

НОРМАЛЬНИЙ ЗАКОН РОЗПОДІЛУ В МЕДИЧНИХ ДОСЛІДЖЕННЯХ

М. А. Іванчук, П. Р. Іванчук

Буковинський державний медичний університет, м. Чернівці

Описано основні критерії перевірки закону розподілу на нормальність. Застосовано критерії перевірки на симетричність та ексцес та графічний метод до результатів медичних досліджень. Показано, що в більшості клінічних досліджень закон розподілу не відповідає нормальному. Надано рекомендації щодо проведення статистичної обробки результатів клінічних досліджень.

Ключові слова: нормальний розподіл, гіпотеза, критерій перевірки, медичні дослідження.

НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ В МЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ

М. А. Иванчук, П. Р. Иванчук

Буковинский государственный медицинский университет, г. Черновцы

Описано основные критерии проверки закона распределения на нормальность. Применен критерий проверки на симметричность и эксцесс к результатам медицинских исследований. Показано, что в большинстве клинических исследований (60 %) закон распределения не соответствует нормальному. Даны рекомендации по выбору статистических критериев при обработке результатов клинических исследований.

Ключевые слова: нормальное распределение, гипотеза, критерий проверки, медицинские исследования.

NORMAL DISTRIBUTION LAW IN MEDICAL RESEARCH

M. A. Ivanchuk, P. R. Ivanchuk

Bukovynian State Medical University, Chernivtsi

The main methods for assessing normality were described. As an example, multiple samples from clinical research were tested for normality using graphical (the histogram and the normal probability plot), and statistical methods. The majority of clinical samples were not normally distributed (60 %). The practical recommendations were provided.

Key words: normal distribution, hypothesis testing, checking criteria medical research.

Вступ. Будь-яке медичне дослідження ґрунтується на порівнянні вибірок. Порівнюються дослідні та контрольні вибірки, результати лікування різними медичними препаратами, показники у хворих та здорових, у чоловіків та жінок та ін. Питання про достовірність різниці вирішують за допомогою статистичних критеріїв, які бувають параметричними та непараметричними. При нормальному розподілі ознаки параметричні критерії є більш потужними, ніж непараметричні, тобто дають меншу ймовірність здійснити помилку другого роду – прийняти нульову гіпотезу про те, що вибірки взяті з однієї генеральної сукупності, коли насправді вибірки взяті з різних сукупностей. Тому в усіх випадках, коли вибірки, що

порівнюються, взяті із нормально розподілених сукупностей, необхідно застосовувати параметричні критерії. Але при розподілах, відмінних від нормального, параметричні критерії застосовувати не можна. В цьому випадку використовують непараметричні критерії або модифікують вибірку до нормального закону.

Тому одним з початкових кроків при проведенні статистичного аналізу медичних досліджень є перевірка на нормальність закону розподілу сукупностей. Для перевірки на нормальність закону розподілу існує міжнародний стандарт ISO 5479-97 [1]. В стандарті розглядається графічний метод перевірки з використанням «ймовірнісного паперу», критерії перевірки на

симетричність та на значення ексцесу, критерії Шапіро-Уїлка та Еппса-Паллі. Розглянемо ці методи та деякі інші, які не ввійшли в цей стандарт.

Критерій перевірки на симетричність використовують при об'ємі вибірки $8 \leq n \leq 5000$. Але оскільки крім нормального закону розподілу існують інші симетричні закони (експоненціальний, Лапласа, логістичний), симетричність розподілу є необхідною, але не достатньою умовою для того, щоб стверджувати, що закон розподілу є нормальним. **Критерій перевірки на ексцес** призначений для перевірки ступеня крутизни кривої розподілу ознаки у порівнянні з крутизною нормального при об'ємі вибірки $8 \leq n \leq 5000$. Потужність цього критерію зростає при збільшенні об'єму вибірки. В міжнародному стандарті розглядається також **сумісний критерій перевірки коефіцієнтів асиметрії та ексцесу** для об'ємів вибірок $8 \leq n \leq 5000$.

