Міністерство освіти та науки

Дніпропетровський національний університет імені О. Гончара

Кафедра математичного забезпечення ЕОМ

Звіт до Лабораторної роботи № 2

з курсу «Випадкові процеси»

на тему «Регресійний аналіз»

Виконав: студент групи ПМ-13-1

Кривоносов Олександр

Перевірила: асистент Ризоль О.О.

Дніпро

2016

# Зміст

[Зміст 2](#_Toc469392632)

[Постановка задачі 3](#_Toc469392633)

[Аналіз результатів 4](#_Toc469392634)

[Завантажимо файл під назвою 04\_1.txt 4](#_Toc469392635)

[Первинний статистичний аналіз. 5](#_Toc469392636)

[Кореляційний аналіз. 5](#_Toc469392637)

[Регресійний аналіз. 6](#_Toc469392638)

[Завантажимо файл під назвою 04\_2.txt 7](#_Toc469392639)

[Первинний статистичний аналіз. 8](#_Toc469392640)

[Кореляційний аналіз. 8](#_Toc469392641)

[Регресійний аналіз. 10](#_Toc469392642)

# Постановка задачі

У **програмі**, створеній в рамках лабораторної роботи 1, додати можливість відновлення регресії. Передбачити відновлення двох моделей регресії – лінійної та нелінійної. Модель нелінійної регресії має відповідати індивідуальному варіанту.

Відновлення кожної з регресій має передбачати:

* + - знаходження точкових та інтервальних оцінок параметрів регресії та перевірку значущості параметрів;
    - побудову відновленої регресії на тому ж графіку, де відображене кореляційне поле;
    - побудову довірчих інтервалів для лінії регресії та прогнозних значень;
    - визначення коефіцієнта детермінації;
    - перевірку значущості відновленої моделі за *F*-тестом;
    - побудову діагностичної діаграми.

Оформити звіт за результатами виконання лабораторної роботи.

Індивідуальний варіант виконавця - №4, **.**

**Вимоги до програми**

* 1. Програма повинна бути незалежна від даних. Вхідний файл має обиратися в діалозі з користувачем. Передбачається, що вхідні дані знаходяться в текстовому файлі.
  2. Результатом знаходження оцінок параметрів кожної регресії має бути така таблиця:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Значення оцінки | Дисперсія | Статистика | Квантиль | Значущість | Дов.інтервал |
| *a* |  |  |  |  |  |  |
| *b* |  |  |  |  |  |  |
| *c* |  |  |  |  |  |  |

* 1. Як результат перевірки адекватності відтворених регресійних залежностей слід вивести на екран статистику критерію, квантиль та висновок.
  2. Графіки ліній регресії та довірчих інтервалів необхідно відобразити разом із графіком кореляційного поля.
  3. Для кожного графіка слід виконати автоматичне масштабування, зобразити шкалу й показати одиниці виміру.
  4. Відображення результатів повинне відповідати точності обчислень (кількість знаків після коми не більша 4).

# Аналіз результатів

## Завантажимо файл під назвою 04\_1.txt

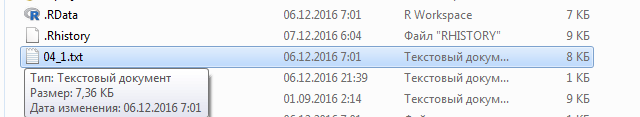


Рис. 1 04\_1.txt.

Об’єм вибірки становить значення 298.

Кореляційне поле, лінії регресії та їх довірчі інтервали мають такий вигляд:

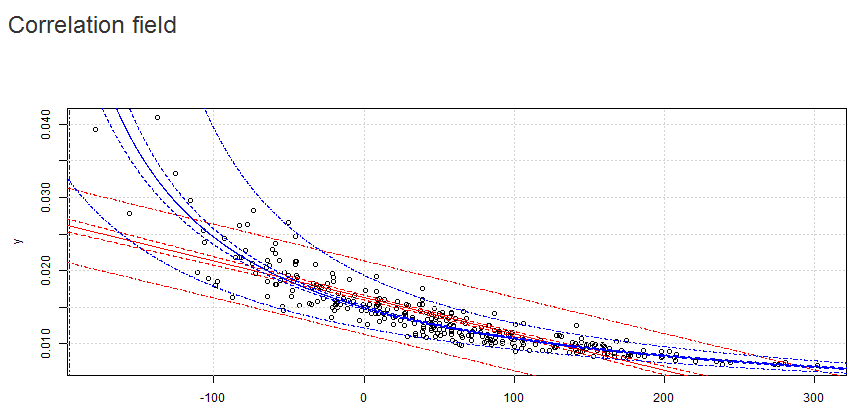


Рис.2 Кореляційне поле для файлу 04\_1.txt.

