

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/326271449>

Посібник з біостатистики аналіз результатів медичних досліджень у пакеті EZR (R-STATISTICS)

Book · July 2018

CITATIONS

0

READS

2,799

1 author:



[Yaroslav Tsekhmister](#)

Bogomolets National Medical University

199 PUBLICATIONS 54 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



MEDICAL EDUCATION [View project](#)



Psychologic, law [View project](#)

Національний медичний університет імені О. О. Богомольця

**ПОСІБНИК З БІОСТАТИСТИКИ
АНАЛІЗ РЕЗУЛЬТАТІВ МЕДИЧНИХ
ДОСЛІДЖЕНЬ У ПАКЕТІ
EZR (R–STATISTICS)**

Для магістрів, інтернів, клінічних ординаторів
та аспірантів галузі знань «Охорона здоров'я»

2018

УДК 61:001.891]:311(075.8)

П61

Посібник з біостатистики. Аналіз результатів медичних досліджень у пакеті EZR (R-statistics) / В. Г. Гур'янов, Ю. Є. Лях, В. Д. Парій, О. В. Короткий, О. В. Чалий, К. О. Чалий, Я. В. Цехмістер : Навчальний посібник. – К. : Вістка, 2018. – 208 с.

ISBN 978-617-7157-67-9

Посібник відповідає новим навчальним програмам з курсу «Методологія наукового дослідження» для аспірантів та пошукувачів учених ступенів за галуззю знань «охорона здоров'я», спеціальність «медицина». Призначений для підготовки магістрів, клінічних ординаторів та аспірантів вищих медичних навчальних закладів МОЗ України та може бути корисним для всіх науковців, які використовують біостатистику в своїх дослідженнях.



*Схвалено та рекомендовано до друку
Вченою радою Національного медичного університету імені О. О. Богомольця
(протокол № 7 від 20.12.2017 р.)*

Авторський колектив:

Гур'янов В. Г. – доцент кафедри медичної і біологічної фізики та інформатики НМУ імені О. О. Богомольця, к. ф.-м. наук, доцент;

Лях Ю. Є. – завідувач кафедри здоров'я людини та фізичної терапії СНУ імені Лесі Українки, д. біол. наук, професор;

Парій В. Д. – завідувач кафедри менеджменту охорони здоров'я НМУ імені О. О. Богомольця, д. мед. наук, професор, Заслужений лікар України, академік АНВО України;

Короткий О. В. – аспірант кафедри менеджменту охорони здоров'я;

Чалий О. В. – завідувач кафедри медичної і біологічної фізики та інформатики НМУ імені О. О. Богомольця, д. ф.-м. наук, професор, член-кореспондент НАПН України, Заслужений діяч науки і техніки України;

Чалий К. О. – професор кафедри медичної і біологічної фізики та інформатики, д.-ф.-м. наук, Ph. D. в інженерії, професор, академік АНВШ України;

Цехмістер Я. В. – перший проректор з науково-педагогічної роботи НМУ імені О. О. Богомольця, професор кафедри медичної і біологічної фізики та інформатики, д. пед. наук, професор, член-кореспондент НАПН України, Заслужений працівник освіти України.

Рецензенти:

Яворовський О. П. – завідувач кафедри гігієни та екології №2 НМУ ім. О. О. Богомольця, академік НАМН України, д. мед. наук, професор, Заслужений діяч науки і техніки України;

Голубчиков М. В. – завідувач кафедри медичної статистики НМАПО імені П. Л. Шупика, д. мед. наук, професор.

УДК 61:001.891]:311(075.8)

ISBN 978-617-7157-67-9

© Гур'янов В. Г., Лях Ю. Є., Парій В. Д., Короткий О. В.,
Чалий О. В., Чалий К. О., Цехмістер Я. В., 2018

ЗМІСТ

ПЕРЕДМОВА	8	4.1.1. Точкова оцінка міри положення кількісної змінної	29
РОЗДІЛ 1. РОЛЬ І МІСЦЕ БІОСТАТИСТИКИ В НАУКОВИХ ДОСЛІДЖЕННЯХ З БІОМЕДИЦИНИ ...	10	4.1.2. Точкова оцінка міри розсіювання кількісної змінної	30
РОЗДІЛ 2. РЕЗУЛЬТАТИ МЕДИЧНОГО ДОСЛІДЖЕННЯ	19	4.1.3. Стандартна похибка оцінки середнього кількісної змінної ...	30
2.1. ТИПИ ДАНИХ	19	4.2. ПРЕДСТАВЛЕННЯ КІЛЬКІСНИХ ДАНИХ. ТЕОРЕТИЧНІ ВІДОМОСТІ. ІНТЕРВАЛЬНА ОЦІНКА..	31
2.1.1. Кількісні дані	20	4.2.1. Інтервальна оцінка міри положення кількісної змінної ...	31
2.1.2. Якісні дані	20	4.3. ПРЕДСТАВЛЕННЯ КІЛЬКІСНИХ ДАНИХ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS).....	33
РОЗДІЛ 3. ГЕНЕРАЛЬНА СУКУПНІСТЬ	22	4.3.1. Приклад 1. Нормальний закон розподілу	33
3.1. ПРЕДСТАВЛЕННЯ КІЛЬКІСНИХ ДАНИХ	22	4.3.2. Приклад 2. Закон розподілу, відмінний від нормального	38
3.1.1. Представлення кількісних даних. Закон розподілу	22	РОЗДІЛ 5. ВИБІРКОВА СУКУПНІТЬ. ЯКІСНІ ДАНІ	42
3.1.2. Міра положення кількісної змінної	24	5.1. ПРЕДСТАВЛЕННЯ ЯКІСНИХ ДАНИХ. ТЕОРЕТИЧНІ ВІДОМОСТІ. ТОЧКОВА ОЦІНКА	42
3.1.3. Міра розсіювання кількісної змінної	25	5.1.1. Точкова оцінка частоти для якісної змінної	42
3.2. ПРЕДСТАВЛЕННЯ ЯКІСНИХ ДАНИХ	27	5.1.2. Стандартна похибка частоти якісної змінної	43
3.2.1. Міра положення якісної змінної	27	5.2. ПРЕДСТАВЛЕННЯ ЯКІСНИХ ДАНИХ. ТЕОРЕТИЧНІ ВІДОМОСТІ. ІНТЕРВАЛЬНА ОЦІНКА ДЛЯ ДИХОТОМІЧНОЇ ЗМІННОЇ	44
3.3. РЕКОМЕНДАЦІЇ З ПРЕДСТАВЛЕННЯ РЕЗУЛЬТАТІВ	28	5.3. ПРЕДСТАВЛЕННЯ ЯКІСНИХ ДАНИХ. ПРИКЛАД РЕАЛІЗАЦІЇ В	
РОЗДІЛ 4. ВИБІРКОВА СУКУПНІТЬ. КІЛЬКІСНІ ДАНІ	29		
4.1. ПРЕДСТАВЛЕННЯ КІЛЬКІСНИХ ДАНИХ. ТЕОРЕТИЧНІ ВІДОМОСТІ. ТОЧКОВА ОЦІНКА	29		

ПАКЕТІ EZR (R-STATISTICS)	45	закон розподілу	59
5.3.1. Приклад 3. Дихотомічна змінна	45	7.2.2. Приклад 5. Закон розподілу, відмінний від нормального	63
РОЗДІЛ 6. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ	48	7.3. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ	65
6.1. СТАТИСТИЧНІ ГІПОТЕЗИ. ОСНОВНІ ПОНЯТТЯ	48	7.4. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS).....	67
6.1.1. Статистичні гіпотези. Критерії перевірки	48	7.4.1. Приклад 6. Нормальний закон розподілу	67
6.1.2. Статистичні гіпотези. Похибки при перевірці	49	7.4.2. Приклад 7. Закон розподілу, відмінний від нормального	69
6.1.3. Статистичні гіпотези. Приклад	51	РОЗДІЛ 8. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ДВІ ГРУПИ ПОРІВНЯННЯ. ДИХОТОМІЧНІ ЯКІСНІ ЗМІННІ	73
6.1.4. Статистичні гіпотези. Критичний рівень значимості... 53		8.1. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ	73
6.1.5. Статистичні гіпотези. Параметричні і непараметричні критерії	54	8.2. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS).....	74
6.2. СТАТИСТИЧНІ ГІПОТЕЗИ. СТАТИСТИЧНА ЗНАЧИМІСТЬ ПРОТИ КЛІНІЧНОЇ ЗНАЧИМОСТІ	54	8.2.1 Приклад 8. Порівняння двох незалежних сукупностей. Дихотомічні якісні змінні	74
РОЗДІЛ 7. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ДВІ ГРУПИ ПОРІВНЯННЯ. КІЛЬКІСНІ ЗМІННІ	57	8.3. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ	77
7.1. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ.	58		
7.2. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS).....	59		
7.2.1. Приклад 4. Нормальний			

8.4. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS).....	78	10.2.1. Приклад 11. Параметричний критерій аналізу	94
8.4.1. Приклад 9. Порівняння двох пов'язаних сукупностей. Дихотомічні якісні змінні	78	10.2.2. Приклад 12. Непараметричний критерій аналізу	98
РОЗДІЛ 9. КІЛЬКІСНА ОЦІНКА КЛІНІЧНОГО ЕФЕКТУ. ДВІ ГРУПИ ПОРІВНЯННЯ. ДИХОТОМІЧНІ ЯКІСНІ ЗМІННІ	81	10.3. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШЕ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ.	101
9.1. КІЛЬКІСНА ОЦІНКА КЛІНІЧНОГО ЕФЕКТУ. ТЕОРЕТИЧНІ ВІДОМОСТІ	81	10.4. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШЕ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS).....	103
9.2. КІЛЬКІСНА ОЦІНКА КЛІНІЧНОГО ЕФЕКТУ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS).....	84	10.4.1. Приклад 13. Параметричний критерій аналізу	103
9.2.1. Приклад 10. Кількісна оцінка клінічного ефекту	84	10.4.2. Приклад 14. Непараметричний критерій аналізу	105
РОЗДІЛ 10. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ТРИ І БІЛЬШЕ ГРУП ПОРІВНЯННЯ. КІЛЬКІСНІ ЗМІННІ	90	РОЗДІЛ 11. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ТРИ І БІЛЬШЕ ГРУП ПОРІВНЯННЯ. ДИХОТОМІЧНІ ЯКІСНІ ЗМІННІ	109
10.1. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШЕ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ	91	11.1. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШЕ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ	109
10.2. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШЕ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS).....	94	11.2. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШЕ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS).....	110
		11.2.1. Приклад 15. Порівняння більше двох незалежних сукупностей. Дихотомічні якісні	

змінні	110	13.1. ПОБУДОВА Й АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ. ТЕОРЕТИЧНІ ВІДОМОСТІ	128
11.3. ПЕРЕВІРКА СТАТИСТИЧ- НИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШЕ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ	113	13.1.1. Розрахунок коефіцієнтів лінійної моделі регресії	128
11.4. ПЕРЕВІРКА СТАТИСТИЧ- НИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШЕ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTCS).....	114	13.1.2. Оцінка якості та адекват- ності моделі регресії	130
11.4.1. Приклад 16. Порівняння більше двох пов'язаних су- купностей. Дихотомічні якісні змінні	114	13.1.3. Відбір незалежних ознак	131
РОЗДІЛ 12. ПЕРЕВІРКА НАЯВНО- СТІ КОРЕЛЯЦІЙНОГО ЗВ'ЯЗКУ. КІЛЬКІСНІ ЗМІННІ	118	13.1.4. Підтвердження практич- ного використання математичної моделі на контрольній множині	133
12.1. ПЕРЕВІРКА НАЯВНОСТІ КОРЕЛЯЦІЙНОГО ЗВ'ЯЗКУ. КІЛЬКІСНІ ЗМІННІ. ТЕОРЕТИЧНІ ВІДОМОСТІ	119	13.2. ПОБУДОВА Й АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTCS).....	133
12.2. ПЕРЕВІРКА НАЯВНОСТІ КОРЕЛЯЦІЙНОГО ЗВ'ЯЗКУ. КІЛЬКІСНІ ЗМІННІ. ПРИКЛА- ДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTCS).....	121	13.2.1. Приклад 19. Побудова моделей множинної лінійної регресії	133
12.2.1. Приклад 17. Параметрич- ний критерій аналізу	121	РОЗДІЛ 14. ПОБУДОВА Й АНА- ЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. ЛОГІСТИЧНА МОДЕЛЬ РЕГРЕСІЇ	141
12.2.2. Приклад 18. Непараме- тричний критерій аналізу	123	14.1. ПОБУДОВА Й АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. ЛОГІСТИЧНА МОДЕЛЬ РЕГРЕСІЇ. ТЕОРЕТИЧНІ ВІДОМОСТІ	142
РОЗДІЛ 13. ПОБУДОВА Й АНА- ЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ	126	14.1.1. Розрахунок коефіцієнтів логістичної моделі регресії	142
		14.1.2. Оцінка якості та адекват- ності моделі	142
		14.1.3. Оцінка якості моделі. Побудова та аналіз кривих опера- ційних характеристик	146

14.1.4. Відбір незалежних ознак	148	РОЗДІЛ 16. ОЦІНКА ОБ'ЄМУ ВИБІРКИ	177
14.1.5. Підтвердження практичного використання математичної моделі на контрольній множині	149	16.1. ОЦІНКА ОБ'ЄМУ ВИБІРКИ. ТЕОРЕТИЧНІ ВІДОМОСТІ	177
14.2. ПОБУДОВА Й АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. ЛОГІСТИЧНА МОДЕЛЬ РЕГРЕСІЇ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTCS).....	149	16.2. ОЦІНКА ОБ'ЄМУ ВИБІРКИ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTCS).....	177
14.2.1. Приклад 20. Побудова логістичних моделей регресії ..	149	16.2.1. Приклад 24. Порівняння частоти для двох незалежних вибірок	177
14.2.2. Приклад 21. Побудова ROC-кривої тесту	155	16.2.2. Приклад 25. Порівняння середніх для двох незалежних вибірок	180
РОЗДІЛ 15. АНАЛІЗ ВИЖИВАНOSTІ	163	16.2.3. Приклад 26. Порівняння середніх для двох пов'язаних вибірок	181
15.1. АНАЛІЗ ВИЖИВАНOSTІ. ТЕОРЕТИЧНІ ВІДОМОСТІ	164	СПИСОК ЛІТЕРАТУРИ	184
15.1.1. Представлення даних ..	164	ДОДАТКИ. ТАБЛИЦІ ДАНИХ ..	186
15.1.2. Порівняння цензурованих даних	166	ПРЕДМЕТНИЙ ПОКАЖЧИК	202
15.1.3. Вплив факторних ознак на виживання. Модель пропорційних інтенсивностей Кокса	167		
15.2. АНАЛІЗ ВИЖИВАНOSTІ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTCS)	168		
15.2.1. Приклад 22. Побудова й аналіз кривих виживання	168		
15.2.2. Приклад 23. Модель пропорційних інтенсивностей Кокса	172		

ПЕРЕДМОВА

Якщо проаналізувати медичні публікації за останні 5–10 років, то привертає увагу те, що в більшості із них обов'язково присутні посилання на основні положення доказової медицини (Evidence Based Medicine). На жаль, найчастіше такі дослідження виконані зовсім не за визнаними у більшості країн правилами проведення клінічних досліджень, що викладені в стандарті GCP (Good Clinical Practice), а також правилами виробництва лікарських засобів (стандарт GMP – Good Manufacturing Practice) та виконання лабораторних досліджень (стандарт GLP – Good Laboratory Practice). Таким чином, реальне використання стандартів доказової медицини підміняється їхньою формальною декларацією, що, на думку, авторів збільшує наукову цінність такої публікації. Слід зауважити, що принципові помилки, які присутні в таких публікаціях, дуже часто пов'язані з некоректним використанням методів біостатистики, нерозумінням суті математичних конструкцій, вірою в те, що застосування якогось екзотичного математичного методу або статистичного пакету може повністю скорегувати всі недоліки дослідження.

Зі свого боку наукові стандарти GCP та GLP містять жорсткі вимоги до використання статистичних методів у біомедичному дослідженні та до якості статистичних пакетів, які повинні бути обов'язково сертифіковані та відомими науковій спільноті. До таких широко відомих статистичних пакетів відносяться, наприклад, SPSS, Statistica, MedCalc, які використовуються в багатьох біомедичних дослідженнях. Слід зауважити, що все це вимагає від дослідника мати в своєму розпорядженні ліцензійний статистичний пакет, який коштує досить дорого (найнижча ціна – 400\$). Складний матеріальний стан вітчизняної науки не завжди дозволяє дослідникам активно використовувати ліцензійні статистичні пакети що, своєю чергою, утруднює їм доступ до розміщення публікацій у рейтингових міжнародних наукових журналах. Одним із варіантів виходу із такої ситуації є використання статистичних пакетів, які є у вільному доступі та, разом з тим, відповідають вимогам стандартів.

Важливими проблемами для користувача (біомедика) є вибір відповідного для розв'язання певної задачі статистичного пакета й, головне, –

вміння його використовувати. Інша перешкода полягає у відсутності необхідної методичної літератури з використання таких пакетів. Розв'язати ці проблеми, хоча б частково, покликаний цей посібник.

Багаторічний досвід викладання основ біостатистики для різних категорій науковців і практичних спеціалістів свідчить, що принциповим положенням використання статистичних методів у медицині є постановка задачі та інтерпретації результатів статистичних розрахунків. Саме виходячи із цієї проблеми був задуманий і створений посібник.

Книга не призначена для систематичного вивчення біостатистики, вона не містить строгих математичних визначень та викладок. Основне завдання посібника полягає в демонстрації переходу від біологічної задачі до її математичного формулювання і, далі, біомедична інтерпретація результатів

статистичних розрахунків і методів їхнього представлення. Розв'язання практичних задач, аналіз даних проводиться на прикладі **статистичного пакету вільного доступу Easy R (EZR для R-statistics)**.

Окремі розділи посібника були апробовані на заняттях зі студентами, магістрами і аспірантами на навчальних курсах з «Організації наукових досліджень» НМУ імені О.О.Богомольця, їх зауваження були використані для удосконалення подання матеріалу, за що хотілося б висловити подяку слухачам. Також висловлюємо подяку колегам Вишнівецькому Івану Івановичу за ідею використання пакету EZR та Міхалеву Кирилу Олексійовичу за цікаві запитання, для відповіді на які були написані й уточнені цілі розділи посібника.

Сподіваємося на критичні зауваження та побажання читачів, що будуть ураховані для удосконалення курсу.

*В. Г. Гур'янов
Ю. Є. Лях*

РОЗДІЛ 1.

РОЛЬ І МІСЦЕ БІОСТАТИСТИКИ В НАУКОВИХ ДОСЛІДЖЕННЯХ З БІОМЕДИЦИНИ

Доказова медицина є філософією медичної науки у 21 сторіччі, а основними складовими доказової медицини, у свою чергу, є клінічна епідеміологія і біостатистика.

Клінічну епідеміологію визначають як науку, що розробляє методи клінічних досліджень, які дають можливість фахівцю робити правильні висновки, контролюючи вплив систематичних та випадкових помилок. Таким чином, клінічна епідеміологія дозволяє реалізувати процедуру прогнозування для кожного конкретного пацієнта на основі аналізу клінічного протікання хвороби в аналогічних випадках із використанням коректних методів вивчення різних груп хворих для забезпечення точності прогнозу, отримати науково обґрунтовану характеристику дієвості й економічної доцільності терапевтичних, хірургічних, реабілітаційних, оздоровчих втручань і різноманітних діагностичних методик. А мету клінічної епідеміології, у свою чергу, можливо сформулювати як розробку та застосування різноманітних лікувальних методів (різних

медикаментозних та фізичних), клінічного спостереження, що дає можливість робити обґрунтовані висновки, уникаючи впливу систематичних і випадкових помилок.

Клінічна епідеміологія вивчає наукові основи лікарської практики – стандартів для прийняття клінічних рішень, а головний постулат клінічної епідеміології (як і доказової медицини в цілому) стверджує, що **кожне клінічне рішення має базуватися на чітко доведених наукових фактах.**

Одним із основних інструментів у наукових дослідженнях є математична статистика, тому що тільки за її допомогою можливо оцінити вплив систематичних і випадкових помилок. **Статистика – це мистецтво і наука збору й аналізу даних.** Якщо ж дані збираються й аналізуються для вирішення специфічних задач у біології, медицині, педагогіці чи психології, або, загалом, там, де вивчається біологічний об'єкт, то така наука називається **біостатистикою**. Оскільки даними називають будь-який вид

zareєстрованої інформації, то біостатистика відіграє важливу роль в усіх сферах діяльності фахівця: лікаря, фізичного терапевта, психолога та ін. У біостатистиці застосовуються різні методи: збір даних, їхнє узагальнення, аналіз і підведення підсумків, заснованих на отриманих спостереженнях.

Біостатистику також визначають як розділ доказової медицини, який займається організацією, описом та аналізом даних дослідження, що дозволяє із заданим рівнем доведеності підтвердити або спростувати гіпотези, які були запропоновані в процесі проведеного дослідження.

Загалом, статистичний аналіз дозволяє добувати інформацію з даних і оцінювати якість цієї інформації.

Що повинен знати про біостатистику науковець, лікар, фізичний терапевт, або інший фахівець, який використовує її в своїй професійній діяльності? Він має обов'язково розуміти основні поняття статистики, чітко оцінювати всі можливості й обмеження статистичних методів, які використовуються, та їхню адекватність біологічним (клінічним) задачам, що вирішуються. Необхідно усвідомлювати **принципову варіа-**

бельність біологічних об'єктів та неоднозначність багатьох аспектів навколишнього світу.

Крім того, необхідно:

- 1) чітко уявляти собі біологічну (клінічну або іншу) сутність проведеного дослідження;
- 2) брати безпосередню участь у біостатистичному дослідженні в процесі збору та аналізу даних;
- 3) розуміти (оцінювати) та використовувати в якості базової інформації для своєї подальшої діяльності результати біостатистичного аналізу.

Для цього не обов'язково вміти виконувати складний статистичний аналіз, але, аби правильно інтерпретувати отримані результати, необхідно орієнтуватися в основних положеннях біостатистики та сучасних комп'ютерних пакетах для статистичної обробки й аналізу інформації.

У свою чергу, біостатистичні результати повинні мати досить просте та логічне пояснення (навіть якщо існуюча теорія набагато складніша), а дослідник, приймаючи рішення, має керуватися такими принципами:

- 1) не приймати беззаперечно до розгляду отримані статистичні результати, довіряти своєму професійному досвіду, у

- першу чергу, враховувати клінічний, біологічний або інший предметний сенс отриманих даних;
- 2) зберігати здоровий професійний скептицизм;
 - 3) не дозволяти ввести себе в оману за допомогою, на перший погляд, оригінального статистичного аналізу, що може спиратися на абсолютно не реальне та некоректне біологічне, клінічне або математичне підґрунтя;
 - 4) треба мати на увазі, що формальне (без розуміння суті математичних процедур) використання навіть найсучасніших (та найдорожчих!) статистичних пакетів не гарантує правильність отриманого результату.

Метою більшості біомедичних, а також інших досліджень, є збір, візуалізація (опис) і аналіз даних, які допомагають отримати інформацію в певній галузі знань.

На **початковій стадії** біостатистичного аналізу дані повністю ще не зібрані, або навіть не прийняте рішення про те, які процеси і як треба вивчати. Ці питання розглядаються на етапі розробки **дизайну дослідження** таким чином, щоб отримати дійсно важливу й корисну інформацію. Якщо дані

вже зібрані, то на першому етапі проводиться перевірка на присутність грубих помилок, промахів і попередній аналіз інформації. Наступний етап – **оцінки** – дозволяє отримати на їхній основі числові значення випадкової величини (процесу, який вивчається). На кінцевій стадії біостатистичного аналізу – **перевірці статистичних гіпотез** – дані використовуються для прийняття рішення про відповідність гіпотези дійсності з обов'язковим урахуванням клінічної або біологічної значимості.

У багатьох випадках біостатистичний аналіз на цьому може бути завершено, а його результати використовуються при прийнятті результативних біомедичних або клінічних висновків. Для цього розраховуються основні **статистичні показники**, які дозволяють звести досить велику кількість даних до кількох числових, так званих узагальнюючих, показників. До них відносяться, наприклад, середнє значення (арифметичне, геометричне, гармонічне, зважене), медіана, мода. Вибір конкретного показника залежить від особливостей даних, що аналізуються.

Важливою і необхідною причиною, яка вимагає використання біостатистики, є принципова (внутрішня) стохастичність бі-

ологічних об'єктів. Ситуація, в якій присутня випадковість, може бути невизначеною, оскільки навіть використання всієї доступної інформації не дозволяє точно спрогнозувати, що ж трапиться з біосистемою під час її функціонування або чи буде ефективним те чи інше лікування та які при цьому можуть виникати ризики.

Для адекватної роботи з ризиками необхідно розуміти їхню природу, що вимагає хоча б елементарних знань з теорії ймовірностей, та вміти оцінювати варіабельність процесу, що вивчається. Для цього використовують різні показники, та найчастіше – це стандартне відхилення або середнє квадратичне відхилення (у випадку нормального розподілу отриманих даних). Такий показник описує, наскільки результат досліджень відрізняється від середнього, а його квадрат має назву дисперсія. Біологічний зміст цього показника – кількісна характеристика варіабельності протікання процесу.

Стандартна помилка середнього відображає точність оцінки середнього значення ознаки в популяції за його вибіркоvim значенням, показуючи таким чином помилку, яка виникає при заміні параметра генеральної сукупності його вибіркоvim значенням.

Вочевидь, стандартна помилка середнього не має ніякого біологічного значення, тому що при суттєвому збільшенні об'єму вибірки вона може досягти нульового значення, що принципово неможливо для біологічного об'єкту, а середнє квадратичне відхилення буде мати певні значення, які характеризують сам процес.

Якщо дані не підпорядковуються нормальному закону розподілу, то розраховуються медіана або мода, а варіабельність результатів може оцінюватись, наприклад, шляхом розрахунку кuartилів.

Досить часто використовується так званий коефіцієнт варіації, який показує, наскільки відрізняється конкретний результат від середнього значення у відсотках. Крім того, він дозволяє порівняти поміж собою варіабельність об'єктів, які мають різну фізичну розмірність, наприклад, масу і довжину тіла.

Вибірка з популяції (вибіркова сукупність) дозволяє **отримати точкову оцінку** параметра, що нас цікавить, та вирахувати стандартну помилку для того, щоб вказати точність оцінки. у біомедичних дослідженнях цього недостатньо і тому необхідно використовувати так званий **вірогідний інтервал (ВІ)**, або 95% ВІ.

Як відомо, ВІ отриманий на вибірковій сукупності, дає оцінку точності результатів дослідження з ціллю отримати висновки про генеральну сукупність усіх пацієнтів. Визначити 95% ВІ можливо таким чином: ВІ – це діапазон значень, у межах якого ми на 95% упевнені, що він включає істинну величину. При використанні ВІ акцент робиться на оцінку кількісного біомедичного ефекту на протизагу широко вживаній величині p , яка розраховується в результаті перевірки статистичної значимості та яку досить часто помилково трактують як **достовірність** отриманого результату. Поняття достовірності в доказовій медицині включає, як мінімум, три основних принципи: рандомізовані вибірки в групах порівнянь; достатній розмір вибірки; незалежний контроль.

Знання p саме по собі не дає ніякої інформації про величину змін та їхній напрямок, тому воно мало інформативне в наукових статтях та дисертаційних дослідженнях. Натомість ВІ вказує на кількісний ефект, наприклад, на ефективність лікування, чутливість діагностичного тесту, підвищення або скорочення відносного ризику (ВР) при лікуванні і таке інше, а також є виміром невизначеності в цій ситуації. Частіше за все 95%

ВІ – це діапазон величин по обидві боки оцінки, в якому імовірно знаходиться істинне (популяційне) значення, і можна бути упевненим в цьому на 95%.

Ідея ВІ заснована на тому, що подібне за дизайном дослідження, виконане на інших вибірках пацієнтів з урахуванням їхніх індивідуальних особливостей, не дало б ідентичних результатів, але вони будуть знаходитися в цьому інтервалі разом з істинною, проте невідомою величиною.

Таким чином, ВІ описує варіабельність, що залежить від вибіркової сукупності. ВІ не відображає додаткову невизначеність, пов'язану з іншими причинами. Так, наприклад, він не включає вплив селективної втрати пацієнтів, невдалого комплайнса, некоректного вимірювання висхідів, відсутність «засліплення» та інше.

Та, незважаючи на різні методичні підходи, ВІ і тести на статистичну значимість пов'язані між собою. Наприклад, якщо величина $p < 0.05$, то їй відповідає 95% ВІ, що свідчить про наявність різниці і, таким чином, відкидається гіпотеза H_0 .

На сьогодні переважає точка зору, що оцінка за допомогою ВІ є більш

адекватною процедурою для оцінки результатів дослідження, але V_1 і p взаємодоповнюють одне одне, і тому в багатьох наукових публікаціях використовуються обидва способи представлення результатів.

У рандомізованих контрольованих дослідженнях (РКД) статистично не значимі відмінності досить часто можуть мати невірну трактовку. У цьому контексті V_1 особливо корисний, тому що він показує, наскільки сумісні отримані результати із клінічно важливим істинним ефектом. Особливо важливий розрахунок V_1 для клінічно значимих порівнянь. Так, при порівнянні двох груп правильним буде той V_1 , що побудований для різниці між групами, а не той V_1 , який можна побудувати для оцінки в кожній групі. Таким чином, некоректно розраховувати тільки окремі V_1 для оцінок у кожній групі.

При аналізі результатів біомедичних і клінічних досліджень дуже важливо і принципово розділяти поняття **статистичної значимості** і **клінічної значимості**. Дуже широко в клінічних дослідженнях помилково використовується теза, що коли відсутня статистична значимість, наприклад, при порівнянні двох способів лікування, то можна вважати ці способи лі-

кування однаково ефективними. На жаль, такий висновок є досить сумнівним. Для цього треба провести більш ретельні та складні статистичні дослідження. Принципово важливо не урівнювати статистичну значимість із клінічно важливим результатом. Клінічну значимість можливо передбачати тільки тоді, коли результат є статистично значимим, та величина оцінки ефективності лікування перевищує деякі наперед встановлені параметри. Широке використання V_1 в біомедичних публікаціях було б дуже корисним для правильного розуміння зовнішніх доказів, які використовуються в практиці доказової медицини.

Біостатистика є обов'язковим атрибутом оцінювання причинно-наслідкових зв'язків в біомедичних дослідженнях. Єдині правильні і переконливі докази причинної ролі того чи іншого фактора при захворюванні, зазвичай, отримують із експериментальних досліджень. Інформацію, отриману з інших досліджень, можливо використовувати тільки в тому випадку, коли вона відповідає досить жорсткому переліку умов, які досить часто не виконують дослідники. Цій проблемі присвячена значна кількість публікацій, та в їхню основу покладені широко відомі критерії О. Б. Хілла:

- 1) Причина обов'язково повинна бути передумовою результату.
- 2) Взаємозв'язок повинен бути правдоподібним, а результати біологічно значимими.
- 3) Результати ряду досліджень, виконаних за єдиним дизайном, повинні бути сумісними.
- 4) Зв'язок причини та результатів має бути досить сильним.
- 5) Співвідношення дози впливу та результатом: більша доза втручання повинна викликати більш тяжке ураження, або викликати більш швидкий початок захворювання.
- 6) Видалення діючого фактору повинно зменшити ризик захворювання або іншого результату.

Найвідомішим методом біомедичних досліджень у виявленні причинно-наслідкових зв'язків є кореляційно-регресійний аналіз даних.

Кореляційний аналіз традиційно використовується для вивчення степені зв'язку поміж двома (або більше) процесами, що вивчаються. Ступінь такого зв'язку зазвичай оцінюється за допомогою коефіцієнту кореляції, який може приймати значення від -1 до $+1$.

Якщо коефіцієнт кореляції дорівнює $+1$, то зв'язок між двома змінними функціональний і пря-

мий, а якщо приймає значення -1 – то зворотній. Якщо ж коефіцієнт кореляції приймає значення 0 , то статистичний зв'язок між двома змінними відсутній.

Звичайно, у статистичних дослідженнях коефіцієнт кореляції не може досягати значення 1 , тому що оцінюється статистичний, а не функціональний зв'язок. Важливо пам'ятати, що **наявність кореляції не означає наявності причинно-наслідкових зв'язків між змінними**, що вивчаються, а зміни значень одного ряду чисел не обов'язково призведуть до зміни значень іншого ряду. Сильна кореляція може бути випадковою і пояснюватися дією третьої змінної, яка залишилася за межами аналізу.

Таким чином, можна стверджувати, **що причинно-наслідкові зв'язки породжують кореляцію, але наявність кореляції не свідчить про наявність причинно-кореляційних зв'язків.**

І головне! Розібратися в причинно-наслідкових зв'язках лише статистичними методами принципово не можливо. Для цього необхідно обов'язково розуміти біологічний механізм, який лежить в основі виявленого зв'язку, а на цій основі розробити дизайн такого біомедичного дослідження. Таким чином, стає зрозумілим,

чому біостатистика стала важливим розділом доказової медицини, а без її використання неможливо спланувати, провести та проаналізувати результати дослідження!

Для ілюстрації важливості й необхідності біостатистики наведемо вимоги до статистики, викладені в стандарті GCP, який є обов'язковим для клінічних досліджень.

- 1) Описати статистичні методи, які плануються використати, включаючи терміни кожного проміжного аналізу.
- 2) Запланувати кількість суб'єктів. У випадку багатоцентрових досліджень необхідно визначити кількість суб'єктів у кожному центрі, що плануються для залучення в дослідження. Обґрунтувати розмір вибірки, включаючи обговорення чи розрахунки для прийняття статистичної потужності дослідження та клінічної ваги дослідження.
- 3) Вказати рівень вибраної статистичної значимості.
- 4) Вибрати критерій зупинки дослідження.
- 5) Передбачити процедуру урахування відсутніх та сумнівних даних, які не підлягатимуть аналізу.
- 6) Усі відхилення від первинного статистичного плану повинні бути описані й обґрунтовані в

протоколі чи фінальному звіті про дослідження.

- 7) Описати процедуру відбору суб'єктів для аналізу (наприклад, усі рандомізовані суб'єкти, усі суб'єкти, які отримували хоча б одну дозу досліджуваного препарату, усі суб'єкти, які відповідають критеріям відбору та дані яких придатні для оцінки).

У стандарті також підкреслюється, що вибір статистичних процедур проводиться на етапі планування дослідження, а не після його завершення!

Для реалізації методів аналізу результатів біомедичних досліджень використовується ряд широко вдомих комерційних статистичних пакетів, а ми, в свою чергу, пропонуємо використати статистичний пакет EZR (Saitama Medical Center, Jichi Medical University, Saitama, Japan) [1]. Цей пакет представляє собою надбудову з графічним інтерфейсом програмного середовища для статистичних обчислень R-statistics (The R Foundation for Statistical Computing, Vienna, Austria). R розповсюджується безкоштовно за ліцензією GNU General Public License для більшості операційних систем: Linux, Microsoft Windows, Mac OS та інших.

У пакеті EZR додані статистичні методи та критерії [2], які часто вико-

ристовують при проведенні аналізу медичних досліджень. За реалізованими функціями аналізу результатів клінічних досліджень пакет EZR може бути поставлений поряд із загальновідомими комерційними статистичними пакетами, як SAS (SAS Institute Inc., Cary, NC, USA), SPSS (SPSS Inc., Chicago, IL, USA), Stata (Stata Corporation, College Station, TX, USA). Пакет EZR вільно завантажується та легко інсталується [3]. На рисунку 1.1 наведено інтерфейс програми та виведено меню статистичних процедур і методів пакету.

Наведене вище дозволяє рекомендувати використання статистичного пакету EZR для проведення аналізу

результатів медичних досліджень та навчання реалізації методів біостатистики для широкого кола користувачів, як студентам медичних закладів, клінічним ординаторам, так і пошукачам ступені доктора філософії у галузі медицини та співробітникам наукових установ.

На основі багаторічного досвіду нами запропоновано практикум з використання методів біостатистики при аналізі результатів медичних досліджень у пакеті EZR. В окремих випадках для графічного представлення результатів було використано пакети MedCalc (MedCalc SoftWare bvba) [4] та MedStat (Лях Ю. Є, Гур'янов В. Г.) [5].

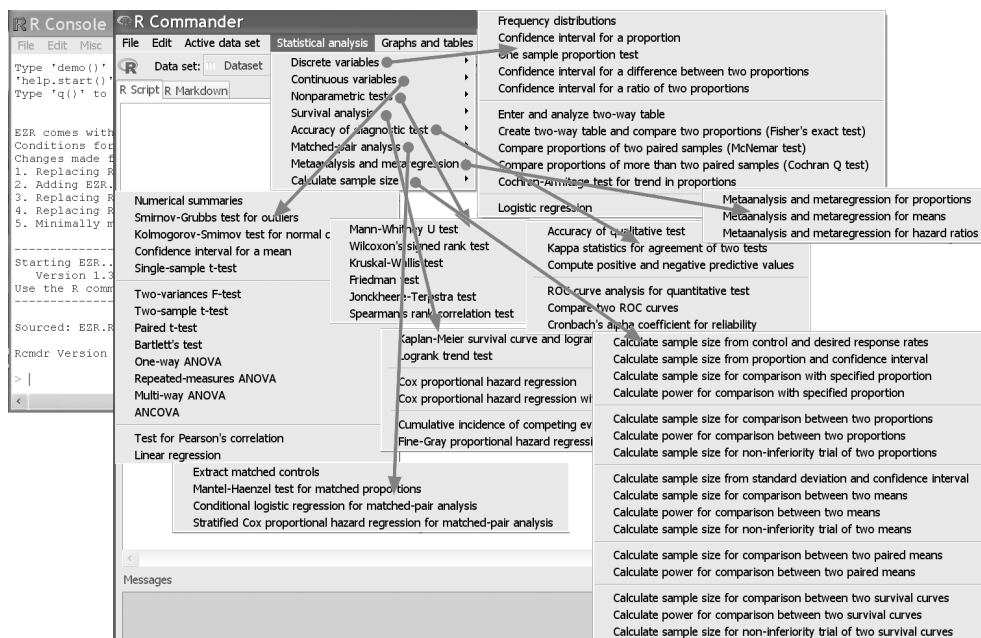


Рис. 1.1. Інтерфейс та основні процедури й методи статистичного пакету EZR

РОЗДІЛ 2.

