

КАМІНСЬКИЙ РОМАННаціональний університет «Львівська політехніка»
<https://orcid.org/0000-0002-6875-8534>
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**Національний університет «Львівська політехніка»
<https://orcid.org/0009-0007-2707-6688>
e-mail: dmhaluna83@gmail.com

АНАЛІЗ МАЛИХ ВИБІРОК БАГАТОВИМІРНИХ ДАНИХ НА ПРИКЛАДІ ТЕРМІНІВ ОДУЖАННЯ ПАЦІЄНТІВ ВІД COVID-19

В даній роботі, на прикладі малої вибірки, здійснено спробу вивчити вплив медико-біологічних ознак на процес одужання 19 пацієнтів з захворюванням на COVID-19. Дані про кожного пацієнта представлені, крім терміну одужання, ще 29-ма медико-біологічними ознаками, розбитими на групи. З метою вивчення впливу ознак на термін одужання було встановлено: дані про одужання мають нормальний розподіл, відсутня гендерна відмінність, приведені ознаки, практично для всіх пацієнтів мають вельми слабкий зв'язок з термінами одужання, а також з віком і вагою пацієнтів. Як результат можна відзначити: по-перше, використання малих вибірок допустиме для прийняття початкових рішень, а по-друге, для визначення впливу на одужання конкретної ознаки чи групи ознак необхідно мати динаміку процесу одужання.

Ключові слова: мала вибірка, багатовимірні дані, нормальний розподіл, гендерні відмінності, COVID-19.

KAMINSKIY ROMAN, SHAKHOVSKA NATALIYA, DMYTRIV HALUNA
L'viv Polytechnic National University

ANALYSIS OF SMALL SAMPLES OF MULTIDIMENSIONAL DATA ON THE EXAMPLE OF TIMES OF RECOVERY OF PATIENTS FROM COVID-19

The analysis of multidimensional data presented in small volumes - small samples is quite often used in publishing, pedagogy, sociology mainly for the purpose of understanding the existing situation and choosing and making appropriate decisions. The main goal of such an analysis is to identify the impact of various factors on the object of attention in order to adjust further steps in one or another type of activity. The authors of this article worked with the material, namely with the medical and biological indicators of 19 patients with the disease of Covid-19, who underwent a course of treatment and at the time of recovery, their condition is presented with relevant data. These data include: duration of recovery of each patient in bed days and 29 indicators of physical and physiological state, divided into five groups. These are the following groups: physical characteristics (age, height, weight) and cardiovascular system, respiratory system, immune and circulatory systems. The data analysis started with the recovery term regarding the normality of the distribution. The following methods were used for this: quartile-quartile graph, as well as Shapiro-Wilk, Kolmogorov-Smirnov and Anderson-Darling tests. Based on the results of the application of these methods, it was concluded that the data are subject to the normal distribution law. The presence of both women and men in the data sample, and the latter are twice as few, required clarification of the gender difference between them. Here, the authors used a visual comparison using a boxplot and Student's T-test. By visual comparison, the differences between the minimums and the first quartiles are greater than the differences between the medians, third quartiles, and maximums, but according to the results of the T-test, the average values of these samples are equal and differ only by chance. The relationship of signs with the terms of recovery was determined in the following way. For each group of features, their individual multivariate averages were determined, and the recovery time interval was divided into four subintervals. According to these subintervals, patients were divided along with their multivariate averages for each group of indicators. Each subinterval is matched with the average value of the individual multivariate averages. The data are presented in a table - subintervals as levels of the main factor and groups of averages from individual multivariate averages as variables. One-factor variance analysis was applied to the data in the table, the equality of the average groups of indicators, as well as the weakness of the influence of diagnostic indicators on recovery time. The work has a scientific and practical nature and can be useful in similar situations.

Keywords: small sample, multivariate data, normal distribution, gender differences, Covid-19.

Постановка проблеми у загальному вигляді

та її зв'язок із важливими науковими чи практичними завданнями

В багатьох, зокрема медико-біологічних, психолого-педагогічних, соціологічних, наукових та професійних дослідженнях мають справу з обмеженою кількістю отриманих даних, тобто малими вибірками. В цьому сенсі малі вибірки, тобто вибірки обсягом меншим за 30 варіант зменшують достовірність, а відтак і точність висновків. В статистиці доведено, що результати, отримані за малими вибірками, також поширюються на генеральну сукупність, що й робить прийнятним результати їх аналізу. Варто зауважити, що арсенал методів аналізу таких вибірок є набагато меншим, порівняно з великими вибірками і досить часто вимагає дотримання певних обмежень. Малі вибірки можуть призвести до великої дисперсії результатів, тому важливо ретельно аналізувати дані та враховувати можливі впливи інших факторів. Тому досить часто інтерес представляє не лише результат аналізу, а також методика чи інформаційна технологія опрацювання таких даних.

Матеріал представлений тут, стосується малої вибірки – групи пацієнтів які одужали від Covid-19 протягом різних термінів часу причому на момент одужання їхній стан описується набором ознак – різних медико-біологічних показників. Припускається, що всі вони пройшли лікування за одним і тим самим протоколом і група була виписана на протязі кількох тижнів.

