|  |  |  |  |
| --- | --- | --- | --- |
| МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені Тараса Шевченка  ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ  **Кафедра програмних систем і технологій**        Дисципліна  **«Емпіричні методи програмної інженерії»**      **Практична робота № 4**  **“**КОРЕЛЯЦІЙНИЙ ТА РЕГРЕСІЙНИЙ АНАЛІЗ РЕЗУЛЬТАТІВ ВИМІРЮВАНЬ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ**”** | | | |
| **Виконав:** | Гоша Д. О | **Перевірила**: | Юрчук Ірина Аркадіївна |
| Група | ІПЗ-23 | Дата перевірки |  |
| Форма навчання | денна | Оцінка |  |
| Спеціальність | 121 |
| 2022 | | | |

Мета: Навчитися визначати та будувати залежності між метриками програмного забезпечення..

# Завдання

1. Побудувати кореляційні поля пар прямих та непрямих метрик.
2. Побудувати лінію регресії та визначити її функцію.

# Варіант 4

# Виконання лабораторної роботи

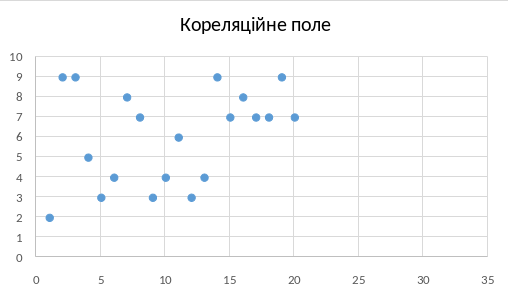
Згідно варіанту, метрики, які будуть досліджуватись: 

Прямі: NOM, NOC, NOP, NDD, CALL.

Непрямі: WMC, TCC, PNAS.

