

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ   
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

ФАКУЛЬТЕТ БІОМЕДИЧНОЇ ІНЖЕНЕРІЇ

КАФЕДРА БІОМЕДИЧНОЇ КІБЕРНЕТИКИ

**Практична робота №2**

## з дисципліни «Вступ до інтелектуального аналізу даних»

# на тему: «Відпрацювання навичок застосування методів розрахунку

# параметрів регресійних моделей на прикладі задач підбору

# стимуляторів імунної системи онкохворих»

**Виконав:**

студент гр. БС-03

Затуловський Г.А.

**Перевірив:**

доц. Павлов В.А.

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

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

(підпис викладача)

Київ-2022

**Теоритичні відомості:**

**Спрощений вивід формули МНК у матричному вигляді**

Нагадаємо

1. **Множення** – **А\*В=С** –

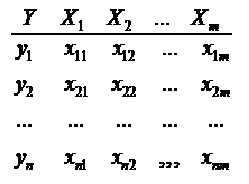
: для одержання jk-того елементу **С** - в **А** беруть j-ту строку, в **В** k-тый столбець та їх скалярно перемножують.

2. **Траспонування**

– столбці робимо строками, строки столбцями

3.Обернена матриця А-1 матриці А це така матриця

для якої виконуєтся

**А-1 \*А=Е**, де  одинична матриця

**Отже** маємо матрицю Х та вектор виходу У



Необхідно побудувати модель лінейної структури

При цьому необхідно максимально наблизити за допомогою вектора а співвідношення 

(\*)

Далі будемо виходити з (\*) як з рівності і будемо з нього шукати а. Більш обгрунтовано висновок проведемо на лекції. Отже маємо

Для того щоб вивільнити *а*, необхідно помножити матрицю при *а* її обернену. Проте ця матриця для виконання цієї операції повинна бути квадратною. Щоб досягти такого, помножимо зліва та справа рівняння (\*) на ХТ , одержимо



Тепер помножимо зліва та спрва на обернену (Х ХТ)-1

и тоді кінцево маємо для *а*



и для моделі



Все

**Оценка качества модели**

**Рассматриваем**

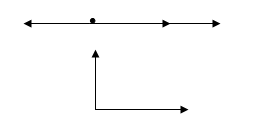
**Коэффициент корреляции, детерминации, СКО, ОСКО и НОСКО и связь между ними**

1.В продолжении лаб1 введем своб. член в модель (введем Хо ) и найдем регрессию ….

2. Как же сравнить полученные модели - какая лучше?

**Наиболее часто применяемые оценки качества модели – как правило достаточно: коэффициент корреляции, детерминации, СКО, ОСКО и НОСКО**

А.Степень связи модели с реальностью - то есть коэффициент **корреляции между моделью и табл. данными**

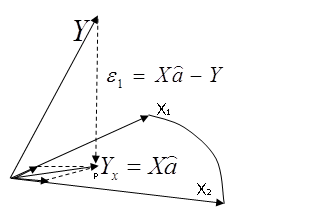
и коэффициент детерминации 

----------------

Но учтем, что функционал, в соответствии с которым мы находили параметры – работал на минимум КО - мин

а это не совсем одно и тоже, что корреляция

– корреляция связана с углом между векторами (в данном случае У и Ум) и от **масштаба векторов**  не зависит а ошибка (невязка)

– **зависит** .Напомним что корреляция 

максимальна (минимальна):

когда угол

между векторами равен 0 (180) град,

когда вектора коллинеарны и

минимальна по модулю - когда вектораортогональны.

На рисунке видим что

корреляция в обоих

случаях одна и та же (**углы**

**между векторами Y и равны**)

А **невязки ε** и **различны**

– ошибка модели во втором

случае меньше.

Поэтому сравнивать качество моделей в рамках **одной и той же задачи** корректнее на основании

Б/ σ – **среднеквадратичной ошибки СКО**

( **-** табличные значения**,** - модельные значения):

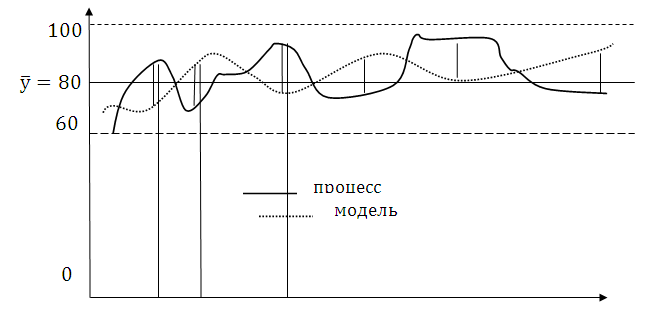
Однако это неудобно (хотя спсс и использует) – в результате имеем оценку в **абсолютных единицах** и сравнить качество моделей из разных задач нельзя – (например градусы с литрами).