Саме в такому плані була вибрана група використаних методів, а в якості матеріалу – реальні дані стосовно одужання групи пацієнтів на захворювання Covid-19.

Аналіз останніх досліджень

Загальні поняття вибіркового дослідження, а також перевірки гіпотез приведені в [1]. Докладно розглянуті приклади аналізу даних, зокрема, стосовно параметрів описової статистики, наведені критерії перевірки гіпотез в [2]. В [3] розглянуто застосування аналізу вибірок в освітніх та гендерних дослідженнях. Задача встановлення статистичних зв'язків між властивостями об'єктів поданих вибірками їх аналізу та діагностики розглянуто в [4]. Використання належних статистичних методів для дослідницьких досліджень із малими вибірками у вигляді аналітичного огляду приведено [5], де обговорено відповідність різних статистичних методів для використання їх при аналізі малих вибірок разом з припущеннями та обмеженнями цих методів і надання рекомендацій. Автор статті [6] вважає, що розмір вибірки є не лише ключовим фактором, який впливає на статистичну потужність і точність дослідження, але також визначає здатність дослідження робити значущі висновки та поширювати результати на більш широкий контекст, а також розглядає, які чинники можна вважати «малими за обсягом» у контексті біомедичних досліджень. Автор статті [7] розглядає особливості клінічного перебігу коронавірусної хвороби COVID-19, зосереджуючись на таких моментах: вплив віку, статі, гендерних відмінностей. Колектив авторів статті [8] проаналізував та оцінив взаємозв'язок між перебігом коронавірусної інфекції та статтю пацієнта і в результаті вказує на те, що гостре респіраторне вірусне захворювання SARS CoV2-19 у жінок та чоловіків перебігає в основному однаково, хоча можуть існувати деякі відмінності. Автор статті [9] приводить техніку визначення того, чи різні набори даних мають спільний розподіл за допомогою квантильно-квантильних графіків, яка може бути використана для визначення, чи набір даних нормально розподілений, Для візуального узагальнення та порівняння вибірок даних автори статті [10] розглядають метод коробкової діаграми або діаграми розмаху, який може покращити наші міркування щодо кількісної інформації. В публікації [11] представлено огляд традиційних методів вираження конкретних статистичних характеристик даних з допомогою візуального представлення, зокрема коробкової діаграми. Автор роботи [12] розглядає два основних види нормалізації, а саме: перший тип нормалізації походить від лінійна алгебра і розглядає дані як вектор у багатовимірному просторі, а другий – від статистики та усуває одиницю вимірювання шляхом перетворення даних у нові бали із середнім значенням 0 і стандартним відхиленням 1. Багатовимірний аналіз соціально-економічних систем розглянуто в [13] разом з відповідними математичними методами та їх описами. Застосування математичних методів для розв'язання широкого класу прикладних задач багатовимірному аналізу даних приведено в [14], для формування вибірових досліджень та оцінювання їх якості. В статті [15] автор розглядає сутності багатовимірному аналізу соціологічної інформації в сенсі ідеї багатовимірності, а також аналізує можливості методів і технологій багатовимірному аналізу соціологічних даних. Автори [16] ставили перед собою за мету викласти питання планування і обробки результатів спостережень в доступній та зрозумілій формі, приділивши увагу використанню табличного процесора Ms Excel. В [17] подано достатньо інформації про опрацювання та особливості редагування даних, а також розглянуто різні способи використання засобів обчислень у таблицях із використанням стандартних функцій.

Формулювання цілей статті

Метою роботи є: дослідження малих вибірок багатовимірних даних для прийняття попереднього рішення щодо одужання пацієнтів, які перенесли захворювання COVID-19.

Виклад основного матеріалу

Представлені для аналізу дані являють собою малу вибірку обсягом 19 варіант, які характеризують терміни одужання пацієнтів з захворюванням COVID-19. Ці дані подають інформацію про пацієнтів а саме: про 13 жінок і 6 чоловіків. Термін одужання кожного пацієнта характеризують 29 фізичними та медико-фізіологічними показниками їхнього стану після одужання. Таким чином, приведені дані характеризують пацієнтів, з медичної точки зору, а сам термін одужання означений як кількість «ліжкоднів», а його значення приведені в табл. 1.

Таблиця 1

Тривалість госпіталізації хворих на Ковід-19

№ пацієнта	Дні лікування	№ пацієнта	Дні лікування	№ пацієнта	Дні лікування	№ пацієнта	Дні лікування
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		

Середня тривалість одужання від цього захворювання у складних випадках госпіталізації, почерпнута з різних джерел, становила в середньому 20 ліжкоднів. З точки зору математичної статистики цю величину можна розглядати як середнє генеральної сукупності. Отримані дані дають такі значення цього параметра: середнє значення кількості днів перебування в стаціонарі становить 19.8 для усієї групи. Загалом, отримані дані можна подати в такий спосіб:

- кожен пацієнт характеризується індивідуальним часом одужання та загальними фізичними характеристиками – вік, вага, ріст, крім того, подано також узагальнений показник, такий як індекс маси тіла (ІМТ);

- по кожному пацієнту приведені значення медико-фізіологічних показників за такими групами: серцево-судинна (два показники), органи дихання (десять), імунна (дев'ять) та кровеносна (чотири) системи, загалом 26 показників.