* **Побудувати кореляційні поля пар прямих та непрямих метрик:**

NOC - PNAS



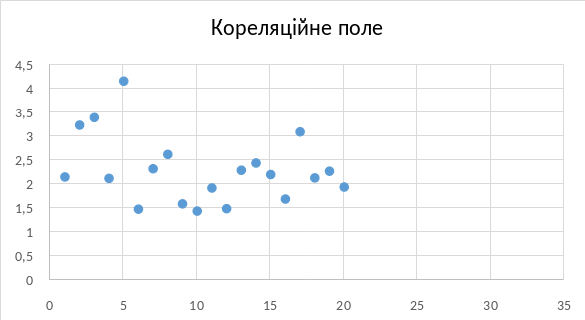
- 0,218

- 190

- 0,93752

Значущий - Так

NOC - ТСС



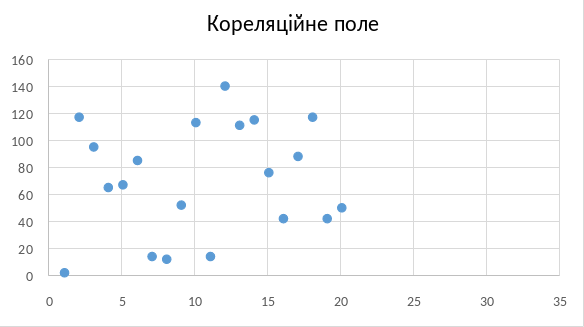
- 0,218

- 190

- 1

Значущий - Так

NOC - WMC



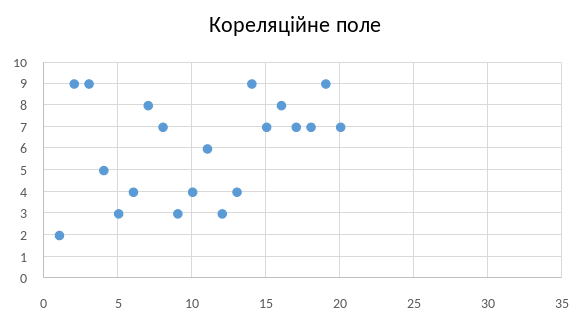
- 0,218

- 190

- 0.99207

Значущий - Так

NOM - PNAS



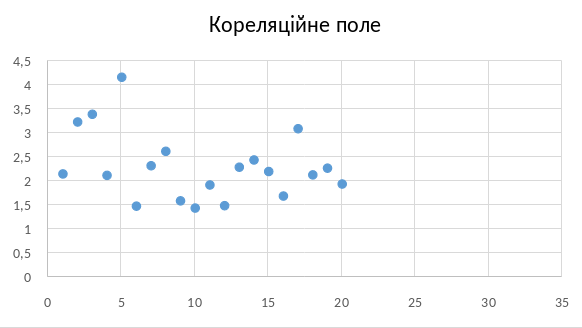
- 0,218

- 190

- 0.93685

Значущий - Так

NOM - TCC



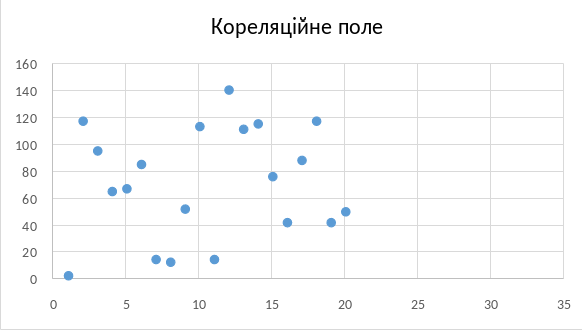
- 0,218

- 190

- 0.99472

Значущий - Так

NOM - WMC



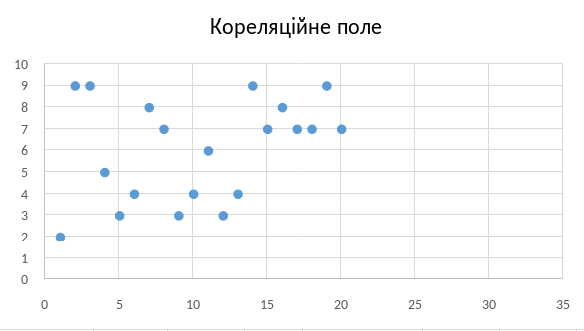
- 0,218

- 190

- 0.98667

Значущий - Так

NDD- PNAS



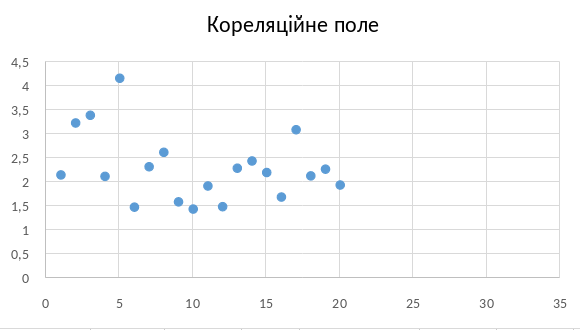
- 0,218

- 190

- 0.9375

Значущий - Так

NDD- TCC



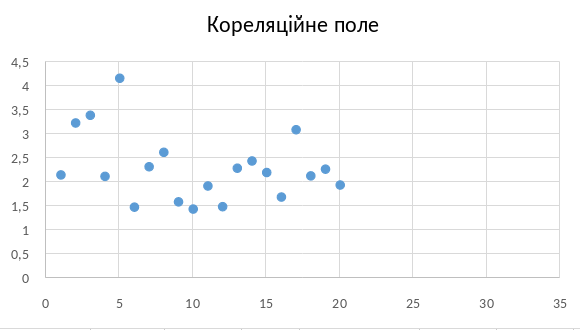
- 0,218

- 190

- 0.12637

Значущий - Так

NDD- WMC



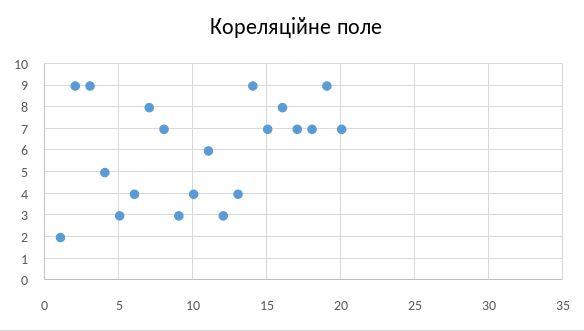
- 0,218

- 190

- 0.9927

Значущий - Так

CALL- PNAS



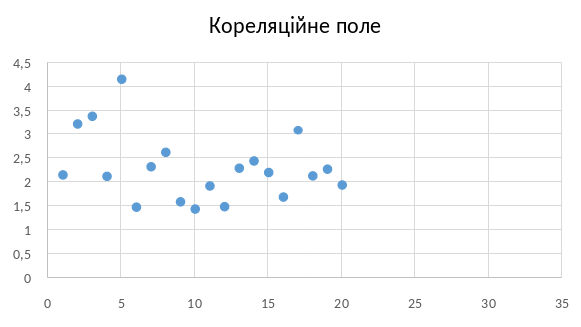
- 0,218

- 190

- 0.93752

Значущий - Так

CALL- TCC



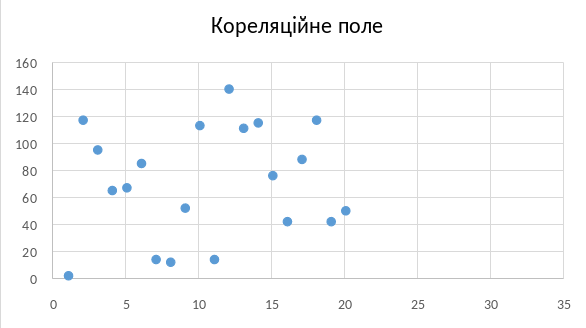
- 0,218

- 190

- 0.00526

Значущий - Так

CALL- WMC



- 0,218

- 190

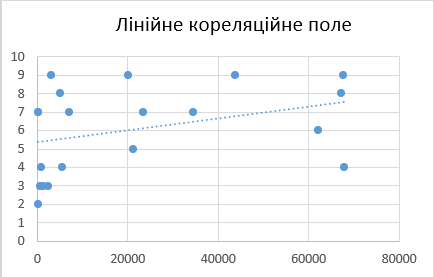