## Первинний статистичний аналіз.

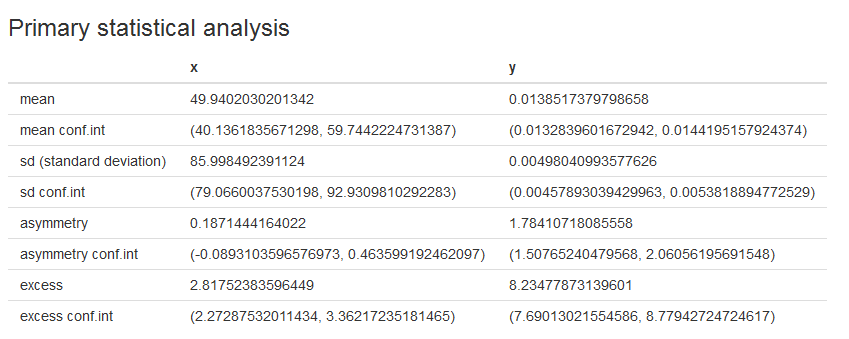


Рис.3 первинний статистичний аналіз (довірчі інтервали взяті для довірчого рівня 0.95) для даних з 04\_1.txt.

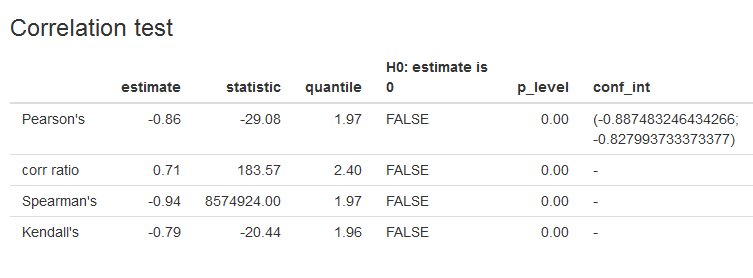
## Кореляційний аналіз.

По кореляційному полю можна сказати що між вибірками є нелінійний монотонний від’ємний стохастичний зв'язок. Про це свідчать такі факти:

1) Точки не впорядковуються в одну лінію.

2) Сукупність точок не має вигляд витягнутого нахиленого овалу.

Після того, як ми висунули припущення, що це нелінійний монотонний стохастичний зв’язок, переходимо до коефіцієнтів.

Рис. 4 Коефіцієнти кореляції для файлу 04\_1.txt (для кореляційного відношення взято 5 рівних за довжиною інтервалів з проміжку [min, max])

Парний коефіцієнт не дасть нам адекватного результату, оскільки ми зробили припущення про нелінійність зв’язку. Кореляційне відношення говорить на, що існує стохастичний зв’язок, який скоріше сильний, ніж слабкий. Коефіцієнти Спірмена та Кендала від’ємні, що підтверджує наші припущення і говорять нам про наявність стохастичного зв’язку.

## Регресійний аналіз.

Для лінійної регресії

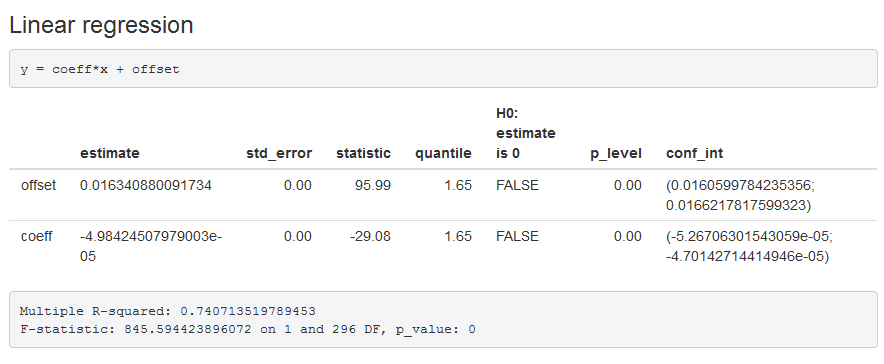
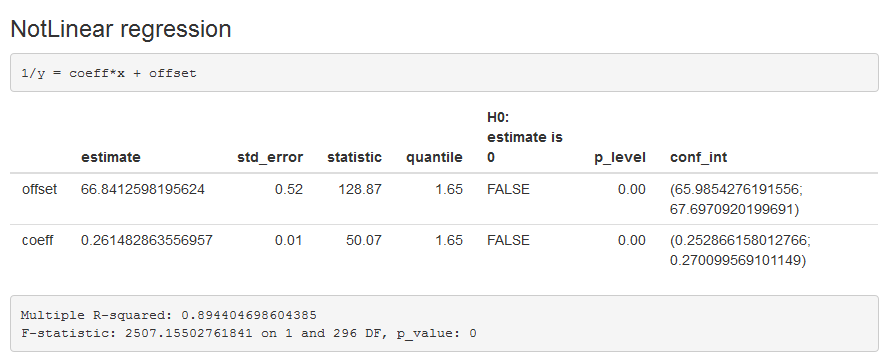