Сравнивать же качесто моделей различных задач- корекнее по

относительному показателю

3/ **относительная СКО - ОСКО**

Однако объективной оценке динамики модели здесь мешает среднее. Действительно, если полоса основной динамики изменения процесса и модели находятся в некоторых пределах вдали от 0, то оценка модели будет явно искажена:





Мы будем оценивать не качество, как модель воспроизводит **динамику процесса,** а качество относительно 0-ля процесса что необъективно завышает качество модели.

Проблему решает вычитание среднего табличного в знаменателе – НОСКО:

Это наиболее распространенная формула **оценки модели**

**Связь между КD и НОСКО: КD=1-**

**Завдання:**

***Відпрацювання навичок застосування методів розрахунку параметрів регресійних моделей на прикладі задач підбору стимуляторів імунної системи онкохворих***

***Частина 1***. Розробка та дослідження поведінки регресійних моделей прогнозу відгуку імунної системи онкохворих для різних структур моделей

***Завдання*** Розрахувати відповідно до додатку 2 параметри моделі складності 3 для значень кількості Т-клітин імунної відповіді на 4-й годині експерименту. Побудувати графік моделі. Підвищити складність моделі до 4-х та розрахувати її параметри. Побудувати графік моделі. Порівняти одержані результати.

***Частина 2***. Додаткове завдання: Дослідження поведінки окремих мір якості регресійних моделей при оцінці моделей різних структур.

***Завдання*** Розрахувати міри якості розроблених моделей згідно до додатку 2. Визначити особливості значень застосувань окремих мір. Зробити висновки про доцільність ускладнення моделей з метою підвищення їх точності.

*Додаткове завдання виконується за бажанням для отримання оцінки «respect»*

**Виконання:**

**ЧАСТИНА 1**

Спочатку розрахуємо параметри моделі складності 3:

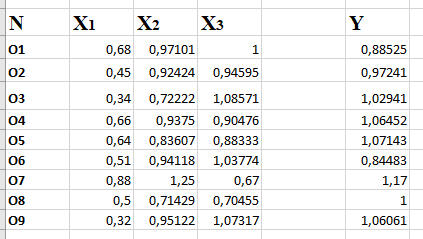


Рис.1 – Матриця з початковими даними

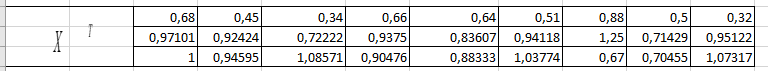


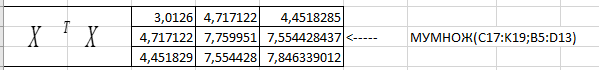
Рис.2 – Транспортована початкової матриці 