- 0.059418

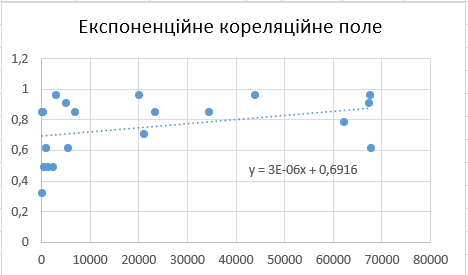
Значущий - Так

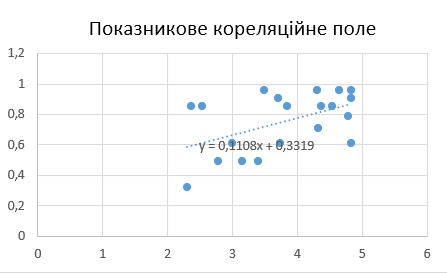
* **Побудувати лінію регресії та визначити її функцію:**

NOC – PNAS

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 21777,8 |
| |  | | --- | |  | | 6,05 |
| |  | | --- | |  | | 3,789197535 |
|  | 0,751849165 |
|  |  |
|  |  |
| Sxy | 408898,2000 |
| Sxlogy | 35347,3995 |
| Slogxlogy | 1,6139 |
|  |  |
| Sxx | 12778367559 |
| Slogxlogx | 14,5618 |
|  |  |
| b1 |  |
| Лінійний | 0,0000 |
| Експонен. | 0,0000 |
| Показник. | 0,1108 |
|  |  |
| b0 |  |
| Лінійний | 5,3531 |
| Експонен. | 0,6916 |
| Показник. | 0,3319 |

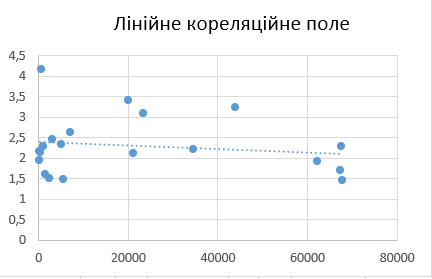


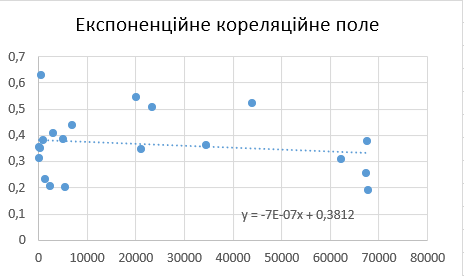




NOC - ТСС

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 21777,8 |
| |  | | --- | |  | | 2,308 |
| |  | | --- | |  | | 3,789197535 |
|  | 0,365279155 |
|  |  |
|  |  |
| Sxy | -54102,6580 |
| Sxlogy | -9315,3174 |
| Slogxlogy | -0,0632 |
|  |  |
| Sxx | 12778367559 |
| Slogxlogx | 14,5618 |
|  |  |
| b1 |  |
| Лінійний | 0,0000 |
| Експонен. | 0,0000 |
| Показник. | -0,0043 |
|  |  |
| b0 |  |
| Лінійний | 2,4002 |
| Експонен. | 0,3812 |
| Показник. | 0,3817 |