Рис. 5 Коефіцієнти регресії для файлу 04\_1.txt

Під таблицею можна побачити коефіцієнт детермінації, та статистику F-тесту.  
Можна бачити, що імовірність помилки 1-го роду (p\_value) на стільки мала, що програма не виводить жодного порядку цього числа відмінного від нуля; через це можна стверджувати, що імовірність відхилити гіпотезу H0:F=0, коли насправді є так, практично неможливо. Це означає, що між змінними та є залежність, але виходячи з коефіцієнтів кореляції та коефіцієнту детермінації можна стверджувати лише те, що або залежна змінна має ще деякий предикат, або те, що залежність між змінними не є лінійною.



Можна зробити аналогічні висновки, як і для лінійної регресії, про значущість регресії. Аля коефіцієнт детермінації для цього випадку значно більший. Це означає, що ця модель краще описує зв’язок між змінними. Це можна помітити й на графіку регресії (лінія синього кольору).

## Завантажимо файл під назвою 04\_2.txt

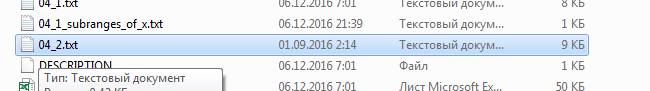
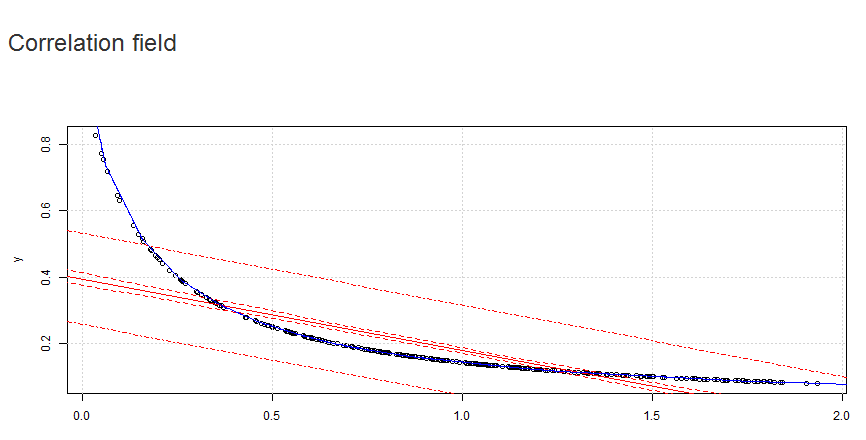


Рис.5 04\_2.txt.

Об’єм вибірки становить значення 300.

Кореляційне поле, лінії регресії та їх довірчі інтервали мають такий вигляд:

Рис.6 Кореляційне поле для файлу 04\_2.txt.

Підмітимо те, що для нелінійної регресії довірчий інтервал та інтервал для прогнозування значень збігаються з самою лінією регресії (лінія синього кольору).

## Первинний статистичний аналіз.

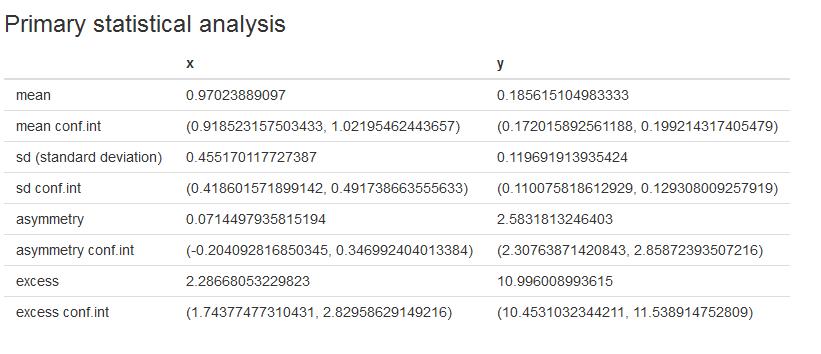


Рис.7 первинний статистичний аналіз (довірчі інтервали взяті для довірчого рівня 0.95) для даних з 04\_2.txt.

