DOI 10.31891/2307-5732-2025-347-61
УДК 519.25:616.98:578.834

КАМІНСЬКИЙ РОМАН
Національний університет «Львівська політехніка»
https://orcid.org/0000-0002-8083-4288
e-mail: kaminsky.roman@gmail.com
ШАХОВСЬКА НАТАЛІЯ
Національний університет «Львівська політехніка»
https://orcid.org/0000-0002-6875-8534
e-mail: nataliya.b.shakhovska@lpnu.ua
ДМИТРІВ ГАЛИНА
СЗШ №96 МЖК-1
https://orcid.org/0009-0007-2707-6688
e-mail: dmhaluna83@gmail.com

ПОБУДОВА РЕГРЕСІЙНОЇ МОДЕЛІ ДЛЯ МАЛОЇ ВИБІРКИ БАГАТОВИМІРНИХ ДАНИХ НА ПРИКЛАДІ ТЕРМІНІВ ОДУЖАННЯ ПАЦІЄНТІВ ВІД COVID-19

В даній роботі, на прикладі малої вибірки, здійснено спробу виявити вплив медико-біологічних ознак на процес одужання 19 пацієнтів з захворюванням на COVID-19. Дані про кожного пацієнта представлені термінами одужання та медико-біологічними ознаками. З метою вивчення впливу ознак на термін одужання було встановлено, що приведені ознаки мають вельми слабий зв'язок з термінами одужання та значну колінеарність. Як результат можна відзначити: побудована модель має високу тісноту зв'язку з оригіналом, проте коефіцієнти моделі виявились незначущими. Першопричиною цієї ситуації можна вважати невідповідність кількості ознак обсягу вибірки.

Ключові слова: мала вибірка, багатовимірні дані, кореляційний аналіз, регресійна модель, COVID-19.

KAMINSKIY ROMAN
National University "Lviv Polytechnic"
SHAKHOVSKA NATALIA
National University "Lviv Polytechnic"
DMYTRIV GALYNA
School № 96

CONSTRUCTION OF A REGRESSION MODEL FOR A SMALL SAMPLE OF MULTIDIMENSIONAL DATA ON THE EXAMPLE OF PATIENTS' RECOVERY TIMES FROM COVID-19

Building a mathematical model, in particular, a regression model of the influence of individual indicators of the patient's condition on the recovery period is of utmost importance for making treatment decisions in various situations. This includes the recovery process in the event of a coronavirus infection. The material for analysis and modeling was presented by data on the recovery period of 19 patients with Covid-19 and a description of the physical and medical-physiological indicators of their health at the time of recovery, that is, a small sample of multidimensional data. This medical data was neatly formatted by medical personnel into a 19 × 30 table, in which the first column contains the duration of recovery (hospital stay in bed days), and the remaining columns are filled with data - individual examination results of each of these patients. The results of this study are presented by solutions to the following three problems. The first task concerned the determination of the correspondence of the parameters of a univariate sample of recovery terms to the terms of the general population – generally accepted recovery terms. In the second task, based on the correlation matrix, multicollinearity was eliminated by removing individual indicators. In addition, it was found that the closeness of the relationship between the indicators of medical and biological indicators and the recovery terms is insignificant. The third task consisted in constructing a regression model, the factors of which were the indicators that remained after the elimination of collinearity. The constructed model has a high closeness of connection with the original data, but the coefficients of the model turned out to be insignificant. The reasons for this situation can be considered both the selection of medical and physiological indicators themselves and the discrepancy in the number of signs of the sample size. In general, it can be stated that the presented indicators do not have a significant impact on the duration of recovery.

Keywords: small sample, multivariate data, correlation analysis, regression model, COVID-19.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Вдосконалення підходів, методик та інформаційних технологій при виробленні медико-діагностичних висновків на основі малої кількості, особливо багатовимірних даних має важливе науково-практичне значення про хід процесу лікування, оскільки виявляє нові знання в даній галузі.