Рис.3 - Множення транспонтованої матриці на початкову

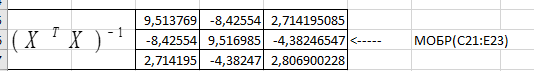


Рис.4 – Обернена матриця

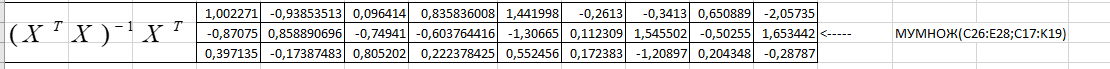


Рис.5 – Множення оберненої на транспортовану матрицю

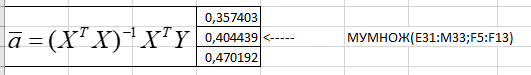


Рис.6 – Розрахуємо вектор а

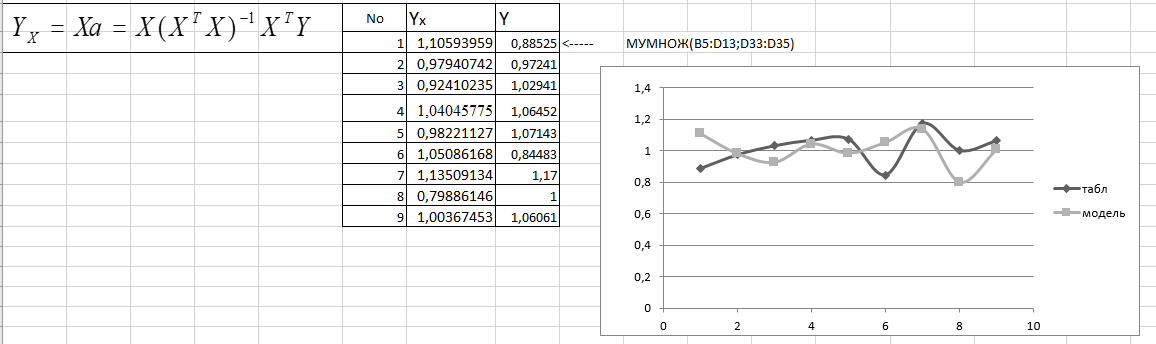


Рис.7 - Побудувати графік моделі складності 3

Зробимо тіж самі операції для моделі складності 4 ( з вільним членом)