## Кореляційний аналіз.

По кореляційному полю можна сказати що між вибірками є нелінійний монотонний від’ємний функціональний зв'язок. Про це свідчать такі факти:

1) Точки впорядковуються в одну лінію.

2) Сукупність точок не має вигляд витягнутого нахиленого овалу.

Після того як ми висунули припущення, що це нелінійний монотонний функціональний зв’язок переходимо до коефіцієнтів.

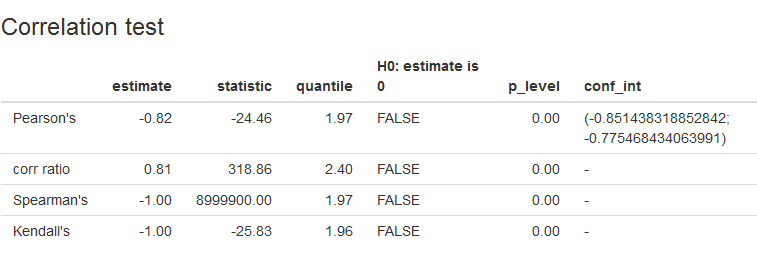
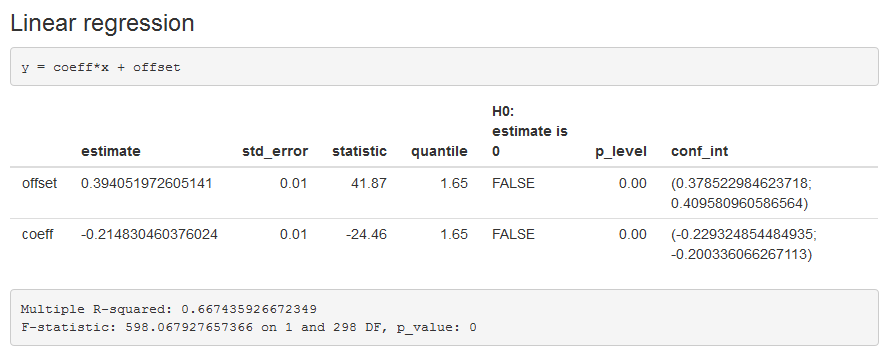


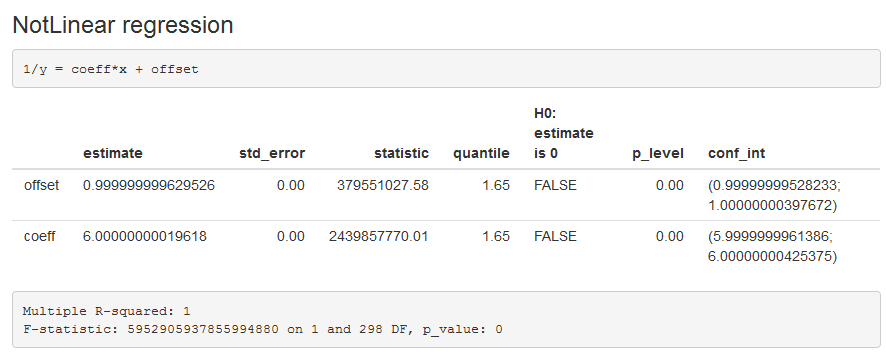
Рис. 8 Коефіцієнти кореляції для файлу 04\_2.txt (для кореляційного відношення взято 5 рівних за довжиною інтервалів з проміжку [min, max]).

Парний коефіцієнт не дасть нам адекватного результату, оскільки ми зробили припущення про не лінійність зв’язку. Коефіцієнти Спірмена та Кендала від’ємні, що підтверджує наші припущення і дорівнюють мінус одиниці і говорять про наявність функціонального зв’язку.

## Регресійний аналіз.



Під таблицею можна побачити коефіцієнт детермінації, та статистику F-тесту.  
Можна бачити, що імовірність помилки 1-го роду (p\_value) на стільки мала, що програма не виводить жодного порядку цього числа відмінного від нуля; через це можна стверджувати, що імовірність відхилити гіпотезу H0:F=0, коли насправді є так, практично неможливо. Це означає, що між змінними та є залежність, але виходячи з коефіцієнтів кореляції та коефіцієнту детермінації можна стверджувати лише те, що або залежна змінна має ще деякий предикат, або те, що залежність між змінними не є лінійною.



Можна зробити аналогічні висновки, як і для лінійної регресії, про значущість регресії. Аля коефіцієнт детермінації для цього випадку дорівнює 1. Це означає, що ця модель повністю описує зв’язок між змінними. Це можна помітити й на графіку регресії: лінія синього кольору проходить точно через крапки кореляційного поля. Отже ми встановили функціональну залежність між змінними.