Аналіз досліджень та публікацій

Дане дослідження є фактично є продовженням досліджень розпочатих в роботі [1], в якій для тих самих даних отримано такий результат – дані про терміни одужання мають нормальний розподіл та відсутня гендерна відмінність між пацієнтами. В роботі [2] Наведено основні показники і процедуру застосування методів множинного кореляційно-регресійного аналізу для дослідження багатофакторних процесів і явищ. Автор роботи [3] розкриває сутність сучасного погляду щодо науково-дослідної праці на основі застосування багатокритеріального кореляційно-регресійного аналізу. Основи економетрики в аспекті понять множинної регресії, багатофакторного кореляційного аналізу та мультиколінеарності розглянуто в [4]. В посібнику [5] приведений алгоритм розв'язання практичної задачі однофакторного кореляційного аналізу та оцінювання параметрів класичної регресійної моделі. Основні припущення

стосовно множинного кореляційно-регресійного аналізу приведені в роботі [6]. Ситуацію з наявністю сильного лінійного зв'язку між всіма або деякими пояснюючим змінними розглянуто в роботі [7], саме при розв'язанні регресійних задач. В статті [8] звернена увага на невірне тлумачення окремих положень при використанні інструменту регресійного аналізу, тобто неправомірне поширення двохфакторних регресійних рівнянь на аналіз багатофакторних економічних явищ та недостатність обсягу кількості значень показників в моделі для забезпечення достовірності математичної вибірки. В [9] приведені таблиці критичних значень різних критеріїв, що використовуються в статистичній обробці даних. Посібник [10] стосується суті та організації вибіркових спостережень. В [11] розглянуто методи кореляційного аналізу та побудови регресійної багатофакторної моделі з використанням різних пакетів прикладних програм. Матеріал посібника [12] призначено для здобувачів ступеня доктор філософії і включає методи встановлення статистичних зв'язків між властивостями досліджуваних об'єктів. В [13] викладено теоретичні основи та приклади застосування методів багатовимірного статистичного аналізу. Основну увага в [14] приділено методам статистичного аналізу результатів медико-біологічних вимірювань, а також наведені приклади з використанням функціональних можливостей MS Excel.

Формулювання цілей статті

Метою даного дослідження є встановлення істотності зв'язку фізичних та медико-фізіологічних показників пацієнтів з захворюванням на Covid-19 з тривалістю часу їх одужання, а також побудови регресійної моделі цього зв'язку.

Виклад основного матеріалу.

Надані для аналізу та моделювання дані були зібрані певним медичним персоналом і представлені в числовій формі – у вигляді таблиці «об'єкт-властивість», де в якості об'єкту виступає результуюча ознака – термін одужання пацієнтів, а в якості факторних ознак – їх індивідуальні медико-фізіологічні показники.

Характеристика даних. До аналізу представлений набір даних, що характеризують групу з 19 пацієнтів на момент одужання, від Covid-19. Група пацієнтів складається з 13 жінок та 6 мужчин. Стан кожного пацієнта поданий 29-ма фізичними та медико-фізіологічними показниками з оригінальними назвами, які розбиті на п'ять груп, а саме:

фізичну ***F*** – чотири показники: x_1^F– Age, x_2^F– Hejght, x_3^F– Weight, x_4^F–IMT;

серцево-судинну ***S*** – два показники: x_5^S– CAT, x_6^S– Pulse);

дихальну ***D*** – десять показників: x_7^D– Test_walk_6min, x_8^D– SaO2, x_9^D– Borg, x_{10}^D– Force_lung_capac, x_{11}^D– Force_exhalation_volume, x_{12}^D– ПОШвид, x_{13}^D– МОШ25, x_{14}^D– МОШ50, x_{15}^D– МОШ75, x_{16}^D– ЖЕЛвдоха;

імунну ***I*** – дев'ять показників: x_{17}^I– CD3, x_{18}^I– CD22, x_{19}^I– limf0, x_{20}^I– CD4, x_{21}^I– CD8, x_{22}^I– CD4_CD8, x_{23}^I– T_в, x_{24}^I– T_plus_B_0, x_{25}^I– CD16);

кровоносну ***K*** – чотири показники: x_{26}^K– ФНПalpha, x_{27}^K– IL8, x_{28}^K– IL4, x_{29}^K– IL10.

Тут верхній індекс визначає групу, а нижній – загальну нумерацію показників. Оскільки, метою аналізу є виявлення зв'язку цих показників з показником ***T*** – часом одужання (*Duration)* останній вважається результативною ознакою, а усі інші – факторними. В сенсі виявлення зв'язків подальший аналіз представлених даних можна розглядати як такі три завдання:

- статистичного аналізу значень ***T*** як окремої вибірки;
- кореляційного аналізу зв'язку ***T*** з усіма показниками;
- регресійного аналізу як побудови регресійної моделі виявлення зв'язків ***T*** з показниками.

Завдання 1. Аналіз значень T як окремої вибірки. Дана вибірка включає $n = 19$ варіант і в статистичному плані відноситься до категорії малих вибірок, оскільки її обсяг $n < 30$. За попередніми дослідженнями [1] її значення підпорядковані нормальному закону розподілу і, крім того, в них відсутня гендерна відмінність.

