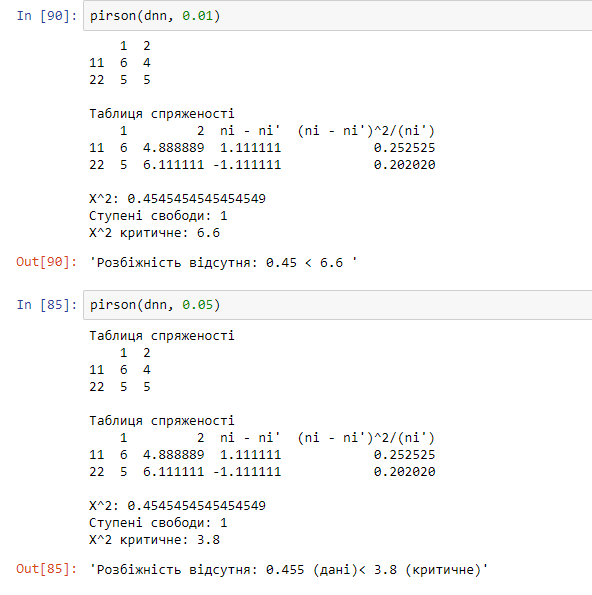


Будуємо графік, що відображає розподіл ознак в групах

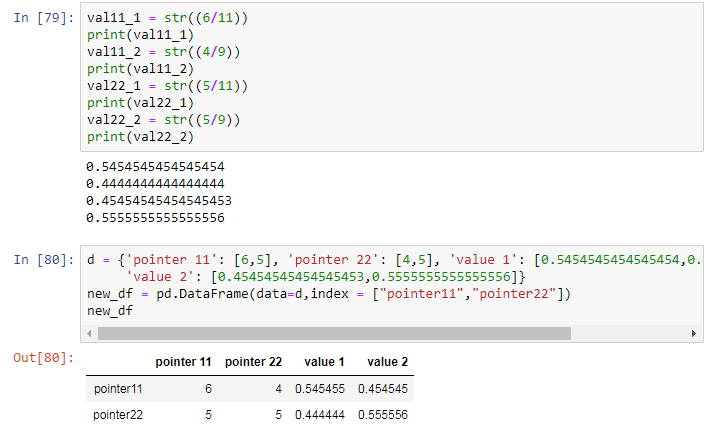


Створення функції, що оцінює зв'язок між ознакою та групою за критерієм Пірсона

Випробовування роботи функції



Формування таблиці спряженності



Висновки: Була створена функція, що оцінює зв'язок між ознакою та групою за критерієм Пірсона. Вона була випробувана на даних бех зв'язку, і дана функція це виявила, тому її можна вважати вірною.

# 6. Комп'ютерний практикум №6 "Парний t-критерій Стьюдента"

Парний t-критерій Стьюдента - одна з модифікацій методу Стьюдента, використовувана для визначення статистичної значущості відмінностей парних (повторних) вимірювань.

t-критерій був розроблений Вільямом Госсетом для оцінки якості пива в компанії Гіннес. У зв'язку із зобов'язаннями перед компанією щодо нерозголошення комерційної таємниці, стаття Госсета вийшла в 1908 році в журналі «Біометрика» під псевдонімом «Student» (Студент).