Застосування **критерію Шапіро-Уїлка** в міжнародному стандарті передбачається при об'ємі вибірки $8 \leq n \leq 50$, **критерій Еппса-Паллі** в міжнародному стандарті пропонується застосовувати для великих об'ємів вибірок ($8 \leq n \leq 200$). Дані критерії є більш потужними, ніж непараметричні критерії згоди, але за їх допомогою важко відрізнити експоненціальний закон розподілу від нормального.

Модифікація Д'Агостіно критеріїв перевірки на симетричність та ексцес хоча і не потрапила до міжнародного стандарту, при малих об'ємах вибірки є більш потужною, ніж критерії Шапіро-Уїлка та Еппса-Паллі [2]. Крім того, за її допомогою можна відрізнити нормальний закон розподілу від експоненціального.

В міжнародному стандарті відмовляються також від використання **критерію Пірсона (критерій χ^2)**, **критерію Колмогорова-Смірнова** та подібних до них критеріїв, оскільки вони підходять лише для згрупованих даних, а згрупування призводить до втрати інформації. Але використання цих критеріїв є доцільним як додаткове до описаних вище критеріїв.

Критерій Колмогорова-Смірнова та критерій Шапіро-Уїлка є реалізованими в комерційних програмах STATISTICA® від компанії StatSoft, Inc та STATGRAPHICS® Centurion XV від компанії StatPoint, Inc. Так, для перевірки вибірки на нормальність в програмі STATISTICA® необхідно вибрати пункт Distribution Fitting в меню Statistics, а в програмі STATGRAPHICS® для цього вибираємо меню Analyze -> Variable Data -> Distribution Fitting. Крім того, в цих програмах можна здійснювати візуальну перевірку закону розподілу на нормальність –

за допомогою гістограми (Statistics -> Distribution Fitting -> Plot of observed and expected distribution в програмі STATISTICA® або Analyze -> Variable Data -> Distribution Fitting в програмі STATGRAPHICS®) або ймовірнісного паперу (Analyze -> Variable Data -> Probability Plots в програмі STATGRAPHICS®).

При перевірці закону розподілу вибірки на нормальність варто користуватися наступним правилом [3]. Якщо об'єм вибірки малий ($n < 30$), необхідно розглядати вибірку як розподілену ненормально і використовувати непараметричні критерії перевірки статистичних гіпотез. При середньому об'ємі вибірки ($30 \leq n \leq 100$) розподіл рідко прямує до нормального, тому якщо вибраний метод показав нормальний розподіл, необхідно перевірити нормальність розподілу ще одним методом. Якщо результат аналогічний, розподіл вибірки можна вважати нормальним. При великих об'ємах вибірки ($n > 100$) розподіл найчастіше відповідає нормальному закону. Тому, якщо вибраний метод показав ненормальний розподіл, необхідно застосувати ще один метод.

Мета дослідження. Перевірити, як часто зустрічається нормальний закон розподілу в медичних дослідженнях.

Матеріали і методи дослідження. Проведена перевірка на нормальність закону розподілу 173 вибірки різних об'ємів (від 18 до 201) результатів медичних обстежень у практично здорових (32 вибірки) та хворих осіб (141 вибірка). Досліджували наступні параметри: інструментальні методи дослідження (ЧСС, КДР, КСР, ФВ, розміри шлуночків та передсердь); рівень гормонів в крові (альдостерон, тестостерон, натрійуретичний гормон, вазопресин, альфа-фетопротеїн); іонограма; біохімічний аналіз крові; загальний аналіз крові; антропометрія (зріст, вага, обвід грудної клітки, голови, голілки, плеча, стегна); коагулограма; об'єктивне обстеження (пульс, САТ, ДАТ, частота дихання).

Гіпотезу про нормальність закону розподілу перевіряли на рівні значущості $\alpha = 0,05$ за допомогою критеріїв перевірки на симетричність та ексцес в програмі Microsoft Excel та графічним методом в програмі STATGRAPHICS®.

Результати дослідження та їх обговорення.

В результаті перевірки розподілів вибірок було виявлено, що 69 (39,88 %) вибірок розподілені за нормальним законом, а розподіл 104 (60,12 %) вибірок не є нормальним (рис.1).