Статистичне оцінювання результуючої ознаки. Представлені для аналізу дані, приведені в табл. 1.

Таблиця 1

Тривалість госпіталізації хворих на Covid-19							
№ з/п	Дні	№ з/п	Дні	№ з/п	Дні	№ з/п	Дні
1	24	*6*	23	*11*	20	*16*	25
2	20	*7*	14	*12*	20	*17*	23
3	19	*8*	16	*13*	19	*18*	21
4	22	*9*	16	*14*	19	*19*	20
5	22	*10*	16	*15*	18		

Отже, дана вибірка належить до деякої генеральної сукупності з нормальним законом розподілу. Загальні статистичні оцінки, середнє арифметичне (математичне очікування), дисперсію та середнє

квадратичне відхилення представимо в класі інтервальних оцінок [5]. Виправлені значення цих показників, отримані на згрупованих даних, є такими:

$\bar{x} = 20.944,\ s^2 = 10.089, \sigma = \sqrt{s^2} = 3.176.$

Середнє арифметичне. Вибіркове середнє арифметичне $\bar{x}$ є оцінкою математичного сподівання генеральної сукупності, а його довірчий інтервал має такі межі:

$\bar{x}_н = \bar{x} - t_\gamma \cdot \frac{\sigma}{\sqrt{n}}$ та $\bar{x}_в = \bar{x} + t_\gamma \cdot \frac{\sigma}{\sqrt{n}}$,

де t_γ – визначено з таблиці інтегральної функції Лапласа $\Phi(t)$ для значення надійності $\gamma = 0.95$, тобто $t_\gamma = 1.96$.

Помилка середньо арифметичного для малих вибірок, розраховується за середньо квадратичним відхиленням, враховуючи ступень вільності за формулою:

$\delta_\nu = \frac{\nu}{\sqrt{2n}} \cdot \sqrt{1 + 2 \cdot \left(\frac{\nu}{100}\right)^2} = 0.0243 \times 1.0223 = 0.0248/$

Обчислений довірчий інтервал для математичного сподівання генеральної сукупності за даною вибіркою є таким: $19.516 \le \bar{x} \le 22.373$.

Оцінка $\bar{x}$ є незсуненою, обґрунтованою та ефективною.

Вибіркова дисперсія. Вибіркова дисперсія s^2, обчислена за згрупованими даними, враховуючи малий обсяг вибірки, рівна $s^2 = 10.090$, а середнє квадратичне відхилення, відповідно рівне $\sigma = \sqrt{s^2} = 3.174$.

Для визначення інтервальної оцінки вибіркової дисперсії, як обґрунтованої та незсуненої випадкової нормально розподіленої величини $s^2 = 10.089$ залучають значення розподілу χ^2-квадрат. Формули для знаходження границь довірчого інтервалу є такими:

$\frac{(n-1)\cdot s^2}{\chi_1^2} \le s^2 \le \frac{(n-1)\cdot s^2}{\chi_2^2}$.

Тут, χ_1^2 – значення цієї функції при$(n-1)$ ступені вільності і рівні значущості для нижнього значення $\alpha = 1 - \frac{0.05}{2} = 0.975$, а верхнє значення цього інтервалу χ_2^2 є значенням функції χ^2 -квадрат при $(n-1)$ ступені вільності і рівні значущості для верхнього значення $\alpha = \frac{0.05}{2} = 0.025$.

В результаті обрахунку з використанням таблиць функції χ_1^2-квадрат отримані такі границі довірчого інтервалу для вибіркової дисперсії s^2

$5.761 \le s^2 \le 22.065$.

Звідси довірчий інтервал для середнього квадратичного відхилення є таким:

$2.4 \le s \le 4.69$.

Середнє квадратичне відхилення вказує на варіабельність варіант і не залежить від обсягу спостережень. Середня помилка середнього квадратичного відхилення

$\delta_s = \frac{\sigma}{2n} = 0.0836$.

Коефіцієнт варіації. Коефіцієнт варіації дозволяє порівнювати мінливість ознак, що представлені в різних шкалах, а також оцінити однорідність вибірки. Цей коефіцієнт є відношенням середнього квадратичного відхилення до середньої арифметичної величини ознаки і переважно поданий у відсотковому вимірі:

$\nu = \frac{s}{\bar{x}} \times 100\% = \frac{3.176}{20.944} \times 100\% = 0.1516 \times 100\% \approx 15\%$.

Чим більший коефіцієнт варіації, тим менш однорідна сукупність і тим менш типова середня для даної сукупності. Орієнтовними оцінками варіативності за цим коефіцієнтом можна вважати: низький рівень – до 10 %; середній рівень – 10-20 %, високий рівень – вище 20 %.

