

Лабораторна робота №1 з ММС

Некряч Владислав, ТК-31

## Завдання до роботи

Задано спостереження за деякою системою у вигляді таблиці 1 ( $x_1$ ,  $x_2$ ,  $x_3$  – вхідні (незалежні) змінні,  $y$  – вихідна (залежна змінна)).

Перевірити вхідні змінні на наявність мультиколінеарності за алгоритмом Фаррара-Глобера.

Використовуючи МНК провести ідентифікацію об'єкта моделювання, тобто визначивши невідомі параметри  $b_0$ ,  $b_1$ ,  $b_2$ ,  $b_3$  у моделі

Таблицю 1 наповнити даними самостійно. Функції  $f_1(x_1, x_2, x_3)$ ,  $f_2(x_1, x_2, x_3)$ ,  $f_3(x_1, x_2, x_3)$  вибрати самостійно. Для отриманої апроксимаційної функції  $y(x_1, x_2, x_3)$  (моделі об'єкта спостереження) виконати кореляційно-регресійний аналіз. Проаналізувати результати.

Дослідити отриману модель на гетероскедастичність (тест обираєте самі) та на автокореляцію (тест Дарбіна-Уотсона).

№	$x_1$	$x_2$	$x_3$	$y$
1	1	2	5	10
2	3	7	11	29
3	5	9	13	27
4	9	4	10	25
5	6	11	13	28
6	9	12	17	34
7	19	22	17	37
8	23	25	19	39
9	24	26	27	47

10	26	30	29	51
11	27	32	33	56
12	29	36	34	56
13	32	36	40	63
14	32	39	44	67
15	35	41	46	69
16	37	42	47	65
17	52	52	37	63
18	54	55	39	63
19	55	56	40	64
20	59	57	44	77

## Хід роботи

1) Перевірка вхідних змінних на мультиколінеарність за алгоритмом Фаррара-Глобера

1. Проводимо стандартизацію (нормалізацію) вхідних змінних за формулою:

$$x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{n\sigma_{x_k}^2}}, k = \overline{1, m}, i = \overline{1, n}$$

$$\sigma_{x_k}^2 = \frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}{n} - \text{дисперсія змінної } x_k,$$

$\bar{x}_k$  – середнє значення змінної  $x_k$ ,

$m$  – кількість вхідних змінних,

$n$  – кількість спостережень.

2. Знаходимо кореляційну матрицю:  $r = X^{*T} X^*$ , яка складається з коефіцієнтів кореляції між  $x_i$  та  $x_j$ , взятих попарно;  $X^*$  – стандартизована матриця, отримана у пункті 1, на її діагоналі – одиниці.

3. Визначаємо наявність мультиколінеарності за критерієм Пірсона  $\chi^2$ :

$$\chi^2 = - \left( n - 1 - \frac{2m+5}{6} \right) \ln |r|, |r| = \det r$$

$$\chi_{\text{кр}}^2 = \chi^2 \left( \alpha, \frac{m(m-1)}{2} \right)$$

$$\chi^2 = 99.869$$

$$\chi_{\text{кр}}^2 = 7.8147$$

$\chi^2 > \chi_{\text{кр}}^2 \Rightarrow$  серед змінних  $x_1, x_2, x_3$  існує мультиколінеарність.

4. Визначимо матрицю  $r^{-1} = \begin{pmatrix} X^{*T} & X^* \end{pmatrix}^{-1}$ .

5. Визначимо наявність мультиколінеарності між кожною зі змінних  $x_i$  та рештою змінних за критерієм Фішера:

$$F_i = \left| c_{ii} - 1 \right| \frac{n-m}{m-1}, i = \overline{1, m},$$

$c_{ii}$  – діагональний елемент матриці  $C$  з пункту 4.

$$F_{\text{кр}} = F(\alpha, n - m, m - 1)$$

$$F_{\text{кр}} = 19.437$$

$$F_1 = 456.95$$

$$F_2 = 709.08$$

$$F_3 = 67.999$$

$F_1, F_2, F_3 > F_{\text{кр}} \Rightarrow$  кожна зі змінних мультиколінеарна з рештою.