РЕЗУЛЬТАТИ МЕДИЧНОГО ДОСЛІДЖЕННЯ

При проведенні медичних досліджень проводиться накопичення та аналіз його результатів, що представляють собою сукупність результатів спостереження (вимірювання) одного чи декількох показників (змінних дослідження).

У медичних дослідженнях у розпорядженні дослідника практично ніколи не буває повної групи об'єктів, тобто неможливо провести суцільне спостереження (наприклад, усі хворі зі специфічним захворюванням). Така повна група всіх теоретично можливих об'єктів називається **генеральною сукупністю** або **популяцією** (population). Оскільки повністю вивчити популяцію неможливо, то, для отримання (хоча й не ідеального) уявлення про неї, можна відібрати невелику групу, що складається з деяких (але далеко не всіх) наявних об'єктів – **вибірку** (sample).

Статистичні методи (методи біостатистики) дозволяють на основі **вибірки** зробити висновки щодо властивостей **генеральної сукупності** (звичайно, що для проведення узагальнення вибірка повинна бути **репрезентативною**,

тобто відображати властивості популяції). Процес узагальнення результатів дослідження на вибірці на всю популяцію називається статистичним висновком.

Випадкова (**рандомізована**) вибірка є одним з найкращих способів отримання репрезентативної вибірки. Рандомізація дозволяє гарантувати, що, по-перше, процес вилучення вибірки є неупередженим, тобто всі об'єкти досліджуваної популяції мають рівні шанси бути відібраними. Тому, у середньому, вибірки є репрезентативними для даної популяції (хоча кожна окрема випадкова вибірка зазвичай є репрезентативною тільки наближено, але не ідеально). По-друге, випадковість, введена контрольованим способом на стадії планування, гарантує коректність наступних статистичних висновків.

2.1. ТИПИ ДАНИХ

Нехай проводиться медичне дослідження, наприклад, ми можемо спостерігати перебіг деякого захворювання групи пацієнтів із специфічним захворюванням, при цьо-

му змінними дослідження можуть бути: стать, вік, група крові, стадія гіпертонічної хвороби, рівень глікозильованого гемоглобіну HbA1c крові хворого й т. і. Досліджувані показники (змінні) можуть набувати різних значень для окремого індивідуума, мета ж дослідження полягає у виділенні із цього набору даних закономірностей, що характерні для хворих із досліджуваним специфічним захворюванням.

Для проведення аналізу результатів дослідження, коректного використання методів біостатистики, насамперед, потрібно визначити, до якого типу відносяться показники, що підлягають аналізу. У медичних дослідженнях можна виділити два основні типи даних: кількісні змінні та якісні змінні.

2.1.1. Кількісні дані

Кількісні (числові) змінні визначають вимірювані чи обчислювані величини, які безпосередньо представляють обсяг певної ознаки або кількість її елементарних одиниць. Для таких змінних визначені операції порівняння, додавання/віднімання, множення/ділення. У наведеному вище прикладі до кількісних даних належать: вік хворого та рівень HbA1c крові. Кількісні дані можуть бути поділені на два типи: дискретні та неперервні.

Дискретні числові змінні – показники, що набувають лише визначених числових значень (обмежену чи необмежену кількість), наприклад, кількість зубів у хворого, кількість серцевих скорочень за одну хвилину та т. і.

Неперервні числові змінні – показники, які можуть набувати будь-яких числових значень у певному інтервалі значень. Вік хворого (приймає будь-яке, у тому числі і не ціле, значення в інтервалі від 0 років до 150 років) та рівень HbA1c крові (приймає значення в інтервалі від 0 % до 100 %) відносяться саме до неперервних числових змінних. Варто зазначити, що в практиці медичних досліджень, як правило, неперервні показники зводяться до дискретних (так, для рівня HbA1c вказуються тільки значення з точністю до десятих %, не наводиться, наприклад, значення рівня HbA1c хворого – 6,51 %)

2.1.2. Якісні дані

Якісні змінні вказують, до якого із декількох нечислових категорій належить об'єкт. Вони не вимірюють, а тільки реєструють певну якість, якою володіє об'єкт. Для таких змінних визначені операції відмінності та, інколи, порівняння. У наведеному вище прикладі до якісних даних будуть відноси-

тися стать і група крові хворого, стадія гіпертонічної хвороби. Якісні дані можуть бути номінальними чи порядковими.

Номінальні змінні визначаються в термінах категорій, які не можна впорядкувати, а вони просто мають назви (стать – чоловіча/жіноча, група крові – О, А, В, АВ). Для таких категорій немає чисел, з якими можна робити обчислення, і немає підстав для ранжирування. Для таких змінних визначені тільки операції відмінності.

Порядкові змінні визначаються в термінах категорій, які можуть

бути ранжовані, для яких існує та має змістовний сенс порядок. Це стадії хвороби (відсутність, 1-а стадія, 2-а стадія чи 3-я стадія гіпертонічної хвороби) і т. і. Можна ранжувати дані відповідно до цього порядку й використовувати це ранжирування при виконанні статистичного аналізу. Для таких змінних визначені операції відмінності та порівняння (наприклад, 2-а стадія гіпертонічної хвороби між 1-ою і 3-ою стадіями).

Залежно від того, до якого типу даних відноситься змінна, обираються статистичні методи їхнього аналізу й узагальнення.

РОЗДІЛ 3. ГЕНЕРАЛЬНА СУКУПНІСТЬ

3.1. ПРЕДСТАВЛЕННЯ КІЛЬКІСНИХ ДАНИХ

3.1.1. Представлення кількісних даних. Закон розподілу

При проведенні аналізу результатів медичних досліджень кожна змінна X , що підлягає статистичному аналізу, інтерпретується як

випадкова величина. Повне уявлення про кількісну випадкову величину може дати *функція щільності розподілу ймовірностей* [6] значень цієї величини – $f(x)$ (Probability Density Function – PDF)

На рисунку 3.1 наведено графік функції щільності розподілу показника рівня глікозильованого гемоглобіну HbA1c крові для

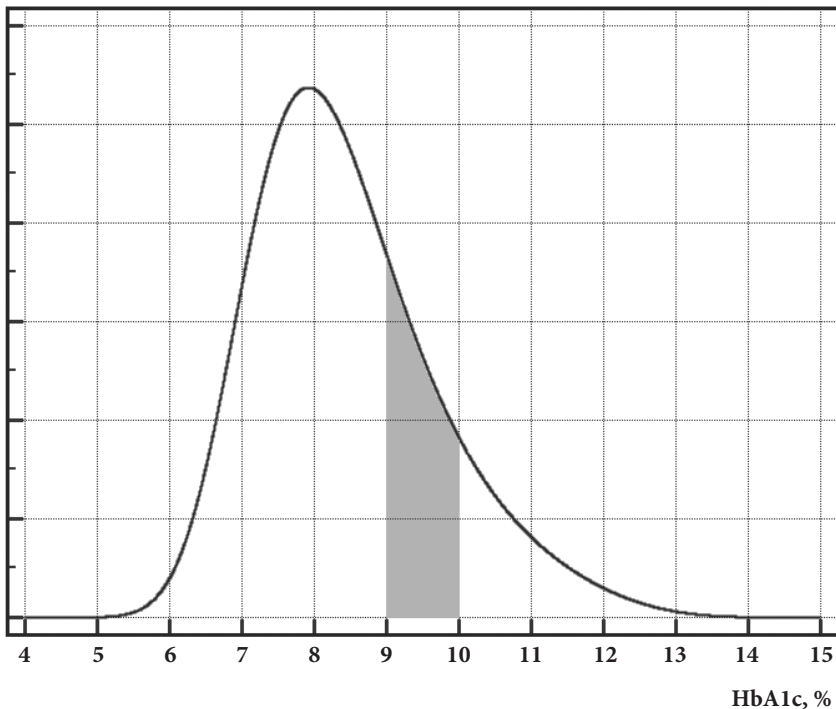


Рис. 3.1. Функція щільності розподілу показника рівня глікозильованого гемоглобіну HbA1c крові для хворих із певним захворюванням. По осі OX вказано рівень HbA1c (%), по осі OY – відповідну функцію щільності ймовірності розподілу цієї величини PDF.

прикладу медичного дослідження (розділ 2.1.1).

За цим графіком можна визначити ймовірність (частоту) того, що значення рівня HbA1c належить до певного інтервалу. За визначенням PDF(x) [6] імовірність зустріти хворого із значенням рівня HbA1c, що лежить у певному інтервалі (x1, x2) чисельно дорівнює площі під графіком функції PDF цього інтервалу (імовірність знайти хворого з $9\% < \text{HbA1c} < 10\%$ при випадковому виборі з генеральної сукупності хворих із певним захворюванням дорівнює площі заштрихованої фігури – див. Рис. 3.1).

З аналізу випливає, що найбільша ймовірність зустріти хворого з рівнем HbA1c близьким до 8%. З відхиленням рівня HbA1c від 8% кількість хворих зменшується, практично не зустрічаються, хворі з рівнем HbA1c < 5% та рівнем HbA1c > 13,5%. При цьому розподіл ймовірностей є асиметричним (див. Рис.1), так із зменшенням показника відносно 8% імовірність спадає швидше, ніж із збільшенням відносно 8%.

Таким чином, отримавши функцію PDF(x) (у вигляді графіка чи в явному аналітичному вигляді) для аналізованої змінної (випадкової величини), ми маємо повну інформацію про цю величину.

Одним з найбільш цікавих у статистиці типів розподілу є нормальний (Гаусів) розподіл функції щільності ймовірності значень кількісної змінної. На рисунку 3.2 представлено PDF(x) значень деякої змінної, що підпорядковується **нормальному закону розподілу**.

Нормальний закон розподілу є спадаючим від центру, симетричним. За визначенням, закон розподілу значень величини X є нормальним (Гаусовим) у випадку, коли функція PDF(x) визначається тільки двома параметрами [6]: **математичним сподіванням** випадкової величини (M(X)) та дисперсією випадкової величини $\sigma^2(X)$, – і задається формулою (3.1):

$$\text{PDF}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-M)^2}{2\sigma^2}\right) \quad (3.1)$$

Математичне сподівання визначає центр розподілу функції PDF, **стандартне відхилення** ($\sigma(X)$) – ширину розподілу функції PDF (див. Рис.2). Треба відзначити, що для нормально розподіленої (з параметрами M(X), $\sigma(X)$) величини X імовірність того, що випадково вибране значення належить інтервалу:

- $M - \sigma < X < M + \sigma$ дорівнює 0.683 (див. Рис. 3.2), тобто 68% всіх значень належать цьому діапазону;

- $M-2\sigma < X < M+2\sigma$ дорівнює 0.954 – 95% усіх значень належать цьому діапазону;
- $M-3\sigma < X < M+3\sigma$ дорівнює 0.997, тобто практично всі значення величини X належать цьому діапазону.

На практиці для представлення результатів медичних досліджень, у випадку кількісних даних, як правило, не представляють функцію розподілу значень аналізованої змінної, а обмежуються числовими характеристиками розподілу [6]. У більшості випадків це: **міра положення** та **міра розсіювання**.

3.1.2. Міра положення кількісної змінної

Для узагальнення результатів дослідження представлення інформації про досліджувану змінну в компактній формі часто розраховують таку числову характеристику, як міра положення, що представляє типовий результат для даної змінної. До часто використаних числових характеристик міри положення кількісної змінної відносяться: математичне сподівання (M) розподілу значень показника, медіана розподілу значень показника (Me), мода/моди розподілу значень показника (Mo).

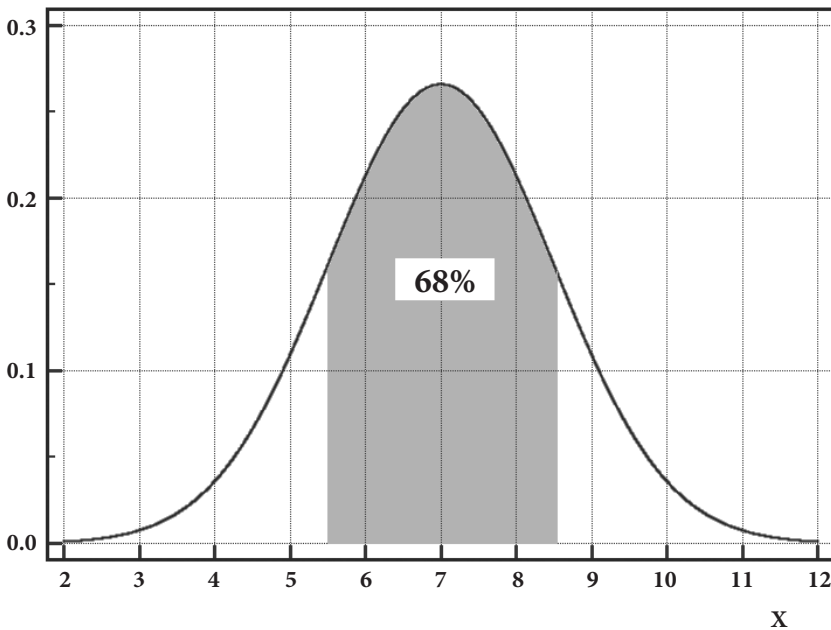


Рис. 3.2. Функція щільності розподілу показника X , що підпорядковується нормальному закону розподілу (параметри розподілу $M(X)=7$, $\sigma^2(X)=2.25$).

Математичне сподівання (середнє значення) кількісної змінної визначається за формулою (для дискретної числової змінної) [6]:

$$M(X) = \sum_{i=1}^n x_i \times p_i, \quad (3.2)$$

де x_i – i -е ($i = 1, \dots, n$) значення, що може мати змінна, p_i – імовірність зустрівти це i -е ($i = 1, \dots, n$) значення в генеральній сукупності.

Слід зауважити, що якщо ми маємо **варіаційний ряд** значень змінних (ряд чисел, що представляють усі значення, які приймає змінна, що зустрічаються в ряду з відповідною частотою), то математичне сподівання змінної X співпадає з простим арифметичним середнім значенням цього варіаційного ряду.

Медіана розподілу значень кількісної змінної представляє собою значення змінної, що лежить у центрі розподілу (імовірність зустрівти значення менше/більше Me дорівнює 0.5).

Мода розподілу значень кількісної змінної представляє собою значення змінної, яке зустрічається з найбільшою частотою (розподіл може мати більш ніж одну моду – бути полімодальним).

Розглянемо ці характеристики показника рівня глікозильованого гемоглобіну HbA1c крові для прикладу

медичного дослідження (PDF(x) наведено на Рис.3.1). У цьому випадку $M = 8,25\%$, $Me = 8,2\%$, $Mo = 7,9\%$.

Для змінної X , що підпорядковується нормальному закону розподілу (PDF наведено на Рис.3.2), розрахунки дають такі значення $M = 7$, $Me = 7$, $Mo = 7$.

Звернімо увагу, що для нормального закону розподілу рівність $M = Me = Mo$ обов'язково виконується; у випадку ж закону розподілу, що відрізняється від нормального, значення показників M , Me та Mo не обов'язково співпадають.

Звідси випливає практичний висновок, що міра положення для кількісної змінної в медичних дослідженнях представляється:

- **показником M (середнім значенням) у випадку нормального закону розподілу;**
- **показником Me (положенням центру) у випадку одномодального закону розподілу, відмінного від нормального;**
- **показником Me та значеннями Mo у випадку полімодального закону розподілу.**

3.1.3. Міра розсіювання кількісної змінної

Кількісна змінна в медичних до-

слідженнях не приймає тільки одне типове значення, вона варіюється у різних об'єктів. Тому для узагальнення результатів цікавою є міра розсіювання (варіабельності) кількісного показника. До часто використаних числових характеристик міри розсіювання кількісної змінної відносяться: розмах значень змінної (W), дисперсію (V) або стандартне відхилення (σ) розподілу, процентилі розподілу значень змінної ($Q_{p\%}$), квартилі розподілу (Q_p).

Розмах значень кількісної змінної є різниця між максимальним (x_{\max}) та мінімальним (x_{\min}) значеннями показника:

$$W = x_{\max} - x_{\min}, \quad (3.3)$$

Дисперсія (Variance) є середнім значенням квадрату відхилення значень змінної від її середнього значення (міра розсіювання значень змінної навколо її математичного сподівання) і визначається за формулою (для дискретної числової змінної) [6]:

$$V(X) = \sum_{i=1}^n (x_i - M(X))^2 \times p_i, \quad (3.4)$$

де x_i – i -е ($i = 1, \dots, n$) значення, що може мати змінна, p_i – імовірність зустріти це i -е ($i = 1, \dots, n$) значення в генеральній сукупності.

Стандартне відхилення (Standard Deviation) є також мірою розсіювання значень змінної навколо її середнього значення і пов'язано з дисперсією формулою визначається за формулою (для дискретної числової змінної) [6]:

$$\sigma(X) = \sqrt{V(X)}, \quad (3.5)$$

Процентиль (Percentile) представляє собою значення змінної $Q_{p\%}$, меншим від якого буде $p\%$ значень змінної. Так $Q_{0\%}$ – мінімальне, а $Q_{100\%}$ – максимальне значення змінної, $Q_{50\%}$ співпадає з медіаною розподілу.

Квартилі (Quartile) є частинні значення процентилю, що поділяють розподіл кількісної змінної на чотири рівні частини.

Так, **перший квартиль** ($Q_I = Q_{25\%}$) є значенням змінної, меншим від якого буде 25 % (чверть) усіх значень змінної.

Третій квартиль ($Q_{III} = Q_{75\%}$) є значенням змінної, меншим від якого буде 75 % усіх значень змінної (більше від якої буде лише 25 % усіх значень змінної).

Значення **другого квартилю** співпадає з медіаною розподілу ($Q_{II} = Q_{50\%} = Me$).

У практиці медичних досліджень міра розсіювання для кількісної змінної, як правило, представляється:

- показником σ у випадку нормального закону розподілу (в інтервалі $M - \sigma < X < M + \sigma$ знаходиться 68 % усіх значень аналізованої змінної);
- показниками Q_I і Q_{III} у випадку закону розподілу, відмінного від нормального (в інтервалі $Q_I < X < Q_{III}$ знаходиться 50 % усіх значень аналізованої змінної).

Звернімо увагу, що у випадку нормально розподіленої кількісної змінної показники M та σ не просто задають числові характеристики, а дають повну, вичерпну інформацію про цю величину (нормальний закон розподілу повністю визначається тільки цими двома параметрами).

Звідси, узагальнюючи аналіз кількісних змінних, можна зробити висновок, що в медичних дослідженнях результати представляються:

- показниками M та σ у випадку нормального закону розподілу кількісної змінної;
- показниками Me , Q_I і Q_{III} у випадку закону розподілу, відмінного від нормального.

3.2. ГЕНЕРАЛЬНА СУКУПНІСТЬ. ПРЕДСТАВЛЕННЯ ЯКІСНИХ ДАНИХ

Для якісних змінних при проведенні аналізу результатів медичних досліджень повне уявлення про випадкову величину може дати закон розподілу категорій цієї змінної. Закон розподілу в цьому випадку простіше представити у вигляді таблиці, у якій перелічені категорії розподілу змінної (x_i , $i=1, \dots, n$, де n – кількість категорій) та ймовірність, із якою значення змінної зустрічається у генеральній сукупності (p_i , $i=1, \dots, n$). При цьому вказані ймовірності, як правило, можуть бути виражені частотою, з якою вони зустрічаються (у відсотках).

3.2.1. Міра положення якісної змінної

Для представлення типового результату для номінальної якісної змінної використовується мода показника, для порядкової якісної змінної може бути використана мода й медіана показника.

Розглянемо представлення якісної змінної для стадії гіпертонічної хвороби для наведеного вище прикладу медичного дослідження (див. Табл. 3.1).

При аналізі цього захворювання встановлено, що найчастіше зустрічаються хворі з гіпертонічною хворобою II стадії.

Звідси, узагальнюючи аналіз якісних змінних, можна зробити висновок, що в медичних дослідженнях результати представляються: показниками M_0 та частотою ($P\%$), із

якою зустрічається відповідне значення показника.

3.3. РЕКОМЕНДАЦІЇ З ПРЕДСТАВЛЕННЯ РЕЗУЛЬТАТІВ

У Таблиці 3.2 наводяться рекомендації з вибору оптимальних характеристик для представлення результатів у залежності від типу даних.

Таблиця 3.1.

Закон розподілу частоти спостереження стадії гіпертонічної хвороби для групи хворих із специфічним захворюванням.

Стадія гіпертонічної хвороби	Відсутня	I стадія	II стадія	III стадія
Частота, %	0	40	50	10

Таблиця 3.2.

Вибір представлення оптимальних характеристик.

Показник	Кількісні змінні		Якісні змінні	
	Нормальний закон розподілу	Закон розподілу, відмінний від нормального	Номінальні	Порядкові
M	+	-	-	-
M_e	-	+	-	+
M_0	-	-	+	+
σ	+	-	-	-
$Q_I - Q_{III}$	-	+	-	-
$P\%$	-	-	+	+

РОЗДІЛ 4. ВИБІРКОВА СУКУПНІТЬ. КІЛЬКІСНІ ДАНІ

У попередньому розділі було з'ясовано, що для представлення результатів медичного дослідження використовуються окремі характеристики розподілу аналізованої змінної, отримані на генеральній сукупності досліджуваних об'єктів. Однак на практиці в будь-якому дослідженні ми маємо справу тільки з невеликою часткою досліджуваної генеральної сукупності об'єктів – вибіркою. Тому розрахувати точне значення цих характеристик неможливо, і за результатами експерименту нам потрібно дати (використовуючи методи біостатистики) приблизну їх оцінку. При цьому використовуються два типи цих наближень: точкова оцінка параметрів та інтервальна оцінка параметрів.

4.1. ПРЕДСТАВЛЕННЯ КІЛЬКІСНИХ ДАНИХ. ТЕОРЕТИЧНІ ВІДОМОСТІ. ТОЧКОВА ОЦІНКА

Для представлення результатів дослідження у випадку кількісних даних використовуються показники M та σ ; у випадку нормального за-

кону – розподілу або M_e , Q_1 і Q_{III} у випадку закону розподілу, відмінного від нормального. Розглянемо, як у цьому випадку на вибірці можуть бути розраховані приблизні значення вказаних показників.

4.1.1. Точкова оцінка міри положення кількісної змінної

Нормальний закон розподілу

При аналізі оцінки міри положення кількісної змінної, отриманій на випадковій (рандомізованій) вибірці із її генеральної сукупності, було показано, що у випадку нормального розподілу величини найкращим (ефективним і незміщеним) наближеним значенням математичного сподівання M буде середнє арифметичне вибірових величин [6] – вибірове середнє:

$$\bar{X} = \frac{1}{n} \times \sum_{i=1}^n x_i, \quad (4.1)$$

де x_i – i -е ($i = 1, \dots, n$) значення показника у вибірці, n – кількість значень (*об'єм вибірки*). Величина \bar{X} називається точковою оцінкою середнього значення (центру розподілу).

Закон розподілу відмінний від нормального

У випадку закону розподілу, відмінного від нормального, найкращою оцінкою центру розподілу буде вибіркова медіана Me . Вибіркова медіана ϵ (у випадку непарного n) значення $(n+1)/2$ варіанту впорядкованого ряду результатів експерименту (середина варіаційного ряду, лівіше і правіше якого буде однакове число членів ряду). У випадку парного n вибіркова медіана дорівнює напівсумі $n/2-1$ -го і $n/2+1$ -го членів ранжованого варіаційного ряду значень результатів експерименту.

4.1.2. Точкова оцінка міри розсіювання кількісної змінної**Нормальний закон розподілу**

При аналізі оцінки міри розсіювання кількісної змінної, отриманій на випадковій (рандомізованій) вибірці із її генеральної сукупності, було показано, що у випадку нормального розподілу величини найкращим (ефективним і незміщеним) наближенням значенням дисперсії розподілу V буде вибіркова (виправлена) дисперсія [6] s^2 :

$$s^2 = \frac{1}{n-1} \times \sum_{i=1}^n (x_i - \bar{X})^2, \quad (4.2)$$

де x_i – i -е ($i=1, \dots, n$) значення показника у вибірці, n – кількість значень (*об'єм вибірки*), \bar{X} – середнє арифметичне вибірки.

Закон розподіл, у відмінний від нормального

У випадку закону розподілу, відмінного від нормального, найкращою точковою оцінкою розсіювання будуть вибіркові квартильні оцінки Q_I та Q_{III} .

4.1.3. Стандартна похибка оцінки середнього кількісної змінної

Слід зазначити, що часто для представлення точкової оцінки середнього значення кількісної величини наводиться не тільки значення, а й степінь точності цієї оцінки – стандартна похибка m (Standard Error) оцінки середнього, що розраховується за формулою:

$$m = \frac{s}{\sqrt{n}}, \quad (4.3)$$

стандартна похибка середнього тим менша, чим менша варіабельність показника і чим більший об'єм вибірки.

4.2. ПРЕДСТАВЛЕННЯ КІЛЬКІСНИХ ДАНИХ. ТЕОРЕТИЧНІ ВІДОМОСТІ. ІНТЕРВАЛЬНА ОЦІНКА

Точкова оцінка параметрів розподілу аналізованої змінної розраховується на вибірці та дає наближене значення оцінюваного параметра генеральної сукупності. Для більш повного представлення аналізованої величини розраховується інтервальна (вірогідний інтервал, VI – Confidence Interval, CI) оцінка цієї величини.

Вірогідний інтервал задає межі, у яких із заданою ймовірністю знаходиться точне (на генеральній сукупності) значення оцінюваного параметра.

4.2.1. Інтервальна оцінка міри положення кількісної змінної

При представленні результатів медичного дослідження у випадку кількісних змінних при представленні інтервальної оцінки, як правило, розраховується тільки оцінка міри положення (матема-

тичного сподівання, у випадку нормального закону розподілу або медіани, у випадку відмінності закону розподілу від нормального). Довірча ймовірність, для якої розраховується інтервальна оцінка у медичних дослідженнях, як правило, дорівнює 0.95, розраховується 95% VI (95% CI). При цьому потрібно зазначити, що ширина VI зменшується із зменшенням степені варіабельності показника, із збільшенням об'єму вибірки. Ширина VI збільшується із зростанням довіри до оцінки (зростанням довірчої ймовірності).

Нормальний закон розподілу

На Рис. 4.1 представлено інтервальну оцінку середнього значення (математичного сподівання M) величини X , яка проводилась на основі результатів дослідження на деякій рандомізованій вибірці об'єму n .

На рисунку представлено положення на числовій осі середнього арифметичного \bar{X} , отриманого на вибірці (точкова оцінка), ліва (нижня) $X_{\text{ліва}}$ та права (верхня) $X_{\text{права}}$ границі 95% VI .

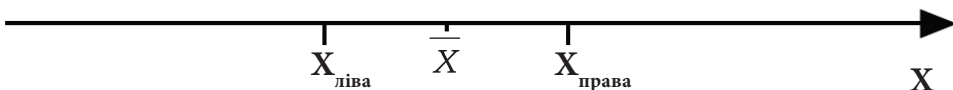


Рис. 4.1. Інтервальна оцінка математичного сподівання (95% VI) величини X , що підпорядковується нормальному закону розподілу.

Можна дати наступні інтерпретації 95% ВІ:

- 1) середнє значення величини X (математичне сподівання) за результатами дослідження приблизно дорівнює \bar{X} і можна стверджувати із 95% впевненістю, що якщо провести безкінечну кількість вимірів (вся генеральна сукупність), то отримане точне середнє значення M не буде менше значення $X_{\text{ліва}}$, але й не буде більше значення $X_{\text{права}}$ ($X_{\text{ліва}} \leq M \leq X_{\text{права}}$);
- 2) якщо провести дуже багато експериментальних досліджень для величин X і у кожній вибірці оцінити середнє арифметичне значення, то 95% усіх середніх значень буде належати 95% ВІ.

Звернімо увагу, що у випадку нормального закону розподілу вірогідний інтервал оцінки середнього є симетричним (див. Рис. 4.1): $\bar{X} - X_{\text{ліва}} = X_{\text{права}} - \bar{X}$.

Закон розподілу, відмінний від нормального

На Рис. 4.2 представлено інтервальну оцінку центру розподілу

(Me) величини Y , яка проводилась на основі результатів дослідження на деякій рандомізованій вибірці об'єму n .

На рисунку представлено положення на числовій осі медіанного значення $Me_{\text{виб.}}$, отриманого на вибірці (точкова оцінка), ліва (нижня) $Y_{\text{ліва}}$ та права (верхня) $Y_{\text{права}}$ границі 95% ВІ.

Можна дати наступні інтерпретації 95% ВІ:

- 1) центр розподілу величини Y (медіана) за результатами дослідження приблизно дорівнює $Me_{\text{виб.}}$, можна стверджувати із 95% впевненістю, що якщо провести безкінечну кількість вимірів (вся генеральна сукупність), то отримане точне середнє значення, Me , не буде менше значення $Y_{\text{ліва}}$, але й не буде більше значення $Y_{\text{права}}$ ($Y_{\text{ліва}} \leq Me \leq Y_{\text{права}}$);
- 2) якщо провести дуже багато експериментальних досліджень і у кожній вибірці оцінити медіанне значення, то 95% всіх вибіркових медіан буде належати 95% ВІ.



Рис. 4.2. Інтервальна оцінка медіани (95% ВІ) величини Y , що не підпорядковується нормальному закону розподілу.

Звернімо увагу, що у випадку відмінності закону розподілу від нормального вірогідний інтервал оцінки Me не обов'язково буде симетричним відносно вибіркової медіани (див. Рис. 4.2).

4.3. ПРЕДСТАВЛЕННЯ КІЛЬКІСНИХ ДАНИХ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

4.3.1. Приклад 1. Нормальний закон розподілу

Дані

У таблиці 1 додатку (Додаток. Приклад 1) наведені результати вимірювання артеріального тиску: значення систолічного (SBP) та діастолічного (DBP) тиску, – хворих, що поступили до реанімаційного відділення. Хворі випадковим чином були розподілені на дві групи, для яких були використані різні підходи реанімаційних засобів: Група 1 і Група 2. Необхідно представити дані з артеріального тиску хворих на момент надходження до відділення.

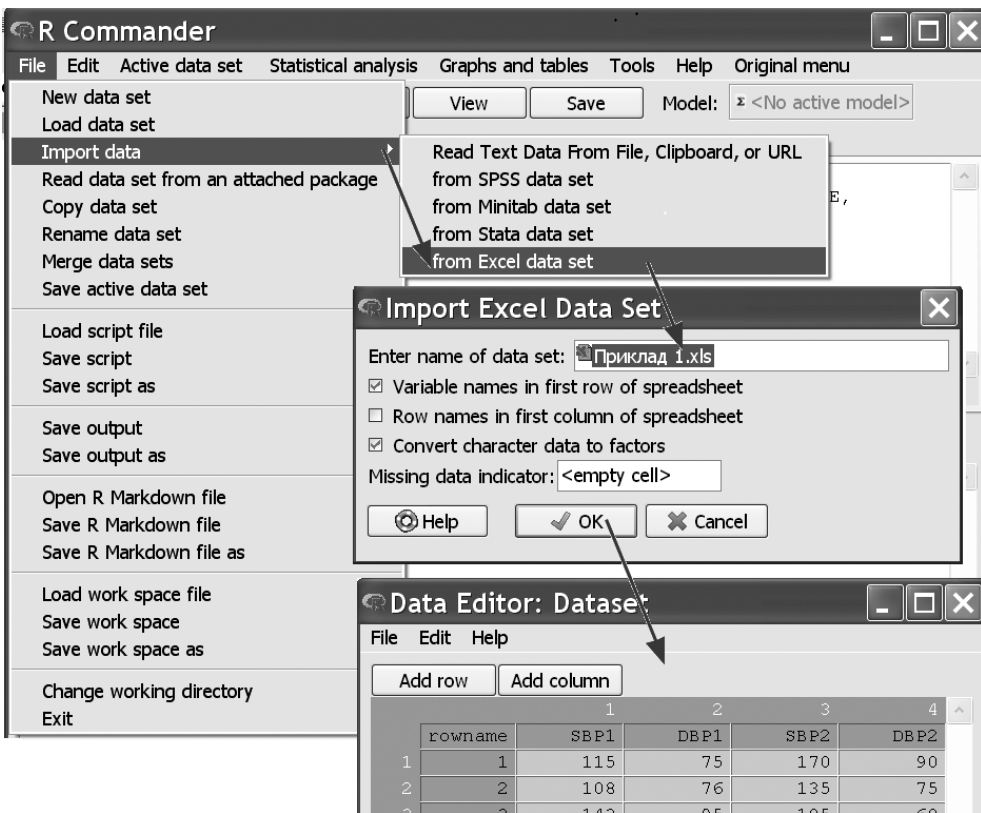


Рис. 4.3. Завантаження даних з «Приклад 1.xls» у пакет EZR.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 1.xls».

Завантажимо файл з даними (Рис. 4.3).

Для представлення кількісних даних потрібно визначитися з видом їх розподілу, необхідно перевірити розподіл змінних на нормальність, для чого використовуємо критерій Shapiro-Wilk. Нижче наведено послідовність вибору пунктів меню пакету EZR, що необхідно вибрати для про-

ведення аналізу (так само наводиться для всіх наступних дій).

EZR (R-statistics)

Original menu →

Statistics →

Summaries →

Shapiro-Wilk test of normality →

Обрати показник для аналізу (DBP1)

Обираючи послідовно змінні (DBP1, DBP2, SBP1, SBP2), отримуємо результат (дивись Рис. 4.4). За результатами перевірки визначимо, що для всіх показників, які

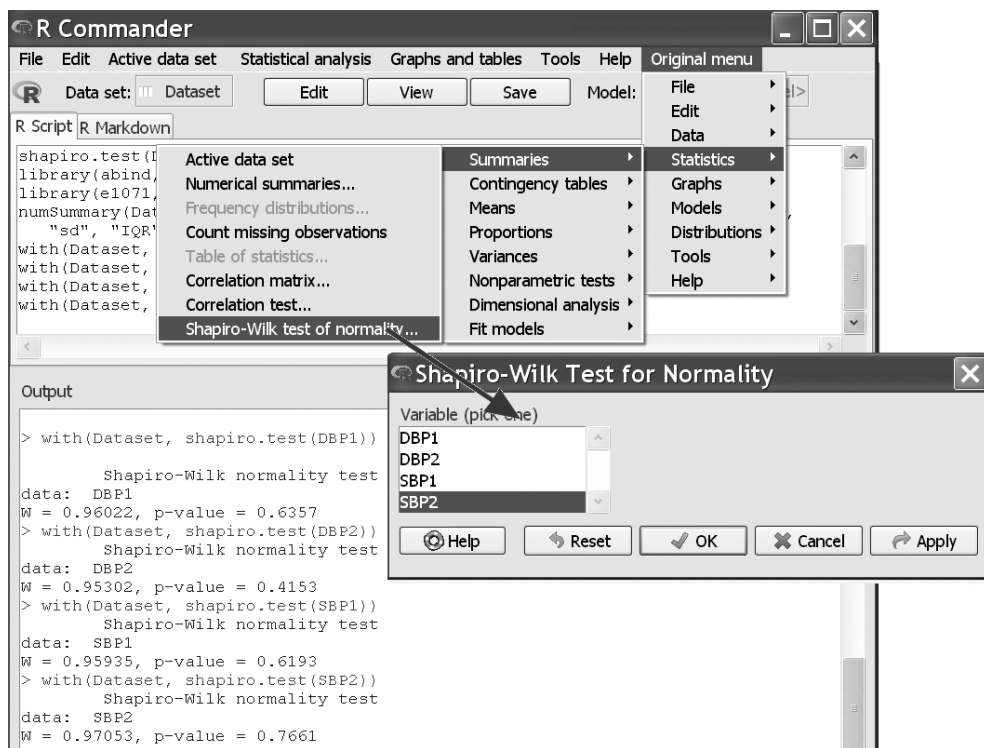


Рис. 4.4. Перевірка даних з «Приклад 1.xls» на нормальність у пакеті EZR.

підлягають аналізу, не виявлено відмінності закону розподілу від нормального (див. Output Рис. 4.4), у всіх випадках рівень значущості $p > 0.05$ ($p = 0.636$ для змінної DBP1, $p = 0.415$ для змінної DBP2, $p = 0.619$ для змінної SBP1, $p = 0.636$ для змінної SBP2).

Таким чином, для представлення точкової оцінки даних потрібно розрахувати вибіркове середнє значення величин (стандартну

похибку SE) та вибіркове стандартне відхилення SD.