Середня помилка коефіцієнта варіації визначається за такою в такий спосіб:

$\delta_\nu = \frac{\nu}{\sqrt{2n}} \cdot \sqrt{1 + 2 \cdot \left(\frac{\nu}{100}\right)^2} = 0.0243 \times 1.0223 = 0.0248$.

Встановлено, що сукупність кількісно однорідна, якщо коефіцієнт варіації не перевищує 33%.

Коефіцієнт детермінації. Цей показник має на меті кількісно подати тісноту зв'язку між даними та їх моделлю. Його переважно використовують в задачах апроксимації даних де він виступає оцінкою «якості придатності» апроксимуючої функції, а його максимальне значення рівне $K_d(x) = 1$. В даному випадку коефіцієнт детермінації обчислений за такою формулою:

$K_d(x) = 1 - \frac{S_\varepsilon^2}{S_x^2}$,

де S_ε^2 – дисперсія відхилень ε варіант x_i від арифметичного середнього $\bar{x}$, а S_x^2 – дисперсія вибірки даних. Отже обчислене значення цього коефіцієнта є таким:

$K_d(x) = 0.49$,

що свідчить про низьку якість лінійної апроксимації.

Коефіцієнт детермінації фактично є мірою лінійної залежності між двома показниками, а його практичне значення в тому, що визначає ту частину загальної варіації для даного показника, яку спричинила варіація іншого показника. Іншими словами, лише p% тісноти їх взаємозв'язку можна пояснити взаємним впливом між ними, а решту (1 – p)% варіативності спричинена іншими факторами.

Завдання 2. Встановлення тісноти зв'язку *T* з приведеними показниками.

Для встановлення тісноти зв'язку між терміном одужання T з кожним з приведених показників здійснено стандартизацію даних і на їх основі побудовано кореляційну матрицю розміром 30×30. Для аналізу результатів загальної кореляційної матриці використано, досить розповсюджену в наукових дослідженнях, шкалу тісноти зв'язку Чеддока. Результати аналізу цієї матриці такі.

Кореляційний аналіз. В результаті аналізу кореляційної матриці виявлено досить слабу тісноту зв'язку терміну одужання *T* з показниками, де в дужках зазначено їх належність до груп. В якості шкали оцінки величини тісноти зв'язку використано шкалу Чеддока, яку часто застосовують в кореляційному аналізі. За цією шкалою, значення коефіцієнтів кореляції між парами «час одужання – конкретний показник» віднесені до таких інтервалів:

$(0 \le r_{ti} \le 0.1)$ – x_5^S; x_8^D, x_{12}^D; x_{18}^I, x_{22}^I; x_{27}^K;

$(0.1 \le r_{ti} \le 0.3)$ – x_2^F, x_3^F; x_6^S; x_7^D, x_9^D, x_{16}^D; x_{19}^I, x_{20}^I, x_{23}^I, x_{24}^I, x_{28}^K, x_{29}^K;

$(0.3 \le r_{ti} \le 0.5)$ – x_1^F, x_4^F; x_{10}^D, x_{11}^D, x_{13}^D, x_{14}^D; x_{17}^I, x_{22}^I, x_{25}^I; x_{26}^K;

$(0.5 \le r_{ti} \le 0.7)$ – x_{15}^D.

Розподіл парних коефіцієнтів кореляції за інтервалами цієї шкали вказує на дуже слабку тісноту шести показників, слабку дванадцяти та помітну десяти показників і лише один показник має помірну тісноту зв'язку з терміном одужання. В даному випадку, шість показників з дуже слабким зв'язком можуть бути вилучені з розгляду, оскільки їх вплив практично відсутній. Таким чином, таблиця даних є зменшеною на шість показників. Проте, все ж можна визнати досить слабий вплив показників на індивідуальні терміни одужання.

Другим важливим моментом є те, що вже візуальний аналіз, отриманої кореляційної матриці показав значну, тобто високу і навіть дуже висок колінеарність між самими показниками.

Зокрема, значна колінеарність виявлена у групах показників дихальної та імунної систем. Такий результат свідчить про необхідність, для подальшого дослідження даних, зменшити розмірність ознакового простору шляхом вилучення окремих показників – тих, для яких тіснота зв'язку є слабкою або дуже слабкою.

Виключення мультиколінеарності в групах проведено в такий спосіб. Серед показників з високою або дуже високою тіснотою зв'язку між собою залишають той показник для якого коефіцієнт кореляції з результуючою ознакою – терміном одужання є найвищим, а іншого, з цієї пари, вилучають. В результаті, в кожній групі залишилися ті показники, для яких значення коефіцієнта кореляції між терміном одужання та іншими показниками не перевищує величину 0.456.