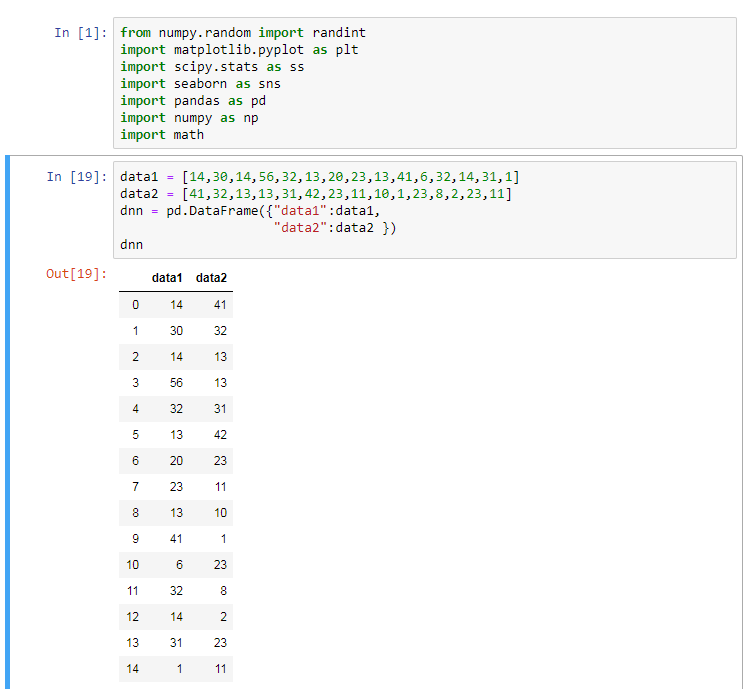
Парний t-критерій Стьюдента використовується для порівняння двох залежних (парних) вибірок. Залежними є вимірювання, виконані у одних і тих же пацієнтів, але в різний час, наприклад, артеріальний тиск у хворих на гіпертонічну хворобу до і після прийому антигіпертензивного препарату. Нульова гіпотеза свідчить про відсутність відмінностей між порівнюваними вибірками, альтернативна - про наявність статистично значущих відмінностей.

Парний t-критерій Стьюдента розраховується за такою формулою:

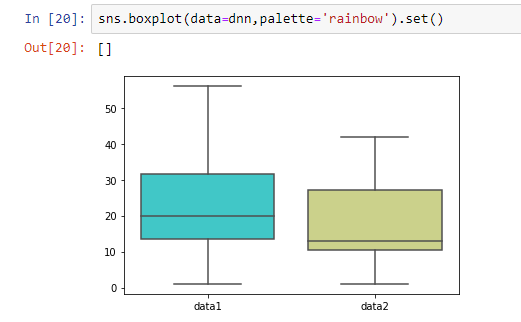


де Мd - середня арифметична різниць показників, виміряних до і після, σd - середньоквадратичне відхилення різниць показників, n - число досліджуваних.