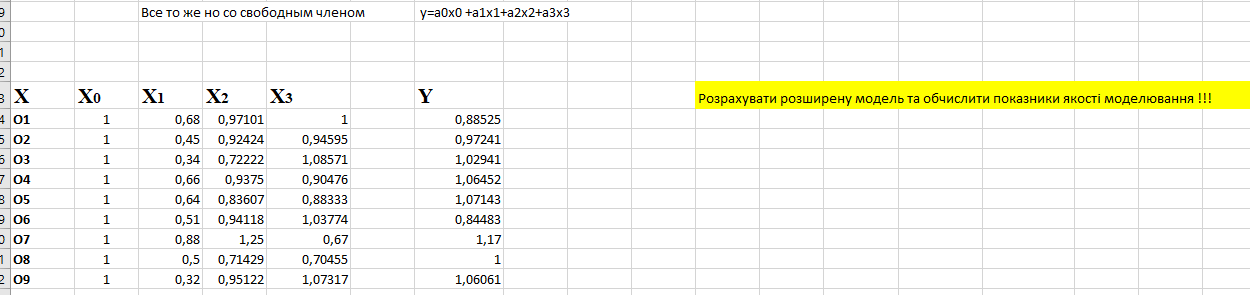


Рис.8 - Матриця з початковими даними

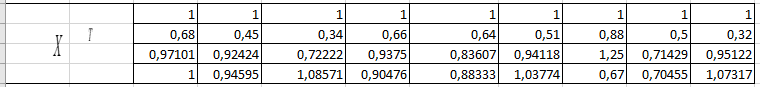


Рис.9 - Транспортована початкової матриці

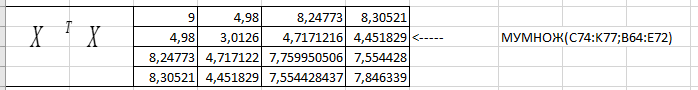


Рис.10 - Множення оберненої матриці на початкову

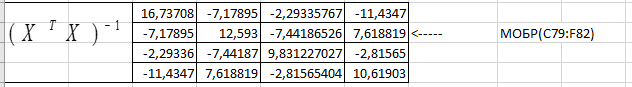


Рис.11 – Обернення матриця

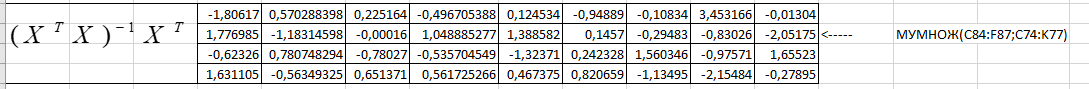


Рис.12 - Множення оберненої на транспортовану матрицю

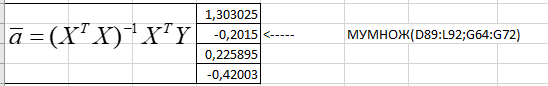


Рис.13 - Розрахуємо вектор а

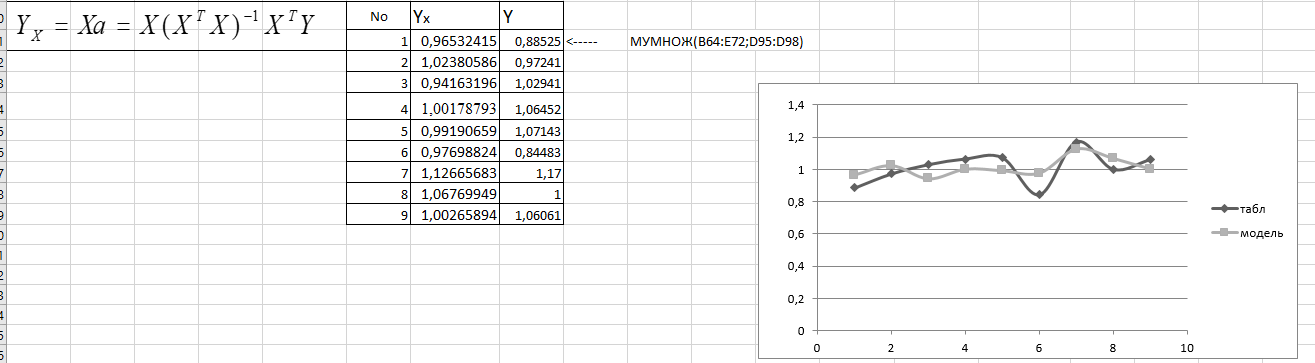
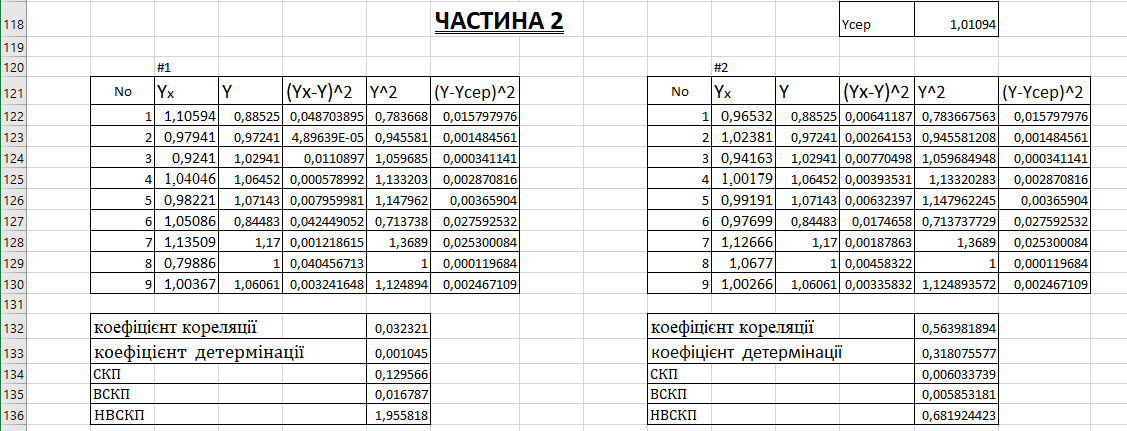


Рис.14 – Побудувати графік моделі складності 4

Розглядаючи графік моделі складності 3 (рис.7) та 4 (рис.14) можно зробити висновок, що на графіку для моделі складності 4 графік таблиці ближчий до графіку табличних значень ніж у моделі складності 3

**ЧАСТИНА 2**

Перейдемо до аналізу отриманих результатів, розрахуємо міри якостей наших розроблених моделей.

 Рис.15 – оцінки якості для моделі складності 3 та 4

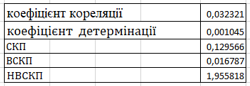
****

Рис.16 – Критерії оцінки для моделі складності 3

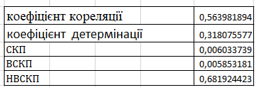


Рис.17 - Критерії оцінки для моделі складності 4

Отже, можна зробити висновок, що модель зі складністю 4 показує більш якісний результат.

Коефіцієнт кореляції, коефіцієнт детермінацій у моделі складності 4 більший (0,032321 < 0,563981894 - коефіцієнт кореляції, 0.001045 < 0,318075577 - коефіцієнт детермінацій) і є ближчим до абсолютного значення. Його СКП (середньоквадратична помилка) , ВСКП (відносна середньоквадратична помилка) у моделі складністю 4 є меншою (0,129566 > 0,006033739 – CКП, 0,016787 > 0,005853181 – ВСКП). НВСКП(Нормована відносна середньоквадратична помилка) моделі зі складністю 3 є більшою за 1, що означає що ця модель є поганою.

**Висновок:** Ми вираховували дані для моделі та знаходили параметри для оцінки моделі та застосували методи для розрахунку параметрів регресійних моделей на прикладі задач підбору стимуляторів імунної системи онкохворих