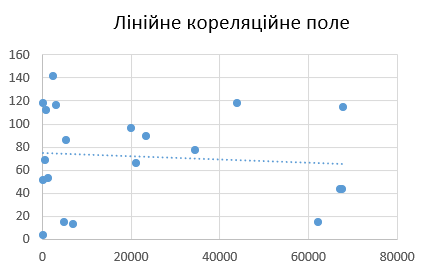


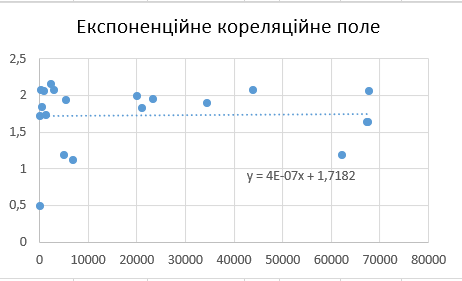




NOC - WMC

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | *21777,8* |
| |  | | --- | |  | | *71,85* |
| |  | | --- | |  | | *3,789197535* |
|  | *1,727140716* |
|  |  |
|  |  |
| *Sxy* | *############* |
| *Sxlogy* | *5232,3078* |
| *Slogxlogy* | *1,1349* |
|  |  |
| *Sxx* | *12778367559* |
| *Slogxlogx* | *14,5618* |
|  |  |
| *b1* |  |
| *Лінійний* | *-0,0001* |
| *Експонен.* | *0,0000* |
| *Показник.* | *0,0779* |
|  |  |
| *b0* |  |
| *Лінійний* | *75,1062* |
| *Експонен.* | *1,7182* |
| *Показник.* | *1,4318* |

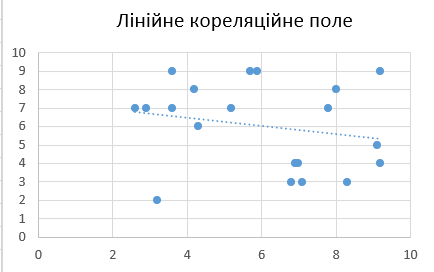


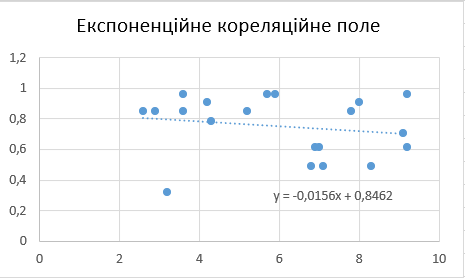


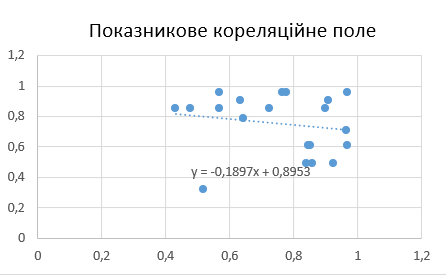


NOM - PNAS

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 6,03 |
| |  | | --- | |  | | 6,05 |
| |  | | --- | |  | | 0,756493087 |
|  | 0,751849165 |
|  |  |
|  |  |
| Sxy | -21,0300 |
| Sxlogy | -1,4656 |
| Slogxlogy | -0,1097 |
|  |  |
| Sxx | 93,662 |
| Slogxlogx | 0,5786 |
|  |  |
| b1 |  |
| Лінійний | -0,2245 |
| Експонен. | -0,0156 |
| Показник. | -0,1897 |
|  |  |
| b0 |  |
| Лінійний | 7,4039 |
| Експонен. | 0,8462 |
| Показник. | 0,8953 |