EZR (R-statistics)

Original menu →

Statistics →

Summaries →

Numerical Summaries →

На закладинці Data: Обрати всі змінні для аналізу

На закладинці Statistics: Обрати числові характеристики для розрахунку

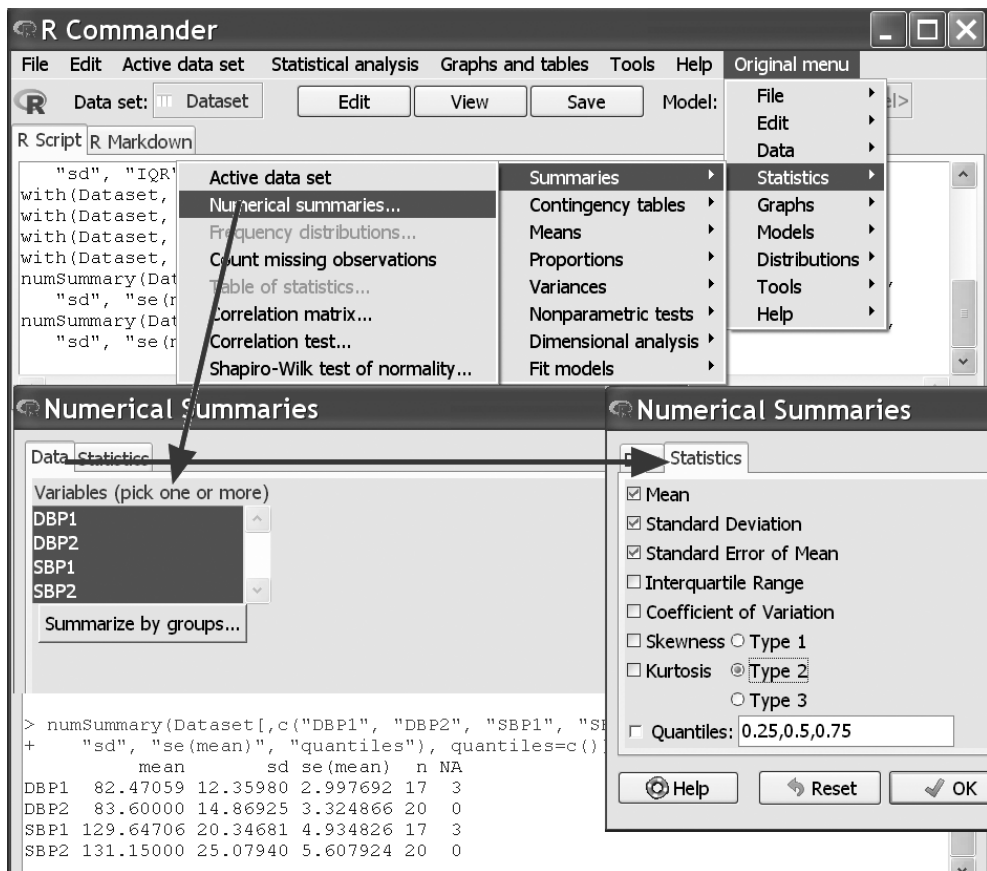


Рис. 4.5. Точкова оцінка даних з «Приклад 1.xls» у пакеті EZR.

На рисунку 4.5 представлені результати розрахунку. Отримані результати можна записати у вигляді таблиці, залишилося тільки питання, з якою точністю (скільки знаків отриманого числа залишати) ці результати представляти.

Звернімо увагу, що за правилами округлення результатів при представленні стандартної похибки (SE) залишається одна (максимум дві) значимих цифри (значимими цифрами величини називаються всі цифри, починаючи від першої, відмінної від нуля). При пред-

Plot of Means

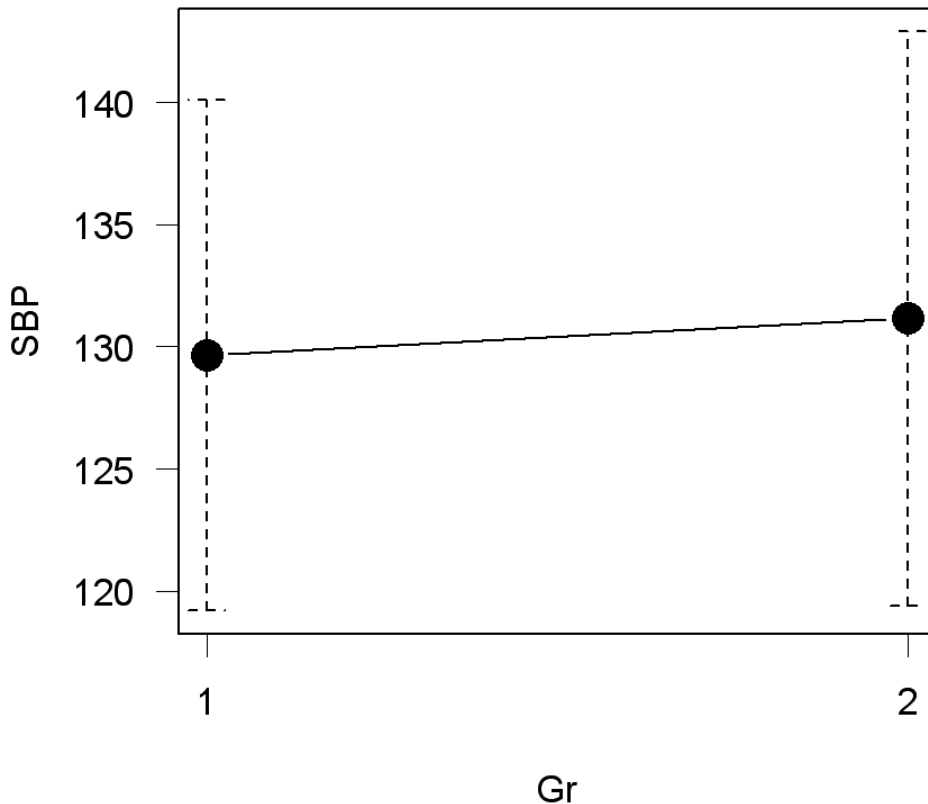


Рис. 4.6. Інтервальна оцінка показника SBP у двох групах з «Приклад 1.xls» у пакеті EZR. Представлено середнє значення показника (коло) та його 95% ВІ («вуса» розподілу).

ставленні ж середнього значення зберігається стільки ж десяткових знаків після коми, скільки їх збережено при представленні SE.

У разі необхідності в пакеті EZR може бути розраховано та представлено (див. Рис. 4.6) й інтервальну оцінку середнього значення показника.

З інтервальної оцінки середнього значення показника систолічного тиску можна записати $\bar{X}_{\text{SBP1}} = 129.6$ мм рт.ст. (95% ВІ 119.2 мм рт.ст. ÷ 140.0 мм рт.ст.) та $\bar{X}_{\text{SBP2}} = 131.2$ мм рт.ст. (95% ВІ 119.5 мм рт.ст. ÷ 142.9 мм рт.ст.). Очевидно, що не можна говорити про відмінність середніх значень (на генеральних сукупностях) цих показників у двох групах пацієнтів (інтервали, у яких лежать точні середні значення SBP у двох групах, практично повністю перекриваються), що є наслідком випадкового вибору пацієнтів у групи з єдиної сукупності хворих

(однієї і тієї ж генеральної сукупності).

Результат

У таблиці 4.1 представлено значення показників артеріального тиску для двох груп хворих на момент надходження до відділення. Закон розподілу показників не відрізнявся від нормального ($p > 0.05$), тому розраховано їхнє середнє значення \bar{X} та стандартне відхилення SD.

Звернімо увагу, що знак \pm використовується в таблиці лише з метою відділення величини середнього значення від величини стандартного відхилення. У деяких посібниках [7] рекомендується використовувати дужки і представляти результати у форматі $\bar{X}(SD)$, наприклад – 129.6 (20.3) мм рт.ст. Таке представлення доволі часто використовується в публікаціях у закордонних (англомовних) виданнях.

Таблиця 4.1.
Значення артеріального тиску в групах дослідження

Показник	$\bar{X} \pm SD$	
	Gr1 (n=17)	Gr2 (n=20)
Систолічний тиск, мм рт.ст.	129.6±20.3	131.2±25.1
Діастолічний тиск, мм рт.ст.	82.5±12.4	83.6±14.9

4.3.2. Приклад 2. Закон розподілу, відмінний від нормального

Дані

У таблиці 2 додатку (Додаток. Приклад 2) наведені результати вимірювання показників: рівень сечовини (U) та альфа-амілаза (A.A), – хворих, що поступили до реанімаційного відділення. Необхідно представити узагальнені показники на момент надходження до відділення.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 2.xls». Завантажимо файл з даними.

Для представлення кількісних даних спочатку визначимося з видом їхнього розподілу, для чого використаємо критерій Shapiro-Wilk.

EZR (R-statistics)

Original menu →

Statistics →

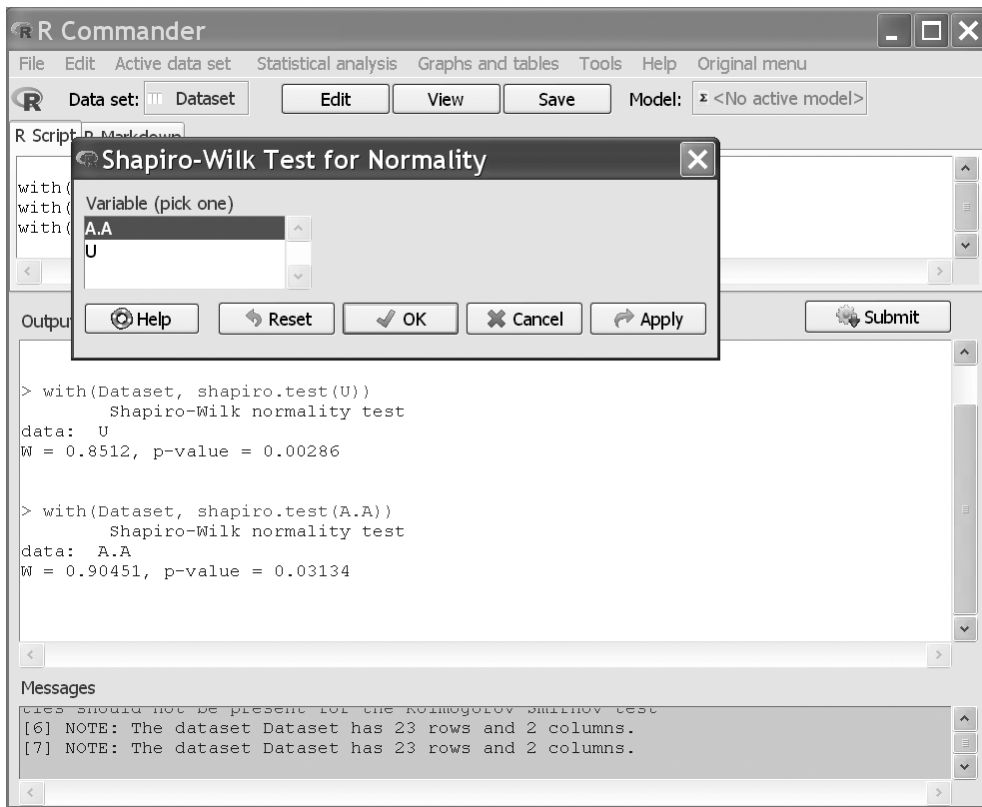


Рис. 4.7. Результати перевірки даних з «Приклад 2.xls» на нормальність у пакеті EZR.

Summaries→

Shapiro-Wilk test of normality→
Обрати показник для аналізу (U)

Обираючи послідовно змінні (U, A.A), отримаємо (Рис. 4.7). За результатами перевірки визначимо, що для показників виявлено відмінності закону розподілу від нормального (див. Output Рис. 4.7), у всіх випадках рівень значимості $p < 0.05$ ($p = 0.003$ для змінної U, $p = 0.031$ для змінної A.A).

Таким чином, для представлення точкової оцінки даних потрібно розрахувати вибіркиму медіану Me, перший (Q_1) та третій (Q_3) квартилі.

EZR (R-statistics)

Original menu→

Statistics→

Summaries→

Numerical Summaries→

На закладинці Data: Обрати всі змінні для аналізу

На закладинці Statistics: Обра-

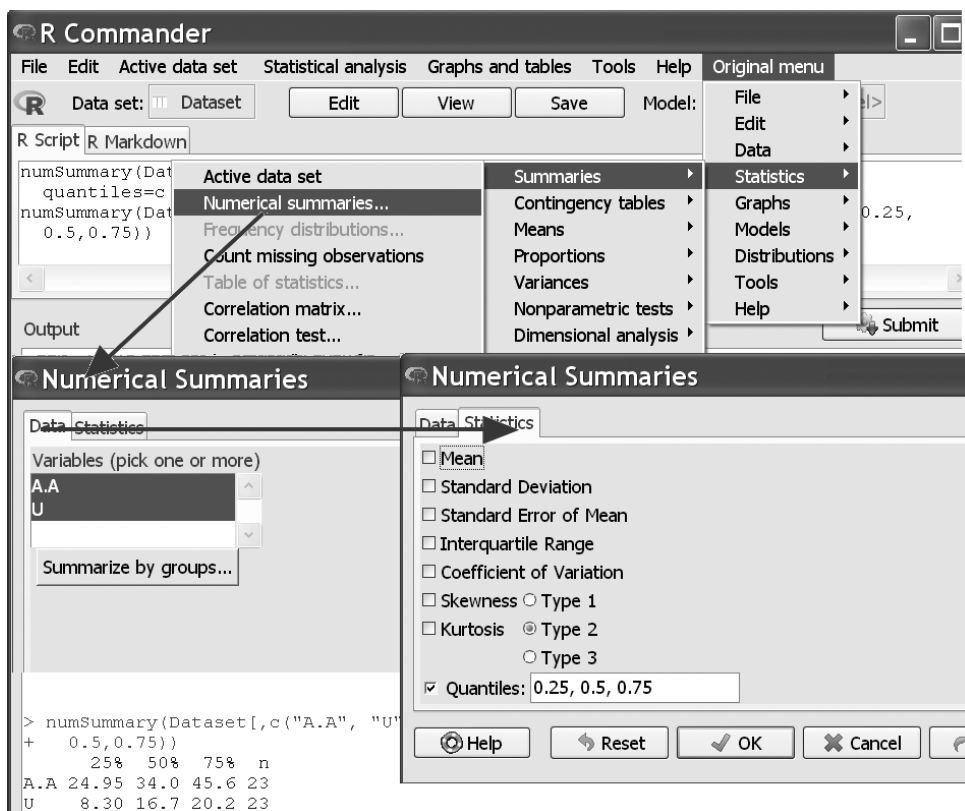


Рис. 4.8. Точкова оцінка даних з «Приклад 2.xls» у пакеті EZR.

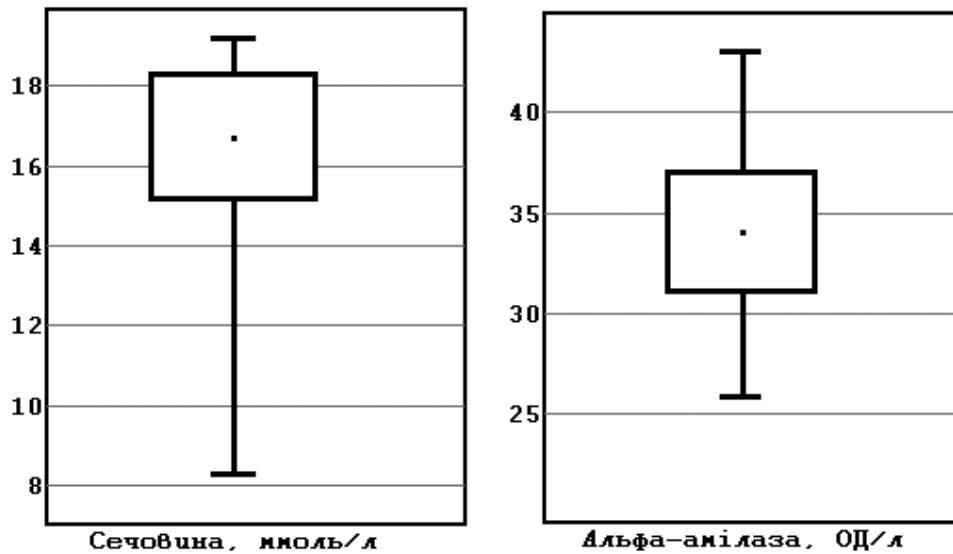


Рис. 4.9. Інтервальна оцінка показників крові хворих, що поступили до реанімаційного відділення (у пакеті MedStat). Представлено медіану (точка), похибку медіани (прямокутник) та 95% ВІ (вуса розподілу).

ти числові характеристики для розрахунку (обираємо проценти-лі 25%, 50% та 75%)

На рисунку 4.8 представлені результати розрахунку.

У разі необхідності може бути розрахована та представлена (див. Рис. 4.9) й інтервальна оцінка медіанного значення показників крові.

У пакеті EZR v.1.32 відсутня можливість такого розрахунку, тому ми навели результати, отримані в пакеті MedStat. Звернемо увагу на суттєву асиметрію інтервальної оцінки медіани у випадку закону розподілу, відмінного від нормального: $Me_U = 16.7$ ммоль/л (95% ВІ 8.3 ммоль/л ÷ 19.2 ммоль/л), $Me_{AA} = 34.0$ ОД/л (95% ВІ 24.95 ОД/л ÷ 45.6 ОД/л).

Таблиця 4.2.

Значення показників крові у хворих на момент надходження до відділення реанімації.

Показник	Me ($Q_I \div Q_{III}$) (n=23)
Сечовина, ммоль/л	16.7 (8.3 ÷ 20.2)
Альфа-амілаза, ОД/л	34.0 (24.95 ÷ 45.6)

Результат

У таблиці 4.2 представлено значення показників крові для хворих на момент надходження у відділення. Закон розподілу по-

казників відрізнявся від нормального ($p < 0.05$), тому розраховано їхнє медіанне значення Me та квартильна оцінка показників ($Q_I \div Q_{III}$).

РОЗДІЛ 5. ВИБІРКОВА СУКУПНІТЬ. ЯКІСНІ ДАНІ

У розділі 3 ми говорили, що для представлення якісних змінних розраховується частота, з якою зустрічається відповідне значення показника. Однак, як і у випадку кількісних змінних, на практиці у будь-якому дослідженні ми маємо справу з вибіркою, тому за результатами експерименту можемо дати приблизну оцінку цієї частоти: точкову або інтервальна оцінки частки.

5.1. ПРЕДСТАВЛЕННЯ ЯКІСНИХ ДАНИХ. ТЕОРЕТИЧНІ ВІДОМОСТІ. ТОЧКОВА ОЦІНКА

5.1.1. Точкова оцінка частоти для якісної змінної

Дихотомічна змінна

Достатньо часто якісна змінна, що аналізується в медичних дослідженнях, може мати лише два значення (чоловік/жінка, є побічні ефекти/відсутні й т. і.).

Нехай проводилося дослідження, у якому аналізувалася частота прояву у хворих деякого показника Q (змінна приймає значення так/ні). Нехай із n (об'єм вибірки) хворих Q виявлено у m випадках, тоді найкращим наближеним значенням частоти прояву ознаки на генеральній сукупності буде частота, отримана на вибірці – f (вибіркова частота):

$$f = \frac{m}{n} \times 100\%, \quad (5.1)$$

Величина f називається точковою оцінкою частоти прояву ознаки.

Гіпергеометричний закон розподілу

Якщо якісна змінна є номінальною (або порядковою, що приймає невелику кількість значень), вона підпорядковується гіпергеометричному закону розподілу.

Нехай проводилася дослідження, у якому аналізувалася частота

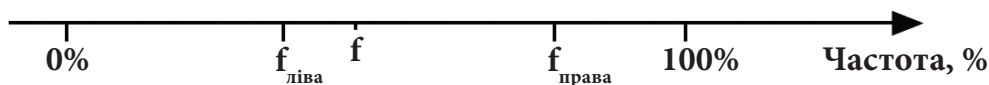


Рис. 5.1. Інтервальна оцінка частоти (95% ВІ) прояву Q (змінна приймає значення так/ні).

прояву у хворих деякого показника G (змінна приймає k різних значень). Нехай із n (об'єм вибірки) хворих у m_i виявлено i -ий ($i=1, \dots, k$) рівень показника G . У цьому випадку точковою оцінкою показника G буде сукупність значень частоти прояву ознаки, отримана на вибірці – f_i (вибіркова частота):

$$f = \frac{m}{n} \times 100\%, \quad (5.2)$$

де m_i – кількість хворих у вибірці із зазначеним i -им ($i= 1, \dots, k$) рів-

нем показника G , n – об'єм вибірки. Зазначимо, що $\sum_{i=1}^k m_i = n, \sum_{i=1}^k f_i = 100\%$.

5.1.2. Стандартна похибка частоти якісної змінної

Як і у випадку кількісної змінної, для частоти прояву ознаки може бути розрахований степінь її точності (на вибірці об'єму n) – стандартна похибка m оцінки частоти прояву дихотомічної ознаки, що розраховується за формулою:

The screenshot shows the R Commander interface. A dialog box titled "Enter and analyze two-way..." is open, allowing the user to enter counts for a 2x4 contingency table. The dialog includes options for computing percentages (Row, Column, or Total) and hypothesis tests (Chi-square with continuity correction, Fisher's exact test, etc.). The console window below shows the resulting R output, including the counts table and column percentages.

	A.	Д.	К.	С.
1	168	141	739	249
2	12591	6504	16762	12933

	A.	Д.	К.	С.
1	1.3	2.1	4.2	1.9
2	98.7	97.9	95.8	98.1
Total	100.0	100.0	100.0	100.0

Рис. 5.2. Точкова оцінка даних для «Приклад 3» у пакеті EZR.

$$m = \sqrt{\frac{f \times (100\% - f)}{n}}, \quad (5.3)$$

знову ж стандартна похибка частоти тим менша, чим більший об'єм вибірки.

5.2. ПРЕДСТАВЛЕННЯ ЯКІСНИХ ДАНИХ. ТЕОРЕТИЧНІ ВІДОМОСТІ. ІНТЕРВАЛЬНА ОЦІНКА ДЛЯ ДИХОТОМІЧНОЇ ЗМІННОЇ

Точкова оцінка частоти P прояву ознаки Q (змінна може мати лише два значення) параметрів розподілу аналізованої змінної розраховується на вибірці і дає наближене значення частоти – f . У випадку, коли це має важливе практичне значення, розраховується інтервальна оцінка цієї величина.

Вірогідний інтервал задає межі, у яких із заданою ймовірністю знаходиться точне значення P час-

тоти прояву оцінюваного параметра на генеральній сукупності.

На Рис. 5.1 представлено інтервальну оцінку частоти прояву величини Q (змінна приймає значення так/ні), яка проводилась на основі результатів дослідження на деякій рандомізованій вибірці об'єму n .

На рисунку представлено положення на осі частот частоти f прояву Q , отриманого на вибірці (точкова оцінка), ліва (нижня) $f_{\text{ліва}}$ та права (верхня) $f_{\text{права}}$ границі 95% ВІ.

Можна дати наступні інтерпретації 95% ВІ:

1) частота прояву ознаки Q за результатами дослідження приблизно дорівнює f , можна стверджувати із 95% впевненістю, що якщо провести безкінечну кількість вимірів (вся генеральна сукупність), то точне значення частоти P не буде менше значення $f_{\text{ліва}}$, але й не

Таблиця 5.1.

Частота прояву генетичної патології у містах D-ої області.

Місто	Загальна кількість новонароджених	Кількість новонароджених з генетичною патологією, абс. (%±m%)
А.	12759	168 (1.3±0.1)
Д.	6504	141 (2.1±0.2)
К.	16762	739 (4.2±0.2)
С.	12933	249 (1.9±0.1)

буде більше значення $f_{права}$ ($f_{ліва} \leq P \leq f_{права}$);

2) якщо провести дуже багато експериментальних досліджень і в кожній вибірці оцінити частоту прояву ознаки Q , то у 95% усіх вибірок вона буде належати 95% ВІ.

Звернімо увагу, що для частоти (закон розподілу відрізняється від нормального) вірогідний інтервал оцінки P не обов'язково буде симетричним відносно вибіркової частоти (див. Рис. 5.1).

5.3. ПРЕДСТАВЛЕННЯ ЯКІСНИХ ДАНИХ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

5.3.1. Приклад 3. Дихотомічна змінна

Дані

У таблиці 3 додатку (Додаток. Приклад 3) наведено результати спостереження прояву деякої генетичної патології в новонароджених для чотирьох міст D-ої області України. Спостереження

The screenshot shows the R Commander interface. The 'Statistical analysis' menu is open, highlighting 'Confidence interval for a proportion'. A dialog box titled 'Confidence interval for a p...' is displayed in the foreground, with the following fields:

- Total number of samples: 12759
- Number of events: 168
- Confidence interval: 95

The background shows the R console with the following output:

```
> #####Confidence interval for a proportion#####
> prop.conf(168, 12759, 95)
[1] Probability : 0.013
[1] 95% confidence interval : 0.011 - 0.015
> #####Confidence interval for a proportion#####
> prop.conf(141, 6504, 95)
[1] Probability : 0.021
[1] 95% confidence interval : 0.018 - 0.025
> #####Confidence interval for a proportion#####
> prop.conf(739, 17501, 95)
[1] Probability : 0.042
[1] 95% confidence interval : 0.039 - 0.045
> #####Confidence interval for a proportion#####
> prop.conf(249, 13182, 95)
[1] Probability : 0.019
[1] 95% confidence interval : 0.017 - 0.021
```

Рис. 5.3. Інтервальна оцінка даних для «Приклад 3» у пакеті EZR.

велися на протязі 10-и років. Дати оцінку частоти прояву патології в цих містах.

Розв'язання в пакеті EZR (R-statistics)

Для представлення точкової оцінки якісних даних потрібно розрахувати частоту прояву патології (%) та стандартну похибку (m%) цієї частоти.

EZR (R-statistics)

Statistical analysis→

Discrete variables→

Enter and analyze two-way table→

Створити таблицю 2x4

Ввести дані спостережень у відповідні колонки

Вибрати необхідні характеристики розрахунку

На рисунку 5.2 представлені результати розрахунку частоти генетичної патології для кожного міста.

Зважимо, що при розрахунку частот не виведено степінь точності результатів (m), однак ця величина може бути легко розрахована для кожної величини за формулою (5.3):

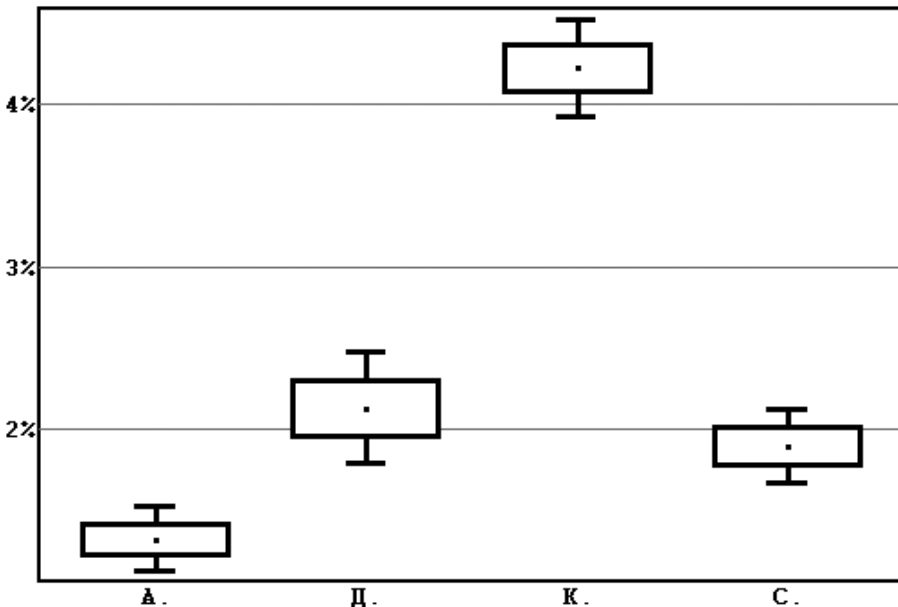


Рис. 5.4. Інтервальна оцінка частоти патології в містах D-ої області (у пакеті MedStat). Представлена частота патології (точка), стандартна похибка (прямокутник) та 95% ВІ (вуса розподілу).

$$m_{A.} = \sqrt{\frac{1.3 \times 98.7}{12759}} = 0.10(\%),$$

$$m_{D.} = \sqrt{\frac{2.1 \times 97.9}{6645}} = 0.18(\%),$$

$$m_{K.} = \sqrt{\frac{4.2 \times 95.8}{17501}} = 0.15(\%),$$

$$m_{C.} = \sqrt{\frac{1.9 \times 98.1}{13182}} = 0.12(\%).$$

Отримані результати можуть бути представлені у таблиці.

У цьому випадку представляється досить важлива величина, тому розрахуємо й інтервальну оцінку (95% ВІ) частоти (імовірності) патології новонароджених для кожного міста.

EZR (R-statistics)

Statistical analysis→

Discrete variables→

Confidence interval for proportion→

Ввести дані для міста (А.)

Далі проведемо аналогічно розрахунки для міст (Д., К., С.) і отримаємо результат (див. Рис. 5.3).

Результат

У таблиці 5.1 представлено час-

тоту генетичної патології новонароджених у містах D-ої області.

Із аналізу інтервальної оцінки імовірності патології (див. Рис. 5.4) слід звернути увагу на проблеми в місті К., де ймовірність патології новонародженого складає 4,2% (95% ВІ 3.9%÷4.5%). У цьому місті навіть нижня границя (точне значення частоти патології не менше 3.9%) перевищує верхню границю оцінки ймовірності генетичної патології новонародженого для міста А. – 1.3% (95% ВІ 1.1%÷1.5%), міста Д. – 2.1% (95% ВІ 1.8%÷2.5%), міста С. – 1.9% (95% ВІ 1.7%÷2.1%), тобто різниця значно перевищує похибку вибірки.

Слід зауважити, що у закордонних (англомовних) публікаціях при представленні якісних даних величину стандартної похибки ($\pm t\%$) у таблицях, як правило, не наводять. Ми пропонуємо все ж таки наводити (у вітчизняних публікаціях) значення похибки оцінки частоти прояву ознаки для того, щоб підкреслити, що розраховано лише наближене її значення.

РОЗДІЛ 6. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

6.1. СТАТИСТИЧНІ ГІПОТЕЗИ. ОСНОВНІ ПОНЯТТЯ

6.1.1. Статистичні гіпотези. Критерії перевірки

Одним із основних методів біостатистики, що використовується для видобування інформації з результатів медичних досліджень (визначення властивостей генеральної сукупності об'єктів дослідження з результатів, отриманих на вибірці), є перевірка статистичних гіпотез. За визначенням [6]: *«Статистична гіпотеза – це деяке твердження (припущення) щодо властивостей даної генеральної сукупності об'єктів».*

У статистиці показники, що характеризують досліджувані об'єкти, розглядаються як випадкові величини, випадкові ж величини вичерпно задаються своїм законом розподілу (функцією щільності розподілу), що описуються параметрами (числовими характеристиками) розподілу чи його формою. Таким чином, під статистичною гіпотезою можна розуміти: «... твердження щодо чис-

лових характеристик розподілу досліджуваного показника» [6].

При перевірці статистичних гіпотез формулюється *нульова гіпотеза H_0* (у біостатистиці це твердження щодо відсутності ефекту (відмінності, зв'язку ..., тощо)) та *альтернативна гіпотеза H_a* – твердження щодо наявності ефекту (відмінності, зв'язку ..., тощо), яка приймається у випадку відхилення нульової гіпотези.

Перевірка гіпотези проводиться на результатах експерименту (вибіркова сукупність) і полягає у відповіді на питання: «Чи може бути відхилена нульова гіпотеза (щодо генеральної сукупності) на основі результатів, отриманих у дослідженні (вибірка)?».

Для перевірки статистичної гіпотези використовують відповідний *критерій K_p* . Перевірка статистичної гіпотези полягає в розрахунку *значення критерію Z_n* (число, що розраховується за відповідним алгоритмом, на основі вибірки). Далі робиться висновок про можливість або неможливість відхилення нульової гіпотези:

якщо Z_n належить області допустимих значень критерію – немає підстав відхилення нульової гіпотези;

якщо Z_n належить до критичної області значень критерію – є підстави відхилення нульової гіпотези.

На рисунку 6.1 представлено приклади критичних (заштрихована область числової осі) і допустимих (незаштрихована область числової осі) областей значень деякого критерію.

Тут (Рис. 6.1) наведено приклади:

- а) односторонньої (правосторонньої) критичної області критерію;
- б) односторонньої (лівосторонньої) критичної області критерію;

- в) двосторонньої критичної області критерію.

Зауважимо, що при аналізі результатів медичних досліджень (як правило) використовуються критерії із двосторонньою критичною областю.

6.1.2. Статистичні гіпотези. Похибки при перевірці

У випадку використання методу перевірки статистичних гіпотез при прийнятті рішення щодо властивостей генеральної сукупності ми робимо висновок про наявність (H_0 відхиляється) або відсутність (H_0 не відхиляється) ефекту. І не обов'язково цей ви-

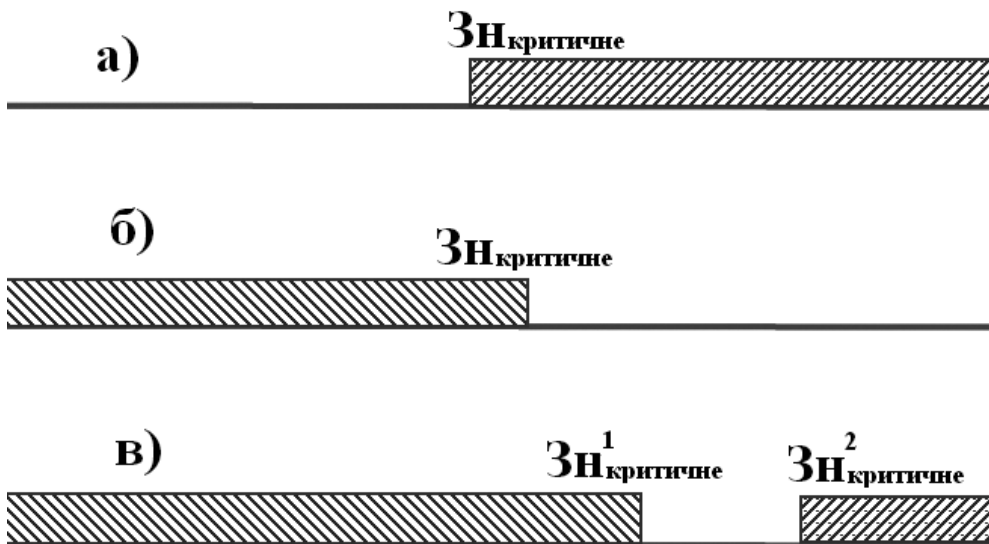


Рис. 6.1. Вигляд односторонньої а), б) та двосторонньої в) критичної області критерію.

сновок буде правильним, оскільки можна припуститися помилки. У таблиці 6.1 наведені типи можливих помилок.

До переваг методу перевірки статистичних гіпотез відноситься не тільки об'єктивізація прийняття рішення, але й можливість розрахунку ймовірності похибки при прийнятті рішення.

У випадку хибного відхилення нульової гіпотези говорять про **похибку першого роду**, ймовірність похибки першого роду позначається α і називається **рівнем значимості критерію**.

У випадку, коли нульова гіпотеза хибно не відхиляється, говорять про **похибку другого роду**, ймовірність похибки другого роду позначається β і називається **оперативною характеристикою критерію**.

Звернімо увагу, що в статистичних пакетах при перевірці статистичних гіпотез, як правило, наводиться тільки ймовірність похибки першого роду.

Величину $1-\alpha$ називають **надійністю критерію**, це ймовірність не допустити помилки першого роду (ймовірність не вказати хибно на наявність ефекту).

Величину $1-\beta$ називають **потужністю критерію**, це ймовірність не допустити помилки другого роду (ймовірність не вказати хибно на відсутність ефекту).

Очевидно, що критерій, який пропонується для перевірки деякої статистичної гіпотези, повинен мати високу надійність і потужність.

Таблиця 6.1.
Помилки при перевірці статистичних гіпотез.

Прийняте рішення	Гіпотези	
	H_0	H_a
H_0 не відхиляється	Правильне рішення	Неправильне рішення, похибка другого роду (ймовірність β)
H_0 відхиляється	Неправильне рішення, похибка першого роду (ймовірність α)	Правильне рішення

6.1.3. Статистичні гіпотези. Приклад

Чому для аналізу результатів медичних досліджень використовують метод перевірки статистичних гіпотез, які переваги він надає при прийнятті рішень? Для чого і як застосовуються критерії перевірки статистичних гіпотез? Для відповіді на ці питання розглянемо приклад.

Приклад

Нехай у нас є 100 великих коробок із чорними кульками і 100 таких самих коробок із білими кульками. Розміри всіх кульок однакові, кількість кульок у кожній коробці велика й однакова.

Дехто вибрав дві коробки (випадковим чином), зсипав їхній вміст у велику бочку, яку закоркував, залишивши тільки вузький отвір, через який можна витягати тільки по одній кульці. Цей дехто потрусив бочку, перемішав її вміст. Задача полягає в тому, щоб, витягаючи з бочки (генеральна сукупність) по одній кульці (експеримент, формує вибірку), визначити, якого ж кольору кульки зсипані в бочку (чи всі кульки одного кольору?).

Розв'язок

Витягаємо із бочки 1-у кульку, і вона білого кольору, витягаємо

з бочки 2-у кульку, і вона знову білого кольору, витягаємо 3-ю кульку, і знову біла... і т. д. Яке рішення може бути прийняте?

Розв'язок із застосуванням методу перевірки статистичних гіпотез

Сформулюємо нульову гіпотезу – немає ефекту, дехто висипав у бочку один ящик із кульками білого кольору, один із кульками чорного кольору (у бочці половина чорних і половина білих кульок). Тоді альтернативна гіпотеза – у бочці кульки тільки одного кольору. При проведенні експерименту витягаємо кульки одну за одною і дивимося на їхній колір.

Тоді статистичні гіпотези стосовно параметрів розподілу генеральної сукупності будуть записані таким чином:

- нульова гіпотеза (у бочці половина чорних і половина білих кульок) – імовірність витягнути при одній спробі білу кульку $p_B = 1/2$;
- альтернативна гіпотеза (у бочці кульки тільки білого кольору) – імовірність витягнути при одній спробі білу кульку $p_B = 1$.

Тобто:

$$H_0: p_B = 1/2$$

$$H_a: p_B = 1.$$

В якості критерію перевірки будемо рахувати кількість білих та чорних кульок у вибірці із m спроб. *Зрозуміло, якщо у вибірці є хоча б одна чорна кулька, то нульова гіпотеза справедлива.*

Розглянемо тепер результати.

Нехай ми вибрали одну кульку, і вона біла. Чи може таке бути за умови, що справедлива H_0 ? Легко порахувати, що ймовірність такої події (за умови справедливості нульової гіпотези, $p_B = 1/2$), $p = 1/2$. А чи можна в цьому випадку відкинути нульову гіпотезу? Відповідь – якщо ми відкинемо нульову гіпотезу, то ймовірність похибки першого роду складе $p = 1/2$. Це дуже велика ймовірність схибити, тому нульову гіпотезу ми не відкинемо і робимо висновок, що немає підстав говорити, що в ящику тільки білі кульки.