Слід зазначити, що вибірки, розподіли яких ми перевіряли, були отримані в різні роки та з різних джерел. Але загальне співвідношення кількості вибірок з

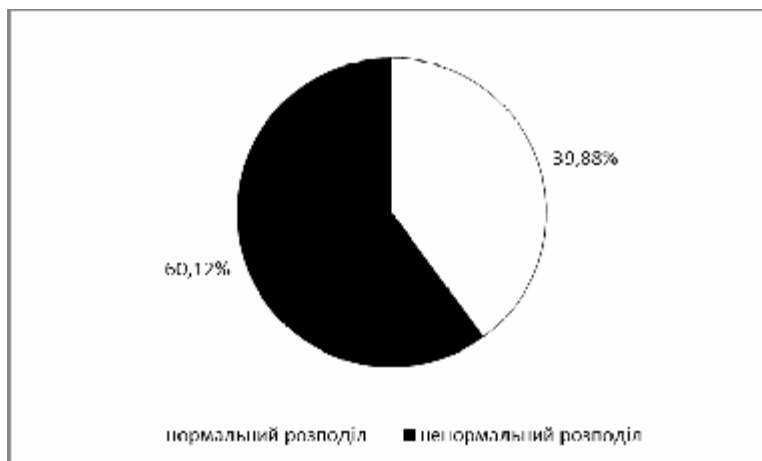


Рис. 1. Співвідношення нормального та ненормального розподілів у досліджуваних вибірках.

нормальним та ненормальним законом розподілу зберігалось незалежно від джерела отриманих даних, тобто не залежало від ретельності збору інформації, присутності суб'єктивних чинників та людського фактора. При цьому розподіл вибірки значною мірою залежить від її об'єму. Дійсно, важко побачити нормальний розподіл в клінічних дослідженнях, оскільки при цьому не вивчається нормальна популяція в цілому, а лише специфічні вибірки з неї. В той час, коли нормальний розподіл, оснований на математичній теорії, відображає лише випадкові варіації, свій внесок в розподіл результатів клінічних вимірів вносить множина інших джерел варіації, особливо

біологічні відмінності між індивідуумами. Отже, подібність реальних розподілів до кривої нормального розподілу в клінічній медицині, як правило, має випадковий характер [4].

Розглянемо приклади перевірки закону вибірки на нормальність.

Приклад 1. Розглянемо вибірку – зріст 127 чоловіків. Спочатку проведемо аналіз графічним методом. Розглянемо гістограму даної вибірки (рис. 2). Гістограма добре «лягає» на криву нормального розподілу. На ймовірнісному папері (рис. 3) також добре видно, що точки (значення відносних накопичених частот у відсотках) розсіяні навколо прямої лінії, що

Рис. 2. Гістограма розподілу вибірки об'ємом $n=127$ в програмі STATGRAPHICS®.

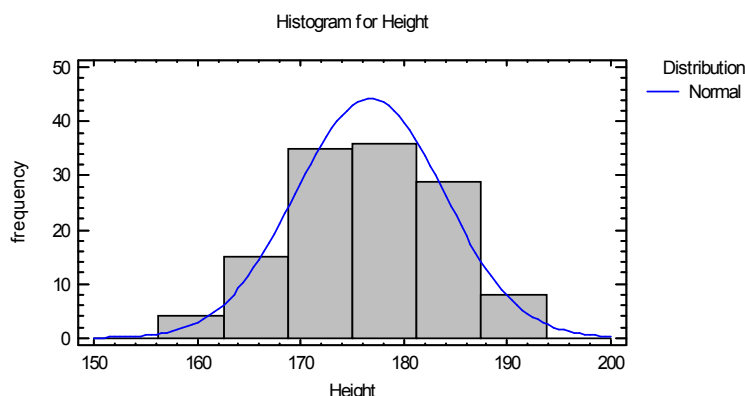
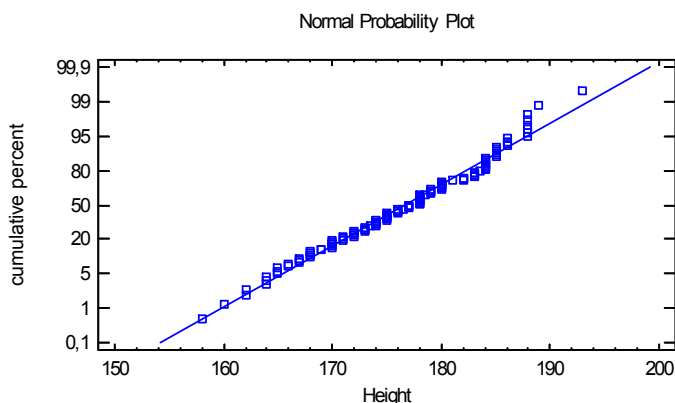