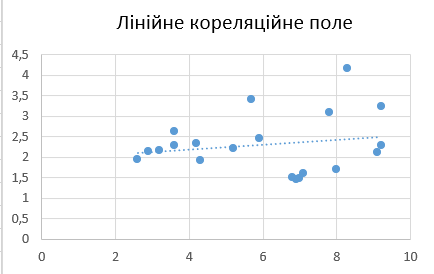


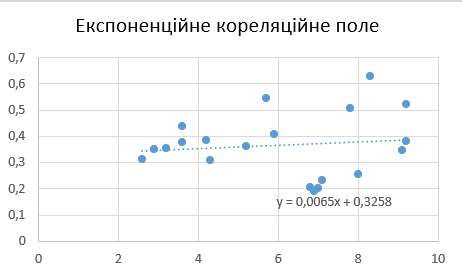


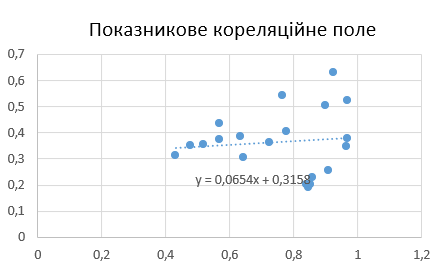


NOM - TCC

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 6,03 |
| |  | | --- | |  | | 2,308 |
| |  | | --- | |  | | 0,756493087 |
|  | 0,365279155 |
|  |  |
|  |  |
| Sxy | 5,6682 |
| Sxlogy | 0,6131 |
| Slogxlogy | 0,0379 |
|  |  |
| Sxx | 93,662 |
| Slogxlogx | 0,5786 |
|  |  |
| b1 |  |
| Лінійний | 0,0605 |
| Експонен. | 0,0065 |
| Показник. | 0,0654 |
|  |  |
| b0 |  |
| Лінійний | 1,9431 |
| Експонен. | 0,3258 |
| Показник. | 0,3158 |



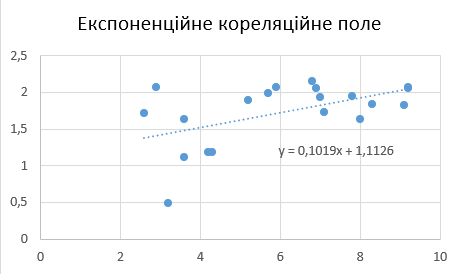


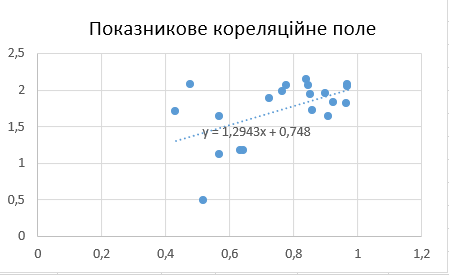


NOM - WMC

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 6,03 |
| |  | | --- | |  | | 71,85 |
| |  | | --- | |  | | 0,756493087 |
|  | 1,727140716 |
|  |  |
|  |  |
| Sxy | 823,2900 |
| Sxlogy | 9,5451 |
| Slogxlogy | 0,7489 |
|  |  |
| Sxx | 93,662 |
| Slogxlogx | 0,5786 |
|  |  |
| b1 |  |
| Лінійний | 8,7900 |
| Експонен. | 0,1019 |
| Показник. | 1,2943 |
|  |  |
| b0 |  |
| Лінійний | 18,8462 |
| Експонен. | 1,1126 |
| Показник. | 0,7480 |

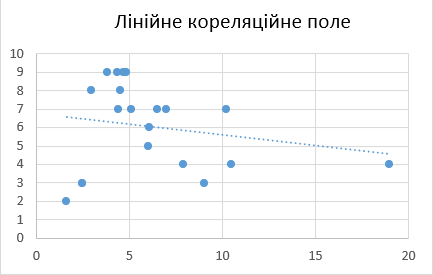


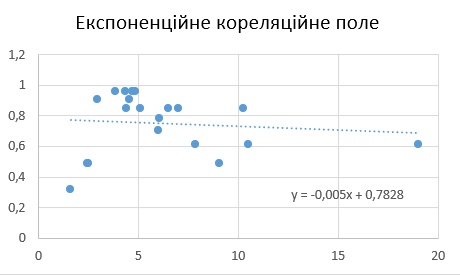


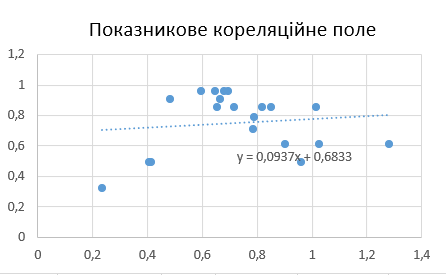


NDD- PNAS

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 6,1815 |
| |  | | --- | |  | | 6,05 |
| |  | | --- | |  | | 0,731776557 |
|  | 0,751849165 |
|  |  |
|  |  |
| Sxy | -33,9915 |
| Sxlogy | -1,4493 |
| Slogxlogy | 0,1059 |
|  |  |
| Sxx | 289,746255 |
| Slogxlogx | 1,1305 |
|  |  |
| b1 |  |
| Лінійний | -0,1173 |
| Експонен. | -0,0050 |
| Показник. | 0,0937 |
|  |  |
| b0 |  |
| Лінійний | 6,7752 |
| Експонен. | 0,7828 |
| Показник. | 0,6833 |