Перевірка даних на відповідність нормальному закону розподілу. В статистичному дослідженні перевірка нормальності отриманого набору даних, щодо відповідності їх нормальному закону розподілу є превалюючою процедурою. В даному ж дослідженні інтерес представляють лише ті методи, які можуть бути застосовані саме до малих вибірок. Саме в цьому плані використано такі методи.

Діаграма квантиль-квантиль (QQ-Plot) [9]. Найпоширенішим використанням діаграми QQ є визначення того, чи відповідають вибіркові дані нормальному розподілу ймовірностей. Ця діаграма забезпечує досить високу візуальну оцінку, яка достатньо точно визначає відхилення між нормальним розподілом та ідентифікуючими точками – відповідними значеннями варіант даної вибірки, що відповідають за них. Порівнюючи положення цих вибіркових варіант з нормальним розподілом ймовірностей, і є змістом перевірки їх приналежності нормальному розподілу. QQ-діаграма порівнює квантилі розподілу ваших даних із квантилями нормального розподілу. Якщо дані нормальні, точки будуть розташовані вздовж прямої лінії.

Результат використання QQ-діаграми зображено на рис. 1. Про приналежність значень варіант до нормального розподілу свідчить те, що зображуючі значення варіант точки досить тісно розміщені до лінії.

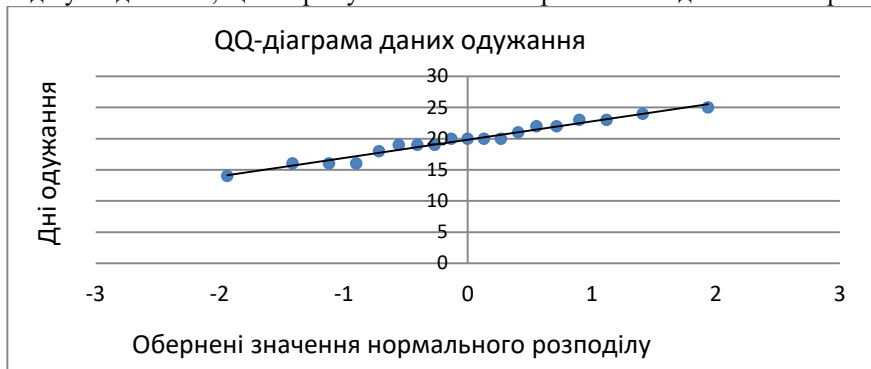


Рис. 1. Перевірка нормальності вибірки

Це вказує на приналежність даних нормальному розподілу, проте, варто зазначити: даний метод є суб'єктивним і можна лише погодитись з тим, що не варто відхиляти гіпотезу відповідності вибірки нормальному закону розподілу.

Тест Шапіро-Уїлка є чутливим тестом для перевірки нормальності розподілу, особливо для малих вибірок. Для цього тесту, нульова гіпотеза полягає в тому, що варіанти вибірки розподілені нормально тільки тоді, коли p -значення (ймовірність відхилити правильну гіпотезу) більші за вибраний альфа-рівень (в даному випадку $\alpha = 0.05$), тобто нульову гіпотезу (що дані надійшли з нормально розподіленої сукупності) не можна відхилити. За результатами цього тесту p -значення рівне 0.4826. Оскільки p -значення більше рівня значущості $\alpha = 0.05$, не має підстав відхиляти нульову гіпотезу – відповідність нормальному розподілу.

Тест Колмогорова-Смірнова також належить до тестів для перевірки нормальності розподілу емпіричних даних і, зокрема, щодо малих вибірок. Критерій Колмогорова-Смірнова для однієї вибірки дозволяє визначити, чи відрізняється розподіл її варіант від нормального. Суть методу полягає у порівнянні розподілу накопичених частот вибірки із теоретичним (очікуваним) нормальним розподілом. Порівняння ґрунтується на вибірковій групі, в якій емпіричний розподіл має найбільше абсолютне відхилення від теоретичного. Далі ця абсолютна різниця порівнюється з критичними (табличним) значенням, з метою визначення, чи може таке відхилення бути випадковим при даному законі розподілу емпіричних даних. Застосування критерію Колмогорова-Смірнова до вибірки отриманих даних дає таке p -значення: $p = 0,65$ для рівня значущості $\alpha = 0,05$. Оскільки $p > \alpha$ чим і підтверджується нормальність емпіричного розподілу отриманих даних.

Критерій Андерсона-Дарлінга також широко використовується для перевірки належності експериментальних даних нормальному закону розподілу. Для цього критерію p -значення становить $p = 0,45$. Таким чином, якщо на рівні значущості $\alpha = 0,05$, p -значення $p > \alpha$, отже, набір даних зі p -значенням більше 0,05 не відхиляє нульову гіпотезу про те, що дані походять з нормально розподіленої популяції.

В підсумку зазначимо, графічний метод QQ-діаграма та статистичні тести Шапіро-Уїлка, Колмогорова-Смірнова, Андерсона-Дарлінга свідчать про те, що немає підстав відхиляти гіпотезу про нормальний розподіл даних у отриманій вибірці. Отже, дані цієї вибірки можна вважати такими, що мають нормальний розподіл.