Для цього застосовано індивідуальний підхід до кожної групи показників та для кожного з показників що залишились встановлено його значущість за ***t***-критерієм Стьюдента на рівні α = 0,05.

Таким чином, здійснено пониження показникового простору, шляхом вилучення мультиколінеарності, в результаті чого залишились такі показники:

Отже, тепер таблиця «термін одужання - пацієнт» для побудови моделі тісноти зв'язку показників стану пацієнтів включає 13 елементів: результат (термін одужання) та 12 факторів (показників фізичного та медико-біологічного стану пацієнтів).

Аналіз цієї нової кореляційної матриці при застосуванні ***t***-критерію Стьюдента на рівні α = 0,05, показав, за винятком показника дихальної системи МОШ75, що коефіцієнти кореляції решти факторів є *не значущими*.

Завдання 3. Регресійна модель тісноти зв'язку.

Для побудови рівняння регресії використано табличний процесор Ms Excel. В результаті, кількість факторів скоротилася до 12-ти показників, приведених в табдиці 2. а в групах залишились такі:

Таблиця 2

Віль-ний Член	**Фізичні та медико-біологічні показники пацієнтів**											
	Age	Hejght	IMT	Pulse	Test_ walk_ 6min	SaO2	МОШ 75	ЖЕЛ вдиху	CD3	CD16	IL4	IL10
	x_1^F	x_2^F	x_4^F	x_6^S	x_1^D	x_2^D	x_9^D	x_{10}^D	x_1^I	x_9^I	x_3^K	x_4^K
-176,7	-0,04	0,3	0,34	0,205	-0,03	1,18	-0,011	-0,12	0,54	0,03	0,02	0,001

Модель (рівняння регресії) для даних, отриманих після усунення мультиколінеарності має такий вигляд:

$$T = -176{,}7 - 0{,}04\,x_1^F + 0{,}3\,x_2^F + 0{,}34\,x_4^F + 0{,}205\,x_6^S - 0{,}03\,x_1^D + 1{,}18\,x_2^D - 0{,}011\,x_9^D - 0{,}12\,x_{10}^D + 0{,}54\,x_1^I + 0{,}03\,x_9^I + 0{,}02\,x_3^K + 0{,}001\,x_4^K.$$

Регресійна статистика. За результатами регресійного аналізу для даної регресійної моделі отримано такі характеристики.

Множинний коефіцієнт кореляції R, який характеризує величину впливу багатьох різних факторів на результуючу ознаку. Його значення $R = 0{,}9127$ є досить близьким до одиниці, що вказує на сильну кореляцію факторів з результатом.

Показник R-квадрат, який часто називають коефіцієнтом детермінації, має досить високе значення R-квадрат = 0,8331, тобто 83,31%. Таке значення свідчить про те, що дана модель пояснює значну частину варіації серед даних, тобто 83.31% варіації залежної змінної пояснюються моделлю.

Нормований R-*квадрат* це скоригований коефіцієнт детермінації, який бере до уваги кількість факторів даної моделі в тому сенсі, що вказує на присутність зайвих, незначущих факторів. Значення нормованого R-квадрат = 0,4993 або 49,93% є значно нижчим від показника R-квадрат, а це означає, що частина факторів у моделі є незначущими, тобто близько 49.93% варіації залежної змінної пояснюються саме моделлю.

Стандартна помилка моделі становить: 2,0999 і, фактично, відображає відхилення значень моделі від фактичних значень.

Спостереження, тобто обсяг даних $n = 19$ відповідає обсягу малих вибірок. В регресійному аналізі рекомендують, для отримання надійних результатів, щоб обсяг вибірки був у 3-5 разів більшим за кількість факторів, що в даному випадку має місце.