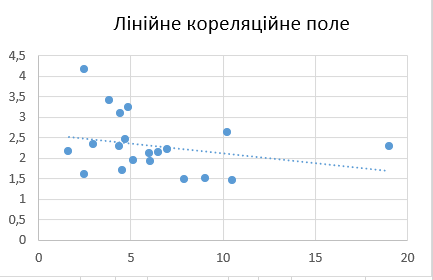


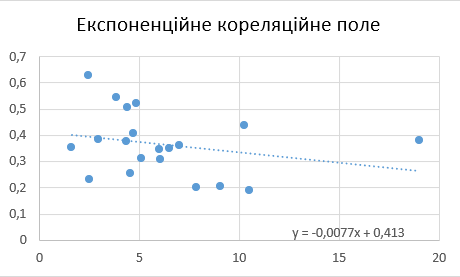


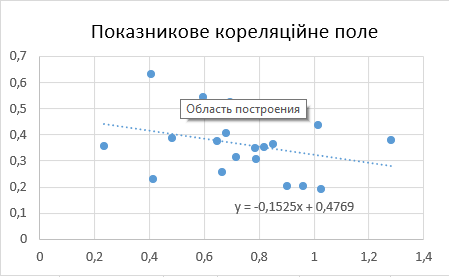


NDD- TCC

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 6,1815 |
| |  | | --- | |  | | 2,308 |
| |  | | --- | |  | | 0,731776557 |
|  | 0,365279155 |
|  |  |
|  |  |
| Sxy | -13,9594 |
| Sxlogy | -2,2387 |
| Slogxlogy | -0,1724 |
|  |  |
| Sxx | 289,746255 |
| Slogxlogx | 1,1305 |
|  |  |
| b1 |  |
| Лінійний | -0,0482 |
| Експонен. | -0,0077 |
| Показник. | -0,1525 |
|  |  |
| b0 |  |
| Лінійний | 2,6058 |
| Експонен. | 0,4130 |
| Показник. | 0,4769 |

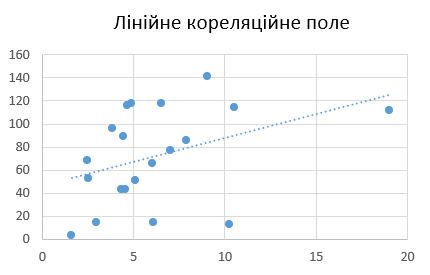


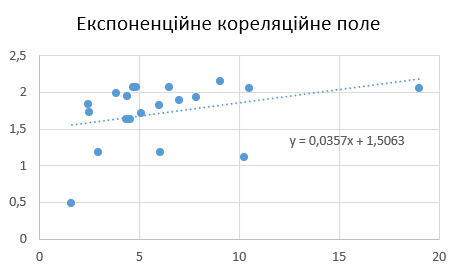


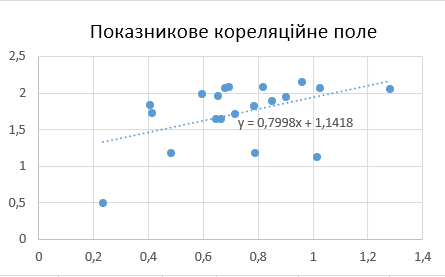


NDD- WMC

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 6,1815 |
| |  | | --- | |  | | 71,85 |
| |  | | --- | |  | | 0,731776557 |
|  | 1,727140716 |
|  |  |
|  |  |
| Sxy | 1197,0645 |
| Sxlogy | 10,3526 |
| Slogxlogy | 0,9042 |
|  |  |
| Sxx | 289,746255 |
| Slogxlogx | 1,1305 |
|  |  |
| b1 |  |
| Лінійний | 4,1314 |
| Експонен. | 0,0357 |
| Показник. | 0,7998 |
|  |  |
| b0 |  |
| Лінійний | 46,3116 |
| Експонен. | 1,5063 |
| Показник. | 1,1418 |