Визначення гендерних відмінностей. Приведені дані, як було зазначено вище включають пацієнтів різної статі. Хоча, їх присутність у даній вибірці схиляє до думки щодо однорідності даних, тим не менш виникає думка чи існує гендерна відмінність цих даних за такого невеликого їх обсягу. Враховуючи величину обсягу факторних ознак побудована діаграма поділу даних на дві окремі групи – чоловіків і жінок. Однак, при порівнянні тривалості лікування між жінками та чоловіками в одному закладі, важливо враховувати інші фактори,

такі як вік, супутні захворювання, стан імунної системи та індивідуальні особливості. Гендерний поділ пацієнтів приведений в табл. 2, а також збережений порядок, який відповідає їхньому порядку в табл. 1.

Таблиця 2

Розбиття вибірки даних за гендерними ознаками

Жінки				Чоловіки	
№ пацієнта	Дні лікування	№ пацієнта	Дні лікування	№ пацієнта	Дні лікування
1	24	8	16	9	16
2	20	10	16	13	19
3	19	11	20	15	18
4	22	12	20	16	25
5	22	14	19	17	23
6	23	18	21	19	20
7	14				

В результаті, за такого поділу виникає слушне питання: чи різняться між собою ці дві вибірки за статистичними показниками, зокрема середніми та дисперсіями.

T-критерій Стьюдента. В статистичному дослідженні встановлення виду закону розподілу отриманого набору даних переважно означає їх відповідність нормальному закону розподілу і є превалюючою процедурою. Для перевірки нормальності використано *t*-критерію Стьюдента, який дав результат стосовно рівності середніх цих двох вибірок, приведені в табл. 3.

В результаті його застосування можна зробити висновок про те, що не виявлено статистично значущих доказів підтвердження того, що обидві ці вибірки мають різні середні значення, як при порівнянні *t*-статистик з критичними табличними значеннями так і при порівнянні *p*-значень з рівнем значущості $\alpha = 0,05$. Іншими словами, відмінність значень цих вибірок зумовлена випадковістю, а не їх фізичною природою. Для візуалізації та виявлення відношень між параметрами варіаційної статистики обох вибірок використано зображення і характеристики діаграм розмаху.

Діаграми розмаху. Ці діаграми (часто можна зустріти інші терміни, як от коробкова діаграма, коробка з вусами) використовуються для показу характерних значень числових даних, особливо коли треба порівняти їх між кількома подібними групами. Вони створені для отримання початкової інформації про дані на перший погляд і дають загальні відомості про симетрію даних, варіаційні характеристики, викид. Легко побачити, де знаходиться основна маса даних та порівняти різні групи. Цей метод узагальнює дані, особливо при порівнюванні груп за допомогою положень їхніх компонент.

Таблиця 3

Двохвибірковий t-тест з різними дисперсіями

	Жінки	Чоловіки
Середнє	19,30769	21
Дисперсія	9,397436	6,8
Обсяг	13	6
Гіпотетична різниця середніх	0	
Df	11	
t-статистика	-1,242	
P(T<=t) одностороння	0,120	
t критичне одностороннє	1,796	
P(T<=t) двостороння	0,240	
t критичне двостороннє	2,201	

Візуальний аналіз гендерної діаграми на рис. 1 вказує на відсутність суттєвих відмінностей, а це означає, що опрацювання приведених даних можна здійснити без відповідного гендерного поділу. Проте, для переконливого твердження що це так необхідно попередньо здійснити кількісний аналіз, в даному випадку, за допомогою методу коробкових діаграм.

Таблиця 4

Квартилі	min	Q1	Q2	Q3	max
Чоловіки	18	19,25	20,5	22,5	25
Жінки	14	16	20	22	24

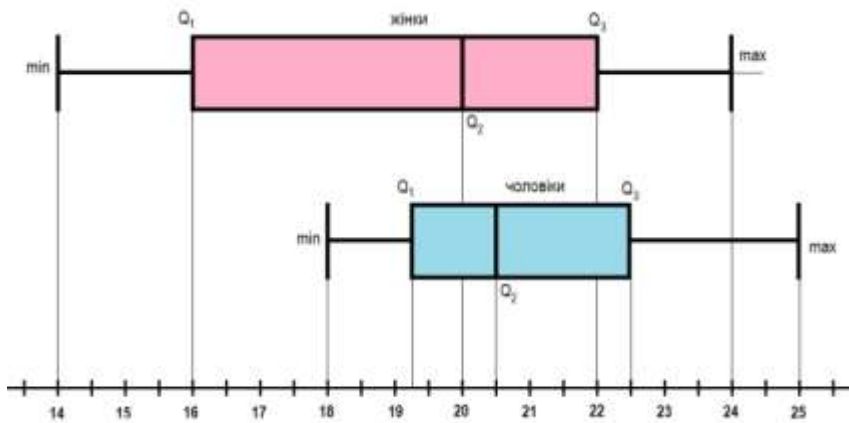


Рис. 2. Гендерне порівняння боксплот діаграми

Побудова цієї діаграми базується на визначенні кватилів з впорядкованого за зростанням набору даних, тобто за значеннями, які ділять цей впорядкований набір даних на рівні чверті. Перший кватиль Q_1 охоплює 25% даних. Другий кватиль Q_2 знаходиться посередині цього набору і ділить дані навпіл. Він також відомий як медіана. Третій кватиль Q_3 охоплює 75% даних. Дані для побудови цієї діаграми наведені в таблиці 3.