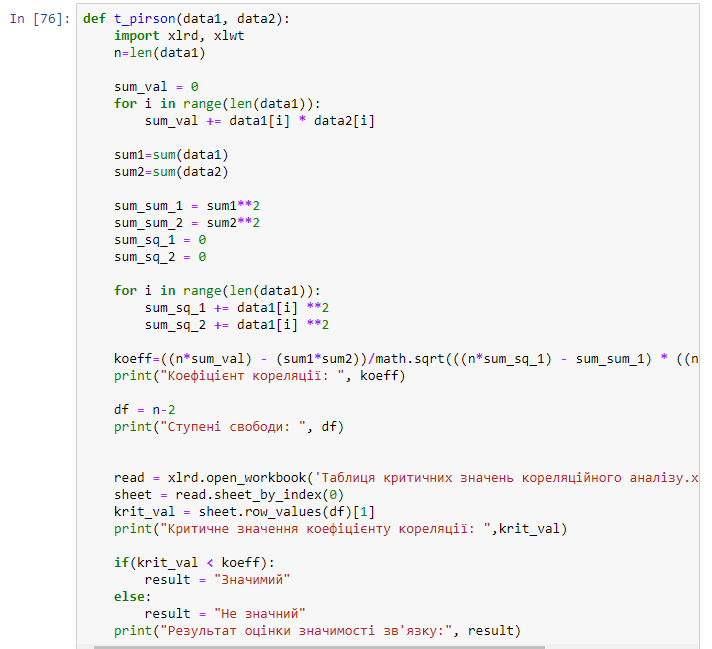
Створення даних

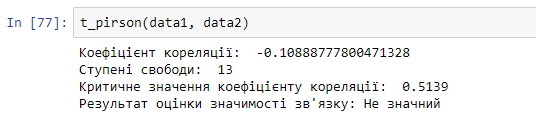


Побудова графіку розкиду даних в групах

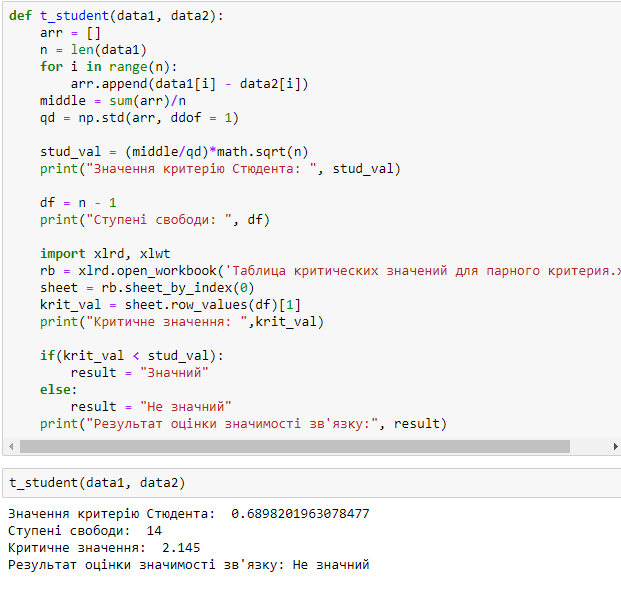


Реалізація функції для оцінки зв'язку між групами за коефіцієнтом кореляції Пірсона.





Реалізація функції для визначення відмінності між групами за критерієм Стюдента



Висновки: Були згенеровані випадкові дані та реалізовано дві функції: для оцінки зв'язку між групами за коефіцієнтом кореляції Пірсона та для визначення відмінності між групами за критерієм Стюдента. Обидві функції не показали якого зв'язку між даними, тому їх можна вважати працюючими.

# 7. Комп'ютерний практикум №7 "t-критерій Стьюдента для незалежних вибірок"

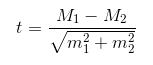
t-критерій Стьюдента - загальна назва для класу методів статистичної перевірки гіпотез (статистичних критеріїв), заснованих на розподілі Стьюдента. Найбільш часті випадки застосування t-критерію пов'язані з перевіркою рівності середніх значень у двох вибірках.

t-критерій Стьюдента використовується для визначення статистичної значущості відмінностей середніх величин. Може застосовуватися як у випадках порівняння незалежних вибірок (наприклад, групи хворих на цукровий діабет і групи здорових), так і при порівнянні пов'язаних сукупностей (наприклад, середня частота пульсу у одних і тих же пацієнтів до і після прийому антиаритмічного препарату). В останньому випадку розраховується парний t-критерій Стьюдента

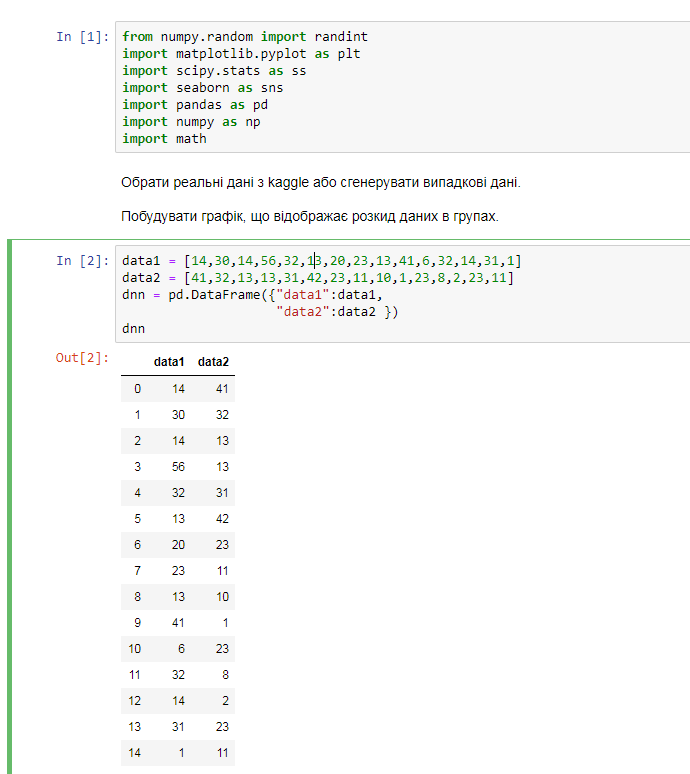
Для застосування t-критерію Стьюдента необхідно, щоб вихідні дані мали нормальний розподіл. Також має значення рівність дисперсій (розподілу) порівнюваних груп (гомоскедастичність). При нерівних дисперсіях застосовується t-критерій в модифікації Уелча (Welch's t).