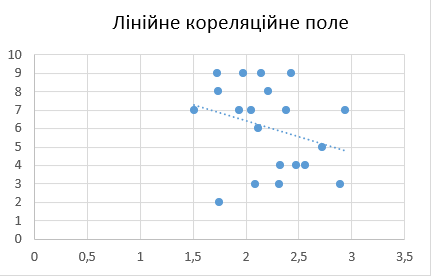


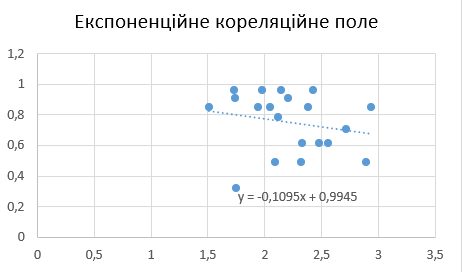


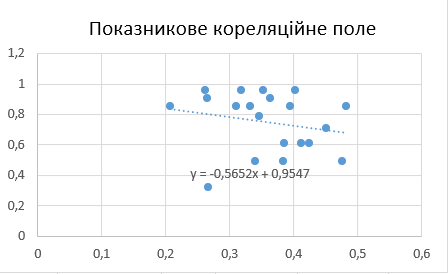


CALL- PNAS

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 2,216 |
| |  | | --- | |  | | 6,05 |
| |  | | --- | |  | | 0,358813957 |
|  | 0,751849165 |
|  |  |
|  |  |
| Sxy | -4,9860 |
| Sxlogy | -0,3158 |
| Slogxlogy | -0,0588 |
|  |  |
| Sxx | 2,88428 |
| Slogxlogx | 0,1041 |
|  |  |
| b1 |  |
| Лінійний | -1,7287 |
| Експонен. | -0,1095 |
| Показник. | -0,5652 |
|  |  |
| b0 |  |
| Лінійний | 9,8808 |
| Експонен. | 0,9945 |
| Показник. | 0,9547 |

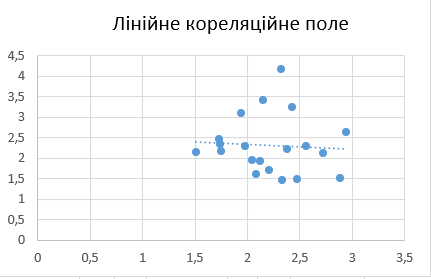




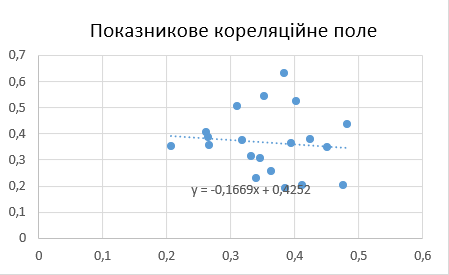


CALL- TCC

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | 2,216 |
| |  | | --- | |  | | 2,308 |
| |  | | --- | |  | | 0,358813957 |
|  | 0,365279155 |
|  |  |
|  |  |
| Sxy | -0,3281 |
| Sxlogy | -0,0976 |
| Slogxlogy | -0,0174 |
|  |  |
| Sxx | 2,88428 |
| Slogxlogx | 0,1041 |
|  |  |
| b1 |  |
| Лінійний | -0,1137 |
| Експонен. | -0,0338 |
| Показник. | -0,1669 |
|  |  |
| b0 |  |
| Лінійний | 2,5600 |
| Експонен. | 0,4403 |
| Показник. | 0,4252 |

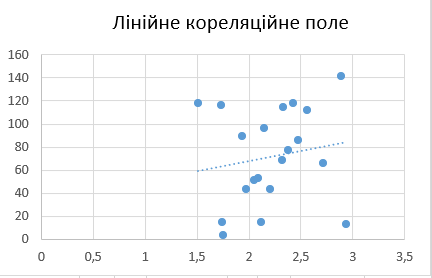


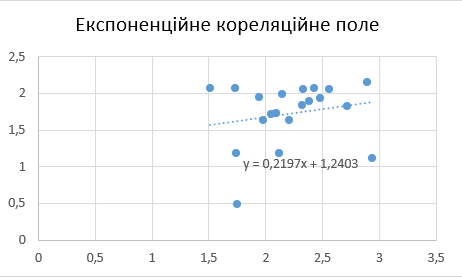


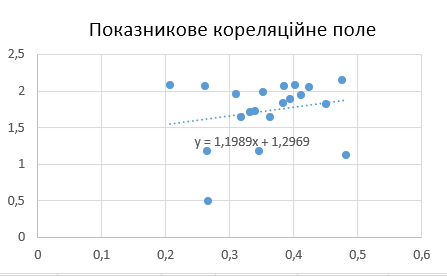


CALL- WMC

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | *2,216* |
| |  | | --- | |  | | *71,85* |
| |  | | --- | |  | | *0,358813957* |
|  | *1,727140716* |
|  |  |
|  |  |
| *Sxy* | *50,7780* |
| *Sxlogy* | *0,6337* |
| *Slogxlogy* | *0,1248* |
|  |  |
| *Sxx* | *2,88428* |
| *Slogxlogx* | *0,1041* |
|  |  |
| *b1* |  |
| *Лінійний* | *17,6051* |
| *Експонен.* | *0,2197* |
| *Показник.* | *1,1989* |
|  |  |
| *b0* |  |
| *Лінійний* | *32,8371* |
| *Експонен.* | *1,2403* |
| *Показник.* | *1,2969* |