Рис. 3. Перевірка на нормальність закону розподілу вибірки об'ємом $n=127$ за допомогою ймовірнісного паперу в програмі STATGRAPHICS®.



дає перше підтвердження того, що дані розподілені нормально. Перевіримо значення показників асиметрії та ексцесу. $As=-0,278$, $Es=-0,47$. Ці значення менше критичних для даного об'єму вибірки. Отже, значення коефіцієнтів асиметрії та ексцесу також

відповідають нормальному закону. Здійснимо додатково перевірку за критерієм Колмогорова-Смірнова на рівні значущості $p<0,05$ в програмі STATGRAPHICS® (рис. 4). Отже, і за критерієм Колмогорова-Смірнова розподіл ознаки є нормальним.

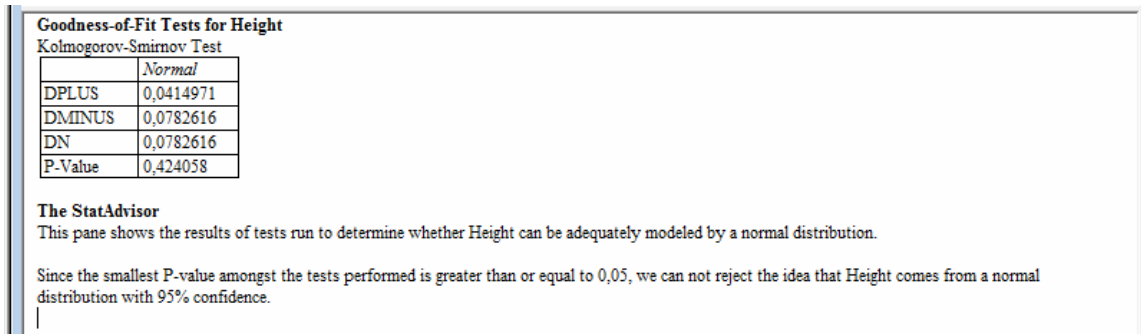


Рис. 4. Перевірка на нормальність закону розподілу вибірки об'ємом $n=127$ за допомогою критерію Колмогорова-Смірнова в програмі STATGRAPHICS®.

Приклад 2. Розглянемо вибірку – рівень альдостерону в крові у хворих до лікування. Об'єм вибірки дорівнює 29. Гістограма частот «не лягає» на криву нормального розподілу (рис. 5). Ще краще відмінність розподілу від нормального видно на ймовірнісному папері – точки відносних накопичених частот зовсім не лягають на пряму (рис. 6). Коефіцієнти ексцесу та асиметрії цієї вибірки не відповідають нормальному закону ($As=4,673$, $Es=37,1$). За критерієм Колмогорова-Смірнова також відкидаємо нульову гіпотезу про нормальність розподілу на рівні значущості $p<0,05$ (рис. 7). Отже, закон розподілу даної вибірки не є нормальним.

ють на пряму (рис. 6). Коефіцієнти ексцесу та асиметрії цієї вибірки не відповідають нормальному закону ($As=4,673$, $Es=37,1$). За критерієм Колмогорова-Смірнова також відкидаємо нульову гіпотезу про нормальність розподілу на рівні значущості $p<0,05$ (рис. 7). Отже, закон розподілу даної вибірки не є нормальним.

