ДНІПРОВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ІМЕНІ ОЛЕСЯ ГОНЧАРА

Механіко-математичний факультет

Кафедра статистики й теорії ймовірностей

**КУРСОВА РОБОТА**

другий (магістерський) рівень вищої освіти

спеціальність 112 Статистика

**РАНДОМІЗОВАНИЙ КРИТЕРІЙ ПОРІВНЯННЯ СЕРЕДНІХ ДВОХ ГРУП**

Виконавець

студент групи МС-23м-1

Роман ЩЕРБАК

Керівник

доцент каф. МСТ, к. ф.-м. н.,

Євген КАРНАУХ

Завідувач випускової кафедри

к.ф.-м.н. Валерій ТУРЧИН

Дніпро

2024

# ВСТУП

## Актуальність теми дослідження

Гострий лімфобластний лейкоз (Acute lymphoblastic leukemia, ГЛЛ, ALL) – онкологічне захворювання клітин крови, за тип і перебіг якого відповідає генетична інформація організму. Серед типів захворювання виділяють два основні: BCR/ABL (…) та NEG (…). Їхня відмінність полягає в тому, що…

У роботі розглянуто

Об'єкт дослідження –.

Предмет дослідження –.

Мета роботи – .

**Огляд літератури**

## Завдання роботи

1. .
2. .
3. .
4. .
5. .

Курсова робота складається з таких частин:

* вступу, який обґрунтовує актуальність роботи; визначає цілі проведення наукового дослідження;
* аналітичного огляду літературних джерел, який містить постановку задачі та;
* програмної реалізації;
* висновків;
* списку використаної літератури;
* додатку А. Код програми мовою R.

# ТЕОРЕТИЧНА ЧАСТИНА

## Постановка задачі

Задано набір даних із бібліотеки ALL для мови R, що містить інформацію про 128 пацієнтів та рівень експресії в них 12 625 генів. Необхідно порівняти середній рівень експресії релевантних генів між пацієнтами з захворюванням типу BCR/ABL та NEG та визначити, відсутня (гіпотеза ) чи наявна (гіпотеза ) в них відмінність. Усього в наборі даних наявно 37 пацієнтів із типом лейкемії BCR/ABL та 42 – з NEG.

Для відбору релевантних генів було застосовано процедуру, запропоновану Gentleman et al. [], що вилучає гени, рівень експресії яких у кров’яних тільцях низький що у хворих на лейкемію типу BCR/ABL або NEG, що у здорових. Також було вилучено гени з низькою мінливістю, з використанням міжквартильного розмаху (IQR).

Стандартно для порівняння середніх застосовують критерій Готелінґа, однак його сфера застосування не включає ситуацію, коли , тобто розмірність даних значно перевищує розмір вибірки. Для цього Bai and Sarandasa запропонували обчислення статистики , а Chen та Qin розширили її до статистики .

Рандомізований критерій…

# Практична частина