Візуальний аналіз діаграм розмаху приведених на рис. 2 показує, що найменша різниця є для компонент Q_2 , Q_3 та максимальних значень, значно більшою є різниця для мінімальних значень та кватилів Q_1 . Крім того, розподіли вибірок мають протилежні асиметрії. Проте ці відмінності, можна вважати, є скоріш за все зумовлені і випадковими значенням даних і їх обсягом.

Групування факторних ознак за фізичними та фізіологічними показниками. В цьому плані ознаки були поділені так: в якості результуючої ознаки виступає тривалість лікування, як правило, в ліжкоднях, а решта ознак прийняті за факторні.

Групування даних здійснено в такий спосіб. За змістом і суттю та попереднім ознайомленням з даною предметною областю (власне бачення) дані можуть бути розподілені на п'ять таких груп.

1. *Фізичні показники.* До них віднесено чотири таких показники: вік, зріст, вага та ІМТ (індекс маси тіла).
2. *Серцево-судинна система.* Ця система представлена лише двома показниками, а саме: САТ (систолический артеріальний тиск) та Pulse (частота пульсу).
3. *Система дихання.* Дана група представлена десятьма показниками, до яких включено такі: Test Walk 6 min (шести-хвилинний тест ходьби), SaO₂ (тест на астму), Borg (навантажувальні проби для хворих з ішемічною хворобою серця), Forge Lung saras (об'єм видихнутого повітря за певний час, що описує стан легень), ПОШвид (пікова об'ємна швидкість видиху), МОШ₂₅ (максимальна об'ємна швидкість повітряна на рівні видиху 25%), МОШ₅₀ (максимальна об'ємна швидкість повітряна на рівні видиху 50%), МОШ₇₅ (максимальна об'ємна швидкість повітряна на рівні видиху 75%), ЖЕЛВдоху (життєва ємність легень).
4. *Імунна система.* В цю групу входять такі біохімічні показники; CD3 (білки, присутні на поверхні т-лейкоцитів), CD22 (білок на поверхні В-лімфоцитів), limf0 (не знайшли визначення), CD4 (білок на поверхні т-лімфоцитів), CD8 (білок антиген), CD4_CD8 (співвідношення між різними групами лімфоцитів), T_v (не знайшли визначення), T_plus_V_0 (не знайшли визначення), CB16 (клітини кілери), ФНПalpha (hsdtyt wbnjrsye).
5. *Кровоносна система.* Сюди входять такі три показники: ІЛ8 (рівень інтерлейкоцину-8), ІЛ4 (рівень інтерлейкоцину-4), ІЛ10 (рівень інтерлейкоцину-10).

Час одужання «Duration» ми вважаємо окремим показником. Саме відносно цього показника мають бути зроблені висновки цього дослідження.

Для встановлення зв'язку кожної групи показників індивідуальним значенням часу одужання пацієнтів використано багатовимірні середні, в результаті кожному пацієнту поставлено його багатовимірний середній показник кожної з груп. Приведені індивідуальні значення часу одужання розбито на чотири інтервали, а саме: 14-16, 17-19, 20-22, 23-25 днів. Для кожного такого інтервалу визначають середнє значення відповідних багатовимірних середніх значень індивідуальних ознак, які подані відповідно в табл. 3. Групи показників-ознак в цій таблиці позначені великими латинськими літерами: **F** – фізичні характеристики та показники **S** – серцево-судинної, **D** – дихальної, **I** – імунної і **K** – кровоносної систем.

Зв'язок ознак з термінами одужання приведений в табл. 3, поданий також діаграмою на рис. 3, в якому кольором позначені групи ознак. Візуалізація цих результатів дає загальне представлення про рівень медико-біологічних показників відносно вказаних вище інтервалів часу одужання пацієнтів.

Зв'язок термінів одужання з групами ознак

Групи ознак	Терміни одужання			
	14-16	17-19	20-22	23-25
F	0,973	0,972	1,040	0,985
S	1,016	1,115	0,935	0,984
D	1,045	0,993	0,953	0,951
I	0,966	0,992	1,047	0,961
K	1,021	1,163	0,999	1,026

Аналіз цієї діаграм дає таку загальну картину зв'язку груп ознак з терміном одужання.

1. Усі групові показники в напрямку зростання тривалості часу одужання вказують на U-подібну форму цієї залежності, тобто для першого і четвертого інтервалу є вищими ніж для другого і третього інтервалів.

2. Найменша зміна належить групі фізичних показників, що можна пояснити повільними змінами таких характеристик як вік, вага, ріст.

3. Відмінність форми залежності для дихальної системи в тому, що вона практично симетрична, причому для першого і четвертого інтервалів величина зв'язку майже у два рази більша ніж для другого і третього інтервалів.