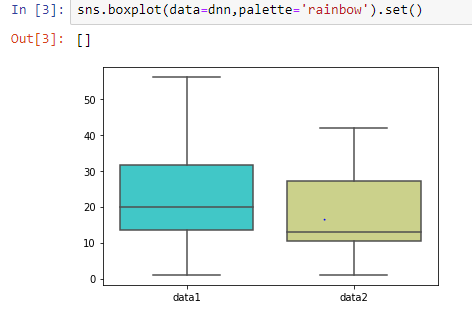
При відсутності нормального розподілу порівнюваних вибірок замість t-критерію Стьюдента використовуються аналогічні методи непараметричної статистики, серед яких найбільш відомими є U-критерій Манна - Уїтні.

Для порівняння середніх величин t-критерій Стьюдента розраховується за такою формулою:



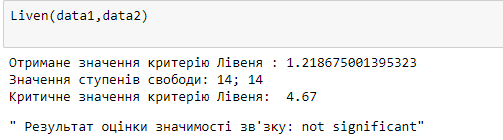
де М1 - середня арифметична першої порівнюєш сукупності (групи), М2 - середня арифметична другий порівнюєш сукупності (групи), m1 - середня помилка першої середньої арифметичної, m2 - середня помилка другий середньої арифметичної.

Створення даних 

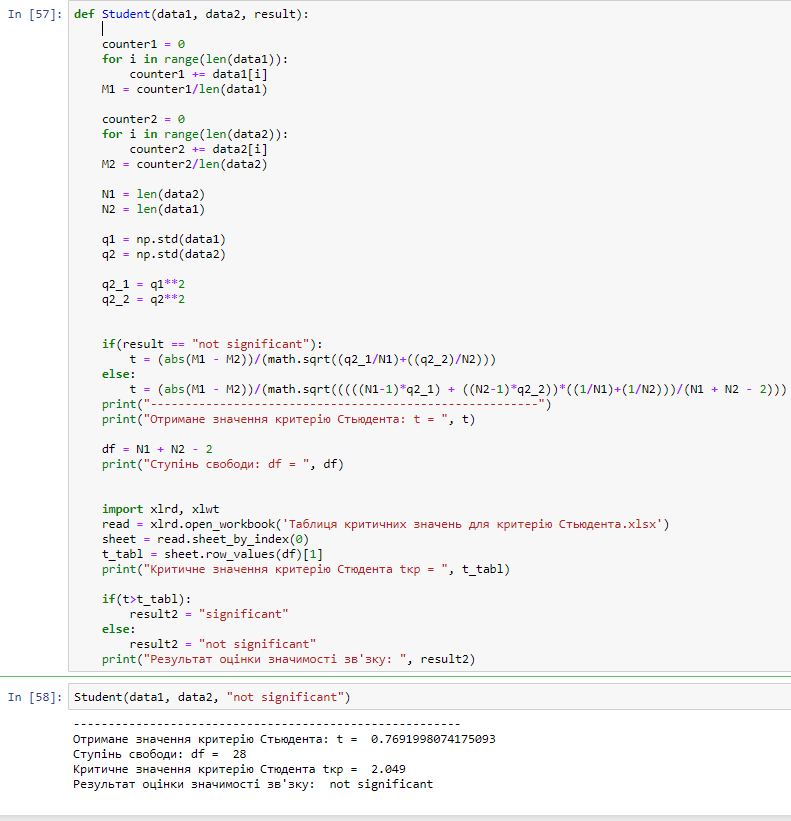
Графік, що відображає розкид даних в групах