2) Нехай ми вибрали дві кульки, і вони білі. Ймовірність такої події (при умові $p_B = 1/2$), $p = (1/2)^2 = 0.25$. Якщо ми відкинемо нульову гіпотезу, то ймовірність похибки першого роду складе $p = 0.25$. Ця ймовірність також велика – нульову гіпотезу ми не відкинемо і робимо висновок, що немає підстав говорити, що в ящику тільки білі кульки.

Нехай тепер ми вибрали 10 кульок, і всі вони білі. Ймовірність такої події (при умові $p_B = 1/2$), $p = (1/2)^{10} \approx 0.001$. Якщо ми відкинемо нульову гіпотезу, то ймовірність похибки першого роду складе $p \approx 0.001$. Якщо вважати цю ймовірність дуже малою (якою можна знехтувати) і вважати $p \approx 0$ (подія неймовірна), то робимо висновок, що в нас відбулася неймовірна подія (10 підряд білих кульок). Але неймовірна подія відбутися не може, це означає, що наше перше твердження про справедливість нульової гіпотези (звідки випливає $p = 0$) було несправедливим – нульова гіпотеза відкидається. Робимо висновок: є підстави говорити, що у ящику тільки білі кульки. Ми зробили висновок, але залишилося питання: «Це абсолютно точна відповідь?». Статистика дає відповідь і на це питання: «На основі результату експерименту є підстави говорити, що в ящику тільки білі кульки, ймовірність того, що ми при цьому хибимо $p=0.001$ ». Ми довели, що в ящику тільки білі кульки, бо ймовірність хибного висновку дуже мала.

4) Якщо ж для деякого $p=0.001$ – велика ймовірність похибки

першого роду, будемо вимагати вибору 20 кульок, і якщо всі вони білі, то ймовірність помилки при відкиданні нульової гіпотези складе усього $p=10^{-6}$.

На цьому прикладі можна зрозуміти процес перевірки статистичних гіпотез і дати відповідь на питання, що були поставлені на початку задачі.

6.1.4. Статистичні гіпотези. Критичний рівень значимості

Наведений вище приклад показує, що для перевірки статистичних гіпотез потрібно вибрати критерій перевірки й розрахувати значення критерію (Z_n). Далі потрібно задати деяке значення ймовірності (критичне значення), і якщо ймовірність (p) отриманого і більш екстремального Z_n (яке розраховано за умови справедливості нульової гіпотези) менше критичної, вважаємо це неможливим і відхиляємо нульову гіпотезу. Якщо ж ця ймовірність не менше критичної, то підстав відхилити нульову гіпотезу ми не маємо. Критичне значення ймовірності називають критичним рівнем значимості ($\alpha_{кр.}$), ймовірність p називають досягнутим рівнем значимості.

Досягнутий рівень значимості розраховується (у статистичних

пакетах), тоді як критичний рівень значимості обирається дослідником до проведення аналізу результатів експерименту. У медичних дослідженнях, як правило, критичний рівень значимості обирають рівним 0.05 (дослідник може обрати й менший критичний рівень значимості, наприклад, $\alpha_{кр.}=0.001$), значення $\alpha_{кр.}$ обов'язково вказується. При перевірці статистичних гіпотез обов'язково вказується досягнутий рівень значимості, при якому відхиляється або не відхиляється нульова гіпотеза.

Звернімо увагу, що якщо в результаті перевірки статистичної гіпотези отримано $p < \alpha_{кр.}$, гіпотеза про відсутність ефекту відхиляється, при цьому ймовірність помилкового рішення дорівнює p . Якщо ж у результаті перевірки статистичної гіпотези отримано $p \geq \alpha_{кр.}$, гіпотеза про відсутність ефекту не відхиляється, але ймовірність помилкового рішення (ймовірність помилки другого роду) невідома. У деяких випадках аналізу результатів медичних досліджень для того, щоб прийняти рішення, необхідно знати саме β , у цьому випадку можна рекомендувати спеціальні статистичні пакети, які розраховують цю величину (наприклад, GPower [8]).

6.1.5. Статистичні гіпотези. Параметричні й непараметричні критерії

Як показано вище, для того, щоб прийняти рішення про відхилення або невідхилення нульової гіпотези, потрібно вибрати критерій перевірки. Якщо розподіл аналізованої змінної відомий (відомий вид функції щільності ймовірності розподілу), математики пропонують для перевірки використовувати так звані **параметричні критерії**. Якщо ж розподіл аналізованої змінної невідомий (невідомо функція щільності ймовірності розподілу), математики для перевірки пропонують так звані **непараметричні критерії** (можуть бути застосовані при будь-якому законі розподілу).

У практиці аналізу результатів медичних досліджень параметричні критерії, як правило, застосовуються тільки у випадку нормального закону розподілу аналізованої величини. У випадку, коли закон розподілу відрізняється від нормального (невідомий – не важливо який), використовуються загальні непараметричні критерії.

Виникає питання про необхідність існування (застосування) параметричних критеріїв (при-

датних тільки у випадку нормального закону змінних), якщо для перевірки тієї ж гіпотези можуть бути використані непараметричні (придатні у випадку будь-якого, у тому числі й нормального, закону розподілу величин). Справа в тому, що непараметричні критерії (при тому ж $\alpha_{кр.}$) мають меншу потужність, ніж параметричні. Тобто, при тій самій ймовірності помилки першого роду непараметричні критерії мають більшу ймовірність пропустити ефект там, де він насправді існує (більша ймовірність помилки другого роду).

Звідси випливає практичний висновок:

- *якщо є основи застосування параметричних критеріїв – обов'язково використовуйте їх;*
- *якщо застосування параметричних критеріїв неможливе – використовуйте тільки непараметричні критерії.*

6.2. СТАТИСТИЧНІ ГІПОТЕЗИ. СТАТИСТИЧНА ЗНАЧИМІСТЬ ПРОТИ КЛІНІЧНОЇ ЗНАЧИМОСТІ

Розглянемо приклад аналізу результатів медичного дослідження.

Приклад

Дослідника цікавило питання про вплив деякого навантаження на

частоту пульсу студентів. Було проведено вимірювання частоти пульсу в 1600 студентів до й після навантаження, отримано результати вимірювання. Дати відповідь на питання про наявність ефекту.

Розв'язок

Проаналізуємо отримані результати, критичний рівень значимості $\alpha_{кр} = 0.05$. Після перевірки розподілу отриманих даних було з'ясовано, що закон розподілу аналізованих величин не відрізнявся від нормального. Для перевірки гіпотези про наявність ефекту в цьому випадку необхідно використати критерій Стьюдента для пов'язаних вибірок. Розрахунки показали, що відмінність статистично значима (частота пульсу, у середньому, зростає), досягнутий рівень значимості $p < 0.001$.

Висновок

Встановлено, що вказане навантаження призводить до зростання частоти пульсу студентів (відмінність статистично значима, $p < 0.001$).

Який практичний висновок зробить із цього замовник аналізу? Можливо такий: «Обережніше з цим навантаженням, воно призводить до підвищення частоти пульсу студентів (це доведено точно,

імовірність помилки набагато менша 0.05)».

Давайте прискіпливіше зупинимось на отриманих результатах.

При проведенні аналізу закон розподілу не відрізнявся від нормального – результати наведемо в форматі $\bar{X} \pm SD$:

- до навантаження середня частота пульсу студентів становила $\bar{X}_{до} = 71.96 \pm 7.14$ уд./хв.;
- після навантаження середня частота пульсу студентів становила $\bar{X}_{після} = 72.06 \pm 7.14$ уд./хв.

Для перевірки гіпотези про наявність ефекту в цьому випадку використаний критерій Стьюдента для пов'язаних вибірок. Відмінність статистично значима ($p < 0.001$), частота пульсу, у середньому, зростає на 0.10 ± 0.10 уд./хв. (95% ВІ 0.09 уд./хв. $\div 0.11$ уд./хв.).

Звідси стає зрозумілим, що хоч, дійсно, після навантаження частота пульсу зростає (**відмінність є статистично значимою**), але таке (у середньому, не більше ніж на 0.11 уд./хв.) зростання ніяк не позначиться на стані студента (**відмінність не є клінічно значимою**), і немає підстав заперечувати його застосування.

Таким чином, представленням тільки висновку про статистичну значимість (незначимість) відмінності, результати аналізу медичного дослідження не можна обмежуватись. Нас цікавить клінічний аналіз отриманого результату.

Повний (правильний) висновок має бути таким:

Встановлено, що вказане навантаження призводить до статистично значимого ($p < 0.001$) зростання частоти пульсу. Указане зростання є незначним (у середньому, на 0.10 ± 0.10 уд./хв.) від 71.96 ± 7.14 уд./хв. (до навантаження) до 72.06 ± 7.14 уд./хв. (після на-

вантаження). Таке підвищення частоти пульсу не призводить до помітної (клінічно значимої) зміни стану студента.

Звідси випливають практичні висновки:

- *при проведенні аналізу результатів медичних досліджень застосування методів біостатистики є обов'язковим (необхідним);*
- *аналіз результатів медичних досліджень не може обмежуватись тільки статистикою (не достатньо), обов'язково потрібно додати клінічну інтерпретацію отриманих результатів.*

РОЗДІЛ 7. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ДВІ ГРУПИ ПОРІВНЯННЯ. КІЛЬКІСНІ ЗМІННІ

Розглянемо приклади застосування методів перевірки статистичних гіпотез для порівняння кількісної змінної у двох групах. При цьому можна розглянути дві основні задачі, які найчастіше зустрічаються при проведенні аналізу результатів медичних досліджень:

- *порівняння двох незалежних сукупностей,*
- *порівняння двох пов'язаних сукупностей.*

При розв'язанні кожної із задач застосовуються свої критерії перевірки статистичних гіпотез.

При *порівнянні двох незалежних сукупностей* є дві різні групи об'єктів (пацієнтів), що характеризуються деякою кількісною змінною. Задача, яку ставить перед собою дослідник, полягає у відповіді на питання: «Чи описуються генеральні сукупності, з яких випадковим чином вилучили ці дві вибірки одним і тим же законом розподілу (вибірки, вилучені із однієї і тієї ж генеральної сукупності)?». Як правило, питання

стосується середнього значення сукупностей: «Чи відрізняються середні генеральних сукупностей, з яких випадковим чином вилучили ці дві вибірки?».

При *порівнянні двох пов'язаних сукупностей* є два результати вимірювання показника для об'єктів (пацієнтів), пов'язаних між собою (наприклад, результати вимірювання деякого кількісного показника у тих самих пацієнтів до й після лікування). Задача, яку ставить перед собою дослідник, полягає у відповіді на питання: «Чи змінюються закони розподілу, яким описуються генеральні сукупності, з яких випадковим чином вилучили цю вибірку залежно від того, до або після лікування вилучені об'єкти?». Як правило, питання стосується зміни середнього значення сукупностей: «Чи відрізняється від нуля різниця показників до й після лікування на генеральній сукупності, з якої випадковим чином вилучили цю вибірку?».

7.1. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Нормальний закон розподілу

Нехай дві вибірки вилучені з генеральних сукупностей, що мають нормальний закон розподілу. Усі спостереження у вибірках не пов'язані попарно (незалежні вибірки). Нульова гіпотеза полягає в тому, що середні значення генеральних сукупностей не відрізняються одна від одної.

Для перевірки нульової гіпотези в цьому випадку використовується ***t*-критерій Стьюдента для незалежних вибірок**. Цей критерій має багато застосувань і виглядає доволі простим для розрахунків. Однак, насправді, при розрахунку критерію виникає доволі багато нюансів, які потрібно враховувати.

Тому для проведення коректного аналізу рекомендується використовувати спеціалізовані статистичні пакети (наприклад, не можна рекомендувати проводити розрахунки в табличному процесорі MO Excel – це не є спеціалізованим статистичним пакетом).

У цьому разі статистика критерію t (у випадку, коли дисперсії сукупностей не відрізняються), розраховується за формулою [9]:

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{SD \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

$$SD^2 = \frac{SD_1^2(n_1 - 1) + SD_2^2(n_2 - 1)}{n_1 + n_2 - 2}, \quad (7.1)$$

де \bar{X}_1 та \bar{X}_2 – вибіркові середні арифметичні; SD_1 та SD_2 – вибіркові стандартні відхилення; n_1 та n_2 – об'єми першої та другої вибірок відповідно. Розподіл значень критерію підпорядковується розподілу Стьюдента [9] із числом ступенів вільності, $v = n_1 + n_2 - 2$.

У разі різних дисперсій для порівняння використовується критерій Уелча, d -статистика критерію розраховується за формулою [9]:

$$d = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}},$$

$$v = \frac{\left(\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}\right)^2}{\frac{SD_1^2}{n_1 - 1} + \frac{SD_2^2}{n_2 - 1}}, \quad (7.2)$$

де \bar{X}_1 та \bar{X}_2 – вибіркові середні арифметичні; SD_1 та SD_2 – вибіркові стандартні відхилення; n_1 та n_2 – об'єми першої та другої вибі-

рок, відповідно. Розподіл значень d -критерію приблизно описується законом розподілу Стюдента [9] із ν ступенями вільності.

Закон розподілу відмінний від нормального

У випадку порівняння незалежних сукупностей, коли хоча б одна з них не підпорядковується нормальному закону розподілу, для порівняння центрів сукупностей використовується **критерій Вілкоксона-Манна-Уїтні** (W -статистика рангового критерію Вілкоксона і U -статистика критерію Манна-Уїтні пов'язані між собою простою формулою) [9]. Для розрахунку статистики критерію використовується спільне ранжування двох порівнюваних вибірок (у випадку однакових значень їм призначається середній ранг). Далі проводиться обчислення статистики критерію W за формулою [9]:

$$W = \min\left(\sum_{i=1}^{n_1} R_i, \sum_{i=1}^{n_2} S_i\right), \quad (7.3)$$

де R_i – ранги вибірки, що має найменшу суму рангів, $i = 1, \dots, n_1$ (n_1 – об'єм цієї вибірки), S_i – ранги другої вибірки, $i = 1, \dots, n_2$ (n_2 – об'єм другої вибірки). Якщо менша із сум рангів менше від верхнього критичного значення, відмінність між сукупностями є статистично

значимою при заданому критичному рівні значимості.

На практиці для розрахунку можуть використовуватись інші наближені апроксимації, таким чином, щоб звести розподіл статистики критерію до стандартного нормального закону розподілу [9]. Тому для проведення коректного аналізу рекомендується використовувати спеціалізовані статистичні пакети.

7.2. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

7.2.1. Приклад 4. Нормальний закон розподілу

Дані

Вивчався вплив наявності вірусного ураження на рівень цитокінів при деякому захворюванні. У таблиці 4 додатку (Додаток. Приклад 4) наведені результати вимірювання рівня ІЛ-1 для хворих без вірусного ураження (Група 1) та при наявності такого ураження (Група 2). Необхідно з'ясувати, чи відрізняється рівень ІЛ-1 у хворих цих груп. При проведенні аналізу критичний рівень значущості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Уведемо отримані результати в таблицю Excel та збережемо у файлі «Приклад 4.xls». Для проведення

	A	B	C
1	Gr	IL1	
2	Gr1	27.77	
3	Gr1	30.42	
4	Gr1	20.47	
5	Gr1	27.26	
6	Gr1	27.72	
7	Gr1	34.58	
8	Gr1	34.17	
9	Gr1	22.49	
10	Gr1	30.47	
11	Gr1	32.16	
12	Gr1	27.17	
13	Gr1	28.39	
14	Gr1	27.26	
15	Gr1	30.68	
16	Gr1	20.47	
17	Gr1	30.16	
18	Gr1	20.06	
19	Gr1	21.6	
20	Gr1	25.92	
21	Gr1	30.47	
22	Gr2	26.14	
23	Gr2	27.54	

Рис. 7.1. Формат введення даних з «Приклад 4.xls» для проведення аналізу в пакеті EZR.

аналізу введемо дані в таблицю, як це показано на рисунку 7.1.

Для проведення аналізу кількісної змінної, на перше потрібно визначитися з видом розподілу, для чого необхідно перевірити розподіл змінних на нормальність, для чого використаємо критерій Shapiro-Wilk.

EZR (R-statistics)

Statistical analysis →

Continuous variables →

Kolmogorov-Smirnov test for normal distribution →

Обрати показник для аналізу (IL1) та вказати групу, для якої проводиться аналіз (Gr1)

Обираючи перевірку на нормальність для Gr1, а далі Gr2, отримаємо (див. Рис. 7.2), що відмінності закону розподілу змінної IL-1 для обох груп від нормального не виявлено ($p=0.132$ і $p=0.561$ для Групи 1 та Групи 2, відповідно).

Закон розподілу змінних не відрізняється від нормального, тому для проведення аналізу обираємо параметричні критерії порівняння, критерій Стьюдента для порівняння незалежних вибірок.

EZR (R-statistics)

Statistical analysis →

Continuous variables →

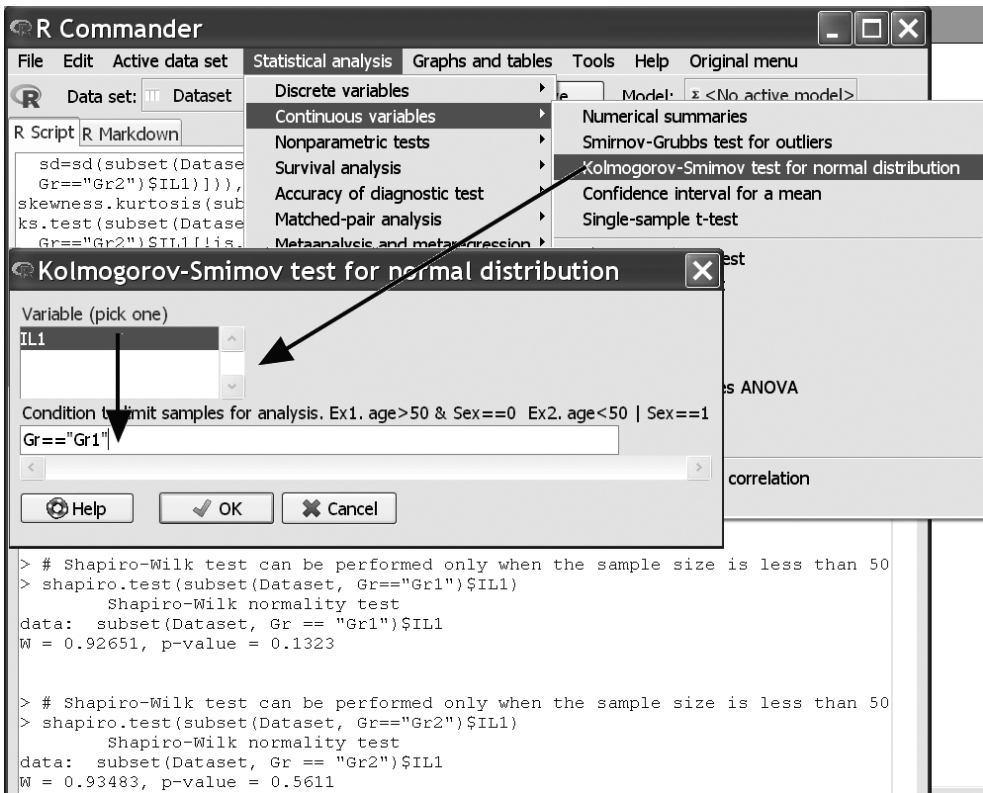


Рис. 7.2. Перевірка даних з «Приклад 4.xls» на нормальність у пакеті EZR.

Two-sample t-test →
 Обираємо змінну порівняння (IL1)
 Обираємо змінну групування (Gr)
 Обираємо двосторонню критичну область перевірки (Two-tailed)
 Обираємо надійність перевірки ($1-0.05=0.95$)
 Перевірку проводимо при умові рівних дисперсій у двох групах

Результати розрахунку представлені на рисунку 7.3. Таким чином,

отримаємо, що для Групи 1 значення рівня ІЛ-1 складає 27.5 ± 4.5 пг/мл, для Групи 2 – 23.6 ± 3.2 пг/мл. Нульова гіпотеза про рівність середніх сукупностей відкидається на рівні значимості $p=0.036$.

Результат

У таблиці 7.1 представлено значення показників рівня ІЛ-1 для двох груп хворих. Закон розподілу показників не відрізнявся від нор-

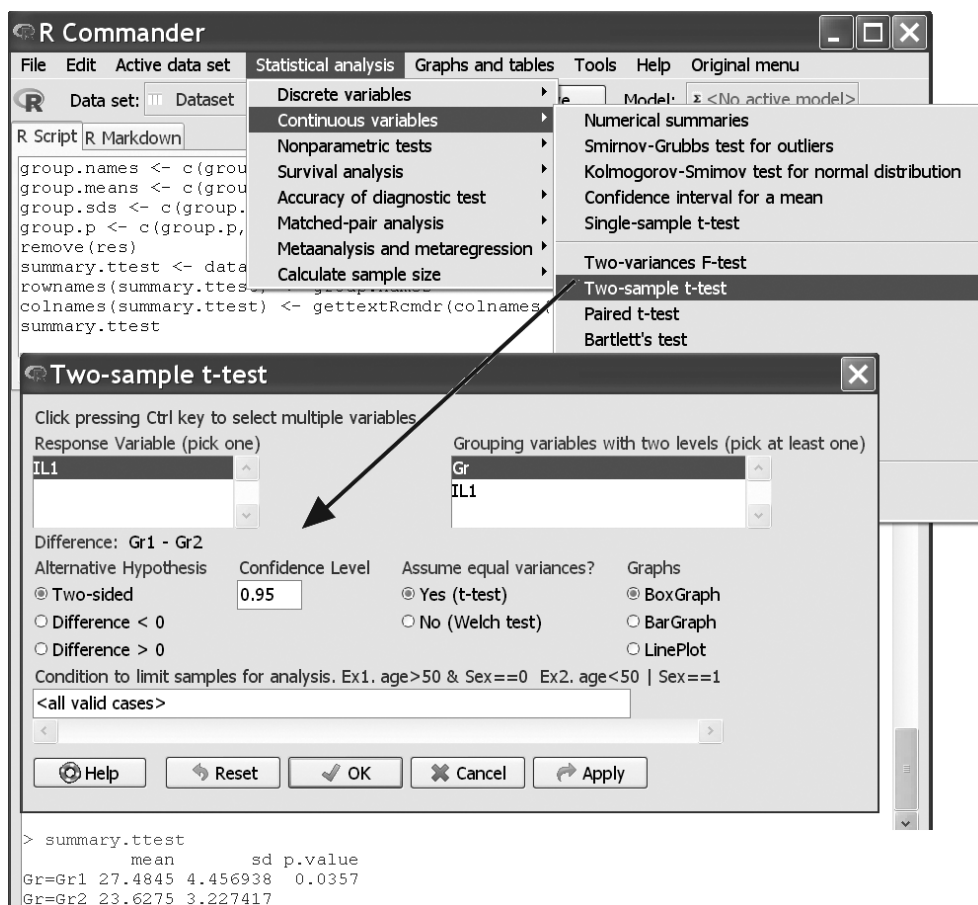


Рис. 7.3. Порівняння двох незалежних сукупностей з «Приклад 4.xls» у пакеті EZR.

мального ($p > 0.05$), тому розраховано їхнє середнє значення \bar{X} та стандартне відхилення SD .

При проведенні порівняння виявлено, що в групі хворих без вірусного ураження середнє значення

Таблиця 7.1.
Значення показників у досліджених групах хворих.

Показник	$\bar{X} \pm SD$		Рівень значимості відмінності між групами, p
	Група 1 (n=20)	Група 2 (n=8)	
ІЛ-1, пг/мл	27.5±4.5	23.6±3.2	0.036

рівня ІЛ-1 (27.5 ± 4.5 пг/мл) було статистично значиме ($p=0.036$ за критерієм Стьюдента) вище, ніж у групі хворих з вірусним ураженням (23.6 ± 3.2 пг/мл).

7.2.2. Приклад 5. Закон розподілу, відмінний від нормального

Дані

Вивчався вплив наявності вірусного ураження на рівень цитокінів при деякому захворюванні. У таблиці 5 додатку (Додаток. Приклад 5) наведені результати вимірювання рівня ІЛ-8 для хворих без вірусного ураження (Група 1) та за наявності такого ураження (Група 2). Необхідно з'ясувати, чи відрізняється рівень ІЛ-8 у хворих цих груп. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 5.xls». Для проведення аналізу введемо дані в таблицю в такому ж форматі, як і в попередньому прикладі.

Перевіримо закон розподілу змінних у двох групах на нормальність (див. Приклад 4).

Якщо для Групи 1 (див. Рис. 7.4) відмінності закону розподілу змінної ІЛ-8 від нормального не виявлено ($p=0.181$), то для Групи 2 виявлено відмінність закону розподілу від нормального ($p=0.043$).

Виявлено відмінність закону розподілу від нормального, тому для проведення аналізу обираємо непараметричні критерії порівняння, критерій Манна-Уїтні для порівняння незалежних вибірок.

EZR (R-statistics)

Statistical analysis →

Nonparametric test →

Mann-Whitney U test →

Обираємо змінну порівняння (IL8)

Обираємо змінну групування (Gr)

Обираємо двосторонню критичну область перевірки (Two-tailed)

Обираємо тип тесту (За замовчуванням)

Результати розрахунку представлені на рисунку 7.5. Нульова гіпотеза про рівність центрів сукупностей не відкидається ($p=0.481$).

Результат

У таблиці 7.2 представлено значення показників рівня ІЛ-8 для двох груп хворих. Закон розподілу показників відрізнявся від нормального ($p < 0.05$), тому розраховано медіану (Me), значення першого (Q_I) та третього (Q_{III}) квантилю.

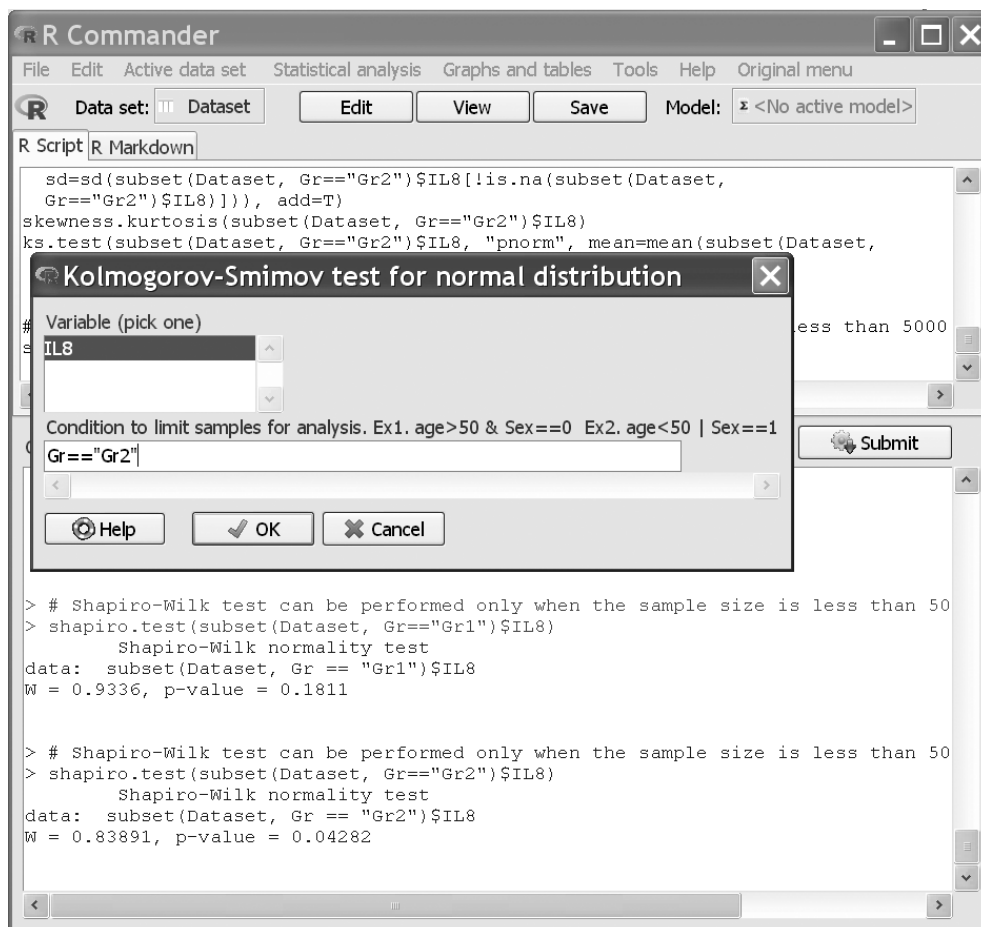


Рис. 7.4. Перевірка даних з «Приклад 5.xls» на нормальність у пакеті EZR.

Таблиця 7.2.

Значення показників у досліджених групах хворих.

Показник	Me (Q _I ÷ Q _{III})		Рівень значимості відмінності між групами, p
	Група 1 (n=20)	Група 2 (n=10)	
ІЛ-8, пг/мл	11.31 (6.53 ÷ 12.56)	10.68 (5.28 ÷ 12.02)	0.481

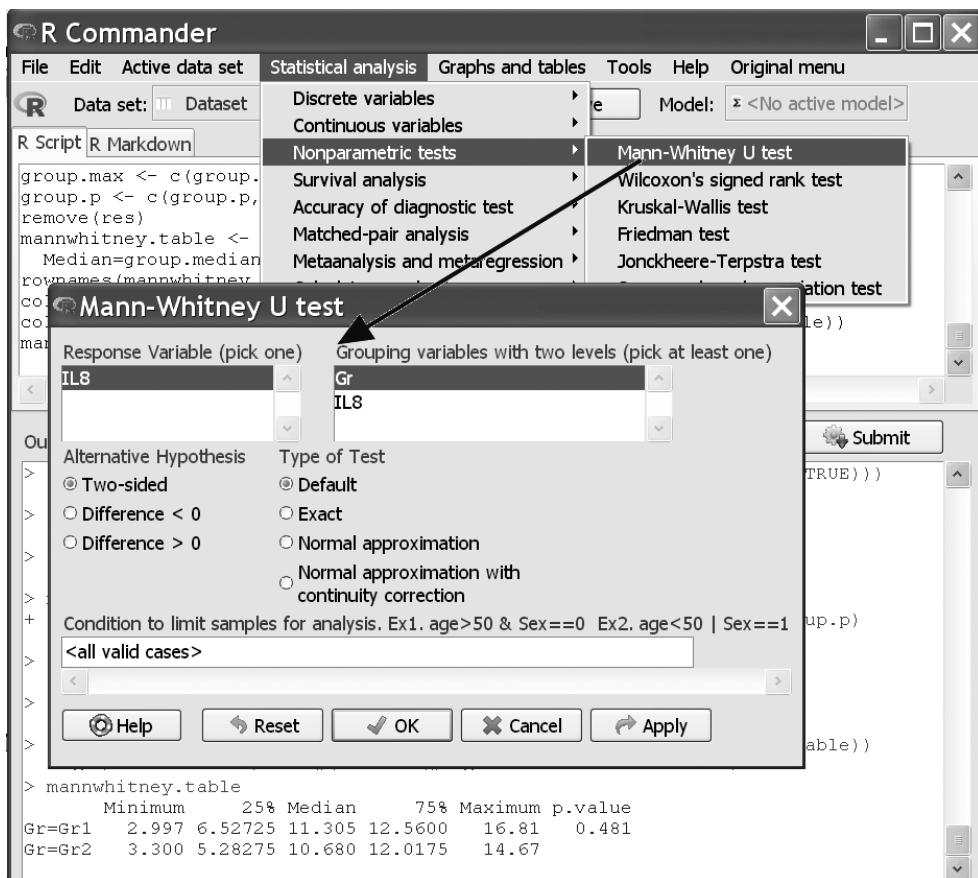


Рис. 7.5. Порівняння двох незалежних сукупностей з «Приклад 5.xls» у пакеті EZR.

При проведенні порівняння не було виявлено відмінностей ($p=0.481$ за критерієм U Манна-Уїтні) центрів (середин) розподілів значень показника ІЛ-8 (11.31 пг/мл (6.53 пг/мл ÷ 12.56 пг/мл) для хворих без вірусного ураження та 10.68 пг/мл (5.28 пг/мл ÷ 12.02 пг/мл) для хворих з ураженням, відповідно.

7.3. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Нормальний закон розподілу

Нехай тепер дві вибірки вилучені з генеральних сукупностей, що мають нормальний закон розподілу, але й спостереження у вибір-

ках попарно пов'язані між собою (пов'язані вибірки, наприклад, до та після дії деякого фактора). Ціль перевірки нульової гіпотези полягає у визначенні наявності ефекту від впливу цього фактора.

Для перевірки нульової гіпотези в цьому випадку використовується так званий одновибірковий **критерій Стьюдента для пов'язаних вибірок**. У цьому випадку нульова гіпотеза полягає в рівності нулю середнього значення різниці варіант двох пов'язаних вибірок.

Статистика критерію t , розраховується за формулою [9]:

$$t = \frac{\sum_{i=1}^n \Delta_i}{\sqrt{\frac{n \times \sum_{i=1}^n \Delta_i^2 - \left(\sum_{i=1}^n \Delta_i\right)^2}{n}}}, \quad (7.4)$$

де Δ_i – попарні різниці варіант вибірок; n – об'єм кожної із вибірок. Розподіл значень критерію підпорядковується розподілу Стьюдента [9] із числом ступенів вільності, $\nu = n - 1$.

Закон розподілу, відмінний від нормального

У випадку порівняння пов'язаних сукупностей, коли хоча б одна з них не підпорядковується нормального закону розподілу, для

перевірки нульової гіпотези застосовується **критерій знаків** Вілкоксона [9]. Нульова гіпотеза полягає в симетричності розподілу різниці варіант двох пов'язаних вибірок відносно нуля. Критерій застосовується для виявлення ефекту впливу в разі, коли закон розподілу відрізняється від нормального. Статистика критерію обчислюється як кількість позитивних різниць варіант спряжених вибірок [9]:

$$Z_n^+ = \sum_{i=1}^n s(x_i - y_i),$$

$$s(\Delta) = \begin{cases} 1, \Delta > 0 \\ 0, \Delta \leq 0 \end{cases}, \quad (7.6)$$

де x_i та y_i – варіанти вибірок, $i = 1, \dots, n$ (n – об'єм вибірок). Висновок про статистичну значимість відмінності робиться у випадку $Z_n^+ < Z_{кр.}^+$ або $Z_n^+ > -Z_{кр.}^-$ ($Z_{кр.}$ розраховується при заданому критичному рівні значимості).

На практиці для розрахунку можуть використовуватися наближені апроксимації, таким чином, щоб звести розподіл статистики критерію до стандартного нормального закону розподілу.

7.4. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

7.4.1. Приклад 6. Нормальний закон розподілу

Дані

Вивчалася ефективність методу лікування, запропонованого для зниження маси тіла жінок, які мають деяке захворювання. У таблиці 6 додатку (Додаток. Приклад 6) наведені результати вимірювання масо-ростового індексу (ВМІ) хворих на початку курсу та через шість місяців лікування. Необхідно з'ясувати, чи змінюється ВМІ хворих у процесі лікування. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Уведемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 6.xls». Для проведення аналізу введемо дані в таблицю, як це показано на рисунку 7.6.

При проведенні аналізу розподілу даних на нормальність використаємо критерій Shapiro-Wilk.

EZR (R-statistics)

Statistical analysis →

Continuous variables →

Kolmogorov-Smirnov test

for normal distribution →

Обрати показник для аналізу (Before)

На Рис.7.7 наведені результати перевірки.

Відмінності закону розподілу показників ВМІ від нормального не виявлено: $p=0.379$ і $p=0.491$ для показників ВМІ до та після лікування, відповідно.

	A	B	C
1	Before	After	
2	28.7	26.8	
3	26.7	22.2	
4	28.7	30	
5	32.5	32.9	
6	24.2	22.7	
7	28.2	27.1	
8	27.9	25.3	
9	29.4	28.4	
10	27.5	27.1	
11	26.7	24	
12	27.2	27.1	
13	24.7	22.1	

Рис. 7.6. Формат введення даних з «Приклад 6.xls» для проведення аналізу в пакеті EZR.

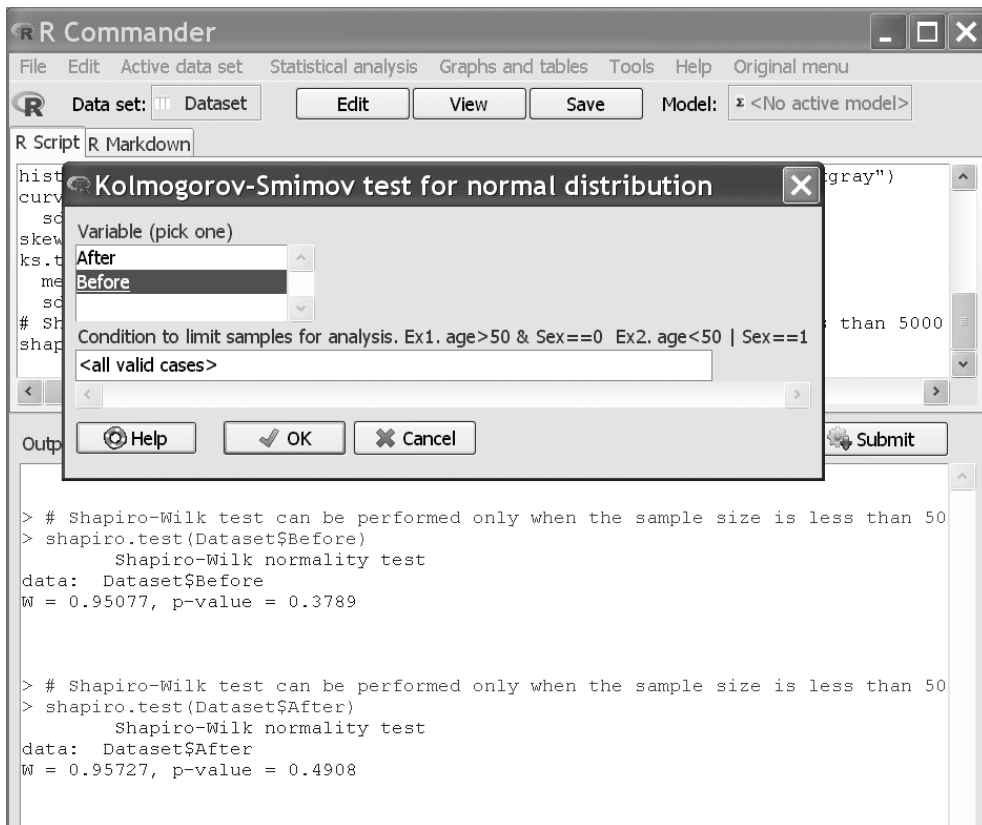


Рис. 7.7. Перевірка даних з «Приклад 6.xls» на нормальність у пакеті EZR.