4. Показники імунної і кровоносної систем дуже схоже змінюються синхронно зростаючи з часом одужання.

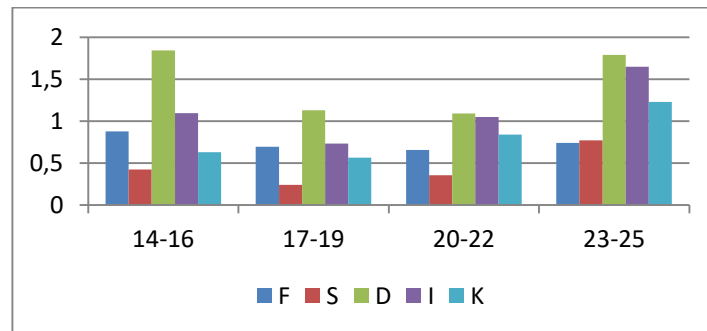


Рис.3. Зв'язок термінів одужання з групами ознак

Таку ситуацію практично неможливо пояснити на підставі отриманих даних. Відмінності в діаграмах, які спостерігаються на рис. 3 скоріше пов'язані з малим обсягом вибірки, а не з існуючою невідомою залежністю, хоча відкидати її існування також не можна. Тому, виникає потреба вяснити чи існує різниця між інтервалами, яку можна подати як задачу однофакторного дисперсійного аналізу. Однофакторний дисперсійний аналіз використовують для визначення наявності або відсутності статистично значущої різниці між середніми значеннями трьох або більше незалежних вибірок.

Тут фактором є інтервал одужання, який можна позначити в такий спосіб – прийняти в якості фактора, часу одужання, що змінюється на чотирьох рівнях, на кожному з яких приймає середнє значення багатовимірних середніх груп ознак.

Тут однофакторний дисперсійний аналіз використано для виявлення статистично значущості різниці між середніми значеннями цих чотирьох груп, які відрізняються між собою. Іншими словами, треба здійснити перевірку нульової гіпотези H_0 : середні значення чотирьох інтервалів є рівні між собою.

В даному випадку результат застосування однофакторного дисперсійного аналізу дає таке. За результатами ANOVA – F-тесту маємо: значення $F = 1.587$, а критичне (табличне) значення цього показника $F_{cr}=3.239$. Оскільки статистика F є меншою за критичне значення F_{cr} немає достатньої впевненості для відхилення нульової гіпотези про те, що середні значення чотирьох груп відрізняються між собою. Спостережувану візуальну різницю можна вважати випадковою. Рівність середніх також може бути підтверджена використанням p -значення, яке рівне $p=0.232$ і перевищує рівень значущості $\alpha = 0.05$. Отже, немає підстав для відхилення нульової гіпотези про рівність середніх цих інтервалів.

Зв'язок віку і ваги пацієнтів з їх медико-біологічними показниками. Для визначення зв'язку показників медико-фізіологічних груп з віком і вагою використано ті ж самі багатовимірні середні також усереднені щодо кожного пацієнта. Для перевірки значущості цього зв'язку використаний однофакторний дисперсійний аналіз. При цьому включено чотири групи показників: серцево-судинну, дихальну, імунологічну та кровоносну.

Віковий та ваговий інтервали розбиті в такий спосіб:

- за віком: 30 – 40, 41 – 51, 52 – 62, 63 – 73;
- за вагою: 59 – 71, 72 – 84, 85 – 97, 98 – 110.

Таблиця 6					Таблиця 7				
Зв'язок віку з медичними показниками					Зв'язок ваги з медичними показниками				
	30-40	41-51	52-62	63-73		59-71	72-84	85-97	98-110
S	0,841	1,239	1,019	0,953	S	1,175	1,131	0,829	0,935
D	0,943	1,037	0,993	1,043	D	1,056	0,974	0,981	0,988
I	1,044	0,939	1,017	0,956	I	0,995	1,036	0,99	0,985
K	0,858	1,163	0,999	1,026	K	1,101	0,933	0,937	1,069

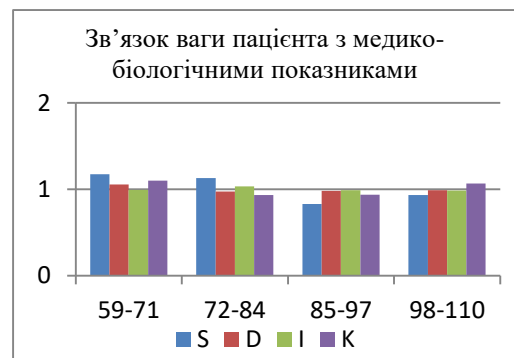
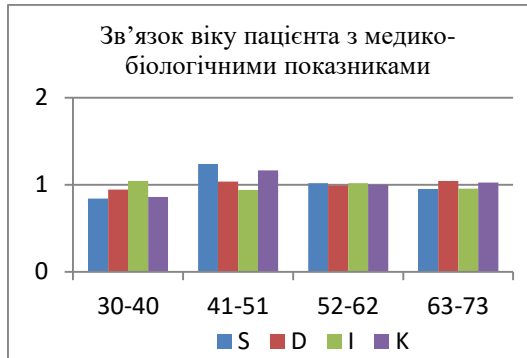


Рис. 4. Зв'язок віку і ваги пацієнтів з їх індивідуальними медико-біологічними показниками

Застосування в цьому випадку однофакторного дисперсійного аналізу дало такі результати:

- для віку $F = 2.805$, $p = 0.085$, $F_{cr} = 3.490$, $F > F_{cr}$, проте $p > 0.05$;
- для ваги $F = 2.746$, $p = 0.089$, $F_{cr} = 3.490$, $F > F_{cr}$, проте $p > 0.05$.