6. Знаходимо частинні коефіцієнти кореляції для кожної пари змінних  $x_i, x_j, i \neq j; i, j = \overline{1, m}$ :

$$r(x_i, x_j) = \frac{-c_{ij}}{\sqrt{c_{ii}c_{jj}}},$$

$c_{ij}, c_{ii}, c_{jj}$  – елементи матриці  $C$  з пункту 4.

7. Перевіряємо наявність мультиколінеарності між парами змінних  $x_i, x_j, i \neq j$  за критерієм Стюдента:

$$t(x_i, x_j) = \left| \frac{r(x_i, x_j) \sqrt{n-m}}{\sqrt{1-r^2(x_i, x_j)}} \right|$$

$$t_{\text{кр}} = t\left(\frac{\alpha}{2}, n - m\right)$$

$$t_{\text{кр}} = 2.458$$

$$t(x_1, x_2) = 14.723$$

$$t(x_1, x_3) = 2.8146$$

$$t(x_2, x_3) = 4.6284$$

$t(x_1, x_2), t(x_1, x_3), t(x_2, x_3) > t_{\text{кр}} \Rightarrow$  змінні  $x_1, x_2, x_3$  попарно мультиколінеарні.

2) Формування гіпотези про вид функціональної залежності  
Критерій МНК:

$$\Phi(b) = \sum_{i=1}^n \left( f(x_{i1}, x_{i2}, \dots, x_{im}, b) - y_i \right)^2 \rightarrow \min$$

Необхідні умови екстремуму критерію:

$$\left\{ \frac{\partial \Phi}{\partial b_0} = 0 \quad \dots \quad \frac{\partial \Phi}{\partial b_m} = 0 \right.$$

Цю систему нормальних рівнянь зручно подати у матричній формі:

$$X^T X b = X^T Y$$

Тоді розв'язком цієї системи відносно невідомих  $b$  буде:

$$b = (b_0 \ b_1 \ : \ b_m) = (X^T X)^{-1} X^T Y$$

$$b_0 = -1.3334$$

$$b_1 = 7.3352$$

$$b_2 = 9.9107$$

$$b_3 = 0.3564$$

Маємо:

$$f(x_1, x_2, x_3) = -1.3334 + 7.3352 \ln(x_1) + 6.4076 \cos(\cos(x_2)) + 0.3564(x_1 + x_3)$$

Значення відгуку моделі та значення відгуку моделі, розраховані за функцією:

No	Y	Y*
1	10	9.8699586 4252882
2	29	21.438370 4927416
3	27	26.405678 3930146
4	25	33.350703 9283343
5	28	32.171903 7159606
6	34	34.564328 848296
7	37	41.240274 2054158
8	39	44.047781 5855065
9	47	49.829788 5104965

10	51	53.274524 0860203
11	56	51.964096 067027
12	56	56.248861 2997419
13	63	59.429118 9709739
14	67	60.585593 1694383
15	69	58.146202 0470552
16	65	62.759187 4821717
17	63	63.447415 4605848
18	63	64.688567 8848611
19	64	61.847768 5552202
20	77	84.689876 6546111

3) Значення критерію найменших квадратів та кореляційно-регресійний аналіз

Стандартні похибки параметрів  $b_i$ :



$$S_{b_0} = 6.4673, S_{b_1} = 1.1142, S_{b_2} = 6.9673, S_{b_3} = 0.0816$$

Коефіцієнт детермінації:  $R^2 = 0.9297$  – близький до 1, побудована залежність є високоякісною.

Коефіцієнт кореляції:  $R = 0.9642$  – згідно зі шкалою Чеддока, маємо дуже сильний зв'язок  $y = f(x)$ .

$F$ -статистика:

$$F = \frac{R^2/m}{(1-R^2)/(n-m-1)},$$

$$F_{\text{кр}} = F(\alpha, m, n - m - 1)$$

$F = 70.575, F_{\text{кр}} = 2.2756, F > F_{\text{кр}} \Rightarrow$  значення коефіцієнта кореляції істотне, модель адекватна.