Для проведення аналізу обираємо параметричні критерії порівняння, критерій Стюдента для порівняння пов'язаних вибірок.

EZR (R-statistics)

Statistical analysis →

Continuous variables →

Paired t-test →

Обираємо змінні порівняння

Обираємо двосторонню

критичну область

перевірки (Two-tailed)

Обираємо надійність перевірки
($1-0.05=0.95$)

Результати розрахунку представлені на рисунку 7.8. Таким чином, отримаємо, що до лікування значення ВМІ складає 27.8 ± 1.9 кг/м², після лікування – 26.7 ± 2.9 кг/м². Нульова гіпотеза про рівність нулю середніх значень різниці відкидається на рівні значимості $p=0.010$.

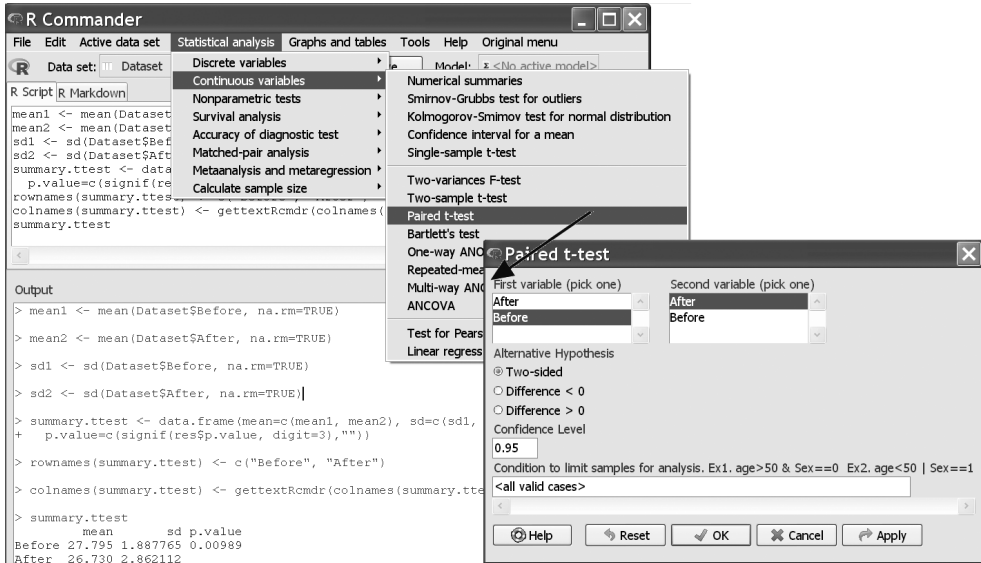


Рис. 7.8. Порівняння двох пов'язаних сукупностей з «Приклад 6.xls» у пакеті EZR.

Результат

У таблиці 7.3 представлено значення показників рівня ВМІ для хворих жінок до та після лікування. Закон розподілу показників не відрізнявся від нормального ($p > 0.05$), тому розраховано їхнє середнє значення \bar{X} та стандартне відхилення SD.

При проведенні аналізу виявлено у середньому, зниження (відмінність статистично значима,

$p = 0.010$ за критерієм Стюдента для пов'язаних вибірок) рівня ВМІ хворих у процесі лікування (від 27.8 ± 1.9 кг/м² до лікування – до 26.7 ± 2.9 кг/м² після лікування).

7.4.2. Приклад 7. Закон розподілу, відмінний від нормального

Дані

Вивчалися фактори ризику розвитку абдомінального компарт-

Таблиця 7.3.
Значення показників у досліджених групах хворих.

Показник	$\bar{X} \pm SD$		Рівень значимості відмінності між групами, p
	До лікування (n=20)	Через 6 місяців (n=20)	
ВМІ, кг/м ²	27.8±1.9	26.7±2.9	0.010

мент-синдрому (АКС) у хворих з хірургічною патологією органів черевної порожнини. Проводився аналіз зміни показника оксиду азоту (NO) хворих у процесі лікування. У таблиці 7 додатку (Додаток. Приклад 7) наведені результати вимірювання рівня NO хворих на момент надходження у відділення і на момент виписки. Необхідно з'ясувати, чи змінюється рівень NO хворих у процесі лікування. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Уведемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 7.xls» (формат – як і в попередньому прикладі).

При перевірці закону розподілу змінних на нормальність було

встановлено, що, якщо на момент надходження закон розподілу значень NO відрізнявся від нормального ($p=0.013$), то на момент виписки відмінності не виявлено ($p=0.190$).

Для однієї із змінних закон розподілу відрізняється від нормального, тому для проведення аналізу обираємо непараметричні критерії порівняння.

Проведемо описову статистику змінних (як це було вище). Для проведення порівняння обираємо непараметричні критерії, критерій знаків для порівняння пов'язаних вибірок.

EZR (R-statistics)

Statistical analysis →

Nonparametric test →

Wilcoxon signed rank test →

```

Output
Submit

> # Shapiro-Wilk test can be performed only when the sample size is less than 5000. (Sample size = 21)
> shapiro.test(Dataset$Before)
      Shapiro-Wilk normality test
data:  Dataset$Before
W = 0.87772, p-value = 0.01326

> # Shapiro-Wilk test can be performed only when the sample size is less than 5000. (Sample size = 21)
> shapiro.test(Dataset$After)
      Shapiro-Wilk normality test
data:  Dataset$After
W = 0.93701, p-value = 0.1901
  
```

Рис. 7.9. Результат перевірки даних з «Приклад 7.xls» на нормальність у пакеті EZR.

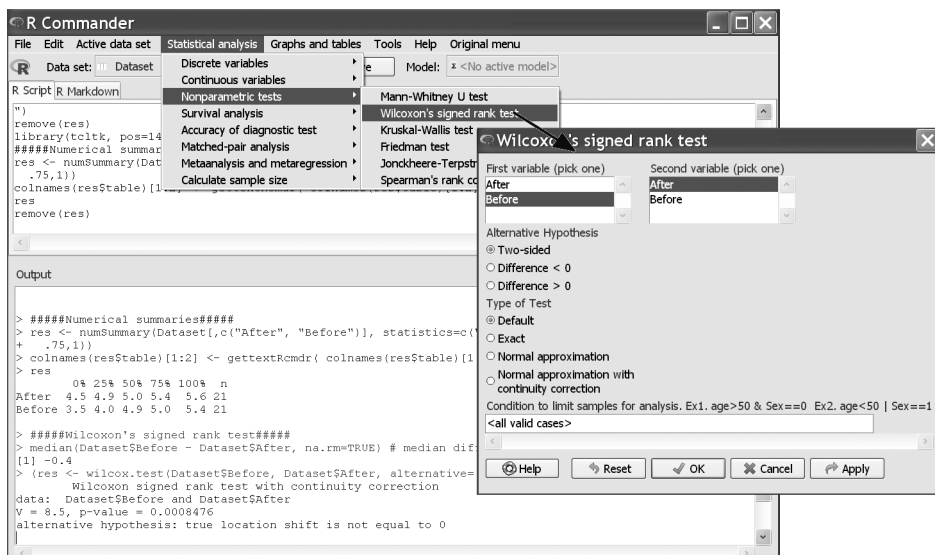


Рис. 7.10. Порівняння двох пов'язаних сукупностей з «Приклад 7.xls» у пакеті EZR.

Обираємо змінні порівняння
Обираємо двосторонню критичну область перевірки (Two-tailed)
Обираємо тип тесту (за замовчуванням)

хворих відносно нуля відкидається, відмінність статистично значима ($p < 0.001$).

Результат

У таблиці 7.4 представлено значення показників рівня NO хворих на момент надходження до відділення і на момент виписки. Закон розподілу показників відрізнявся від нормального ($p < 0.05$), тому

Результати розрахунку представлені на рисунку 7.10. Нульова гіпотеза про симетричність розподілу зміни показників NO для

Таблиця 7.3.
 Значення показників у досліджених групах хворих.

Показник	Me ($Q_I \div Q_{III}$)		Рівень значимості відмінності між групами, p
	На момент надходження до відділення (n=21)	На момент виписки (n=21)	
NO, ммоль/л	5.0 (4.9 ÷ 5.4)	4.9 (4.0 ÷ 5.0)	<0.001

розраховано медіану (Me), значення першого (Q_I) та третього (Q_{III}) кватилей.

При проведенні аналізу було виявлено зниження ($p < 0.001$ за критерієм знаків), у середньому,

значення показника NO у процесі лікування: від 5.0 ммоль/л (4.9 ммоль/л ÷ 5.4 ммоль/л) на момент надходження до відділення – до 4.9 ммоль/л (4.0 ммоль/л ÷ 5.0 ммоль/л) на момент виписки.

РОЗДІЛ 8. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ДВІ ГРУПИ ПОРІВНЯННЯ. ДИХОТОМІЧНІ ЯКІСНІ ЗМІННІ

Як і у випадку із кількісними змінними, при порівнянні якісної змінної у двох групах можна розглянути дві основні задачі:

- *порівняння двох незалежних сукупностей,*
- *порівняння двох пов'язаних сукупностей.*

При порівнянні двох незалежних сукупностей є дві різні групи об'єктів (пацієнтів), що характеризуються якісною змінною, яка може мати тільки одне із двох значень (так/ні). Дослідник намагається відповісти на питання: «Чи з однаковою частотою указана ознака проявляється в двох генеральних сукупностях, з яких випадковим чином вилучили ці дві вибірки?».

При порівнянні *двох пов'язаних сукупностей* є два результати вимірювання для групи об'єктів (пацієнтів), пов'язаних між собою. Вимірювані показники характеризуються частотою прояву якісної ознаки (наприклад, результати визначення «так/ні» у тих самих пацієнтів наявності деякого за-

хворювання за допомогою двох різних методів). Дослідник має відповісти на питання: «Чи одночасно проявляються ці ознаки в генеральній сукупності, з якої випадковим чином вилучили цю вибірку?».

8.1. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Нехай є дві дихотомічні вибірки, спостереження у вибірках не пов'язані попарно (незалежні вибірки). Нульова гіпотеза полягає в тому, що частоти прояву ознаки в генеральних сукупностях не відрізняються одна від одної.

Для перевірки нульової гіпотези у цьому випадку може бути використаний *точний метод Фішера*. Критерій використовується для аналізу чотирипольних таблиць, або таблиць «2x2» (див. Табл. 8.1).

Імовірність похибки першого роду при відкиданні нульової гіпотези при використанні цього критерію розраховується за формулою [9]:

$$\alpha = \frac{(a+b)! \times (c+d)! \times (a+c)! \times (b+d)!}{n! \times a! \times b! \times c! \times d!}, \quad (8.1)$$

У випадку, коли кількість спостережень у кожній із клітинок таблиці перевищує 5, для перевірки нульової гіпотези може бути використаний **критерій χ^2 (хі-квадрат)**, значення статистики критерію (з урахуванням поправки на неперервність – поправки Йейтса) розраховується за формулою [9]:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(|x_{ij} - e_{ij}| - 1/2)^2}{e_{ij}}, \quad (8.2)$$

де x_{ij} – значення у i -ому стовпчику та j -ому рядку таблиці, r – кількість рядків таблиці, c – кількість стовпчиків таблиці, e_{ij} – очікуване (теоретичне) значення у i -ому рядку та j -ому стовпчику таблиці. Розподіл значень критерію підкорюється розподілу хі-квадрат

із числом ступенів вільності, $\nu = (r-1) \times (c-1)$.

8.2. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

8.2.1. Приклад 8. Порівняння двох незалежних сукупностей. Дихотомічні якісні змінні

Дані

Вивчалася частота розвитку безпосередніх ускладнень після проведення оперативного втручання з приводу деякого захворювання. Операція проводилася з використанням двох методів: традиційного та пропонованого. Для проведення рандомізації вибірок метод втручання визначався за допомогою генератора випадкових чисел. У таблиці 8 додатку

Таблиця 8.1.
Частоти прояву ознаки у вибірках (групах).

Ознака	Група 1	Група 2	Усього
Так	a	b	a+b
Ні	c	d	c+d
Усього	$n_1 = a+c$	$n_2 = b+d$	$n = a+b+c+d$

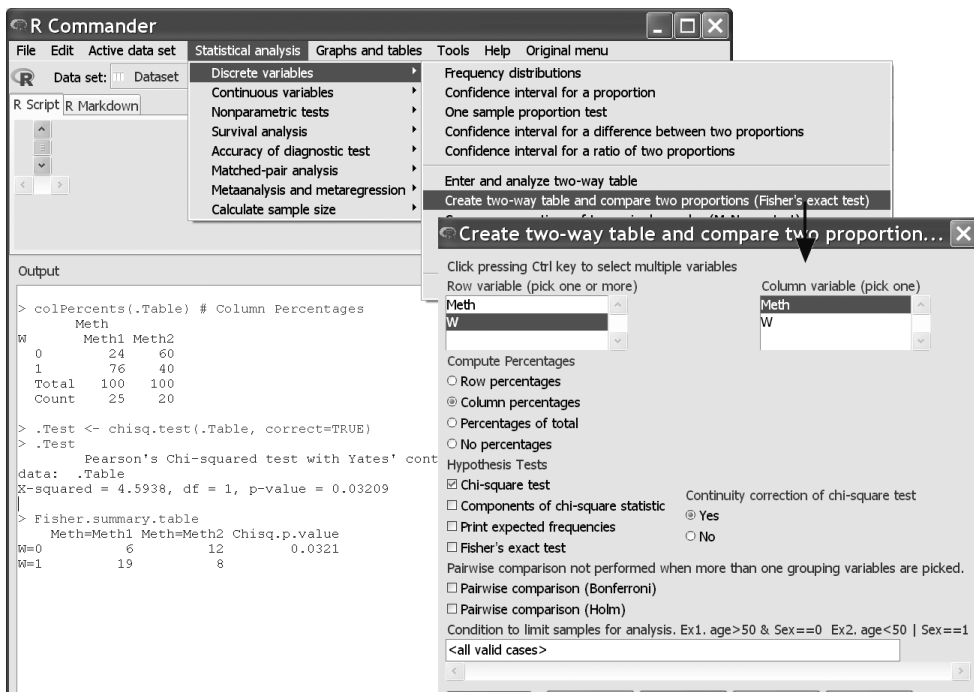


Рис. 8.2. Порівняння двох незалежних сукупностей з «Приклад 8.xls» у пакеті EZR.

Результати розрахунку представлені на рисунку 8.2. Для представлення точкової оцінки розраховуємо також похибку частоти за формулою (5.3). Отримаємо, що для хворих, прооперованих традиційним методом, частота післяопераційних ускладнень скла-

ла 76.0±8.5%, для хворих, оперованих пропонованим методом – 40.0±11.0%. Нульова гіпотеза про рівність частот ускладнень відкидається на рівні значимості p=0.032.

Таблиця 8.2.

Значення частоти післяопераційних ускладнень показників у групах дослідження.

Ускладнення		Традиційний метод (n=25), абс. (%±m%)	Пропонований метод (n=20), абс. (%±m%)	Рівень значимості відмінності між групами, p
Післяопераційне ускладнення	Так	19 (76.0±8.5)	8 (40.0±11.0)	0.032
	Ні	6 (24.0±8.5)	12 (60.0±11.0)	

Результат

У таблиці 8.2 представлені значення частоти післяопераційних ускладнень у двох групах хворих: частота появи ускладнень (%) та стандартна похибка частоти ($\pm m\%$).

При проведенні аналізу було виявлено зниження ($p=0.032$ за критерієм χ^2 -квадрат з урахуванням поправки Йейтса) частоти післяопераційних ускладнень від $76.0 \pm 8.5\%$ (для хворих, прооперованих з використанням традиційного методу) до $40.0 \pm 11.0\%$ (для хворих, прооперованих з використанням пропонованого методу).

8.3. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Нехай є дві дихотомічні вибірки, при цьому спостереження у ви-

бірках пов'язані попарно. Нульова гіпотеза полягає в симетричності розподілу відмінності прояву ознаки двох пов'язаних вибірок відносно нуля.

Нехай є якась вибірка об'єктів, що можуть знаходитись у двох станах ($C1$ і $C2$). Для кожного із станів визначається наявність/відсутність деякої ознаки (O) у кожного з елементів вибірки (див. Табл. 8.3).

У цьому випадку n – об'єм вибірки, w – кількість співпадінь одночасного прояву ознаки у обох станах, z – кількість співпадінь одночасної відсутності ознаки у обох станах, y – кількість випадків прояву ознаки в стані $C1$ і відсутності її в тому ж об'єкта в стані $C2$ і, навпаки, x – кількість випадків прояву ознаки в стані $C2$ і відсутності її в тому ж об'єкта в стані $C1$.

Для перевірки нульової гіпотези в цьому випадку може бути вико-

Таблиця 8.3.
Частоти прояву ознаки O у двох станах вибірки.

C2	C1		
	Так	Ні	Усього
Так	w	x	$w+x$
Ні	y	z	$y+z$
Усього	$w+y$	$x+z$	$n= w+x+y+z$

	A	B
1	Before	After
2	1	1
3	1	1
4	1	0
5	1	0
6	1	0
7	1	0
8	1	0
9	1	0
10	1	0
11	1	0
12	1	0
13	1	0
14	1	0
15	1	0
16	1	1
17	1	1
18	1	0
19	1	0
20	1	1
21	0	1
22	0	1
23	0	1
24	0	1
25	0	1
26	0	1

Рис. 8.3. Формат введення даних з «Приклад 8.xls» для проведення аналізу в пакеті EZR.

ристаний *критерій Мак-Немара*, значення статистики критерію (з урахуванням поправки на неперервність – поправки Йейтса) розраховується за формулою:

$$\chi^2 = \frac{(|x - y| - 1)^2}{x + y}, \quad (8.3)$$

Розподіл значень критерію підкорюється розподілу хі-квадрат із числом ступенів вільності, $\nu = 1$.

8.4. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

8.4.1. Приклад 9. Порівняння двох пов'язаних сукупностей. Дихотомічні якісні змінні

Дані

Вивчалася ефективність препарату від головного болю при деякому захворюванні. Наявність симптому була визначена до прийому препарату й за дві години після його прийому. Визначити, чи впливає препарат на прояв цього симптому. Результати дослідження наведені в таблиці 9 додатку (Додаток. Приклад 9). При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Уведемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 9.xls». Для проведення аналізу введемо дані в таблицю, як це показано на рисунку 8.3.

При аналізі результатів зазначимо, що із 25 хворих до прийому препарату 19 (76.0±8.5%) пацієнтів скаржилися на головний біль, після прийому препарату симптом спостерігався в 11 (44.0±9.9%) пацієнтів. Для визначення впливу препарату на зміну частоти прояву симптому застосуємо критерій Мак-Немара для пов'язаних вибірок.

EZR (R-statistics)

Statistical analysis →

Discrete variables →

Compare proportions of two paired samples (McNemar test) →

Обрати показники

для аналізу (Before) та (After)

З поправкою на неперервність

Результати розрахунку представлені на рисунку 8.4. Нульова гіпотеза про відсутність ефекту не відкидається на рівні значимості $p=0.118$.

Звернемо увагу, що при виборі критерію порівняння для неза-

The screenshot shows the R Commander interface. The 'Statistical analysis' menu is open, and the 'Compare proportions of two paired samples (McNemar test)' option is selected. A dialog box titled 'Compare proportions of two paired samples (McNemar...)' is open, showing 'After' selected for both row and column variables, 'Yes' for continuity correction, and 'all valid cases' for the condition to limit samples. The R console shows the execution of the test, resulting in a p-value of 0.118.

```

> #####Compare proportions of two paired samples
> .Table <- xtabs(~Gr1+Gr2, data=Dataset)
> .Table
      After
Before 0  1
      0  0  6
      1 14  5

> (res <- mcnemar.test(.Table, correct=TRUE))
McNemar's Chi-squared test with continuity correction
data:  .Table
McNemar's chi-squared = 2.45, df = 1, p-value = 0.1175
> cat(gettextRcmdr( "McNemar's test"), "p.value = 0.118
+ ")
McNemar's test p.value = 0.118
  
```

Рис. 8.4. Порівняння двох пов'язаних сукупностей з «Приклад 9.xls» у пакеті EZR.

лежних вибірок для перевірки нульової гіпотези ми б отримали $p=0.042$ і зробили б помилковий висновок про зниження ($p<0.05$) частоти прояву симптому від $76.0\pm 8.5\%$ пацієнтів (до прийому препарату) до $44.0\pm 9.9\%$ (після прийому препарату).

Результат

При аналізі результатів лікування було встановлено, що до прийому препарату $76.0\pm 8.5\%$ пацієнтів скаржилися на головний

біль, за дві години після прийому препарату симптом виявлений у $44.0\pm 9.9\%$ пацієнтів, статистично значимої зміни частоти прояву симптому не виявлено ($p=0.118$ за критерієм Мак-Немара). При цьому зазначимо, що, хоча в 14 ($56.0\pm 9.9\%$) пацієнтів головний біль зник, але в 6 ($24.0\pm 8.5\%$) пацієнтів симптом, навпаки, виник. Не можна стверджувати, що ці зміни зумовлені впливом прийому препарату, а не випадковими причинами ($p=0.118$).

РОЗДІЛ 9. КІЛЬКІСНА ОЦІНКА КЛІНІЧНОГО ЕФЕКТУ. ДВІ ГРУПИ ПОРІВНЯННЯ. ДИХОТОМІЧНІ ЯКІСНІ ЗМІННІ

У попередньому розділі ми обговорювали порівняння двох незалежних сукупностей, що характеризуються якісною змінною, що може приймати тільки одне із двох значень (так/ні). Для таких змінних у медичних дослідженнях досить часто виникає задача порівняння ефективності двох методів (препаратів), де оцінюється ймовірність уникнення негативного результату (ризик «Випадку») у двох групах. При цьому дослідник ставить задачу не тільки дати відповідь на питання: «Чи відрізняються ризики «Випадку» при використанні двох різних методів (препаратів) на генеральних сукупностях?», – але й розрахувати величину ефекту, дати **кількісну оцінку величини клінічного ефекту**.

9.1. КІЛЬКІСНА ОЦІНКА КЛІНІЧНОГО ЕФЕКТУ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Розглянемо основні поняття, які використовуються для розрахунку кількісної оцінки величини клінічного ефекту. Порівнюємо ймовірність «Випадку» для оцінки ефективності двох груп: група контролю і група дослідження.

Для проведення оцінки використовується аналіз таблиць «2x2» (див. Табл. 9.1).

Для кількісної оцінки ризику негативного результату при аналізі результатів медичних досліджень найчастіше використовують показник **абсолютного ризику**

Таблиця 9.1.
Частоти «Випадку» для двох вибірок (методів).

Результат	Група контролю	Група дослідження
«Випадок»	a	b
«Не випадок»	c	d
Усього	$n_1 = a + c$	$n_2 = b + d$

(АР) «Випадку» або показник **шансу** (Ш) «Випадку». Показники обчислюються за різними формулами, дещо по-різному інтерпретуються, але обидва є рівноправними для проведення оцінки.

Абсолютний ризик або просто **ризик** – це ймовірність того, що станеться «Випадок» (як правило – негативний результат) у генеральній сукупності.

Для оцінки абсолютного ризику показник АР (як правило, виражається у відсотках) розраховується на вибірці за формулою:

$$AP = \frac{a}{n} \times 100\%, \quad (9.1)$$

де a – кількість «Випадків», що спостерігалася для даної групи, n – загальна кількість хворих у групі (об'єм вибірки). Нагадаємо, що показник АР, розрахований на вибірці, задає тільки приблизну оцінку **ризик**у для генеральної сукупності, тому оцінка обов'язково повинна супроводжуватися розрахунком вірогідного інтервалу (як правило, **95% ВІ**).

Шанс – це співвідношення ймовірності того, що станеться «Випадок», до ймовірності того, що «Випадок» не станеться, у генеральній сукупності.

Для оцінки шансів показник Ш розраховується на вибірці за формулою:

$$Ш = \frac{a}{c} \times 100\%, \quad (9.2)$$

де a – кількість «Випадків», що спостерігалася для даної групи, c – кількість «Не випадків», що спостерігалася для даної групи. Для показника Ш розраховується **95% ВІ**.

Звернемо увагу на те, що показники Шансів та Абсолютного Ризику мають досить різні значення для оцінки однієї і тієї ж сукупності. Так, наприклад, якщо в групі спостерігалася однакова кількість «Випадків» і «Невипадків», ризик «Випадку» складе $AP=50\%$, шанси «Випадку» складуть $Ш=1/1=1$.

Для кількісної оцінки величини зміни ризику негативного результату при аналізі результатів медичних досліджень найчастіше використовують: показник **зниження абсолютного ризику** (ЗАР), показник **відношення ризиків** (ВР), показник **кількості хворих, яких потрібно лікувати** (КХПЛ), показник **відношення шансів** (ВШ).

Кожен із показників може бути використаний для кількісної оцінки величини клінічного ефекту, який

з них обрати для більш наочного представлення величини ефекту – залежить від поставленої задачі.

Зниження абсолютного ризику (*ARR – Absolute Risk Reduction*) – показує, наскільки знизилася ймовірність «Випадку» у групі дослідження, є різниця між абсолютним ризиком «Випадку» у групі контролю і групі дослідження.

Для оцінки зниження абсолютного ризику показник ЗАР (як правило, виражається у відсотках) розраховується на вибірці за формулою:

$$\text{ЗАР} = \left(\frac{a}{n_1} - \frac{b}{n_2} \right) \times 100\%, \quad (9.3)$$

де показники (див. Табл. 9.1).

При ЗАР > 0 можна говорити про зменшення ризику «Випадку» у групі дослідження, ЗАР < 0 свідчить про збільшення ризику «Випадку» у групі дослідження (ЗАР = 0 – ризик не змінюється). Для показника ЗАР розраховується 95% ВІ.

Відношення ризиків (*RR – Risk Ratio*) – показує, у скільки разів знизилась ймовірність «Випадку» у групі дослідження, є відношенням абсолютних ризиків «Випадку» в групі дослідження і групі контролю.

Для оцінки відношення ризиків показник ВР розраховується на вибірці за формулою:

$$\text{ВР} = \frac{b \times n_1}{a \times n_2}, \quad (9.4)$$

де показники (див. Табл. 9.1).

При ВР < 1 можна говорити про зменшення ризику «Випадку» у групі дослідження, ВР > 1 свідчить про збільшення ризику «Випадку» у групі дослідження (ВР = 1 – ризик не змінюється). Для показника ВР розраховується 95% ВІ.

Кількість хворих, яких потрібно лікувати (*NNT – Number Needed to Treat*) – представляє кількість пацієнтів, яких потрібно в середньому пролікувати для попередження рівно одного «Випадку» у групі дослідження в порівнянні з групою контролю.

Для оцінки кількості хворих, яких потрібно лікувати, показник КХПЛ розраховується за формулою:

$$\text{КХПЛ} = \frac{100\%}{\text{ЗАР}}. \quad (9.4)$$

Чим ближче значення КХПЛ до одиниці, тим ефективніший метод, що використовується в групі дослідження в порівнянні з групою контролю (високе значення,

КХПЛ= ∞ свідчить про відсутність ефекту). Так, наприклад, КХПЛ=2 показує, що, якщо пролікувати по два пацієнта кожним із методів, то в групі дослідження буде, у середньому, рівно на одного вилікуваного («Не випадку») більше, ніж у групі контролю. Для показника КХПЛ розраховується 95% ВІ.

Відношення шансів (*OR – Odds Ratio*) – показує, у скільки разів знизилася шанси «Випадку» у групі дослідження, є відношення шансів «Випадку» у групі дослідження і групі контролю.

Для оцінки відношення шансів показник ВШ розраховується на вибірці за формулою:

$$\text{ВШ} = \frac{b \times c}{a \times d}, \quad (9.4)$$

де показники (див. Табл. 9.1).

При ВШ<1 можна говорити про зменшення ризику «Випадку» у групі дослідження, ВШ>1 свідчить про збільшення ризику «Випадку» у групі дослідження (ВШ=1 – ризик не змінюється). Для показника ВШ розраховується 95% ВІ.

Звернемо увагу на те, що показники ВР та ВШ мають досить різні значення для оцінки однієї і тієї ж сукупності. Так напри-

клад, якщо у групі контролю спостерігалася однакова кількість «Випадків» і «Невипадків», ризик «Випадку» складе $AP=50\%$, шанси «Випадку» складуть $Ш=1/1=1$. А в групі дослідження на кожні 10 пролікованих було лише 2 «Випадки» і 8 «Невипадків», ризик «Випадку» складе $AP=20\%$, шанси «Випадку» складуть $Ш=2/8=0.25$. Для цього прикладу $ВШ=0.25$ (шанси знизилася у 4 рази) а $ВР=0.4$ (ризик знизився у 2.5 рази).

9.2. КІЛЬКІСНА ОЦІНКА КЛІНІЧНОГО ЕФЕКТУ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

9.2.1. Приклад 10. Кількісна оцінка клінічного ефекту

Дані

Звернемося до Прикладу 8. Вивчалася частота розвитку безпосередніх ускладнень після проведення оперативного втручання з приводу деякого захворювання. Операція проводилася з використанням двох методів: традиційний (Група контролю) та пропонуваній (Група дослідження). Для проведення рандомізації вибірок метод втручання визначався за допомогою генератора випадкових чисел. У таблиці 10

додатку (Додаток. Приклад 10) наведено частоти появи ускладнень у двох групах пацієнтів. Необхідно дати кількісну оцінку ефективності пропонованого методу в порівнянні із традиційним. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Для аналізу якісних даних дамо оцінку абсолютних ризиків післяопераційних ускладнень (представимо інтервальну оцінку).

EZR (R-statistics)
Statistical analysis →
Discrete variables →
Confidence interval for a proportion →
Введемо об'єм вибірки та кількість випадків у Групи контролю, потім для Групи дослідження
Оберемо 95% VI

Результати розрахунку представлені на рисунку 9.1. Отримаємо, що для хворих, прооперованих традиційним методом, ризик післяопераційних ускладнень склав

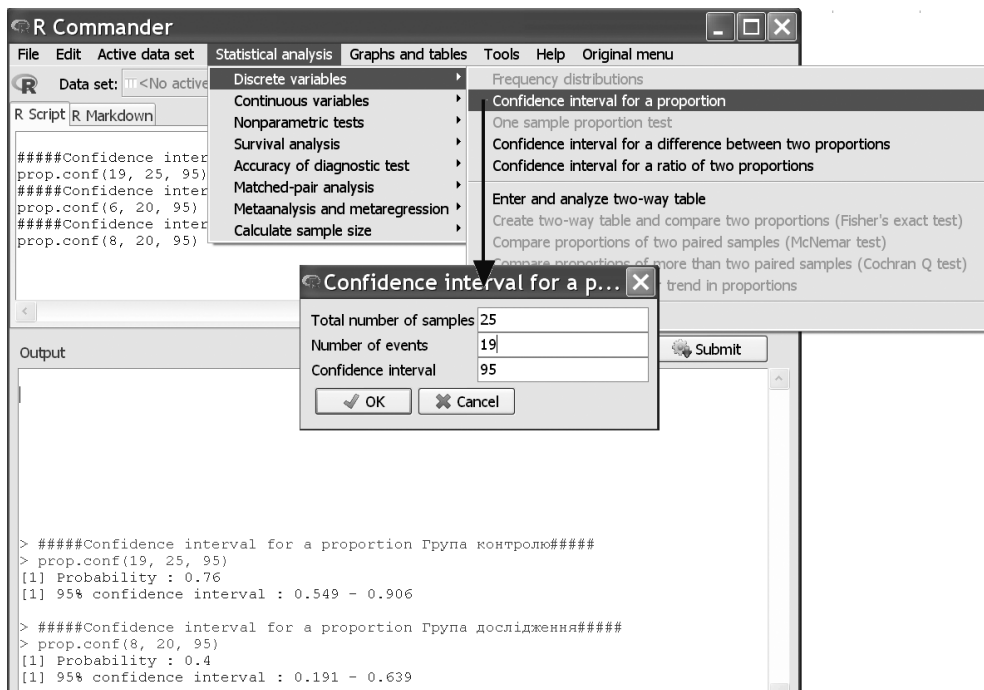


Рис. 9.1. Інтервальна оцінка ймовірності ускладнень для двох сукупностей з Приклад 10 у пакеті EZR.

76.0% (95% ВІ 54.9%÷90.6%), для хворих, оперованих пропонованим методом, – 40.0% (95% ВІ 19.1%÷63.9%).

Для кількісної оцінки клінічного ефекту розрахуємо показник ЗАР.

EZR (R-statistics)

Statistical analysis →

Discrete variables →

Confidence interval for
a difference between

two proportion →

Введемо об'єм вибірки
та кількість випадків

у Групі контролю і Групі
дослідження

Оберемо 95% ВІ

Результати розрахунку представлені на рисунку 9.2. Отримаємо, що для хворих, оперованих про-

понованим методом ризик після-операційних ускладнень знизився, ЗАР=36.0% (95% ВІ 8.8%÷63.2%). Розрахунки 95% ВІ показують, що на генеральній сукупності зниження складе не менше ніж 8.8%.

Розрахуємо показник ВР.

EZR (R-statistics)

Statistical analysis →

Discrete variables →

Confidence interval for a ratio
of two proportion →

Введемо об'єм вибірки

та кількість випадків у Групі
дослідження і Групі контролю

Оберемо 95% ВІ

Результати розрахунку представлені на рисунку 9.3. Отримаємо, що для хворих, оперованих пропонованим методом, ризик післяопера-

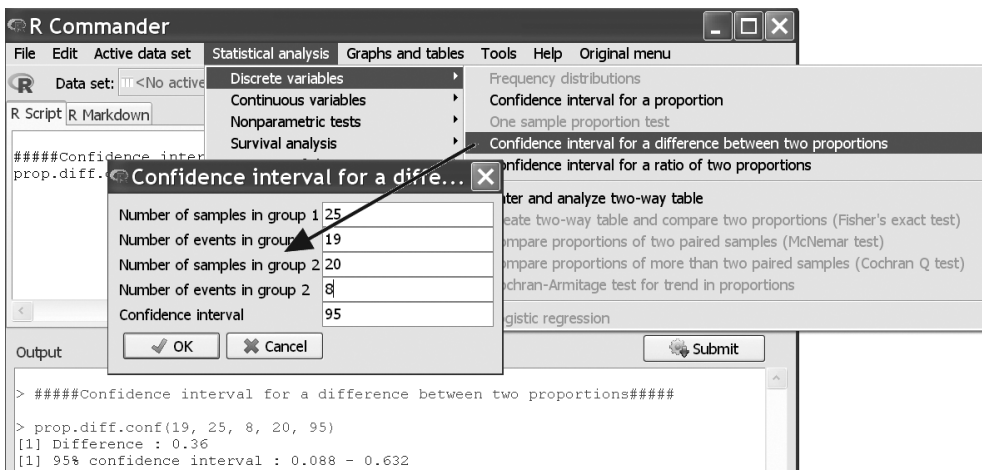


Рис. 9.2. Розрахунок показника ЗАР для двох сукупностей з Приклад 10 у пакеті EZR.

ційних ускладнень знизився (майже у два рази), $VP=0.53$ (95% ВІ 0.30÷0.94). Розрахунки показують, що навіть права границя 95% ВІ не перевищує 1 (тобто відмінність є статистично значимою, $p<0.05$).

У даній версії статистичного пакету розрахунок показника КХПЛ не реалізований, однак його легко можна обчислити, знаючи показник ЗАР, за формулою (9.4). Розрахунки дають значення $КХПЛ=2.7$ (95% ВІ 1.7÷14.0). Таким чином, на кожні 3 проліковані пацієнти застосування пропонуваного методу попереджує на одне ускладнення більше, ніж традиційний метод оперативного втручання.

Розрахуємо показник ВШ.

EZR (R-statistics)

Statistical analysis →

Discrete variables →

Enter and analyze two-way table →

Ведемо кількість випадків

та не випадків у Групі

дослідження і Групі контролю

Оберемо точний критерій Фішера

Результати розрахунку представлені на рисунку 9.4.

Для хворих, прооперованих пропонуваним методом, шанси післяопераційних ускладнень знизилися (майже у п'ять разів), $ВШ=0.22$ (95% ВІ 0.05÷0.89). Відмінність

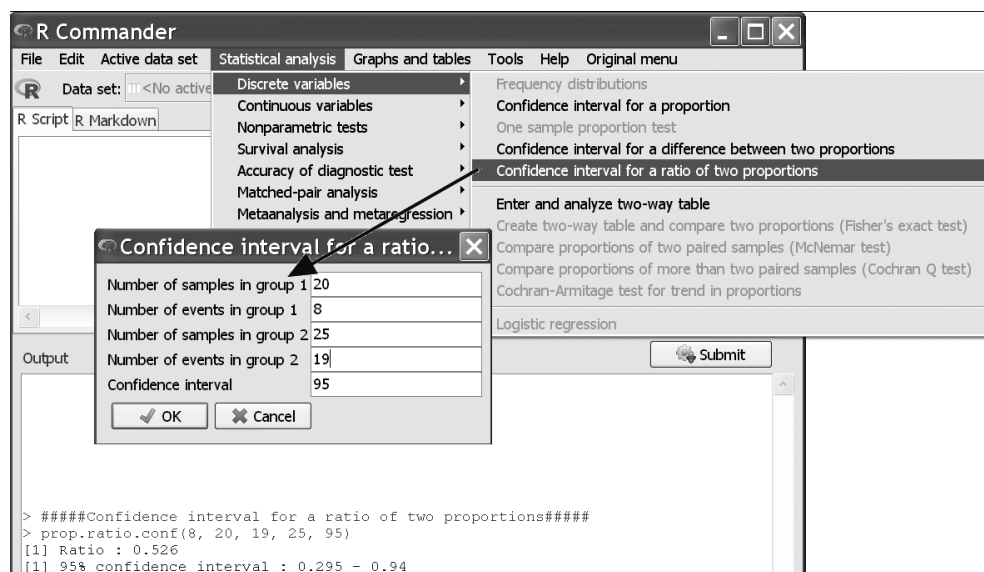


Рис. 9.3. Розрахунок показника ВР для двох сукупностей з Прикладу 10 у пакеті EZR.