Функція для оцінки вімінностей дисперсій між групами за критерієм Лівеня





Функція для визначення відмінності між групами за критерієм Стьюдента в залежності від рівності/відмінності дисперсій



Висновки: Створені функції не знайшли значного зв'язку в даних. Частково це є через те, що дані були згенеровані випадково. Створені функції можна вважати вірними

# 8. Комп'ютерний практикум №8 "Дисперсійний аналіз"

Дисперсійний аналіз - (ANOVA) застосовується для дослідження впливу однієї або декількох якісних змінних (факторів) на одну залежну кількісну змінну (відгук).

В основі дисперсійного аналізу лежить припущення про те, що одні змінні можуть розглядатися як причини (фактори, незалежні змінні): , а інші як слідства (залежні змінні). Незалежні змінні називають іноді регульованими чинниками саме тому, що в експерименті дослідник має можливість варіювати ними і аналізувати виходить результат.

Основною метою дисперсійного аналізу (ANOVA) є дослідження значущості відмінності між середніми за допомогою порівняння (аналізу) дисперсій. Поділ загальної дисперсії на кілька джерел, дозволяє порівняти дисперсію, викликану відмінностями між групами, з дисперсією, викликаної внутрішньогрупової мінливістю. При істинності нульової гіпотези (про рівність середніх в декількох групах спостережень, вибраних з генеральної сукупності), оцінка дисперсії, пов'язаної з внутрішньогрупової мінливістю, повинна бути близькою до оцінки груповий дисперсії. Якщо ви просто порівнюєте середні в двох вибірках, дисперсійний аналіз дасть той же результат, що і звичайний t-критерій для незалежних вибірок (якщо порівнюються дві незалежні групи об'єктів або спостережень) або t-критерій для залежних вибірок (якщо порівнюються дві змінні на одному і тій же множині об'єктів або спостережень).

Сутність дисперсійного аналізу полягає в розчленуванні загальної дисперсії досліджуваного ознаки на окремі компоненти, зумовлені впливом конкретних факторів, і перевірці гіпотез про значущість впливу цих факторів на досліджуваний ознака. Порівнюючи компоненти дисперсії один з одним за допомогою F-критерію Фішера, можна визначити, яка частка загальної варіативності результативної ознаки зумовлена ​​дією регульованих факторів.

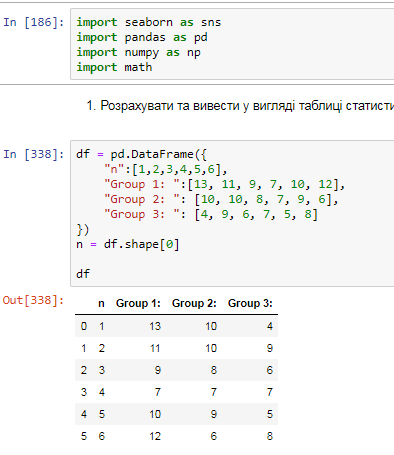
\

Вихідним матеріалом для дисперсійного аналізу є дані дослідження трьох і більше вибірок: , які можуть бути як рівними, так і нерівними за чисельністю, як зв'язковими, так і незв'язними. За кількістю виявлених регульованих факторів дисперсійний аналіз може бути однофакторний (при цьому вивчається вплив одного фактора на результати експерименту), двофакторна (при вивченні впливу двох чинників) і багатофакторним (дозволяє оцінити не тільки вплив кожного з факторів окремо, але і їх взаємодія).