За критерієм F не має підстав відхилити нульову гіпотезу про рівність середніх, проте імовірність p дуже незначно перевищує рівень значущості $\alpha = 0.05$ не підтверджує рівність середніх. Очевидно це потребує додаткового дослідження. Таким чином, групи медико-біологічних показників, значення яких представлені індивідуальними багатовимірними середніми, пов'язані з віком та вагою пацієнтів можна вважати такими, що пов'язані з одужанням практично однаково.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

Результати проведеного дослідження доводять практичну цінність застосування аналізу малих вибірок багатовимірних даних для розуміння стану ситуації, а в даному випадку одужання 19 пацієнтів з захворюванням на Covid-19. Застосування методів перевірки приналежності до нормального розподілу дало підстави стверджувати що дані мають нормальний розподіл, а гендерні відмінності відсутні. Приведені для кожного пацієнта 29 медико-біологічних показників були розділені на п'ять груп. За допомогою багато вимірних середніх встановлено досить слабкий їх зв'язок із індивідуальними термінами одужання, а також виявився досить слабкий зв'язок фізіологічних показників з віком і вагою пацієнтів. Таким чином, використання малих вибірок багатовимірних даних є цілком правомірним щодо прийняття, принаймні початкових, рішень в процесі лікування. Даний підхід може бути безпосередньо використаний в процесі формування та інтерпретації діагностичних результатів та відповідної звітності, а також з науковою метою.

Література

1. Бахрушин В.С. Методи аналізу даних : навчальний посібник для студентів / В.С. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
2. Грицюк П.М., Остапчук О.П. Аналіз даних: Навчальний посібник.– Рівне: НУВГП, 2008. – 218 с.
3. Лупан І.В. Вибіркове обстеження у психології, соціології та педагогіці: навчальний посібник. – Кропивницький: Видавець Лисенко В.Ф., 2019. – 192 с.
4. Куц Ю. В., Лисенко Ю. Ю. Статистичні методи визначення залежностей між випадковими величинами: навчальний посібник. – Київ: КПІ ім. Ігоря Сікорського, 2022. – 115 с.
5. Морган Сі Джей. Use of proper statistical techniques for research studies with small samples./ Am J Physiol Lung Cell Mol Physiol 313: L873–L877, 2017. Електронний ресурс: <https://journals.physiology.org/doi/pdf/10.1152/ajplung.00238.2017>
6. Yang S, Berdine G. Small sample size. The Southwest Respiratory and Critical Care Chronicles 2023; 11(49): 52–55.
7. Ковальова О.М. Вікові та гендерні особливості клінічного перебігу коронавірусної хвороби COVID-19. / О.М. Ковальова. Східноєвропейський журнал внутрішньої та сімейної медицини, 2021, № 2, С. 4 – 9.

8. Турияница С.М. Гендерні відмінності перебігу гострого респіраторного вірусного захворювання SARS-CoV2-19 / С.М. Турияница, О.О. Корчинська, М.А. Поляк, Н.М. Гема-Багіна, В.Д. Симулик, О.А. Балога, О.В. Золина, І.В. Поляк. УКРАЇНА. ЗДОРОВ'Я НАЦІЇ. 2021. № 4 (66).
9. Garcia E. A Tutorial on Quantile-Quantile Plots / Електронний ресурс: https://www.researchgate.net/publication/308631899_A_Tutorial_on_Quantile-Quantile_Plots
10. Williamson D.F. The Box Plot: A Simple Visual Method to Interpret Data / David F. Williamson, Robert A. Parker, Juliette S. Kendrick. *Annals of Internal Medicine*. 1989; 110: 916-921.
11. Potter K. Methods for Presenting Statistical Information: The Box Plot / Електронний ресурс: <http://www.sci.utah.edu/~kpotter/publications/potter-2006-MPSI.pdf>
12. Herv'e Abdi. Normalizing Data. In Neil Salkind (Ed.), *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage. 2010 1 – 4. Електронний ресурс: <https://personal.utdallas.edu/~herve/abdi-Normalizing2010-pretty.pdf>
13. Пономаренко В. С. Багатовимірний аналіз соціально-економічних систем : навчальний посібник / В. С. Пономаренко, Л. М. Малярець. – Харків : Вид. ХНЕУ, 2009. – 384 с.
14. Бізнес-аналітика багатовимірних процесів : навчальний посібник [Електронний ресурс] / Т. С. Клебанова, Л. С. Гур'янова, Л. О. Чаговець та ін. – Харків : ХНЕУ ім. С. Кузнеця, 2018. – 272 с.
15. Багатовимірний аналіз соціологічної інформації: сутність, класичні методи та інтелектуальні технології / О.М. Кислова. *Вісник Харківського національного університету імені В.Н. Каразіна* '2009, № 881. С. 98 – 102.
16. Горват А.А., Молнар О.О., Мінкович В.В. Методи обробки експериментальних даних з використанням MS Excel: Навчальний посібник. Ужгород: Видавництво УжНУ “Говерла”, 2019. – 160 с.
17. Excel 2013–2016 : навчальний посібник / Укладач: Дячук С.Ф. – Тернопіль : Вид-во ТНТУ імені Івана Пулюя, 2021. – 308 с.