Статистичну значущість коефіцієнтів  $b_0, b_1, \dots, b_m$  перевіримо за критерієм Стьюдента:

$$t(b_i) = \left| \frac{b_i}{S_{b_i}} \right|, i = \overline{0, m}$$

$$t_{\text{кр}} = t\left(\frac{\alpha}{2}, n - m - 1\right)$$

$$t_{\text{кр}} = 2.4728$$

$$t(b_0) = 0.2061$$

$$t(b_1) = 6.5828$$

$$t(b_2) = 1.4224$$

$$t(b_3) = 4.3645$$

$t(b_0), t(b_2) < t_{\text{кр}}, t(b_1), t(b_3) > t_{\text{кр}} \Rightarrow b_0, b_2$  статистично незначущі,  $b_1, b_3$  статистично значущі.

Довірчий інтервал для коефіцієнта  $b_i$ :

$$b_i \in \left[ b_i - t_{\text{кр}} S_{b_i}; b_i + t_{\text{кр}} S_{b_i} \right]$$

$$b_0 \in [-17.3263, 14.6594]$$

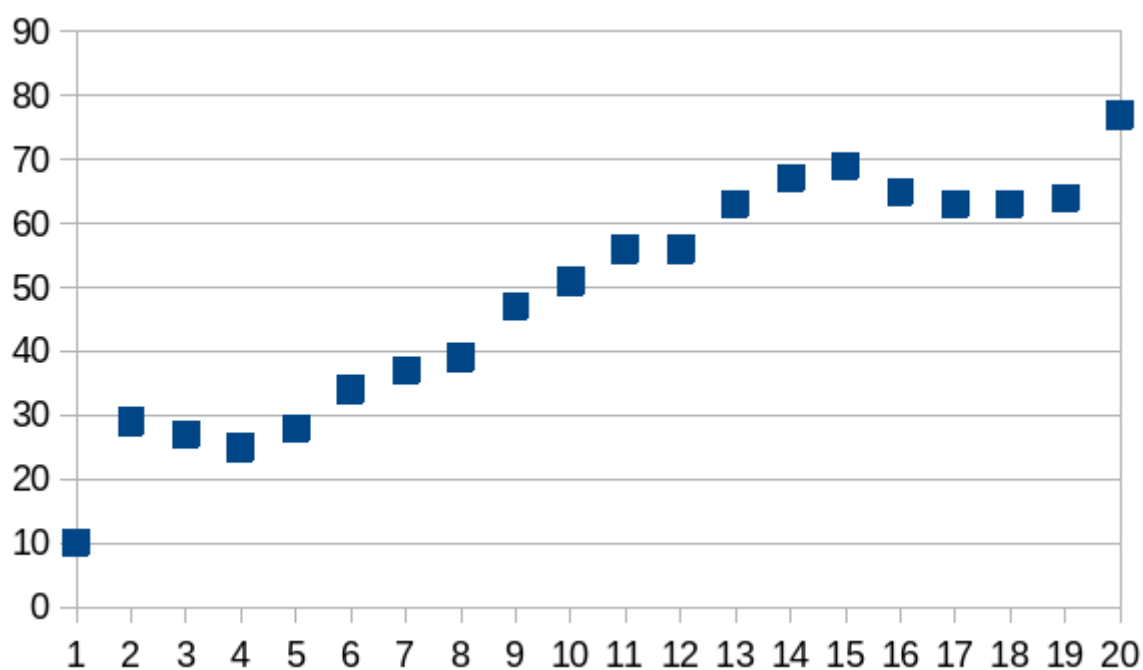
$$b_1 \in [4.5797, 10.09]$$

$$b_2 \in [-7.3187, 27.1401]$$

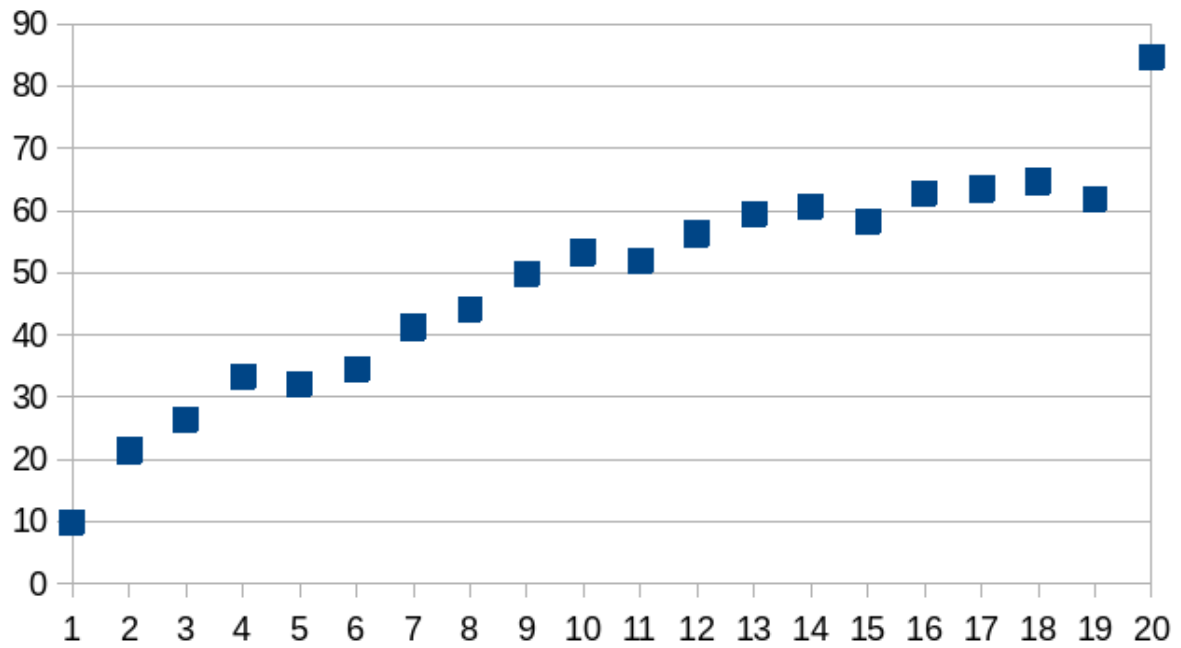
$$b_3 \in [0.1544, 0.5583]$$

Графіки функцій:

*Зафіксовані дані:*



*Апроксимація:*



Графік функції за апроксимацією доволі близький до графіку функції за спостережуваними даними.

### 3) Гетероскедстичність (тест кореляції Спірмена)

Розрахуємо коефіцієнти рангової кореляції Спірмена між значеннями незалежних змінних та залишками  $u_i$ , використовуючи наступну формулу:

$$r_{xu} = 1 - 6 \frac{\sum_{i=1}^n d_i^2}{n(n^2 - 1)},$$

де  $d_i = rang(x_i) - rang(u_i)$ ,  $n$  – розмір вибірки.

Corr X1 ↔ y	Corr X2 ↔ y	Corr X3 ↔ y
0.1098570	0.0518992	0.2958225