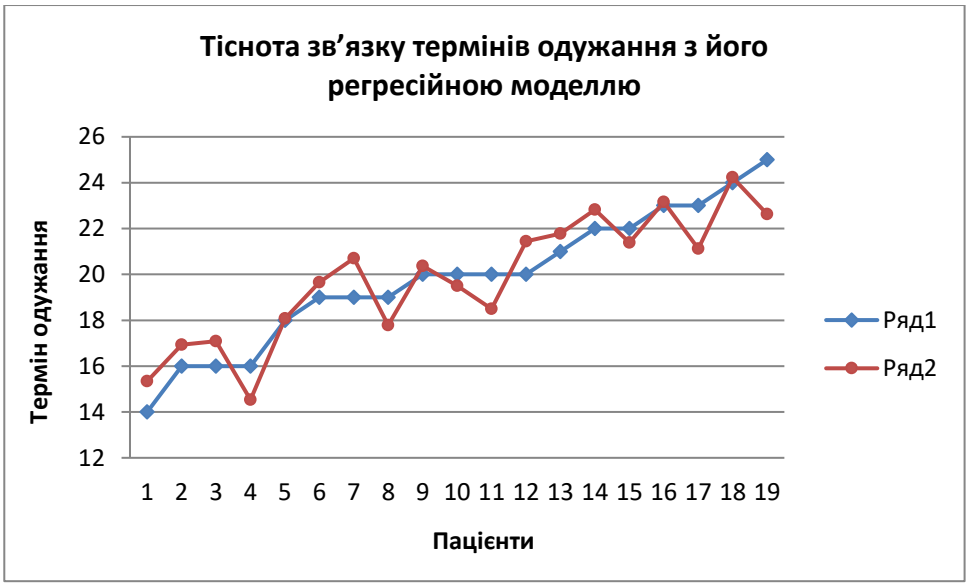


Рис. 1. Тіснота зв'язку термінів одужання з його моделлю: ряд 1 – оригінал, ряд 2 – модель.

Дисперсійний аналіз. Для цього, в регресійному аналізі, використовують ***F***-критерій Фішера, який перевіряє, чи дана модель є статистично значущою. Іншими словами, мова йде про те, чи хоча б один із факторів моделі впливає на результат – термін одужання.. Для моделі множинної регресії, нульова гіпотеза H_0 для F-критерію формулюється як «усі коефіцієнти моделі рівні нулю»:

H_0: $\beta_1 = \beta_2 = \ldots = \beta_m = 0$.

Альтернативна гіпотеза

H_1: хоча б один з $\beta_i \neq 0$

означає, що принаймні один з коефіцієнтів є відмінним від нуля, і, отже, має вплив на залежну змінну.

Щодо значущості ***F***-критерію проведено порівняння існуючої варіації з варіацією, яку пояснює модель. Для даної ситуації мають місце такі значення:

- обраховане значення $F = 2{,}4958$;
- обрахована значущість F (тобто p-value) = 0,1353;
- обсяг даних $n = 19$;
- кількість факторів в моделі $m = 12$;
- табличне значення F для рівня значущості $\alpha = 0.05$ і числа ступенів свободи $df_1 = 6$ та $df_2 = 12$, становить $F = 3.00$.

Отже, порівняння обрахованого значення $F = 2{,}4958$ з табличним $F = 3.00$, а саме $2{,}4958 < 3.00$, тобто вказує на те, що немає підстав відхиляти нульову гіпотезу – усі коефіцієнти моделі є *незначущими*.

Значущість F (тобто p-value) = 0,1353 перевищує загальноприйнятий 5% - відсотковий рівень значущості, тобто F(p-value) = 0,1353 > 0,05, означає, що ймовірність отримати таке значення F при справедливій нульовій гіпотезі (коли всі коефіцієнти регресії рівні нулю) становить 13,53%.

За результатами третього завдання можна стверджувати, що побудована модель загалом не є статистично значущою, тому, що обраховане значення F-критерію не перевищує відповідне його критичне (табличне) значення. Крім того, значущість F (p-value) перевищує стандартний рівень

значущості α = 0,05, чим також підтверджено незначущість впливу як окремих показників так і моделі в цілому.

Висновки з даного дослідження
і перспективи подальших розвідок у даному напрямі

Запропоновані для дослідження впливу медико-фізіологічних та фізичних показників, розбитих на п'ять груп: фізичної, серцево-судинної, дихальної, на їх тісноту зв'язку з термінами одужання, у відповідності з поставленими завданнями, дало такі результати.

1. Аналіз термінів одужання 19 пацієнтів обох статей, розглянутих як окрема вибірка, підтвердив відповідність різним генеральним сукупностям величиною інтервальних оцінок для середнього арифметичного та дисперсії, за різними літературними даними, в середньому, термін одужання становить 20 ліжкоднів.

2. Розгляд та вивчення кореляційної матриці, розміром 30 × 19, впливу на тісноту зв'язку між терміном одужання T і 29-ма показниками стану пацієнтів показав, що за шкалою Чеддока не перевищують помірного рівня тісноти зв'язку (це < 0.5), крім одного показника МОШ 75, коефіцієнт кореляції якого $r_9^{TD} = -0.518$, що відповідає інтервалу 0.5 ÷ 0.7 і характеризується як помітний. Проте, було виявлено значну мультиколінеарність між ознаками, зокрема дуже сильну в групах дихальної та імунної та інших систем. В результаті вилучення корельованих ознак, для побудови моделі використано 12.