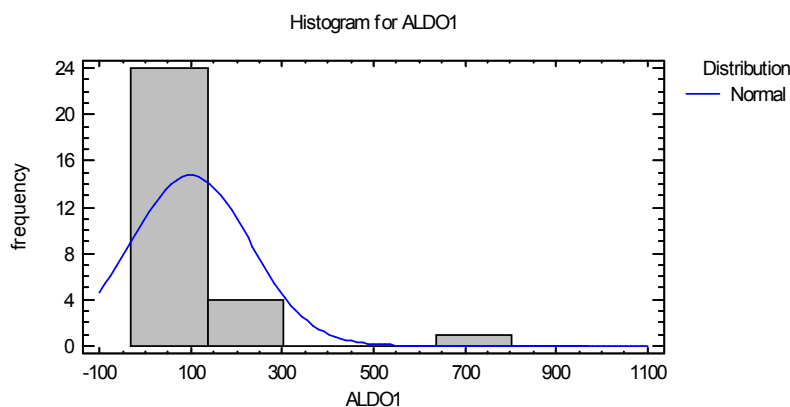


Рис. 5. Гістограма розподілу вибірки об'ємом $n=29$ в програмі STATGRAPHICS®.

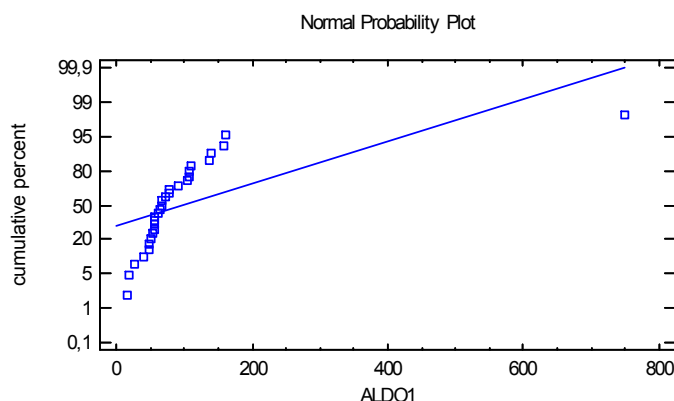


Рис. 6. Перевірка на нормальність закону розподілу вибірки об'ємом $n=29$ за допомогою ймовірнісного паперу в програмі STATGRAPHICS®.

Goodness-of-Fit Tests for ALDO1	
Kolmogorov-Smirnov Test	
	Normal
DPLUS	0,29749
DMINUS	0,260424
DN	0,29749
P-Value	0,0117975
The StatAdvisor	
This pane shows the results of tests run to determine whether ALDO1 can be adequately modeled by a normal distribution.	
Since the smallest P-value amongst the tests performed is less than 0,05, we can reject the idea that ALDO1 comes from a normal distribution with 95% confidence.	

Рис. 7. Перевірка на нормальність закону розподілу вибірки об'ємом $n=29$ за допомогою критерію Колмогорова-Смірнова в програмі STATGRAPHICS®.

Висновки. 1. На початку обробки результатів медичних досліджень необхідним кроком є перевірка закону розподілу всіх вибірок на нормальність.

2. Знання про закон розподілу вибірок дає можливість досліднику правильно вибрати критерій перевірки висунутих гіпотез і отримати статистично

достовірні дані. При нормальному законі розподілу вибірок, що порівнюються, використовують параметричні критерії (наприклад, критерій Стюдента). Якщо закон розподілу хоча б однієї з вибірок, що порівнюються, не є нормальним, використовують непараметричні критерії (критерії знаків, Уїлкоксона, Манна-Уїтні, Розенбаума та ін.)

Література.

1. ГОСТ Р ИСО 5479-2002. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения. – М. : Изд-во стандартов. 2002. – 30 с.
2. Сравнительный анализ критериев проверки нормальности одномерных величин – Режим доступа: URL [http://ami.nstu.ru/~headrd/seminar/Kontrol_Q/](http://ami.nstu.ru/~headrd/seminar/Kontrol_Q/new_normal_red.htm)

[new_normal_red.htm](http://ami.nstu.ru/~headrd/seminar/Kontrol_Q/new_normal_red.htm)

3. Jitendar Vij. Statistical tests in medical research. – Jaypee Brothers Medical Publishers, 2011. – 182 p.
4. Справочник заболеваний. Распределение значений лабораторных показателей – Режим доступа: URL <http://medhealth.ru/klinicheskaya-epidemiologiya/raspredelenie-znachenij-laboratornykh-pokazatelej>.