66464056	138996467	44605529

Знаходимо для кожної змінної відповідне значення t-критерію за наступною формулою:

$$t = \left| \frac{r_{xu} \sqrt{n-2}}{\sqrt{1-r_{xu}^2}} \right|.$$

0.494	0.2261	1.4956
-------	--------	--------

Критичне значення:  $t_{кр} = t(\alpha/2, n - 2) = 2.445$

Гетероскедстичність відсутня для кожної з змінних, т.я. значення критерію менше за критичне значення -> гіпотеза про наявність гетероскедстичності відкидається.

#### 4) Автокореляція залишків (тест Дарбіна-Уотсона)

Вираховується критерій Дарбіна-Уотсона, який тісно пов'язаний з автокореляцією:  $DW \sim 2(1 - r(u_i - u_{i-1}))$ .

$$DW = \frac{\sum_{i=2}^n (u_i - u_{i-1})^2}{\sum_{i=1}^n u_i^2}.$$

В нашому випадку  $DW = 1.314$ .

Табличні межі критерію для 3 змінних та 20 спостережень для рівня значущості 0.05:  $DL = 1$ ,  $DU = 1.68$

Т.я.  $DL < DW < DU$ , то висновків про наявність або відсутність автокореляції зробити не можна.

#### **Висновок**

Незважаючи на те що серед змінних наявна мультиколінеарність, побудована модель є адекватною та має високі показники якості зв'язку, більшість її коефіцієнтів є статистично значущими, тому підсумковий графік функції, побудований за апроксимацією, є близьким до графіку функції, побудованого за спостережуваними даними. Гетероскедастичність відсутня, що є добре, а про наявність автокореляції висновків зробити не можна.