3. Побудована моделі за значеннями показників які залишились після усунення мультиколінеарності має такі характеристики:

- за регресійною статистикою множинний коефіцієнт кореляції та коефіцієнт детермінації вказують на дуже високу тісноту зв'язку використаних показників з термінами одужання, тобто вказує на те, що модель добре пояснює залежну змінну. Проте, з другої сторони, корегований коефіцієнт детермінації є досить низьким, а отже, фактори даної моделі є незначущими. Іншими словами, значне зниження цього коефіцієнта свідчить про надмірність факторів або декотрі з них, можуть мати малий вплив або заважають моделі.

- за дисперсійним аналізом можна стверджувати, що побудована модель загалом не є статистично значущою, оскільки обраховане значення F-критерію не перевищує відповідне його критичне значення, а також, значущість – величина p-value перевищує стандартний рівень значущості α = 0,05, що підтверджує незначущість впливу і окремих показників, і моделі в цілому.

В загальному, отриманому результату має місце таке підтвердження: оскільки 19 пацієнтів – це мала вибірка, а тому, за рекомендаціями побудови регресійних моделей кількість багатовимірних даних має перевищувати кількість факторів принаймні у 3 – 5 разів, щоб уникнути низької точності та нестабільності оцінок. Іншими словами, в перспективі подальших досліджень це співвідношення має бути дотримане. Тим не менш, відповідно до мети дослідження можна вказати на те, що приведені показники практично не впливають на тривалість одужання.

Література

1. Камінський, Р., Шаховська, Н., & Дмитрів, Г. (2024). Аналіз малих вибірок багатовимірних даних на прикладі термінів одужання пацієнтів від covid-19. Herald of Khmelnytskyi National University. Technical Sciences, 341(5), 304-311. https://doi.org/10.31891/2307-5732-2024-341-5-44

2. Степанишин В.М., Тисовський Л.О. Побудова моделі кореляційного аналізу для дослідження багатофакторних процесів і явищ // Вісник Нац. ун-ту «Львівська політехніка». – 2011. – № 707. – С. 42–45.

3. Павлюк К. В. Методичні підходи до розроблення нормативів і оцінки науково-дослідної праці на основі багатофакторного кореляційно-регресійного аналізу // Наукові праці НДФІ. – 2020. – № 3 (92). – С. 5–19.

4. Економетрика в електронних таблицях : навч. посіб. / Васильєва Н. К., Мироненко О. А., Самарець Н. М., Чорна Н. О. ; за заг. ред. Н. К. Васильєвої. – Дніпро : Біла К. О., 2017. – 149 ***с***.

5. Економетрія. Частина 2 : навчальний посібник / [Азарова А. О., Сачанюк-Кавецька Н. В., Роїк О. М., Міронова Ю. В.] – Вінниця : ВНТУ, 2011. – 118 с.

6. Економетрія. Частина 2 : навчальний посібник / [Азарова А. О., Сачанюк-Кавецька Н. В., Роїк О. М., Міронова Ю. В.] – Вінниця : ВНТУ, 2011. – 118 с.

7. Літнарович Р.М.Побудова і дослідження математичної моделі за джерелами експериментальних даних методами регресійного аналізу. Навчальний посібник, МЕГУ, Рівне, 2011.-140 с.

8. Захарченко В. І. Коректність використання інструменту регресійного аналізу в економічних дослідженнях. / В. І. Захарченко // Економічний журнал Одеського політехнічного університету. – 2019. – № 4 (10). – С. 68-75.

9. Таблиці функцій критичних точок розподілів. Розділи: Теорія ймовірностей. Математична статистика. Математичні методи в психології / Укладач: М.М. Горонескуль. – Х.: УЦЗУ, 2009. – 90 с.

10. Василик, О. І. Лекції з теорії і методів вибіркових обстежень : навчальний посібник / О. І. Василик, Т. О. Яковенко. – К. : Видавничо-поліграфічний центр "Київський університет", 2010. – 208 с.

11. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.

12. Куц, Ю. Ю. Статистичні методи визначення залежностей між випадковими величинами / Ю. В. Куц, Ю. Ю. Лисенко Навчальний посібник для здобувачів ступеня доктор філософії. Електронне мережне навчальне видання : Київ, КПІ ім. Ігоря Сікорського. – 2022. – 115 с.