між групами статистично значуща, $p=0.031$ за точним методом Фішера.

Результат

Порівнювалися ризики післяопераційних ускладнень у двох групах хворих. На рисунку 9.5 представлено інтервальну оцінку ризику післяопераційних ускладнень при використанні двох методик втручання.

Встановлено, що використання пропонованого методу операційного втручання дозволяє знизити ($p=0.031$ за точним критерієм Фішера) ризик післяпе-

раційного ускладнення, $VP=0.53$ (95% ВІ $0.30\div 0.94$), від 76.0% (95% ВІ $54.9\%\div 90.6\%$) для хворих групи контролю – до 40.0% (95% ВІ $19.1\%\div 63.9\%$) для хворих, оперованих пропонованим методом.

Звернемо увагу, що для кількісної оцінки ефекту було використано показник VP, однак оцінку можна було дати за будь-яким (але достатньо одного) із показників: ЗАР, VP, КХПЛ, ВШ. Рівень значимості відмінності між групами для всіх оцінок, $p=0.031$.

The screenshot shows the R Commander interface. The 'Statistical analysis' menu is open, highlighting 'Enter and analyze two-way table'. A dialog box titled 'Enter and analyze two-w...' is displayed, showing a 2x2 contingency table with counts: 8, 19, 12, and 6. The dialog also includes options for 'Compute Percentages' (No percentages selected), 'Hypothesis Tests' (Fisher's exact test selected), and buttons for Help, OK, and Cancel.

The R console output shows the results of the Fisher's Exact Test:

```
> fisher.test(.Table)
Fisher's Exact Test for Count Data
data: .Table
p-value = 0.03071
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.04775374 0.89159638
sample estimates:
odds ratio
0.2187815
> remove(res)
> summary.table
  X1 X2 Fisher.p.value
1  8 19      0.0307
2 12  6
```

Рис. 9.4. Розрахунок показника ВШ для двох сукупностей з Прикладу 10 у пакеті EZR.

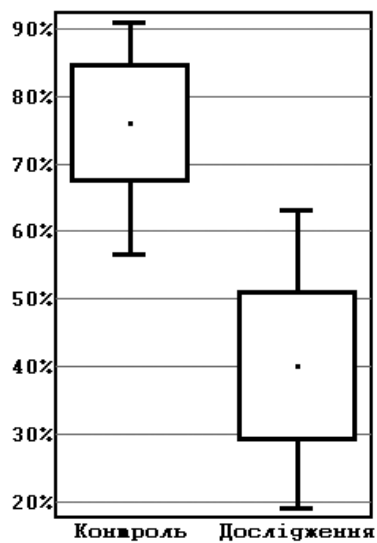


Рис. 9.5. Інтервальна оцінка ризику післяопераційних ускладнень (у пакеті MedStat). Представлена частота патології, стандартна похибка та 95% ВІ.

РОЗДІЛ 10. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ТРИ І БІЛЬШЕ ГРУП ПОРІВНЯННЯ. КІЛЬКІСНІ ЗМІННІ

У попередніх розділах були розглянуті задачі перевірки статистичних гіпотез при порівнянні двох сукупностей. Перейдемо до задач з порівнянням трьох і більше груп. Під час їхнього розв'язання виникають суто математичні проблеми аналізу випадкових величин. Це вимагає застосування додаткових критеріїв, окрім тих, які використовувались для порівняння двох груп. Такий підхід зумовлений виникненням проблеми зростання ймовірності похибки I роду при парних порівняннях кожної із пар – проблеми *множинних порівнянь*. Неврахування цієї проблеми може призвести до некоректних висновків.

Тому для проведення коректного аналізу у випадку порівняння більш ніж двох сукупностей використовуються спеціальні статистичні критерії.

Як і у випадку порівняння двох сукупностей, можуть бути розглянуті дві основні задачі:

- *порівняння більш ніж двох незалежних сукупностей;*

- *порівняння більш ніж двох пов'язаних сукупностей.*

При *порівнянні незалежних сукупностей* є три і більше різні групи об'єктів (пацієнтів), що характеризуються кількісною змінною. Дослідник шукає відповідь на питання: «Чи є однаковими всі середні значення генеральних сукупностей, з яких випадковим чином вилучили ці вибірки, чи вони відрізняються?».

При *порівнянні пов'язаних сукупностей* є група об'єктів (пацієнтів), для кожного з яких проводиться вимірювання кількісного показника в різних (трьох і більше) станах (наприклад, результати вимірювання показника в тих самих пацієнтів у різні моменти часу). Дослідник у цьому випадку відповідає на питання: «Чи відрізняється від нуля різниця показників у процесі лікування на генеральній сукупності, з якої випадковим чином вилучили цю вибірку?».

10.1. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Нормальний закон розподілу

Нехай ми аналізуємо деякий кількісний показник (результуюча ознака). Проведені вимірювання показника для сукупностей, які відрізняються між собою значенням якоїсь іншої ознаки (факторна ознака). Наприклад, аналізуємо значення рівня IgA (*результуюча ознака*) при деякому захворюванні, при цьому хворі поділяються на вікові (*факторна ознака*) групи (до 30 років, від 30 до 60 років, старші ніж 60 років).

Нехай при цьому справедливі такі допущення:

- 1) групи визначаються за рівнями одного фактора,
- 1) результуюча змінна в кожній сукупності підпорядковується нормальному закону розподілу,
- 1) дисперсії у всіх сукупностях є однаковими,

– тоді для проведення порівняння використовується *однофакторний дисперсійний аналіз*. Нульова гіпотеза полягає в тому, що середні значення генеральних сукупностей не відрізняється одна від одної.

Однофакторний дисперсійний аналіз (ANOVA – ANalysis Of VAriance) використовується для перевірки гіпотези про наявність зв'язку між кількісною результуючою і (кількісною або якісною) факторною ознаками (при умові виконання всіх перерахованих вище допущень).

Однофакторний дисперсійний аналіз розділяє загальну варіабельність ознаки на міжгрупову варіабельність (вплив факторної ознаки) та внутрішньогрупову варіабельність (випадкова, нез'ясована варіабельність). Якщо міжгрупову варіабельність перевищує випадкову, то можна зробити висновок про зв'язок між факторною і результуючою ознаками, якщо ж вони не відрізняються між собою, то наявність зв'язку відкидається. Для перевірки нульової гіпотези розраховується статистика критерію F за формулою [6]:

$$F = \frac{SD_{\text{МГ}}^2}{SD_{\text{ВГ}}^2}, \quad (10.1)$$

де $SD_{\text{МГ}}^2$ – вибіркова дисперсія, обчислена за середніми значеннями груп, $SD_{\text{ВГ}}^2$ – середня вибіркова дисперсія у групах. Статистика критерію підпорядковується розподілу Фішера [9] із числом ступенів вільності, $v_1 = k - 1$, $v_2 = n - k$,

де k – кількість груп порівняння, n – загальна кількість варіант у всіх вибірках.

Якщо при проведенні ANOVA виявлений вплив факторної ознаки на результуючу змінну (середні в групах не рівні між собою), виникає питання: «Середні яких саме сукупностей відрізняються одна від одної?». Для відповіді на це питання використовують **апостеріорні порівняння** (*Post hoc tests*).

Як зазначалося вище, у цьому випадку для порівняння кожної із пар звичайні критерії парних порівнянь не можуть бути використані. Для розв'язання проблеми застосовують спеціальні методи множинних порівнянь.

Один з методів урахування проблеми **множинних порівнянь** полягає у використанні **поправки Бонферроні** (*Bonferroni correction*) при проведенні тестів. Цей метод полягає в застосуванні звичайних парних критеріїв для парних порівнянь, але при відкиданні нульової гіпотези розрахунки вести для рівня значимості $\alpha_{кр.}/m$, де $\alpha_{кр.}$ – критичний рівень значимості, m – кількість пар порівняння. Це дозволяє розв'язати проблему зростання ймовірності похибки I роду при множинних порівняннях.

Наприклад, при порівнянні 4-х груп (отримуємо 6 пар порівняння, $m=6$) при обраному $\alpha_{кр.}=0.05$, з урахуванням поправки Бонферроні нульова гіпотеза про рівність середніх із пари відкидається лише при досягнутому рівні значимості $p < 0.05/6 = 0.0083$.

Поправка Бонферроні добре спрацьовує при невеликих кількостях пар порівняння (3–4 пари), але при великих кількостях вона є занадто консервативною і суттєво знижує потужність критерію.

Іншим методом урахування проблеми **множинних порівнянь** є використання спеціальних **критеріїв множинних порівнянь: критерію Шеффе, критерію Тьюкі, критерію Даннета** та ін. [9].

Критерій Шеффе (так само **і критерій Тьюкі**) побудований з урахуванням порівняння не одної, а кількох пар вибірок. Вони застосовуються у випадку порівняння середніх значень для змінних, що підпорядковуються нормальному закону розподілу (і дисперсії в групах не відрізняються), коли потрібно порівняти всі пари між собою.

Критерій Даннета використовується для проведення множин-

них порівнянь у випадку порівнянь з контрольною групою (нас цікавить не порівняння всіх можливих пар між собою, а порівняння пар з одною групою, яка обирається за контрольну).

У пакеті EZR (*R-Statistics*) для проведення *Post hoc* порівнянь можна обрати поправку Бонферроні (чи більш потужний метод Холма-Бонферроні), критерій Тьюкі (для множинних порівнянь) чи критерій Даннета (для порівнянь з контрольною групою).

Закон розподілу, відмінний від нормального

Нехай тепер ми перевіряємо гіпотезу про наявність зв'язку між кількісною результуючою і факторною (кількісною або якісною) ознаками (і маємо більше двох незалежних груп), але допущення використання ANOVA не виконані. У цьому випадку для проведення факторного аналізу потрібно використовувати непараметричні критерії.

Звернемо увагу, що якщо ANOVA, є доволі стійким до порушення умови нормальності законів розподілу значень у групах, та він дає хибні висновки при умові порушення рівності дисперсій. Тому умовою використання непараметричних методів фак-

торного аналізу буде або відмінність закону розподілу змінних від нормального (хоча б в одній групі) або не рівність внутрішньогрупових дисперсій між собою (може бути перевірена за критерієм Бартлета).

Критерій Крускала-Уолліса – непараметричний ранговий аналог однофакторного дисперсійного аналізу, призначений для перевірки гіпотези про рівність ефектів впливу на вибірки (кількість вибірок більше двох). Обчислення статистики H критерію проводиться за формулою:

$$H = \frac{12}{n \times (n-1)} \times \sum_{i=1}^k \frac{R_i^2}{n_i} - 3 \times (n-1), \quad (10.2)$$

де $n = \sum_{i=1}^k n_i$ – загальна кількість варіант у всіх вибірках, n_i ($i=1, \dots, k$) – об'єм i -ої вибірки, k – число вибірок, R_i ($i=1, \dots, k$) – суми рангів i -ої вибірки (при співпадінні рангів у розрахунок вноситься поправка). При великих значеннях k статистика критерію має асимптотичний χ^2 розподіл із числом ступенів вільності $v = k - 1$.

Для проведення апостеріорних порівнянь (у випадку виявлення зв'язку між факторною та ре-

зультуючою ознаками) у випадку використання непараметричного факторного аналізу також використовується *поправка Бонферроні*.

Урахувати проблему *множинних порівнянь* можна за допомогою спеціальних *критеріїв множинних порівнянь: критерію Данна, критерію Стіла-Двасса, критерію Стіла* і т.і. [9].

Критерій Стіла-Двасса (так само і *критерій Данна*) застосовують у випадку порівняння медіанних значень, коли потрібно порівняти всі пари між собою.

Критерій Стіла (так само і *критерій Данна*) застосовують у випадку порівняння з контрольною групою.

10.2. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

10.2.1. Приклад 11. Параметричний критерій аналізу

Дані

Вивчався клінічний перебіг деякого захворювання, що може про-

тікати у трьох формах (А, В та С). Серед 56 хворих 20 пацієнтів мали форму А, 18 пацієнтів – форму В і 18 пацієнтів – форму С перебігу хвороби. У таблиці 11 додатку (Додаток. Приклад 11) наведені результати вимірювання рівня загального білірубіну крові в цих хворих. Необхідно з'ясувати, чи відрізняється рівень загального білірубіну у хворих цих груп. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Уведемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 11.xls». Для проведення аналізу введемо дані в таблицю у тому ж форматі, що й у Прикладі 4 (Рис. 7.1).

Обираючи в пакеті EZR (R-statistics) змінну групування Gr почергово (А, В, С) отримуємо результат перевірки розподілу за критерієм нормальності (Рис. 10.1).

Для визначення, який вид факторного аналізу (параметричний або непараметричний критерій) потрібно використати для проведення аналізу з'ясуємо питання про гомогенність дисперсій для трьох вибірок.

```

Output
> shapiro.test(subset(Dataset, Gr=="A")$B)
      Shapiro-Wilk normality test
data:  subset(Dataset, Gr == "A")$B
W = 0.90907, p-value = 0.06119

> shapiro.test(subset(Dataset, Gr=="B")$B)
      Shapiro-Wilk normality test
data:  subset(Dataset, Gr == "B")$B
W = 0.94122, p-value = 0.3041

> shapiro.test(subset(Dataset, Gr=="C")$B)
      Shapiro-Wilk normality test
data:  subset(Dataset, Gr == "C")$B
W = 0.93112, p-value = 0.2032

```

Messages

Рис. 10.1. Результат перевірки даних з «Прикладу 11.xls» на нормальність у пакеті EZR.

*EZR (R-statistics)**Statistical analysis*→*Continuous variables*→*Bartlett's test*→*Обираємо результуючу змінну (B)**Обираємо факторну змінну (Gr)*

Таким чином, розподіл значень показника загального білірубину у всіх трьох групах не відрізняється від нормального (Рис. 10.1, $p > 0.05$), відмінності дисперсій для груп не виявлено (Рис. 10.2), за критерієм Бартлета $p = 0.894$, тому для проведення порівняння оберемо параметричний критерій ANOVA.

*EZR (R-statistics)**Statistical analysis*→*Continuous variables*→*One-way ANOVA*→*Обираємо результуючу змінну (B)**Обираємо факторну змінну (Gr)**Для проведення апостеріорних порівнянь оберемо критерій Тьюкі*

Результати розрахунку представлені на рисунку 10.3. Таким чином, отримаємо, що нульова гіпотеза про рівність середніх сукупностей відкидається на рівні значимості $p < 0.001$ (за результатами ANOVA). При цьому виявлено статистично значиму

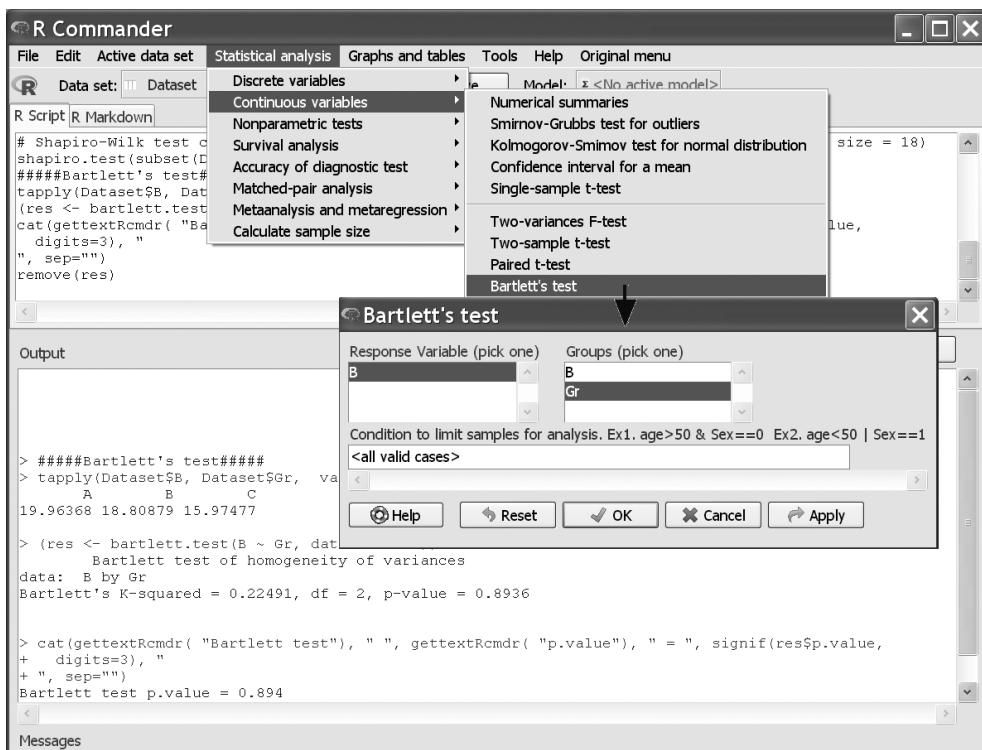


Рис. 10.2. Перевірка гомогенності дисперсій для даних «Приклад 11.xls» у пакеті EZR.

відмінність середнього рівня загального білірубину між групами А і В ($p=0.005$ за критерієм Тьюкі), групами А і С ($p<0.001$ за критерієм Тьюкі), для груп В і С відмінності не виявлено ($p=0.412$ за критерієм Тьюкі).

Результат

У таблиці 10.1 представлено значення показників рівня загального білірубину для трьох груп хворих. Закон розподілу показників не відрізнявся від нормального ($p>0.05$), тому розраховано їхнє середнє значення \bar{X} та стандартне відхилення SD.

При проведенні аналізу виявлено зв'язок загального рівня білірубину з формою протікання захворювання ($p<0.001$ за результатом ANOVA). При цьому в групі хворих з формою А протікання захворювання середнє значення рівня білірубину (11.1 ± 4.5 мкмоль/л) було нижче, ніж у групі хворих з формою В (15.6 ± 4.3 мкмоль/л, $p=0.005$) і групі хворих з формою С (17.4 ± 4.0 мкмоль/л, $p<0.001$). Статистично значимої відмінності середнього рівня білірубину між групами з формою В і С протікання захворювання не виявлено ($p=0.412$).

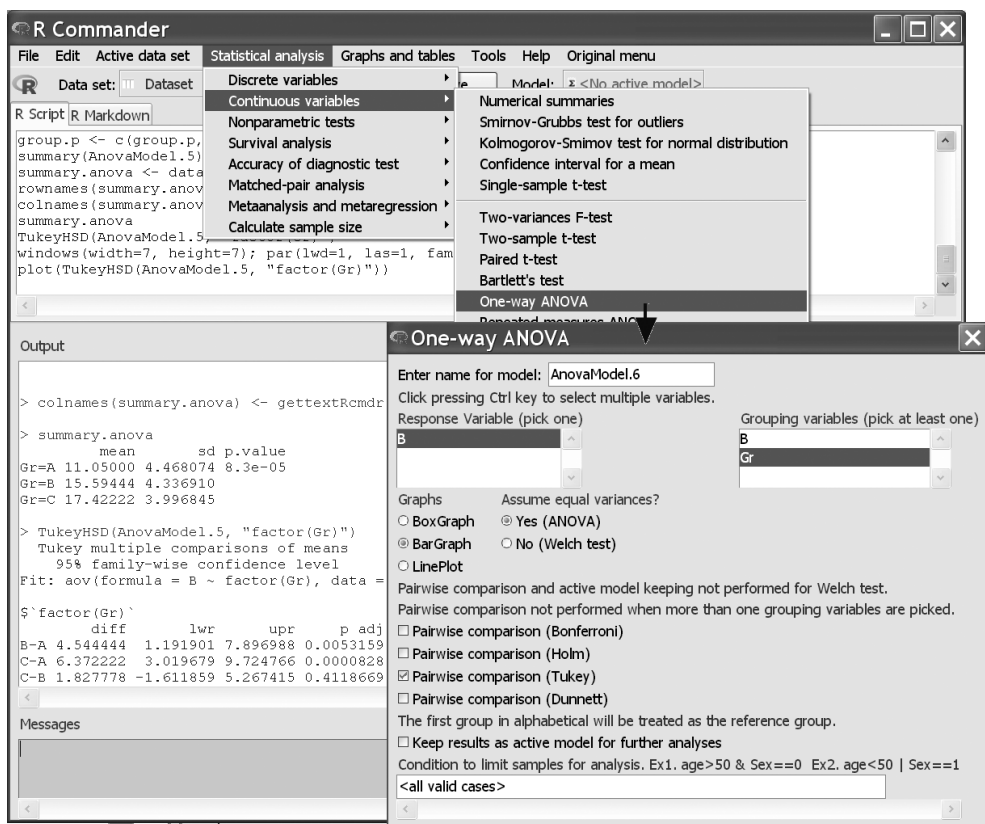


Рис. 10.3. Порівняння трьох незалежних сукупностей з «Прикладу 11.xls» у пакеті EZR.

Таблиця 10.1.

Значення показника загального білірубину для хворих із різними формами перебігу захворювання.

Показник	$\bar{X} \pm SD$			Рівень значимості відмінності між групами, p
	Форма А (n=20)	Форма В (n=18)	Форма С (n=18)	
Загальний білірубін, мкмоль/л	11.1±4.5	15.6±4.3*	17.4±4.0*	<0.001

Примітка: * – відмінність від групи пацієнтів із формою А протікання захворювання статистично значима (використано апостеріорний критерій множинних порівнянь Тьюкі), p<0.05

10.2.2. Приклад 12. Непараметричний критерій аналізу

Дані

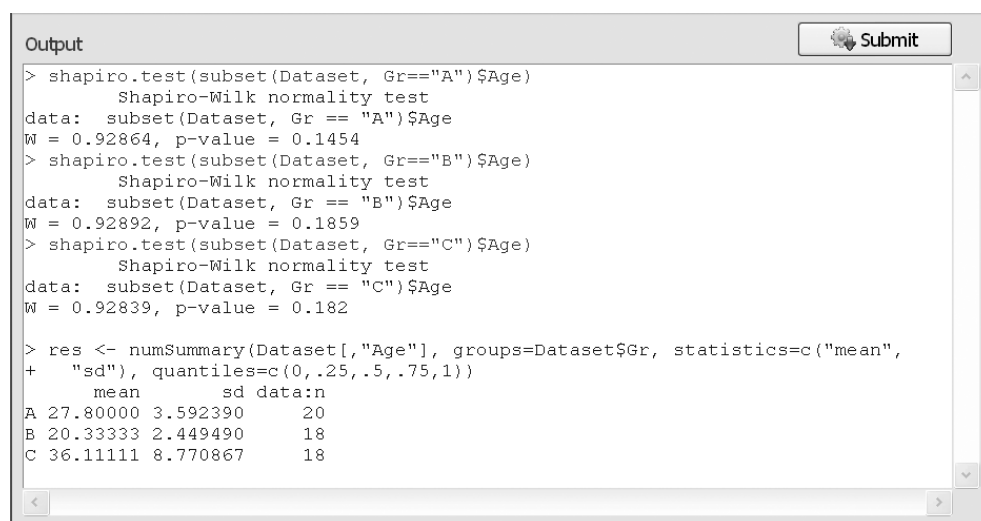
Вивчався клінічний перебіг деякого захворювання, що може протікати в трьох формах (А, В та С). Серед 56 хворих 20 пацієнтів мали форму А, 18 пацієнтів – форму В і 18 пацієнтів – форму С перебігу хвороби. У таблиці 12 додатку (Додаток. Приклад 12) наведено вік хворих. Необхідно з'ясувати, чи відрізняється вік для хворих у цих групах. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Уведемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 12.xls». Для проведення аналізу введемо дані в таблицю у тому ж форматі, що й у попередньому прикладі.

Обираючи в пакеті EZR (R-statistics) змінну групування Gr почергово (А, В, С) проведемо перевірку розподілу на нормальність. Результати перевірки змінної на нормальність і точкової оцінки віку для трьох груп наведено на рисунку 10.4.

Для визначення виду факторного аналізу (параметричний або непараметричний критерій) з'ясуємо питання про гомогенність дисперсій для трьох вибірок.



```

Output
Submit
> shapiro.test(subset(Dataset, Gr=="A")$Age)
  Shapiro-Wilk normality test
data:  subset(Dataset, Gr == "A")$Age
W = 0.92864, p-value = 0.1454
> shapiro.test(subset(Dataset, Gr=="B")$Age)
  Shapiro-Wilk normality test
data:  subset(Dataset, Gr == "B")$Age
W = 0.92892, p-value = 0.1859
> shapiro.test(subset(Dataset, Gr=="C")$Age)
  Shapiro-Wilk normality test
data:  subset(Dataset, Gr == "C")$Age
W = 0.92839, p-value = 0.182

> res <- numSummary(Dataset[, "Age"], groups=Dataset$Gr, statistics=c("mean",
+ "sd"), quantiles=c(0, .25, .5, .75, 1))
      mean      sd data:n
A 27.80000 3.592390    20
B 20.33333 2.449490    18
C 36.11111 8.770867    18
  
```

Рис. 10.4. Результат перевірки даних з «Приклад 12.xls» на нормальність та точкова оцінка для трьох груп у пакеті EZR.

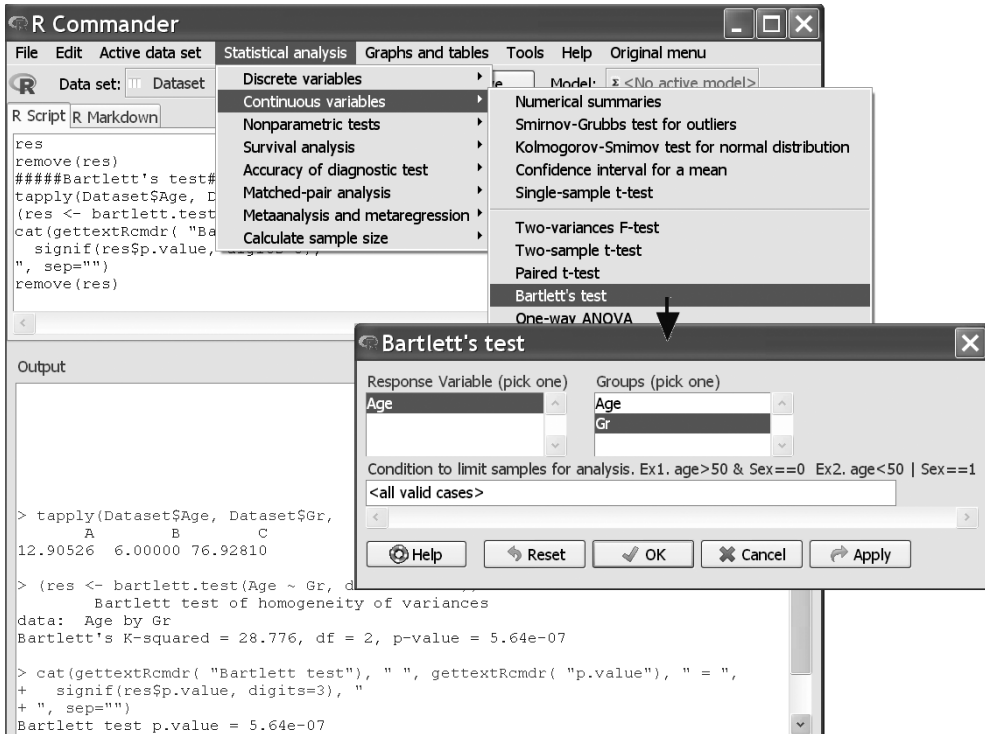


Рис. 10.5. Перевірка гомогенності дисперсій для даних «Приклад 12.xls» у пакеті EZR.

*EZR (R-statistics)**Statistical analysis*→*Continuous variables*→*Bartlett's test*→*Обираємо результуючу змінну (Age)**Обираємо факторну змінну (Gr)*

Таким чином, розподіл значень віку в трьох групах не відрізняється від нормального (Рис. 10.4, $p > 0.05$), але дисперсії в групах статистично значимо (Рис. 10.5, $p < 0.001$) відрізняються, тому для проведення порівняння обираємо непараметричний критерій Крускала-Уолліса.

*EZR (R-statistics)**Statistical analysis*→*Nonparametric tests*→*Kruscal-Wallis test*→*Обираємо результуючу змінну (Age)**Обираємо факторну змінну (Gr)*

Для проведення апостеріорних порівнянь оберемо критерій Стіла-Двасса

Результати розрахунку представлені на рисунку 10.6. Таким чином, отримуємо, що нульова гіпотеза про рівність центрів сукупностей відкидається на рівні значимості $p < 0.001$ (за результатами застосу-

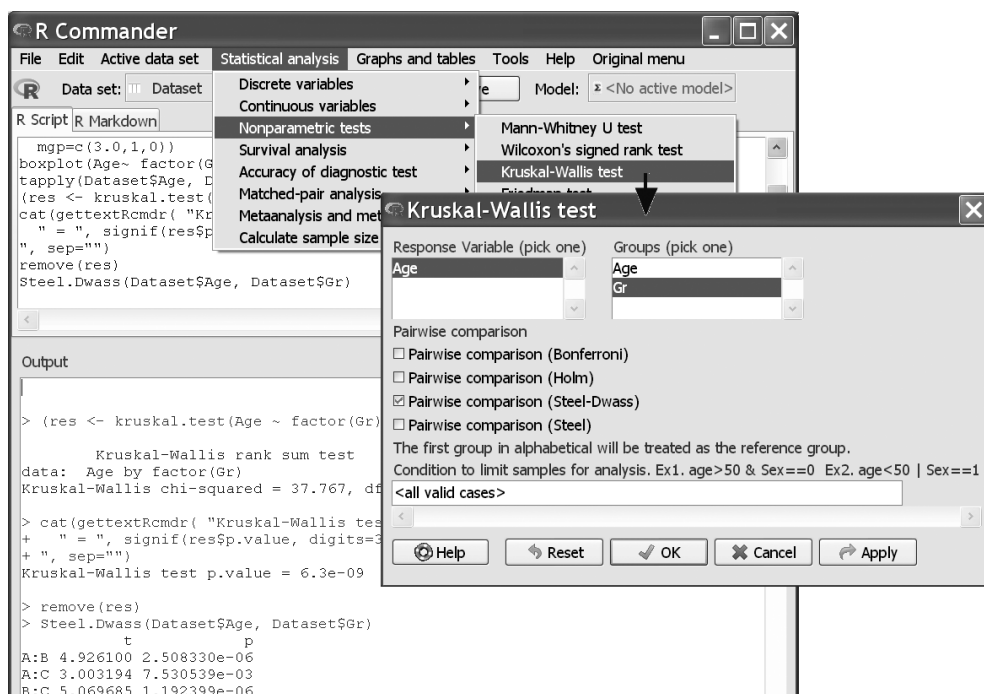


Рис. 10.6. Порівняння трьох незалежних сукупностей з «Приклад 12.xls» у пакеті EZR.

вання критерію Крускала-Уолліса). віку між групами А і В ($p < 0.001$), При цьому виявлено статистично групами А і С ($p = 0.008$), групами значиму відмінність середнього В і С ($p < 0.001$).

Таблиця 10.2.

Середній вік хворих із різними формами перебігу захворювання.

Показник	$\bar{X} \pm SD$			Рівень значимості відмінності між групами, p
	Форма А (n=20)	Форма В (n=18)	Форма С (n=18)	
Вік, років	27.8±3.6 [#]	20.3±2.4 [*]	36.1±8.8 ^{*#}	<0.001

Примітка:

* – відмінність від групи пацієнтів із формою А протікання захворювання статистично значима, $p < 0.05$;

– відмінність від групи пацієнтів із формою В протікання захворювання статистично значима, $p < 0.05$.

Використано апостеріорний критерій множинних порівнянь Стіла-Двасса.

Результат

У таблиці 10.2 представлено середній вік для трьох груп хворих. Закон розподілу не відрізнявся від нормального ($p > 0.05$), тому розраховано середнє значення \bar{X} та стандартне відхилення SD .

При проведенні аналізу виявлено відмінність віку пацієнтів із різними формами протікання захворювання ($p < 0.001$ за критерієм Крускала-Уолліса). При цьому найнижчий, у середньому, вік у пацієнтів з формою А протікання захворювання (20.3 ± 2.4 років), відмінність від пацієнтів з формою В (27.8 ± 3.6 років) і формою С (36.1 ± 8.8 років) протікання захворювання статистично значима ($p < 0.001$ і $p = 0.008$, відповідно). Найвищий, у середньому, вік у пацієнтів з формою С протікання захворювання (відмінність від двох інших груп статистично значима, $p < 0.001$ в обох випадках).

10.3. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ.

Нормальний закон розподілу

Нехай ми аналізуємо деякий кількісний показник (результуюча ознака). Проведені вимірювання

показника для деякої сукупності об'єктів при різних рівнях деякої ознаки (факторна ознака). Наприклад, аналізуємо значення рівня IgA (*результуюча ознака*) при деякому захворюванні, при цьому вимірювання проводиться на тій самій групі пацієнтів: на початку лікування, через тиждень, через місяць, через рік і т.д. (*повторні вимірювання*).

Нехай при цьому справедливі такі допущення:

- 1) групи визначаються за рівнями одного фактора,
 - 2) зміна показника підпорядковується нормальному закону розподілу,
 - 3) дисперсії зміни показника в усіх сукупностях є однаковими,
- тоді для проведення порівняння використовується *однофакторний дисперсійний аналіз для повторних вимірювань*. Нульова гіпотеза полягає в тому, що зміна факторної ознаки не призводить до зміни показника.

Однофакторний дисперсійний аналіз для повторних вимірювань (rANOVA) використовується для перевірки гіпотези про наявність зв'язку між кількісною результуючою і (кількісною або якісною) факторною ознаками (при умові виконання всіх перерахованих вище допущень).

Якщо при проведенні гANOVA виявлено вплив факторної ознаки на результуючу змінну (зміни, у середньому, відрізняються від 0), виникає питання: «При яких саме рівнях факторної ознаки виникають відмінності?». Для відповіді на це питання використовують **апостеріорні порівняння** (*Post hoc tests*).

При порівнянні може бути використаний звичайний критерій Стюдента для пов'язаних вибірок, де для урахування проблеми **множинних порівнянь** використовується поправка Бонферроні, що дозволяє розв'язати проблему зростання ймовірності похибки I роду при множинних порівняннях.

У *накemi EZR (R-Statistics)* для проведення *Post hoc порівнянь* можна обрати *поправку Бонферроні* (чи *більш потужний метод Холма-Бонферроні*).

Закон розподілу, відмінний від нормального

Нехай ми проводимо аналіз для більш ніж двох пов'язаних сукупностей, але допущення використання гANOVA не виконані. У цьому випадку для проведення факторного аналізу потрібно використовувати непараметричні критерії.

Критерій Фрідмана – непараметричний ранговий аналог однофакторного дисперсійного аналізу для повторних вимірювань. Обчислення статистики S критерію проводиться за формулою [9]:

$$H = \frac{12}{n \times k \times (k+1)} \times \left(\sum_{i=1}^k R_i^2 - \frac{1}{k} \left(\sum_{i=1}^k R_i \right)^2 \right), \quad (10.3)$$

де $n = \sum_{i=1}^k n_i$ – загальна кількість варіант у всіх вибірках, n_i ($i=1, \dots, k$) – об'єм i -ої вибірки, k – кількість рівнів факторної ознаки (число вибірок), R_i ($i=1, \dots, k$) – суми рангів i -ого блоку. Статистика критерію обчислюється за χ^2 розподілом із числом ступенів вільності $v = k - 1$.

Для проведення апостеріорних порівнянь (у випадку виявлення зв'язку між факторною та результуючою ознаками) у випадку використання непараметричного факторного аналізу також використовується **поправка Бонферроні**.

10.4. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

10.4.1. Приклад 13. Параметричний критерій аналізу

Дані

Вивчалися результати лікування хворих із діабетичною ретинопатією, що були прооперовані з приводу катаракти. Оцінка результатів проводилася за динамікою зміни гостроти зору на 1-ий день, 7-ий день та через місяць після проведення оперативного втручання. У таблиці 13 додатку (Додаток.

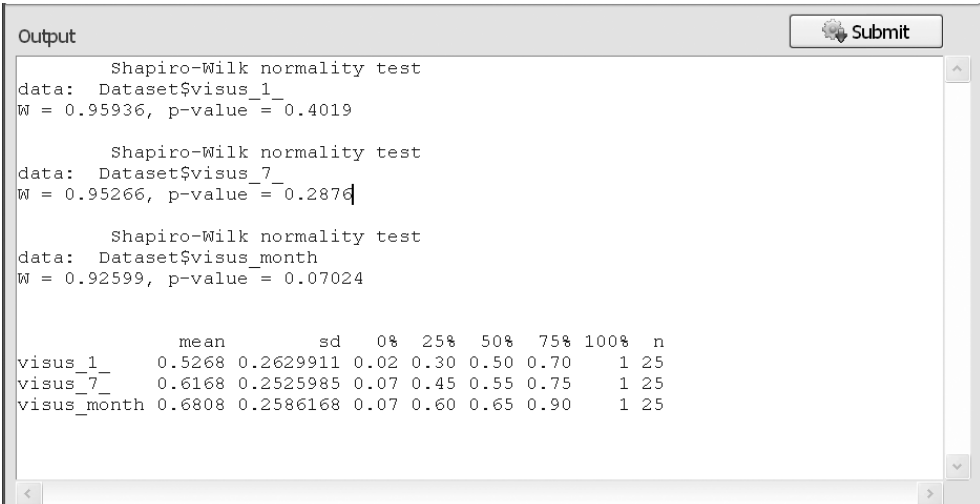
Приклад 13) наведені результати вимірювання гостроти зору в цих хворих. Необхідно з'ясувати, чи змінюється гострота зору у хворих протягом одного місяця після проведення операції. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 13.xls».

Обираючи в пакеті EZR (R-statistics) змінні, отримуємо результат перевірки розподілу на нормальність (Рис.10.7).