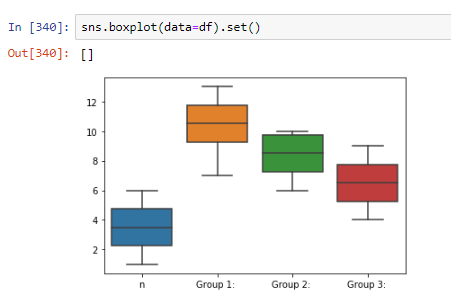
Дисперсійний аналіз відноситься до групи параметричних методів і тому його слід застосовувати тільки тоді, коли доведено, що розподіл є нормальним.

Дисперсійний аналіз використовують, якщо залежна змінна вимірюється в шкалі відносин, інтервалів або порядку, а що впливають змінні мають нечислову природу (шкала найменувань).

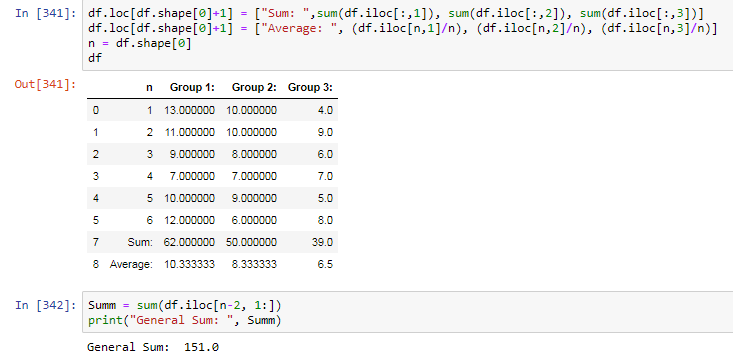
Створення випадкових даних



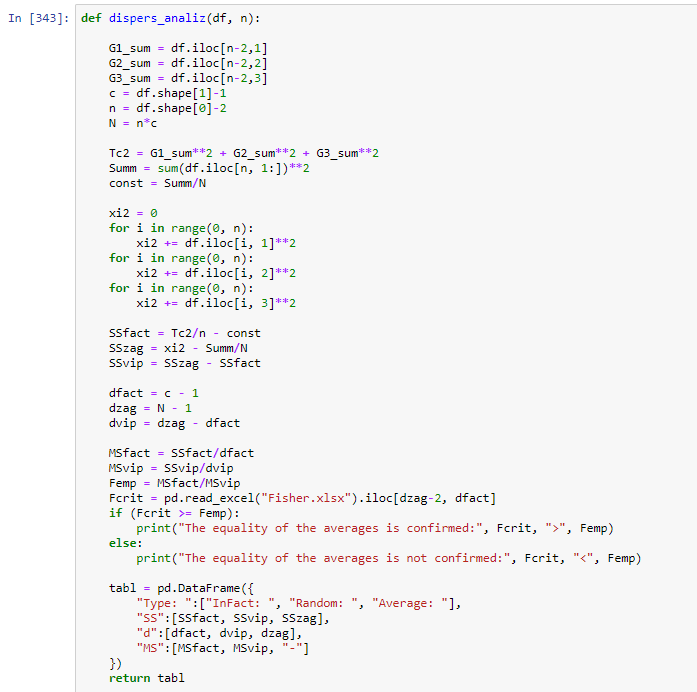
Графік розкиду в групах

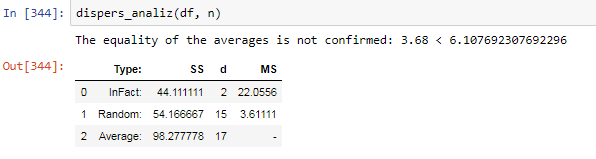


Сума даних



Функція, що реалізовує однофакторний дисперсійний аналіз.





Висновок: Була створена функція, що реалізовує однофакторний дисперсійний аналіз. Дані були пропущені через функцію. Рівність середніх підтверджена не була.