References

1. Bakhrushyn V.Ie. *Metody analizu danykh : navchalnyi posibnyk dlia studentiv / V.Ie. Bakhrushyn. – Zaporizhzhia : KPU, 2011. – 268 s.*
2. Hrytsiuk P.M., Ostapchuk O.P. *Analiz danykh: Navchalnyi posibnyk. – Rivne: NUVHP, 2008. – 218 s.*
3. Lupan I.V. *Vybirkove obstezhennia u psykholohii, sotsiolohii ta pedahohitsi: navchalnyi posibnyk. – Kropyvnytskyi: Vydavets Lysenko V.F., 2019. – 192 s.*
4. Kuts Yu. V., Lysenko Yu. Yu. *Statystychni metody vyznachennia zalezhnosti mizh vypadkovymy velychynamy: navchalnyi posibnyk. – Kyiv: KPI im. Ihoria Sikorskoho, 2022. – 115 s.*
5. Morhan Si Dzhei. Use of proper statistical techniques for research studies with small samples./ *Am J Physiol Lung Cell Mol Physiol* 313: L873–L877, 2017. *Elektronnyi resurs: https://journals.physiology.org/doi/pdf/10.1152/ajplung.00238.2017*
6. Yang S, Berdine G. Small sample size. *The Southwest Respiratory and Critical Care Chronicles* 2023; 11(49): 52–55.
7. Kovalova O.M. *Vikovi ta henderni osoblyvosti klinichnoho perebihu koronavirusnoi khvoroby COVID-19. / O.M. Kovalova. Skhidnoievropeyskyi zhurnal vnutrishnoi ta simeinoi medytyny, 2021, № 2, S. 4 – 9.*
8. Turianytsia S.M. *Henderni vidminnosti perebihu hostroho respiratornoho virusnoho zakhvoriuvannia SARS-CoV2-19 / S.M. Turianytsia, O.O. Korchynska, M.A. Poliak, N.M. Hema-Bahyna, V.D. Symulyk, O.A. Baloha, O.V. Zolina, I.V. Poliak. UKRAINA. ZDOROVIA NATsII. 2021. № 4 (66).*
9. Garcia E. A Tutorial on Quantile-Quantile Plots / *Elektronnyi resurs: https://www.researchgate.net/publication/308631899_A_Tutorial_on_Quantile-Quantile_Plots*
10. Williamson D.F. *The Box Plot: A Simple Visual Method to Interpret Data / David F. Williamson, Robert A. Parker, Juliette S. Kendrick. Annals of Internal Medicine. 1989; 110: 916-921.*
11. Potter K. *Methods for Presenting Statistical Information: The Box Plot / Elektronnyi resurs: http://www.sci.utah.edu/~kpotter/publications/potter-2006-MPSI.pdf*
12. Herv'e Abdi. *Normalizing Data. In Neil Salkind (Ed.), Encyclopedia of Research Design. Thousand Oaks, CA: Sage. 2010 1 – 4. Elektronnyi resurs: https://personal.utdallas.edu/~herve/abdi-Normalizing2010-pretty.pdf*
13. Ponomarenko V. S. *Bahatovymirnyi analiz sotsialno-ekonomichnykh system : navchalnyi posibnyk / V. S. Ponomarenko, L. M. Maliarets. – Kharkiv : Vyd. KhNEU, 2009. – 384 s.*
14. *Biznes-analytika bahatovymirnykh protsesiv : navchalnyi posibnyk [Elektronnyi resurs] / T. S. Klebanova, L. S. Hurianova, L. O. Chahovets ta in. – Kharkiv : KhNEU im. S. Kuznetsia, 2018. – 272 s.*
15. *Bahatovymirnyi analiz sotsiolohichnoi informatsii: sutnist, klasychni metody ta intelektualni tekhnolohii / O.M. Kyslova. Visnyk Kharkivskoho natsionalnoho universytetu imeni V.N. Karazina '2009, № 881. S. 98 – 102.*
16. Horvat A.A., Molnar O.O., Minkovych V.V. *Metody obrobky eksperymentalnykh danykh z vykorystanniam MS Excel: Navchalnyi posibnyk. Uzhhorod: Vydavnytstvo UzhNU “Hoverla”, 2019. – 160 s.*
17. *Excel 2013–2016 : navchalnyi posibnyk / Ukladach: Diachuk S.F. – Ternopil : Vyd-vo TNTU imeni Ivana Puliuia, 2021. – 308 s.*