Таким чином, розподіл значень показника гостроти зору для всіх трьох точок вимірювання не від-



```

Output
Submit

Shapiro-Wilk normality test
data: Dataset$visus_1_
W = 0.95936, p-value = 0.4019

Shapiro-Wilk normality test
data: Dataset$visus_7_
W = 0.95266, p-value = 0.2874

Shapiro-Wilk normality test
data: Dataset$visus_month
W = 0.92599, p-value = 0.07024

      mean      sd  0%  25%  50%  75% 100%  n
visus_1_  0.5268 0.2629911 0.02 0.30 0.50 0.70  1 25
visus_7_  0.6168 0.2525985 0.07 0.45 0.55 0.75  1 25
visus_month 0.6808 0.2586168 0.07 0.60 0.65 0.90  1 25

```

Рис. 10.7. Результат перевірки даних з «Приклад 13.xls» на нормальність у пакеті EZR.

різняється від нормального (Рис. 10.7, $p > 0.05$), для проведення порівняння оберемо параметричний критерій t -ANOVA.

EZR (R-statistics)

Statistical analysis →

Continuous variables →

Repeted-measures ANOVA →

Обираємо змінні порівняння

Для проведення апостеріорних порівнянь оберемо поправку Бонферроні (Рис. 10.8).

Результати розрахунку представлено на рисунку 10.8. Таким чином, отримаємо, що нульова гіпотеза відкидається на рівні значимості $p < 0.001$ (за результатами t -ANOVA). При цьому виявлено статистично значиму відмінність показників гостро-

ти зору (Рис. 10.7, 10.8) на 1-ий (0.53 ± 0.26) і 7-ий дні (0.62 ± 0.25) при $p < 0.001$, на 1-ий день (0.53 ± 0.26) і через місяць після операції (0.68 ± 0.26 при $p < 0.001$), на 7-ий день (0.62 ± 0.25) і через місяць після операції (0.68 ± 0.26 при $p = 0.021$).

Результат

У таблиці 10.3 представлено значення показників гостроти зору для трьох моментів часу. Закон розподілу показників не відрізнявся від нормального ($p > 0.05$), тому розраховано їхнє середнє значення \bar{X} та стандартне відхилення SD.

При проведенні аналізу виявлено зростання гостроти зору з часом (протягом місяця) після операції

The screenshot shows the R Commander interface. The 'Statistical analysis' menu is open, and the 'Repeated-measures ANOVA' dialog box is displayed. The dialog box contains the following information:

- Enter name for model: AnovaModel.4
- Repeatedly measured data (pick at least 2): visus_1, visus_7, visus_month
- Grouping variable (pick 0, 1, or more): visus_1, visus_7, visus_month
- Graph not created when 3 or more grouping variables are picked.
- Line discrimination: Color
- Line type: Line type
- Line width: Line width
- Pairwise comparison not performed when more than one grouping variables are picked.
- Pairwise comparison (Bonferroni)
- Pairwise comparison (Holm)
- Keep results as active model for further analyses
- Condition to limit samples for analysis. Ex1.age>50 & Sex==0 Ex2.age<50 | Sex==1
- <all valid cases>

The 'Apply' button is highlighted in the dialog box.

Рис. 10.8. Порівняння трьох пов'язаних сукупностей з «Приклад 13.xls» у пакеті EZR.

Таблиця 10.3.

Показник	$\bar{X} \pm SD$			Рівень значимості відмінності між групами, p
	1-ий день (n=20)	7-ий день (n=25)	Через 1 місяць (n=25)	
Гострота зору	0.53±0.26 ^{#§}	0.62±0.25 ^{*§}	0.68±0.26 ^{*#}	<0.001

Примітки:

* – відмінність від показників на 1-ий день статистично значима, $p < 0.05$;

– відмінність від показників на 7-ий день статистично значима, $p < 0.05$;

§ – відмінність від показників на 1-ий місяць статистично значима, $p < 0.05$.

Для попарних порівнянь використано критерій Стьюдента для пов'язаних вибірок з урахуванням поправки Бонферроні.

загального ($p < 0.001$ за результатом гANOVA). При цьому найнижче ($p < 0.001$) значення спостерігалося на 1-ий день після операції (0.53 ± 0.26), а найвище ($p = 0.021$) через місяць після проведення операції (0.68 ± 0.26).

10.4.2. Приклад 14. Непараметричний критерій аналізу

Дані

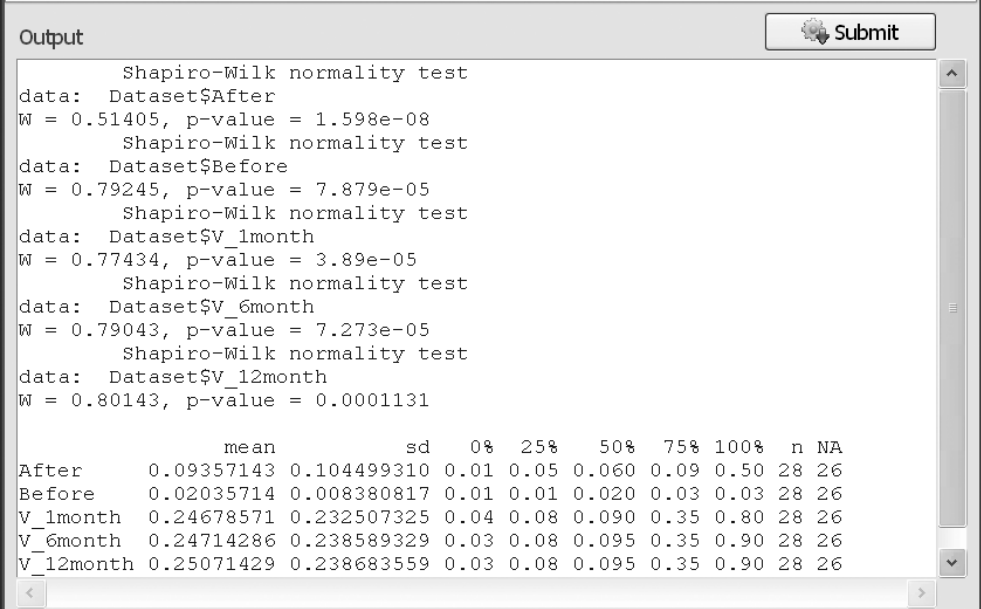
Вивчалися результати лікування при гнійному захворюванні роговиці. Оцінка результатів лікування проводилася за динамікою зміни гостроти зору до (Before) та через тиждень після (After) початку лікування та протягом одного року. У таблиці 14 додатку (Додаток. Приклад 14) наведені результати вимірювання гостроти зору в цих

хворих. Необхідно з'ясувати, чи змінюється гострота зору (і як змінюються) у хворих протягом одного року від початку лікування. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 14.xls». Для проведення аналізу введемо дані в таблицю у тому ж форматі, що й у попередньому прикладі.

Обираючи в пакеті EZR (R-statistics) змінні, отримаємо результат перевірки розподілу на нормальність. Результати перевірки змінної на нормальність і точкової оцінки віку для трьох груп наведено на рисунку 10.9.



```

Output
Submit

Shapiro-Wilk normality test
data: Dataset$After
W = 0.51405, p-value = 1.598e-08
Shapiro-Wilk normality test
data: Dataset$Before
W = 0.79245, p-value = 7.879e-05
Shapiro-Wilk normality test
data: Dataset$V_1month
W = 0.77434, p-value = 3.89e-05
Shapiro-Wilk normality test
data: Dataset$V_6month
W = 0.79043, p-value = 7.273e-05
Shapiro-Wilk normality test
data: Dataset$V_12month
W = 0.80143, p-value = 0.0001131

      mean      sd  0%  25%  50%  75% 100% n NA
After  0.09357143 0.104499310 0.01 0.05 0.060 0.09 0.50 28 26
Before  0.02035714 0.008380817 0.01 0.01 0.020 0.03 0.03 28 26
V_1month 0.24678571 0.232507325 0.04 0.08 0.090 0.35 0.80 28 26
V_6month 0.24714286 0.238589329 0.03 0.08 0.095 0.35 0.90 28 26
V_12month 0.25071429 0.238683559 0.03 0.08 0.095 0.35 0.90 28 26

```

Рис. 10.9. Результат перевірки даних з «Приклад 14.xls» на нормальність та точкова оцінка для п'яти груп у пакеті EZR.

Таким чином, розподіл значень гостроти зору в групах відрізняється від нормального (Рис. 10.9, $p > 0.05$), тому для проведення порівняння обираємо непараметричний критерій Фрідмана.

EZR (R-statistics)

Statistical analysis →

Nonparametric tests →

Friedman test →

Обираємо змінні для аналізу

Для проведення апостеріорних порівнянь оберемо поправку Бонферроні (Рис. 10.10).

ва гіпотеза відкидається на рівні значимості $p < 0.001$ (за результатами застосування критерію Фрідмана). При цьому виявлено, що показники гостроти зору до лікування були статистично значимо ($p < 0.001$) нижче, ніж після лікування протягом усього року спостереження. Також встановлено, що через 1 місяць після початку лікування подальшої зміни показника не виявлено ($p > 0.999$) при попарному порівнянні груп V_1month, V_6month та V_12month між собою.

Результати розрахунку представлені на рисунку 10.10. Таким чином, отримуємо, що нульо-

Результат

У таблиці 10.4 представлено значення гостроти зору хворих

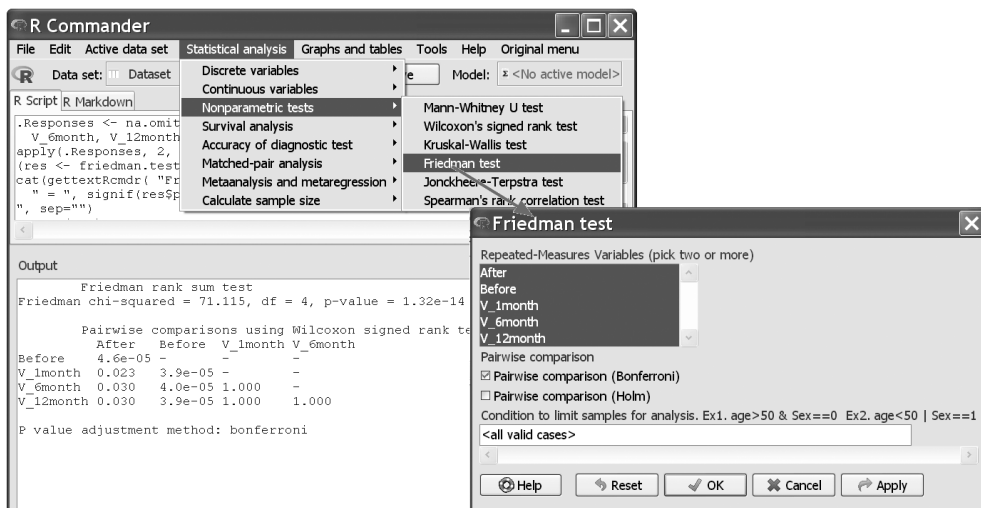


Рис. 10.10. Порівняння п'яти пов'язаних сукупностей з «Приклад 14.xls» у пакеті EZR.

Таблиця 10.4.

Динаміка зміни гостроти зору після проведення лікування.

Час вимірювання	Значення гостроти зору, Me (Q _I ; Q _{III}) (n=25)	Рівень значимості відмінності між групами, p
До лікування	0.02 (0.01; 0.03)	<0.001
За тиждень після початку лікування	0.06 (0.05; 0.09)*	
За місяць після початку лікування	0.09 (0.08; 0.35)*#	
За 6 місяців після початку лікування	0.095 (0.08; 0.35)*#	
За 12 місяців після початку лікування	0.095 (0.08; 0.35)*#	

Примітка:

* – відмінність від показників до лікування статистично значима, $p < 0.05$;# – відмінність від показників за тиждень після початку лікування статистично значима, $p < 0.05$.

Використано апостеріорний критерій Т-Вілкоксона для пов'язаних вибірок з урахуванням поправки Бонферроні.

у кожен час визначення. Закон розподілу показника для всіх груп відрізнявся від нормального ($p < 0.001$), тому розраховано медіанне значення Me , значення першого Q_1 та третього Q_{III} кватилей.

При проведенні аналізу виявлено зростання ($p < 0.001$) показника гостроти зору в усі часи визначення в порівнянні з показником до лікування ($0.02 Q_1 = 0.01; Q_{III} = 0.03$). Встановлено подальше зростан-

ня ($p = 0.023$) показника гостроти зору протягом місяця від початку лікування ($0.09 Q_1 = 0.08; Q_{III} = 0.35$) у порівнянні з показником за тиждень після початку лікування ($0.06 Q_1 = 0.05; Q_{III} = 0.09$). Подальшої зміни показника гостроти зору за шість місяців ($0.095 Q_1 = 0.08; Q_{III} = 0.35$) і за рік ($0.095 Q_1 = 0.08; Q_{III} = 0.35$) після початку лікування в порівнянні з показником за місяць після початку лікування ($0.09 Q_1 = 0.08; Q_{III} = 0.35$) не виявлено ($p > 0.999$).

РОЗДІЛ 11. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ТРИ І БІЛЬШЕ ГРУП ПОРІВНЯННЯ. ДИХОТОМІЧНІ ЯКІСНІ ЗМІННІ

При порівнянні якісної змінної в трьох і більше групах можна розглянути дві основні задачі:

- *порівняння незалежних сукупностей,*
- *порівняння пов'язаних сукупностей.*

Як і у випадку порівняння кількісних змінних для проведення коректного аналізу у випадку порівняння більш ніж двох сукупностей використовуються спеціальні статистичні критерії.

При *порівнянні незалежних сукупностей* є різні групи об'єктів (пацієнтів), що характеризуються якісною змінною, яка може набувати тільки одного із двох значень (так/ні). Дослідник намагається відповісти на питання: «Чи з однаковою частотою вказана ознака проявляється в генеральних сукупностях, з яких випадковим чином вилучили ці вибірки?».

При *порівнянні пов'язаних сукупностей* є група об'єктів (пацієнтів), на які діє деякий фактор (наприклад, наявність захворю-

вання діагностується різними методами в тих самих пацієнтів). Пацієнти характеризуються частотою прояву якісної ознаки (наявність/відсутність захворювання, визначена за допомогою методу діагностики). Дослідник має відповісти на питання: «Чи всі методи дають той самий результат у генеральній сукупності, з якої випадковим чином вилучили цю вибірку?».

11.1. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Нехай є більше двох дихотомічних вибірок, спостереження у вибірках не пов'язані попарно (незалежні вибірки). Нульова гіпотеза полягає в тому, що частоти прояву ознаки в генеральних сукупностях не відрізняються одна від одної (маємо таблицю яка складається із двох рядків та кількох стовпчиків – таблиця $m \times 2$, див. Табл. 11.1).

Таблиця 11.1.
Частоти прояву ознаки у вибірках (групах).

Ознака	Група 1	Група 2	...	Група m
Так	a_1	a_2	...	a_m
Ні	c_1	c_2	...	c_m
Усього	$n_1 = a_1 + c_1$	$n_2 = a_2 + c_2$...	$n_m = a_m + c_m$

Для перевірки нульової гіпотези в цьому випадку може бути використаний **критерій χ^2 (хі-квадрат)**, значення статистики критерію (з урахуванням поправки на неперервність – поправки Йейтса) розраховується за формулою [9]:

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^2 \frac{(|x_{ij} - e_{ij}| - 1/2)^2}{e_{ij}}, \quad (11.1)$$

де x_{ij} – значення в i -ому стовпчику та j -ому рядку таблиці, m – кількість стовпчиків таблиці, e_{ij} – очікуване (теоретичне) значення у i -ому рядку та j -ому стовпчику таблиці. Розподіл значень критерію підкорюється розподілу хі-квадрат із числом ступенів вільності, $v = (2 - 1) \times (m - 1)$.

Якщо при проведенні аналізу за критерієм хі-квадрат виявлено вплив факторної ознаки на результуючу змінну (частоти прояву ознаки в групах не рівні між собою), виникає питання: «Для яких

саме сукупностей є відмінності?». Для відповіді на це питання використовують **апостеріорні порівняння (Post hoc tests)**.

Для урахування проблеми **множинних порівнянь** у цьому випадку, як правило, використовують парні (для таблиць 2×2 , див. Розділ 8.1.1) критерії порівнянь (такі, як точний критерій Фішера або критерій хі-квадрат) з урахуванням при проведенні тестів **поправки Бонферроні** (у пакеті EZR для проведення порівнянь можна обрати більш потужний метод Холма-Бонферроні).

11.2. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШЕ ДВОХ НЕЗАЛЕЖНИХ СУКУПНОСТЕЙ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

11.2.1. Приклад 15. Порівняння більше двох

незалежних сукупностей.**Дихотомічні якісні змінні****Дані**

Вивчалася частота розвитку безпосередніх ускладнень після проведення оперативного втручання з приводу деякого захворювання. Операція проводилася з використанням трьох методів: традиційний метод (M1), новий метод (M2) та новий удосконалений метод (M3). У таблиці 15 додатку (Додаток. Приклад 15) наведені результати появи післяопераційних ускладнень (так – «1», ні – «0»). Необхідно з'ясувати, чи відрізняються частоти післяопераційних ускладнень для цих методів втручання. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 15.xls». Для проведення аналізу введемо дані в таблицю, як це показано на рисунку 11.1.

Для проведення порівняння частоти післяопераційних ускладнень оберемо критерій χ^2 -кватрат.

EZR (R-statistics)
Statistical analysis→

Meth ▼	Result ▼
M1	1
M1	1
M1	1
M1	1
M1	1
M1	1
M1	1
M1	1
M1	0
M1	1
M1	1
M1	1
...	...
M3	0
M3	0
M3	0
M3	0
M3	0
M3	1
M3	0
M3	0
M3	0
M3	0

Рис. 11.1. Формат введення даних з «Приклад 15.xls» для проведення аналізу в пакеті EZR.

Discrete variables→

Create two-way table and compare two proportions (Fisher's exact test)→

Обрати показник для аналізу (W) та вказати змінну групування (Meth)

Вкажемо розрахунок частоти для кожної колонки

Оберемо критерій χ^2 -квадрат з поправкою на неперервність

Оберемо використання поправки Бонферроні для апостеріорних порівнянь

Результати розрахунку представлені на рисунку 11.2. Для представлення точкової оцінки розрахуємо також похибку частоти

за формулою (5.3). Отримаємо, що для хворих, прооперованих традиційним методом, частота післяопераційних ускладнень склала $81.8 \pm 8.2\%$, для хворих, прооперованих новим методом, – $58.3 \pm 10.1\%$, прооперованих новим удосконаленим методом – $16.7 \pm 8.8\%$. Нульова гіпотеза про рівність частот ускладнень відкидається на рівні значимості $p < 0.001$. При цьому встановлено, що використання М3 дає меншу частоту ускладнень, ніж М1 ($p < 0.001$) та М2 ($p = 0.048$).

Результат

У таблиці 11.2 представлено значення частоти післяопераційних

The screenshot shows the R Commander interface. The 'Statistical analysis' menu is open, showing options like 'Discrete variables', 'Continuous variables', 'Nonparametric tests', 'Survival analysis', 'Accuracy of diagnostic test', 'Matched-pair analysis', 'Metaanalysis and metaregression', and 'Calculate sample size'. A dialog box titled 'Create two-way table and compare two proportions (Fisher's exact test)' is open, allowing the user to select variables for the table and specify options for computing percentages and hypothesis tests. The 'Output' window displays the following results:

```

Meth
Result  M1  M2  M3
0       18.2 41.7 83.3
1       81.8 58.3 16.7
Total 100.0 100.0 100.0
Count  22.0 24.0 18.0

Pearson's Chi-squared test
data: .Table
X-squared = 17.164, df = 2, p-value = 0.0001874

Meth=M1 Meth=M2 Meth=M3 Chisq.p.value
Result=0      4      10      15      0.000187
Result=1     18     14       3

Pairwise comparisons using chisq.test
M1
M2 0.47699
M3 0.00046 0.04854
P value adjustment method: bonferroni
  
```

Рис. 11.2. Порівняння трьох незалежних сукупностей з «Приклад 15.xls» у пакеті EZR.

ускладнень у трьох групах хворих. У таблиці представлена частота появи ускладнень (%) та стандартна похибка частоти ($\pm m\%$).

При проведенні аналізу було виявлено зниження частоти прояву ускладнень при використанні нового удосконаленого методу оперативного втручання до $16.7 \pm 8.8\%$, як у порівнянні з традиційною методикою (ускладнення у $81.8 \pm 8.2\%$ пацієнтів, $p < 0.001$), так і в порівнянні з новою методикою без удосконалення (ускладнення у $58.3 \pm 10.1\%$ пацієнтів, $p = 0.048$).

Статистично значимої відмінності частоти прояву ускладнень для пацієнтів, прооперованих за традиційною методикою і новою методикою без удосконалення, не виявлено ($p = 0.477$).

11.3. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШЕ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Нехай є дихотомічні вибірки, при цьому спостереження відбува-

Таблиця 11.2.

Значення частоти післяопераційних ускладнень показників у досліджених групах хворих.

Ускладнення		Традицій-ний метод (n=22), абс. ($\% \pm m\%$)	Новий метод (n=24), абс. ($\% \pm m\%$)	Новий удоско-налений метод (n=18), абс. ($\% \pm m\%$)	Рівень значимості відмінності між групами, p
Після-операційне ускладнення	Так	18 (81.8 \pm 8.2) [§]	14 (58.3 \pm 10.1) [§]	3 (16.7 \pm 8.8) ^{*#}	<0.001
	Ні	4 (18.2 \pm 8.2)	10 (41.7 \pm 10.1)	15 (83.3 \pm 8.8)	

Примітка:

* – відмінність від показників традиційного методу оперативного втручання статистично значима, $p < 0.05$;

– відмінність від показників нового методу оперативного втручання статистично значима, $p < 0.05$;

§ – відмінність від показників нового удосконаленого методу оперативного втручання статистично значима, $p < 0.05$.

Використано апостеріорний критерій χ^2 -квадрат з урахуванням поправки Бонферроні.

ється на тих самих об'єктах, але при різних рівнях факторної ознаки. Нехай для кожного з рівнів факторної ознаки визначається наявність/відсутність (1/0) деякої результуючої ознаки у кожного з елементів вибірки. Нульова гіпотеза полягає у симетричності розподілу відмінності прояву ознаки пов'язаних вибірок відносно нуля.

У випадку, коли кількість рівнів факторної ознаки більше двох для перевірки нульової гіпотези використовується **критерій Кохрена**, значення статистики критерію розраховується за формулою [9]:

$$Q = m \times (m - 1) \frac{\sum_{j=1}^m \left(X_{\cdot j} - \frac{N}{m} \right)^2}{\sum_{i=1}^n X_{i \cdot} \times (m - X_{i \cdot})}, \quad (11.3)$$

де m – кількість рівнів факторної ознаки, n – кількість об'єктів у вибірці, $N = n \times m$ – загальна кількість об'єктів, $X_{\cdot j}$ – сума «одиниць» у j -му стовпчику, $X_{i \cdot}$ – сума «одиниць» у i -му рядку. Розподіл значень критерію відповідно до розподілу χ^2 -квадрат із числом ступенів вільності, $v = m - 1$.

Якщо при проведенні аналізу за критерієм Кохрена виявлено вплив факторної ознаки на ре-

зультуючу змінну, виникає питання: «Для яких саме сукупностей є відмінності?». Для відповіді на це питання використовують **апостеріорні порівняння** (*Post hoc tests*).

Для урахування проблеми **множинних порівнянь** у цьому випадку для кожної пари порівняння може бути використаний критерій Кохрена для двох вибірок з урахуванням **поправки Бонферроні**.

11.4. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ. ПОРІВНЯННЯ БІЛЬШ ДВОХ ПОВ'ЯЗАНИХ СУКУПНОСТЕЙ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

11.4.1. Приклад 16. Порівняння більш двох пов'язаних сукупностей. Дихотомічні якісні змінні

Дані

Вивчалася алергенна дія трьох різних речовин на пацієнтів. Для цього речовина наносилася на шкіру пацієнта й фіксувалася реакція (1 – є дія, 0 – дії не виявлено). Усі три препарати були випробувані на 19 здорових пацієнтах. Результати дослідження наведені в таблиці 16 додатку (Додаток. Приклад 16). При

	A	B	C
1	P1	P2	P3
2	0	0	1
3	0	0	1
4	0	0	1
5	0	0	1
6	0	0	1
7	0	0	1
8	0	0	1
9	1	1	1
10	1	0	0
11	1	0	0
12	1	0	0
13	1	0	0
14	1	0	0
15	1	0	0
16	1	0	0
17	1	0	0
18	1	0	0
19	1	1	0
20	1	1	0
21			

Рис. 11.3. Формат введення даних з «Приклад 16.xls» для проведення аналізу в пакеті EZR.

проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 16.xls». Для проведення аналізу введемо дані в таблицю, як це показано на рисунку 11.3.

При аналізі результатів зазначимо, що з 19 піддослідних у 12 ($63.2 \pm 11.1\%$) виникла реакція на речовину P1, у 3 ($15.8 \pm 8.4\%$) – на речовину P2 і у 8 ($42.1 \pm 11.3\%$) – на речовину P3. Для визначення відмінностей застосуємо критерій Кохрена.

EZR (R-statistics)
 Statistical analysis →
 Discrete variables →
 Compare proportions
 of more than two paired
 samples (Cochran Q test)
 Обрати показники P1, P2, P3

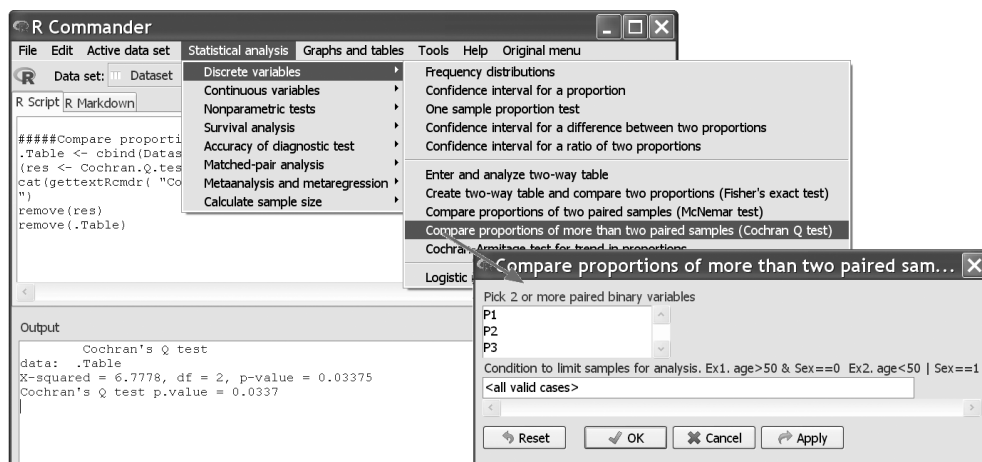


Рис. 11.4. Порівняння трьох пов'язаних сукупностей з «Приклад 16.xls» у пакеті EZR.

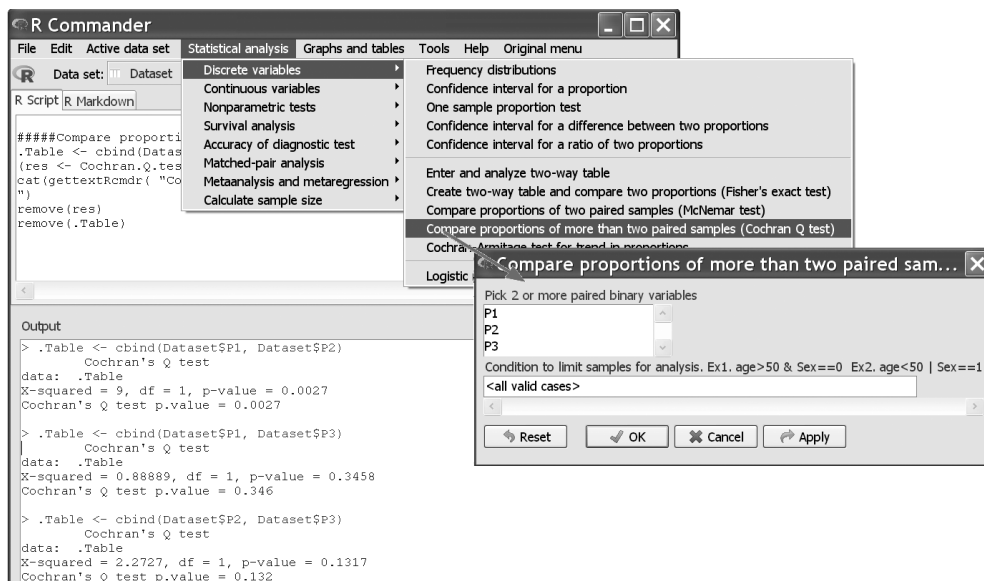


Рис. 11.5. Post hoc порівняння трьох пов'язаних сукупностей з «Приклад 16.xls» у пакеті EZR.

Результати розрахунку представлені на рисунку 11.4. Нульова гіпотеза про відсутність ефекту відкидається на рівні значимості $p=0.034$.

Для проведення порівняння алергенної дії кожної з речовин використовуємо критерій Кохрена для кожної пари вибірок і врахуємо поправку Бонферроні

Таблиця 11.3. Значення частоти алергічних реакцій для досліджених речовин.

Ускладнення		P1 (n=19), абс. (%±m%)	P2 (n=19), абс. (%±m%)	P3 (n=19), абс. (%±m%)	Рівень значимості відмінності між групами, p
Алергічна реакція	Так	12 (63.2±11.1)	3 (15.8±8.4)*	8 (42.1±11.3)	0.034
	Ні	7 (36.8±11.1)	16 (84.2±8.4)	11 (57.9±11.3)	

Примітка:

* – відмінність від речовини P1 статистично значима, $p<0.05$.

Використано апостеріорний критерій Кохрена з урахуванням поправки Бонферроні.

(див. Рис 11.5). На рисунку показані результати розрахунку без урахування поправки Бонферроні. Результати свідчать, що відмінність виявлено між дією препаратів P1 і P2 (з урахуванням поправки Бонферроні – $p = 0.0027 \times 3 < 0.01$). Для інших пар порівняння відмінностей не виявлено ($p > 0.05$).

Результат

У таблиці 11.3 представлено значення частоти алергічних реакцій для трьох речовин. Показана частота появи ускладнень (%) та стандартна похибка частоти ($\pm m\%$). При проведенні аналізу було виявлено відмінність алергенної дії речовин ($p = 0.034$). Встановлено, що алергенна дія речовини P1 ($63.2 \pm 11.1\%$) вища, ніж у речовини P2 ($15.8 \pm 8.4\%$).

РОЗДІЛ 12. ПЕРЕВІРКА НАЯВНОСТІ КОРЕЛЯЦІЙНОГО ЗВ'ЯЗКУ. КІЛЬКІСНІ ЗМІННІ

При проведенні аналізу результатів медичних досліджень доволі часто виникають питання про наявність/відсутність зв'язку між показниками, що характеризують об'єкт (наприклад, чи існує зв'язок між віком пацієнта й рівнем його систолічного тиску, віком і зростом пацієнта та ін.). У загальному випадку дослідника може цікавити виявлення наявності зв'язків як між кількісними, так і між якісними змінними, однак ми зупинимось на застосуванні методів **кореляційного аналізу** для виявлення наявності таких зв'язків (кореляційних зв'язків) тільки для кількісних змінних (у випадку якісних змінних для оцінки степеня зв'язку можуть бути рекомендовані показники ВШ, ВР, що розглянуті в розділі 9).

Таким чином, ми будемо розглядати застосування методів **кореляційного аналізу**, що може бути використаний для виявлення наявності **кореляційного зв'язку** при виконанні таких умов:

- 1) аналіз проводиться для виявлення зв'язку між двома показниками (змінними),
- 2) змінні повинні бути кількісними або порядковими (ранговими, що приймають більше ніж дві упорядковані градації),
- 3) ці змінні повинні відноситися до характеристики того самого об'єкта (пацієнта), або якимось чином пов'язаних між собою об'єктів,
- 4) дані для аналізу включають не більше одного вимірювання по кожному об'єкту,
- 5) відсутні аномальні значення змінних,
- 6) дані для аналізу отримані для однорідної групи об'єктів (не включають підгруп, що мають відмінні середні значення змінних).

У випадку порушення хоча б однієї з умов результати кореляційного аналізу можуть ввести в оману.

При проведенні кореляційного аналізу дослідник шукає відповідь на питання: «Чи є кореляційний зв'язок між значеннями одного й другого показника для генеральної сукупності, з якої випадковим

чином вилучили цю вибірку, чи вони змінюються незалежно?».

Слід зауважити, що при інтерпретації результатів кореляційного аналізу потрібно звернути увагу на два суттєвих моменти:

- 1) *кореляційним називається такий зв'язок між показниками, коли при зміні значення одного показника змінюється середнє значення другого показника,*
- 2) *наявність/відсутність кореляційного зв'язку між двома показниками не може однозначно тлумачитися як наявність/відсутність причинно-наслідкового зв'язку між цими показниками.*

12.1. ПЕРЕВІРКА НАЯВНОСТІ КОРЕЛЯЦІЙНОГО ЗВ'ЯЗКУ. КІЛЬКІСНІ ЗМІННІ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Нормальний закон розподілу

Нехай ми проводимо аналіз з виявлення наявності **лінійного кореляційного зв'язку** для двох кількісних (не рангових) змінних (X та Y). Нехай при цьому справедливі наведені вище допущення і закон розподілу значень хоча б для однієї змінної не відрізняється від нормального. У цьому випадку для виявлення наявності й

оцінки сили кореляційного зв'язку між змінними розраховується **коефіцієнт лінійної кореляції Пірсона**. Значення коефіцієнта кореляції на генеральній сукупності R оцінюється на вибірці за показником r , що розраховується за формулою [6]:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{X}) \times (y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 \times \sum_{i=1}^n (y_i - \bar{Y})^2}}, \quad (12.1)$$

де x_i, y_i – пара значень показників для i -го об'єкту вибірки, та відповідні середні значення, n – кількість об'єктів. У випадку відмінності R від нуля робиться висновок про наявність лінійного кореляційного зв'язку між змінними, у випадку, коли R не відрізняється від нуля – відсутність лінійного кореляційного зв'язку між змінними. Для перевірки нульової гіпотези про $R=0$ використовують статистику T , що розраховується за формулою [6]:

$$T = \sqrt{\frac{r^2 \times (n - 2)}{1 - r^2}}, \quad (12.2)$$

що підпорядковується розподілу Стюдента з числом ступенів вільності $\nu = n - 2$.

Розрахований коефіцієнт лінійної кореляції має значення, що знаходяться на відрізку $[-1; +1]$.

Знак R вказує спрямованість кореляційного зв'язку: при $R > 0$ робиться висновок про наявність позитивного лінійного кореляційного зв'язку між ознаками (при зростанні однієї змінної значення другої, у середньому, зростає), при $R < 0$ робиться висновок про наявність негативного лінійного кореляційного зв'язку між ознаками (при зростанні однієї змінної значення другої, у середньому, спадає).

Величина $|r|$ вказує на силу кореляційного зв'язку, значення r^2 можна інтерпретувати як долю варіабельності змінної Y , що обумовлена лінійним співвідношенням із змінною X . Силу лінійного кореляційного зв'язку умовно поділяють на:

- 1) при $|r| < 0.3$ – слабкий кореляційний зв'язок,
- 2) при $0.3 \leq |r| < 0.7$ – кореляційний зв'язок середнього ступеня вираженості,

- 3) при $|r| \geq 0.7$ – сильний кореляційний зв'язок.

На рисунку 12.1 наведені приклади зображення поля кореляції при проведенні кореляційного аналізу для змінних, які пов'язані між собою лінійним кореляційним зв'язком (показник кореляції у всіх випадках статистично значимо, $p < 0.05$, відрізняється від 0): а) слабкий негативний, б) позитивний, середнього ступеня вираженості, с) сильний позитивний.

Закон розподілу, відмінний від нормального

Нехай ми проводимо аналіз з виявлення наявності зв'язку для двох кількісних змінних (X та Y). Коефіцієнт лінійної кореляції Пірсона не може бути використаний, якщо справедливе хоча б одне з наведеного:

- 1) хоча б одна зі змінних вимірюється у ранговій шкалі,

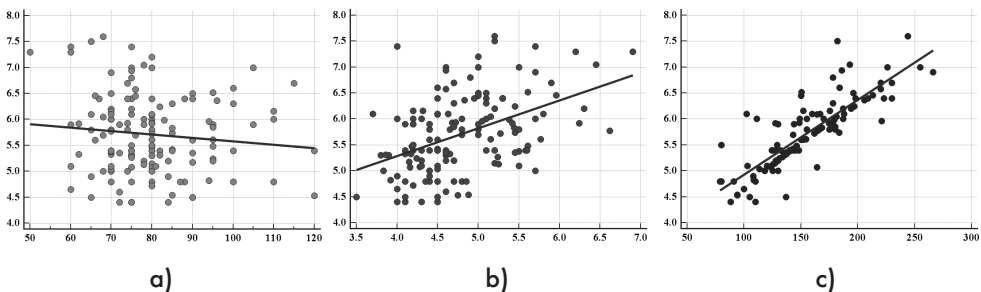


Рис. 12.1. Приклади поля кореляції при різному ступеню вираженості кореляційного зв'язку між змінними: а) для $r = -0.18$, б) для $r = +0.45$, с) для $r = +0.82$.

- 2) розподіл значень для обох змінних відрізняється від нормального,
- 3) заздалегідь передбачається нелінійний (але монотонний) зв'язок між змінними.

У цьому випадку для проведення аналізу використовують непараметричний аналог коефіцієнта лінійної кореляції – *показник рангової кореляції Спірмена*.

Значення показника кореляції на генеральній сукупності R оцінюється на вибірці за показником ρ , що розраховується за формулою [6]:

$$\rho = 1 - 6 \times \frac{\sum_{i=1}^n (r_i - s_i)^2}{n^3 - n}, \quad (12.3)$$

де r_i , s_i – ранги пари значень аналізованих змінних для i -го об'єкту вибірки, n – кількість об'єктів.

У випадку, коли показник R на генеральній сукупності статистично значимо відрізняється від нуля, можна зробити висновок про наявність кореляційного зв'язку між змінними X та Y .