13. А. Т. Яровий, Є. М. Страхов. Багатовимірний статистичний аналіз : начально-методичний посібник для студентів математичних та економічних фахів. – Одеса: Астропринт, 2015. – 132 с.

14. Мулик О.В. Біостатистика / О.В. Мулик, Т.Г. Пригалінська, Л.О. Свистун-Золотаренко. Навчальний посібник. Електронне мережне навчальне видання : Київ. - КПІ ім. Ігоря Сікорського. – 2023. – 364 с.

References

1. Kaminskyi, R., Shakhovska, N., & Dmytriv, H. (2024). Analiz malykh vybirok bahatovymirnykh danykh na prykladi terminiv oduzhannia patsiientiv vid covid-19. Herald of Khmelnytskyi National University. Technical Sciences, 341(5), 304-311. https://doi.org/10.31891/2307-5732-2024-341-5-44

2. Stepanyshyn V.M., Tysovskyi L.O. Pobudova modeli koreliatsiinoho analizu dlia doslidzhennia bahatofaktornykh protsesiv i yavyshch // Visnyk Nats. un-tu «Lvivska politekhnika». – 2011. – № 707. – S. 42–45.

3. Pavliuk K. V. Metodychni pidkhody do rozroblennia normatyviv i otsinky naukovo-doslidnoi pratsi na osnovi bahatofaktornoho koreliatsiino-rehresiinoho analizu // Naukovi pratsi NDFI. – 2020. – № 3 (92). – S. 5–19.

4. Ekonometryka v elektronnykh tablytsiakh : navch. posib. / Vasylieva N. K., Myronenko O. A., Samarets N. M., Chorna N. O. ; za zah. red. N. K. Vasylievoi. – Dnipro : Bila K. O., 2017. – 149 s.

5. Ekonometriia. Chastyna 2 : navchalnyi posibnyk / [Azarova A. O., Sachaniuk-Kavetska N. V., Roik O. M., Mironova Yu. V.] – Vinnytsia : VNTU, 2011. – 118 s.

6. Ekonometriia. Chastyna 2 : navchalnyi posibnyk / [Azarova A. O., Sachaniuk-Kavetska N. V., Roik O. M., Mironova Yu. V.] – Vinnytsia : VNTU, 2011. – 118 s.

7. Litnarovych R.M.Pobudova i doslidzhennia matematychnoi modeli za dzherelamy eksperymentalnykh danykh metodamy rehresiinoho analizu. Navchalnyi posibnyk, MEHU, Rivne, 2011.-140 s.

8. Zakharchenko V. I. Korektnist vykorystannia instrumentu rehresiinoho analizu v ekonomichnykh doslidzhenniakh. / V. I. Zakharchenko // Ekonomichnyi zhurnal Odeskoho politekhnichnoho universytetu. – 2019. – № 4 (10). – S. 68-75.

9. Tablytsi funktsii krytychnykh tochok rozpodiliv. Rozdily: Teoriia ymovirnostei. Matematychna statystyka. Matematychni metody v psykholohii / Ukladach: M.M. Horoneskul. – Kh.: UTsZU, 2009. – 90 s.

10. Vasylyk, O. I. Lektsii z teorii i metodiv vybirkovykh obstezhen : navchalnyi posibnyk / O. I. Vasylyk, T. O. Yakovenko. – K. : Vydavnycho-polihrafichnyi tsentr "Kyivskyi universytet", 2010. – 208 s.

11. Bakhrushyn V.Ie. Metody analizu danykh : navchalnyi posibnyk dlia studentiv / V.Ie. Bakhrushyn. – Zaporizhzhia : KPU, 2011. – 268 s.

12. Kuts, Yu. Yu. Statystychni metody vyznachennia zalezhnostei mizh vypadkovymy velychynamy / Yu. V. Kuts, Yu. Yu. Lysenko Navchalnyi posibnyk dlia zdobuvachiv stupenia doktor filosofii. Elektronne merezhne navchalne vydannia : Kyiv, KPI im. Ihoria Sikorskoho. – 2022. – 115 s.

13. A. T. Yarovyi, Ye. M. Strakhov. Bahatovymirnyi statystychnyi analiz : nachalno-metodychnyi posibnyk dlia studentiv matematychnykh ta ekonomichnykh fakhiv. – Odesa: Astroprynt, 2015. – 132 s.

14. Mulyk O.V. Biostatystyka / O.V. Mulyk, T.H. Pryhalinska, L.O. Svystun-Zolotarenko. Navchalnyi posibnyk. Elektronne merezhne navchalne vydannia : Kyiv. - KPI im. Ihoria Sikorskoho. – 2023. – 364 s.