**Зробити висновок про залежності між метриками, провести інтерпретацію та пояснення отриманих даних:**

**Пояснити вибір способу розрахунку коефіцієнта кореляції:**

Якщо метрики мають нормальний розподіл, то використовував 

Якщо у метрики не є нормальний закон розподілу, то використовують розрахунок парної рангової кореляції Спірмена чи Кендала. Рекомендовано використати розрахунок рангового коефіцієнта Кендала, так як він є простішим у реалізації. Для цього потрібно розрахувати:



Побудувати кореляційні поля для 5-и найбільш ймовірних пар (у яких присутня залежність)

Так як первина оцінка кореляції показала , що 12 пар мають кореляцію більше ніж 0.9, а це свідчить про високий ківень за показником кендела, було побудовано 12 ліній регресій у показниковому, експоненціальному і лінійному полях.

На кореляційному полі побудувано кілька функцій, які найбільш ймовірно показують лінію регресії а саме показникова, експоненціальна і лінійнома.

**Кількість обраних функцій потрібно пояснити (при цьому обрати не менше 3-х функцій регресії):**

До нелінійних регресійних моделей по оцінюваним параметрами відносяться функції:

* + степенева - ;
  + показникова - ;
  + експоненціальна - .

Дані моделі за своєю природою *не є лінійними*, тому що вони не можуть бути представленими у вигляді простої регресійної моделі з деякими перетвореннями незалежних змінних.

Розглянемо деякі нелінійні моделі, для яких можливо зведення до лінійних. Для того, щоб звести нелінійну модель до лінійної (лінеаризувати модель) зазвичай за допомогою деяких перетворень змінних нелінійну модель представляють у вигляді лінійного співвідношення між перетвореними змінними, оцінюють коефіцієнти цього співвідношення і потім, за допомогою зворотного перетворення, знаходять оцінки параметрів вихідної нелінійної моделі. Відразу зауважимо, що не всяка нелінійна модель може бути оцінена таким чином, в ряді випадків неможливо підібрати необхідне перетворення, що лінеаризує модель. У цьому випадку доводиться використовувати методи нелінійної оптимізації.

Оцінка нелінійних моделей проводиться за допомогою пояснюючих змінних, але лінійних – за параметром, що оцінюється та не представляє особливої складності. В цьому випадку зазвичай використовують заміну змінних для зведення моделі до лінійної та оцінки параметрів за допомогою звичайного МНК (метод найменших квадратів, що застосовується до моделі з заміненими змінними).

**Шляхом перевірки обраних функцій визначити найбільш оптимальну:**

Найбільш оптимальною функцією оцінки кореляції стала лінійна, тому що мала найменьший емпіричне відхилення.

Для прикоаду візьмемо кільна найбільш кореляційних пар як CALL -TCC, CALL – PNAS, NOP -TCC і проведемо обрахунок емпіричного відхилення від теоретичної функції регресії. Отримаємо результати:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Лінійна | Експоненційна | Показникова |
| CALL -TCC | 9,53 | 9,55 | 9,59 |
| CALL – PNAS | 96,33 | 99,54 | 98,01 |
| NOP -PNAS | 100,86 | 105,18 | 110,52 |

**Зробити висновок про залежності між метриками, тобто провести інтерпретацію та пояснення отриманих дани:**

Після оцінки кореляції метрик ми отримали такі дані. Деякі з пар метрик чітко корелюють. Для розуміння нам треба проінтерпретувати значення метрик і логічно співставити значення і використанні у коді. Наприклад кореляції метрик NOM-TCC показала непоганий результат. Тому що TCC метрика що містить значення щільності методів у класі. А метрика NOM , саме кількість методів. Тому за збільшеням методів корелює їх ушільненям у програмі.

Також візьмемо приведу вишче пару метрик NOP -PNAS. Вони корельють меньш однозначно, але теж можено бачити закономірності у їх функціях. Так як метрика NOP – це кількість відкритих атрибутів класу а PNAS - кількість відкритих методів класу, що не перевантажені або спеціалізовані відносно попередників, поділено на загальну кількість відкритих методів. Закономірно ми можемо передбачити що відкриті методи класу взаємно пропрорційні з відкритими полями у класі але не завжди.

# Висновки

В цій лабораторній роботі мною було досліджено що таке коефіцієнт кореляції. Які властивості коефіцієнта кореляції. Види коефіцієнтів кореляції. Розрахована парна рангова кореляція. Визначитизалежність між метриками. Всі розрахунки проводилися у MS Excel. Файл з розрахунками додано до звіту.