Інтерпретація показника ρ така сама, що й у випадку нормально-го закону розподілу, за винятком того, що ρ^2 не може бути інтерпретовано як долю варіабельності

змінної Y , що обумовлена співвідношенням зі змінною X . Показник оцінює ступінь вираженості монотонного (а не лінійного) кореляційного зв'язку.

12.2. ПЕРЕВІРКА НАЯВНОСТІ КОРЕЛЯЦІЙНОГО ЗВ'ЯЗКУ. КІЛЬКІСНІ ЗМІННІ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

12.2.1. Приклад 17. Параметричний критерій аналізу

Дані

Вивчався перебіг лікування хворих із псоріазом. У таблиці 17 додатку (Додаток. Приклад 17) наведені результати вимірювання показників цитокінів IL-6 та IL-8 у цих хворих. Необхідно з'ясувати, чи існує зв'язок рівнів цих показників у хворих. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 17.xls».

Результат перевірки розподілу на нормальність наведено на рисунку 12.2.

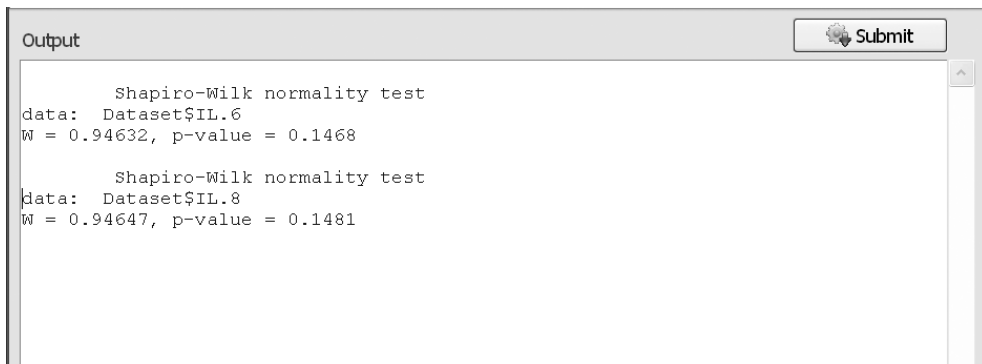


Рис. 12.2. Результат перевірки даних з «Приклад 17.xls» на нормальність у пакеті EZR.

Таким чином, розподіли значень показників IL-6 та IL-8 не відрізняються від нормального (Рис. 12.2, $p > 0.05$), для проведення кореляційного аналізу оберемо параметричні критерії. *EZR (R-statistics)*

Statistical analysis →
Continuous variables →
Test for Pearson's correlation →
Обираємо змінні аналізу

На рисунку 12.4 показане поле кореляції в координатах змінних аналізу.

Результати розрахунку (див. Рис. 12.3) свідчать про відсутність лінійного кореляційного зв'язку ($p = 0.123$). Звернімо увагу на 95% ВІ розрахованого коефіцієнта кореляції, $r = 0.29$ (95% ВІ -0.08 ;

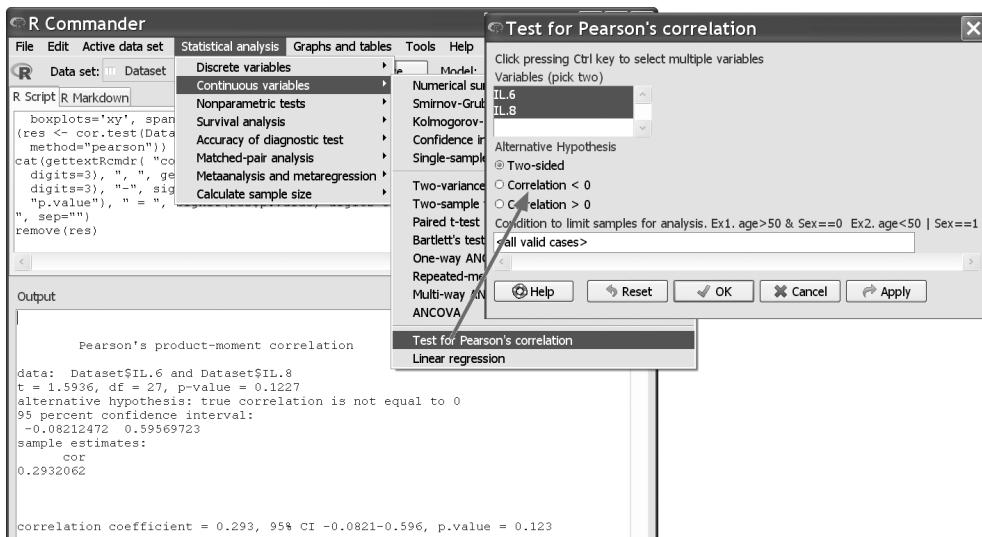


Рис. 12.3. Кореляційний аналіз з «Приклад 17.xls» у пакеті EZR.

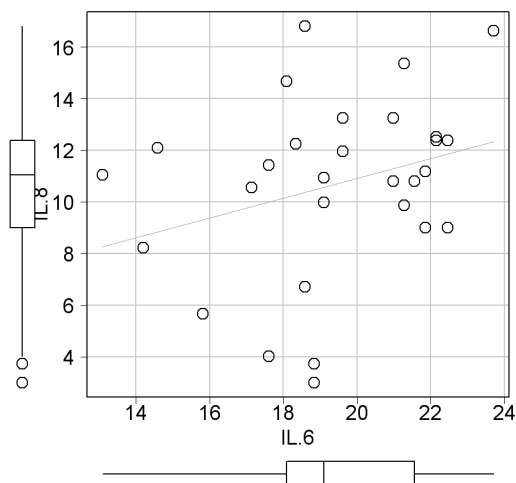


Рис. 12.4. Поле кореляції з «Приклад 17.xls» у пакеті EZR.

+0.60), що включає значення нуль та означає, що відкидати H_0 про рівність значення коефіцієнта лінійної кореляції від нуля не можна.

Результат

При проведенні аналізу зв'язку показників цитокінів ІЛ-6 та ІЛ-8 у хворих із псоріазом було розраховано показник лінійної кореляції Пірсона. Значення коефіцієнту кореляції $r = 0.29$ (95% ВІ $-0.08; +0.60$) статистично значимо не відрізняється від 0 ($p=0.123$). Таким чином, лінійного кореляційного зв'язку між показниками ІЛ-6 та ІЛ-8 у хворих із псоріазом не виявлено ($p=0.123$).

12.2.2. Приклад 18. Непараметричний критерій аналізу

Дані

Вивчалися результати лікування деякого захворювання. В таблиці 18 додатку (Додаток. Приклад 18) наведені результати вимірювання показники цитокінів ударного об'єму (УО) та ваги хворих. Необхідно з'ясувати, чи існує зв'язок УО та ваги при даному захворюванні. При проведенні аналізу критичний рівень значимості взяти рівним 0.05.

Розв'язання в пакеті EZR (R-statistics)

Уведемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 18.xls».

```

Output

      Shapiro-Wilk normality test
data:  Dataset$Bara
W = 0.88574, p-value = 0.009094

      Shapiro-Wilk normality test
data:  Dataset$YO
W = 0.87294, p-value = 0.004958

```

Рис. 12.5. Результат перевірки даних з «Приклад 18.xls» на нормальність у пакеті EZR.

Результат перевірки розподілу на нормальність наведено на рисунку 12.5.

Таким чином, розподіл значень YO і ваги відрізняється від нормально-го (Рис. 12.5, $p < 0.05$), для проведення кореляційного аналізу оберемо непараметричні критерії.

EZR (R-statistics)

Statistical analysis →

Nonparametric tests →

Spearman's rank correlation test →
Обираємо змінні аналізу

На рисунку 12.7 показане поле кореляції у координатах змінних аналізу.

Результати розрахунку (див. Рис. 12.6) свідчать про наявність кореляційного зв'язку між значеннями YO і ваги ($\rho = -0.427$, статистично значимо відрізняється від 0, $p = 0.033$).

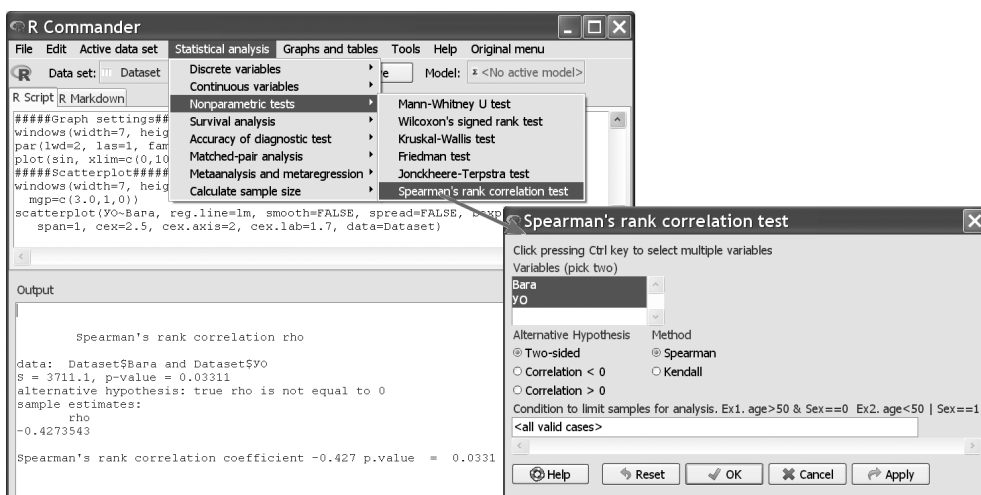


Рис. 12.6. Кореляційний аналіз з «Приклад 18.xls» у пакеті EZR.

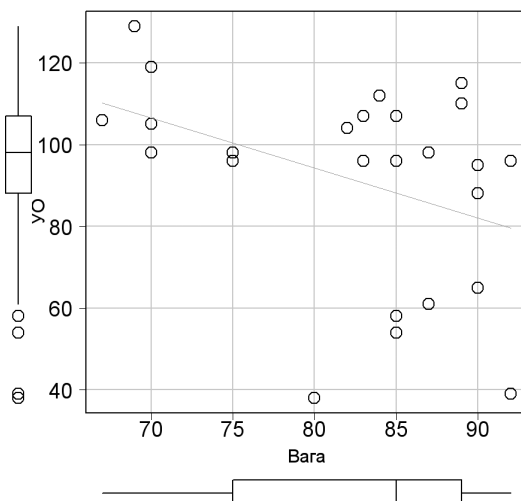


Рис. 12.7. Поле кореляції з «Приклад 18.xls» у пакеті EZR.

Результат

При проведенні аналізу зв'язку показників УО і ваги у хворих було розраховано показник рангової кореляції Спірмена. Значення коефіцієнта кореляції $\rho = -0.427$ (статистично значимо відрізняється від 0, $p=0.033$). Та-

ким чином, виявлено наявність негативного кореляційного зв'язку показників УО і ваги у хворих ($\rho = -0.427$, $p=0.033$), збільшення ваги хворих супроводжується, у середньому, зменшенням викиду крові серцем, УО.

РОЗДІЛ 13. ПОБУДОВА І АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ

Кореляційний аналіз відповідає на питання наявності зв'язку між ознаками X та Y , однак, якщо ми вважаємо, що між ними існує не просто кореляційний зв'язок, а причинно-наслідкова залежність, $y = f(x)$, то виникає задача прогнозування значень Y за результатами вимірювання величини X . Для розв'язання задачі побудови функціональної залежності величини Y від величини X й аналізу побудованої моделі використовують методи **регресійного аналізу**. У цьому випадку змінна Y називається залежною (вихідною) змінною, а змінна X – незалежною (вхідною) змінною, або предиктором.

У випадку, коли функція регресії являється лінійною, залежність Y від X шукають у вигляді математичного рівняння:

$$y = b_0 + b_1 \times x \quad , \quad (13.1)$$

де b_0 та b_1 – коефіцієнти регресії (деякі числа), b_0 – вільний член регресії, b_1 – кутовий коефіцієнт. У цьому випадку для побудови й

аналізу математичних моделей використовують методи однофакторного регресійного аналізу (лінійна модель регресії).

У медичних дослідженнях часто виникають задачі аналізу впливу кількох факторних (вхідних) змінних X_1, X_2, \dots, X_m на результуючу (вихідну) змінну Y . При цьому слід зауважити, як правило, факторні змінні, що характеризують об'єкт медичного дослідження пов'язані між собою, тому дослідник повинен оцінювати вплив i -ої факторної ознаки на результуючу змінну при урахуванні (*on adjustment*) впливу на Y всіх інших $m-1$ факторних ознак. Змінна Y називається залежною (вихідною) змінною, а змінні X_1, X_2, \dots, X_m – незалежні змінні, або предиктори.

Практичні рекомендації

Звернімо увагу, що, хоча змінні X_i і називаються незалежними, це не потрібно інтерпретувати як незалежність одної факторної змінної X_i від іншої X_j , а розуміти як виявлення незалежного (стан-

дартизованого за іншими змінними) впливу факторної змінної X_j на вихідну змінну Y .

Для побудови і аналізу математичних моделей використовують

методи багатофакторного регресійного аналізу.

У випадку, коли функція регресії являється лінійною, залежність Y від X_1, X_2, \dots, X_m шукають у вигляді рівняння:

$$y = b_0 + b_1 \times x_1 + b_2 \times x_2 + \dots + b_m \times x_m, \quad (13.2)$$

змінна Y називається вихідною змінною, а змінні X_j – вхідними змінними ($j=1, 2, \dots, m$), b_0 – вільний член регресії (значення вихідної змінної Y коли всі x_j дорівнюють 0), b_1, b_2, \dots, b_m – коефіцієнти регресії при відповідному предикторі. Коефіцієнт b_j дорівнює величині, на яку змінюється результуюча змінна Y при зміні X_j на 1 одиницю.

При проведенні багатофакторного регресійного аналізу дослідником можуть бути поставлені задачі:

побудова математичної моделі (формули), за якою, знаючи значення факторних змінних, може бути обчислене значення результуючої змінної Y ,

визначення факторних ознак, які значимо впливають на зміну результуючої ознаки,

оцінка величини незалежного впливу факторної ознаки на зміну результуючої ознаки (при урахуванні впливу інших факторів).

Далі ми розглянемо теоретичні і практичні аспекти аналізу із застосуванням побудови *множинної лінійної моделі регресії*, що може бути застосований при виконанні таких умов:

1) між факторними змінними X_1, X_2, \dots, X_m і результуючою змінною Y існує лінійна залежність,

2) спостереження у вибірці незалежні (для кожного об'єкту не більше одного набору змінних),

3) кожному значенню з набору величин X_1, X_2, \dots, X_m відповідає набір значень вихідної величини Y , що підпорядковується нормальному закону розподілу і дисперсії цих розподілів однакові,

4) змінні X_1, X_2, \dots, X_m вимірюються без похибки.

Звернімо увагу, що серед умов немає умови нормального закону розподілу для факторних змінних, більше того, множинна лінійна модель регресії може включати в себе в якості факторних і номінальні (при відповідному кодуванні) ознаки.

13.1. ПОБУДОВА І АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Нехай ми маємо на меті побудувати *лінійну багатофакторну модель регресії* залежності результуючої кількісної змінної Y від факторних змінних X_1, X_2, \dots, X_m . Нехай при цьому справедливі наведені у попередньому розділі допущення про можливість проведення такого аналізу.

$$y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \dots + \beta_m \times x_{mi} + \text{res}_i \quad (13.3)$$

де y_i – значення результуючої ознаки для i -го ($i = 1, 2, \dots, N$) об'єкту, x_j – значення j -ої ($j = 1, 2,$

Побудова і аналіз множинної моделі регресії проводиться за алгоритмом:

розрахунок коефіцієнтів $b_0, b_1, b_2, \dots, b_m$ лінійної моделі регресії на навчальній множині,

оцінка якості і адекватності моделі регресії в цілому,

оцінка впливу для кожної окремої факторної ознаки,

підтвердження практичного використання математичної моделі на контрольній множині.

13.1.1. Розрахунок коефіцієнтів лінійної моделі регресії

Нехай у нас є вибірка з N об'єктів (пацієнтів), і для кожного із об'єктів проведено вимірювання значення факторних X_1, X_2, \dots, X_m і результуючої Y ознак. Будемо вважати, що на генеральній сукупності залежність результуючої ознаки від предикторів описується формулою:

\dots, m) факторної ознаки для i -го об'єкту, $\beta_0, \beta_1, \beta_2, \dots, \beta_m$ – точні зна-

чення коефіцієнтів моделі на генеральній сукупності (тоді

$$y_i^T = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \dots + \beta_m \times x_{mi}$$

– теоретичне, прогнозоване значення вихідної змінної для i -го об'єкту), res_i – залишок (похибка прогнозування, різниця між виміряним та теоретичним значенням вихідної змінної для i -го об'єкту,

$$res_i = y_i - y_i^T.$$

Для проведення розрахунку значень b_j на множині даних, що є оцінкою точних значень β_j для генеральної сукупності (підгонки коефіцієнтів регресії на навчальній

множині), як правило, використовується метод найменших квадратів (МНК). Суть методу полягає в підборі таких коефіцієнтів $b_0, b_1, b_2, \dots, b_m$, що мінімізують суму квадратів залишків SS_{res} :

$$SS_{res} = \frac{1}{N} \times \sum_{i=1}^N res_i^2 \rightarrow \min$$

(13.4)

Математично це зводиться до розв'язання системи $m+1$ лінійних рівнянь з $m+1$ невідомими ($b_0, b_1, b_2, \dots, b_m$):

$$\left\{ \begin{array}{l} b_0 \times N + \sum_{j=1}^m b_j \times \sum_{i=1}^N x_{ji} = \sum_{i=1}^N y_i \\ b_0 \times \sum_{i=1}^N x_{1i} + b_1 \times \sum_{i=1}^N x_{1i}^2 = \sum_{i=1}^N (y_i \times x_{1i}) \\ b_0 \times \sum_{i=1}^N x_{1i} + b_2 \times \sum_{i=1}^N x_{2i}^2 = \sum_{i=1}^N (y_i \times x_{2i}) \\ \dots \\ b_0 \times \sum_{i=1}^N x_{mi} + b_m \times \sum_{i=1}^N x_{mi}^2 = \sum_{i=1}^N (y_i \times x_{mi}) \end{array} \right. , \quad (13.5)$$

Явні формули для розрахунку [6] коефіцієнтів b_j наводити тут немає сенсу, зважаючи на їх складність для розуміння і біологічної

інтерпретації, тим більше, що в різних статистичних пакетах можуть бути реалізовані різні методи розрахунку.

У результаті проведених розрахунків отримуємо точкову оцінку β -коефіцієнтів моделі (b_j), також може бути розрахована стандартна похибка оцінки (m_{b_j}) або 95% ВІ цих оцінок.

13.1.2. Оцінка якості й адекватності моделі регресії

Для оцінки якості побудованої моделі в пакеті EZR може бути розраховано **коефіцієнт множинної кореляції R** (розраховується як коефіцієнт парної кореляції між експериментальними Y та теоретичними значеннями Y^T вихідної змінної). Для оцінки узгодженості побудованої моделі з експериментальними даними розраховується **корегований коефіцієнт детермінації R^2 -adjusted**, що представляє собою долю дисперсії вихідної змінної, що знайшла пояснення в рамках моделі та корегується таким чином, щоб можна було співставляти моделі з різною кількістю факторних змінних.

Якщо значення R^2 -adjusted статистично значимо не відрізняється від нуля ($p > 0.05$), то математична модель, в цілому, не адекватна експериментальним даним і про лінійну залежність від вибраного набору факторних ознак не може

йти мова, **лінійної залежності Y від X_1, X_2, \dots, X_m не виявлено**.

Якщо R^2 -adjusted статистично значимо ($p < 0.05$) відрізняється від нуля, то модель, в цілому, адекватна експериментальним даним, **виявлено лінійну залежність Y від X_1, X_2, \dots, X_m** .

Як і при інтерпретації звичайного парного коефіцієнта кореляції Пірсона, чим більшим є значення R^2 -adjusted, тим в більшій мірі модель узгоджена з експериментальними даними (найкраща модель має R^2 -adjusted = 1 – що відповідає 100% узгодженості).

Часто для оцінки якості моделі використовується також **значення F -статистики ANOVA** (підпорядковується розподілу Фішера). Статистично значима ($p < 0.05$) відмінність F від одиниці свідчить про те, що не всі β -коефіцієнти моделі дорівнюють 0, існує, лінійна залежність між вихідною змінною Y , та, у крайньому разі, одною факторною ознакою.

Для перевірки відмінності від 0 для кожного j -го коефіцієнту моделі – β_j використовується значення t -статистики, що підпорядковується розподілу Стюдента. Статистика розраховується для

кожної незалежної змінної множинної лінійної моделі і відповідає на питання, чи впливає j -а факторна ознака (з урахуванням стандартизації за всіма іншими факторами) на вихідну змінну.

Для оцінки ступеню впливу для кожної j -ої факторної змінної може бути розрахований **частковий коефіцієнт кореляції** r_{part} , що оцінює вплив j -ої факторної ознаки (з урахуванням стандартизації за всіма іншими факторами) на вихідну змінну.

Для оцінки адекватності множинної лінійної моделі регресії експериментальним даним може бути проведений аналіз залишків (res), для чого будують графік залежності res_i від y_i^T (де $i=1, 2, \dots, N$ – i -ий об'єкт у вибірці), та для кожної j -ої факторної ознаки графік залежності res_i від x_{ij} (де $i=1, 2, \dots, N$ – i -ий об'єкт у вибірці, $j=1, 2, \dots, m$ – j -а факторна змінна). У випадку відсутності тенденцій до збільшення (зменшення) значення залишків на цих графіках робиться висновок про адекватність лінійної моделі.

13.1.3. Відбір незалежних ознак

Відповідно до принципу «бритви Оккама», який для нашого випа-

дку можна записати, що «... з набору наявних еквівалентних моделей будь-якого явища слід вибрати найпростішу ...», у дослідника при побудові множинних моделей регресії часто виникає задача відбору мінімального набору факторних ознак, які якісно прогнозують результуючу змінну. При цьому слід зазначити, що із самого методу побудови множинних моделей регресії випливає, що чим більше підгоночних параметрів (змінних) включено у модель, тим точніше буде відбуватися прогнозування. Таким чином, при виборі найпростішої моделі потрібно знайти компроміс між мінімізацією кількості факторних ознак і якістю прогнозування.

Існує багато методів відбору набору ознак для побудови моделі, тут зупинимося на тих, що реалізований у пакеті EZR, це **метод покрокового включення/виключення змінних (stepwise method)**:

- 1) побудова моделі,
- 2) включення змінної у модель при виконанні умов включення,
- 3) перевірка нової моделі, виключення змінних із моделі при виконанні умов виключення,
- 4) перехід до кроку 2) для наступної змінної.

В якості умов включення та виключення в пакеті EZR може бути вибраний:

- **інформаційний критерій Акаїке (Akaike information criterion, AIC),**
- **байєсівський інформаційний критерій (Bayesian information criterion – BIC),**
- **критерій досягнення порогу значимості.**

Інформаційний критерій Акаїке

При виборі цього методу відбору незалежних ознак, що включаються у модель для оцінки цінності моделі розраховується показник AIC:

$$AIC = 2 \times m + N \times \ln(SSRes/N), \quad (13.6)$$

де m – кількість факторних змінних, що включені у модель, N – об'єм вибірки, $SSRes$ – сума квадратів залишків.

При включенні та виключенні змінних до моделі змінюється перший і другий доданок формули (13.6). За кінцеву обирається модель, що має найменше значення AIC.

Байєсівський інформаційний критерій

При виборі методу BIC відбору незалежних ознак, що включаються у модель для оцінки цінності моделі розраховується показник BIC:

$$BIC = m \times \ln(N) + N \times \ln(SSRes/N), \quad (13.7)$$

де m – кількість факторних змінних, що включені у модель, N – об'єм вибірки, $SSRes$ – сума квадратів залишків. За кінцеву обирається модель, що має найменше значення BIC.

Критерій досягнення порогу значимості

При виборі цього методу відбору незалежних ознак, включення j -ої змінної у модель відбувається тоді, коли рівень значимості відмінності відповідного коефіцієнта регресії від нуля менше за поріг включення (p_{enter}), виключення j -ої змінної з моделі відбувається тоді, коли рівень значимості відмінності відповідного коефіцієнта регресії від нуля перевищує поріг виключення (p_{rem}). Після перебору всіх змінних обирається кінцева модель.

13.1.4. Підтвердження практичного використання математичної моделі на контрольній множині

При проведенні аналізу множинних моделей слід не забувати, що процес побудови моделі на навчальній множині даних полягає у підгонці коефіцієнтів моделі до відомих результатів, тому для практичного використання прогнозу, отриманого в рамках статистичної моделі рекомендують перевірити прогностичні якості побудованої моделі на нових даних. Для перевірки (упередження перенавчання) модель аналізують на даних, що не були використані при побудові моделі (контрольній множині) і якщо отримують ту ж якість прогнозування, що і на навчальній множині даних, то можна говорити про відсутність підгонки і про можливість використання результатів прогнозування на практиці. Це особливо стосується побудови моделей, що включають велику кількість коефіцієнтів (параметрів підгонки) багатofакторних, нелінійних моделей.

13.2. ПОБУДОВА І АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ. ПРИКЛАД РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

13.2.1. Приклад 19. Побудова моделей множинної лінійної регресії

Дані

Вивчалися результати тестування курсантів військового ліцею, які виконували завдання із 30 тестів за обмежений час. В таблиці 19 додатку (Додаток. Приклад 19) наведено кількість отриманих правильних відповідей («Відповідь» – змінна Y) для вибірки із 38 курсантів. Поставлено задачу виявлення від яких показників: («Оперативна пам'ять» – змінна X_1), («Реакція на об'єкт» – змінна X_2), («Сенсомоторна реакція» – змінна X_3), («Вербальне мислення» – змінна X_4), («Психоемоційний статус» – змінна X_5), – залежить точність виконання завдання. При аналізі виходили із лінійної залежності результуючої ознаки Y від факторних ознак (X_1, X_2, X_3, X_4, X_5).

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 19.xls».

Задачу будемо розв'язувати в рамках побудови та аналізу багатофакторної лінійної моделі регресії.

EZR (R-statistics)

Statistical analysis→

Continuous variables→

Linear regression →

Обираємо вихідну змінну (Y)

Обираємо факторні змінні

(X1, X2, X3, X4, X5)

Виберемо пункт представлення графіку аналізу залишків

Обираємо метод вибору мінімального набору змінних, що включаються в модель (Stepwise selection based on p-value)

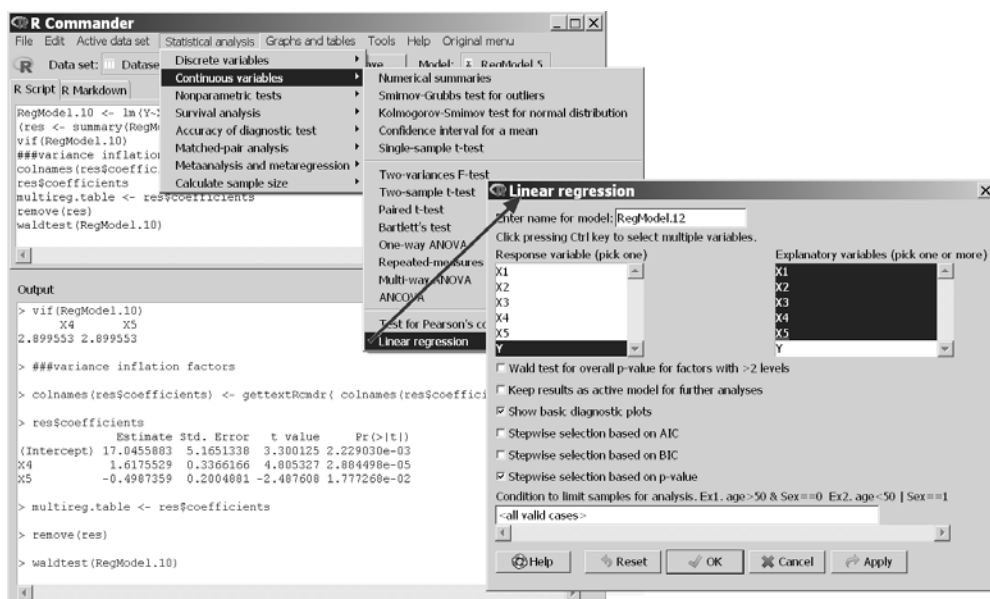


Рис. 13.1. Множинна лінійна регресія з «Приклад 19.xls» у пакеті EZR.

На рисунку 13.2 наведено результати розрахунків, на рисунку 13.3 – аналіз залишків, при побудові множинної лінійної регресії.

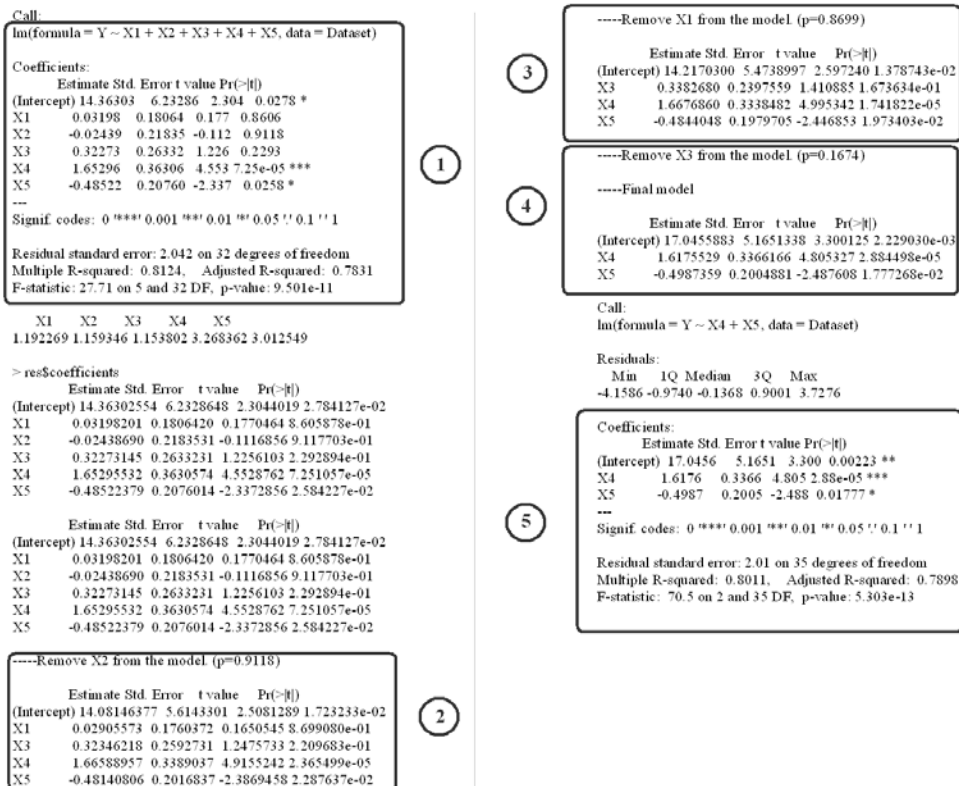


Рис. 13.2. Результати розрахунків для множинної лінійної регресії з «Приклад 19.xls» у пакеті EZR.

На рис. 13.2 наведено результати розрахунку коефіцієнтів та оцінка якості багатофакторної лінійної моделі регресії, побудованої на всіх п'яти факторних ознаках (1). В пунктах (2 – 4) рис. 13.2 наведено результати оцінки коефіцієнтів моделі при покроковому вилученні вхідних змінних. У пункті (5)

наведено результати розрахунку коефіцієнтів і оцінки якості кінцевої моделі прогнозування кількості правильних відповідей (Y), побудованої на двох значимих факторних змінних: «Вербальне мислення» (X4) та «Психоемоційний статус» (X5).

$$\text{lm}(Y \sim X4 + X5)$$

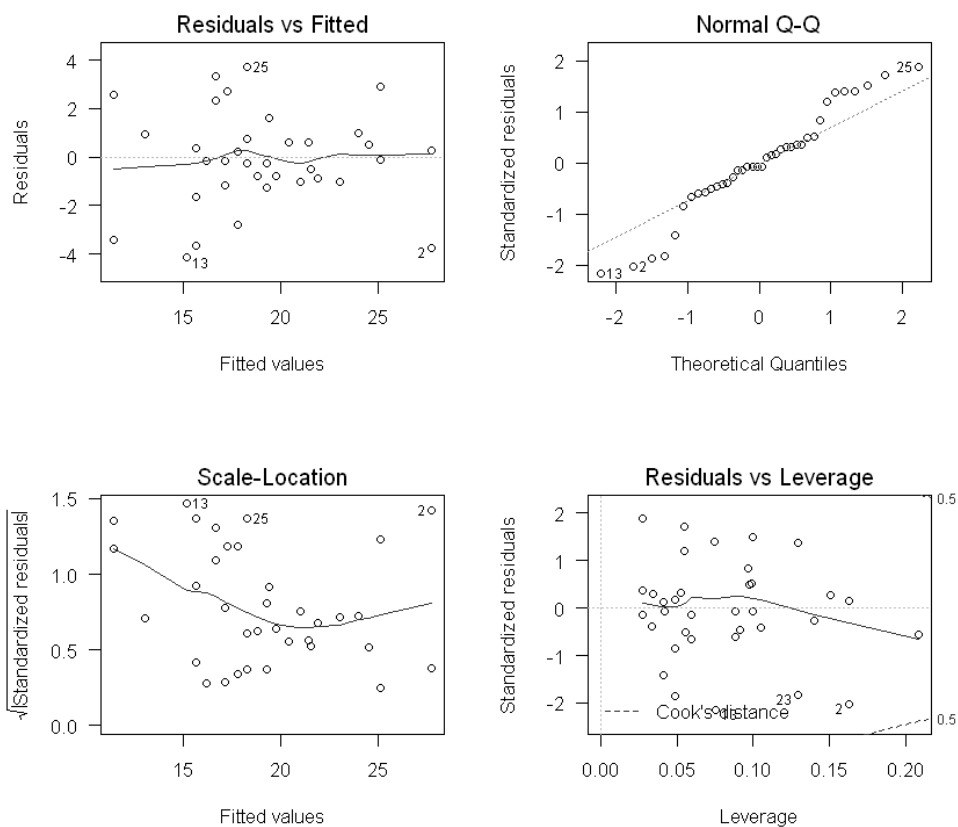


Рис. 13.3. Аналіз залишків для множинної лінійної регресії з «Приклад 19.xls» у пакеті EZR.

Результат

Для аналізу залежності числа отриманих правильних відповідей (Y) від показників: «Оперативна пам'ять» ($X1$), «Реакція на об'єкт» ($X2$), «Сенсомоторна реакція» ($X3$), «Вербальне мислення» ($X4$), «Психоемоційний статус» ($X5$), – що оцінювалися в балах,

використано метод побудови та аналізу багатofакторної лінійної моделі регресії. При аналізі виходили із лінійної залежності результуючої ознаки Y від факторних ознак ($X1, X2, X3, X4, X5$). Аналіз проводився на результатах обстеження 38 курсантів військового ліцею.

При побудові п'ятифакторної моделі виявлено залежність вихідної змінної від факторних ознак (корегований коефіцієнт детермінації $R^2_{adj} = 0.783$, значення $F=27.7$, $p<0.001$), середньоквадратичні залишки $RSE=2.04$ (при 32 ступенях вільності). В таблиці 13.1 наведені результати оцінки коефіцієнтів моделі.

Таблиця 13.1.

Коефіцієнти п'ятифакторної моделі прогнозування числа правильних відповідей курсантів військового ліцею.

Показник	Значення коефіцієнту моделі, $b \pm m_b$	Рівень значимості відмінності коефіцієнту від 0, p
Const	14.4±6.2	0.03*
X1	0.03±0.18	0.86
X2	-0.02±0.22	0.91
X3	0.32±0.26	0.23
X4	1.65±0.36	<0.001*
X5	-0.48±0.21	0.03*

Примітка: * – відмінність параметра моделі від 0 статистично значима, $p<0.05$.

Для відбору мінімального набору факторних ознак, пов'язаних з вихідною змінною було використано метод покрокового відкидання/включення змінних (Stepwise). Було виділено дві факторні ознаки, пов'язані з точністю виконання тесту курсантами: «Вербальне мислення» та «Психоемоційний статус». Модель, побудована на виділених факторних ознаках адекватна (корегований коефіцієнт детермінації $R^2_{adj} = 0.790$, значення $F=70.5$, $p<0.001$), середньоквадратичні залишки $RSE=2.01$ (при 35 ступенях вільності). При порівнянні прогностичних характеристик двофакторної моделі з якістю моделі, побудованої на всіх п'яти змінних не виявлено їх погіршення, таким чином, зменшення числа факторних ознак не призводить до погіршення прогнозу оцінки. У таблиці 13.2 наведені результати оцінки коефіцієнтів моделі.

Таблиця 13.2.

Коефіцієнти двофакторної моделі прогнозування числа правильних відповідей курсантів військового ліцею.

Показник	Значення коефіцієнту моделі, $b \pm m_b$	Рівень значимості відмінності коефіцієнту від 0, p
Const	17.0 ± 5.2	$<0.001^*$
X4	1.62 ± 0.34	$<0.001^*$
X5	-0.50 ± 0.20	0.02^*

Примітка: * – відмінність параметра моделі від 0 статистично значима, $p < 0.05$.

Таким чином, встановлено, що точність виконання тесту курсантами пов'язана ($p < 0.001$) з показником «Вербальне мислення» та «Психоемоційний статус». Виявлено (див. табл. 13.2), що з підвищенням бальної оцінки показника «Вербальне мислення» курсанта число правильних відповідей зростає ($p < 0.001$), в середньому на 1.62 ± 0.34 балу на кожен бал зростання показника «Вербальне мислення». Також встановлено, що з

підвищенням бальної оцінки показника «Психоемоційний статус» курсанта число правильних відповідей зменшується ($p = 0.02$), в середньому на 0.50 ± 0.20 балу на кожен бал зростання показника «Психоемоційний статус».

Отримана математична модель прогнозування числа правильних відповідей може бути виражена формулою:

$$Y = 17.0 + 1.62 \times X4 - 0.5 \times X5 \quad (13.6)$$

На рисунку 13.4 наведено діаграма розсіювання для побудованої моделі.

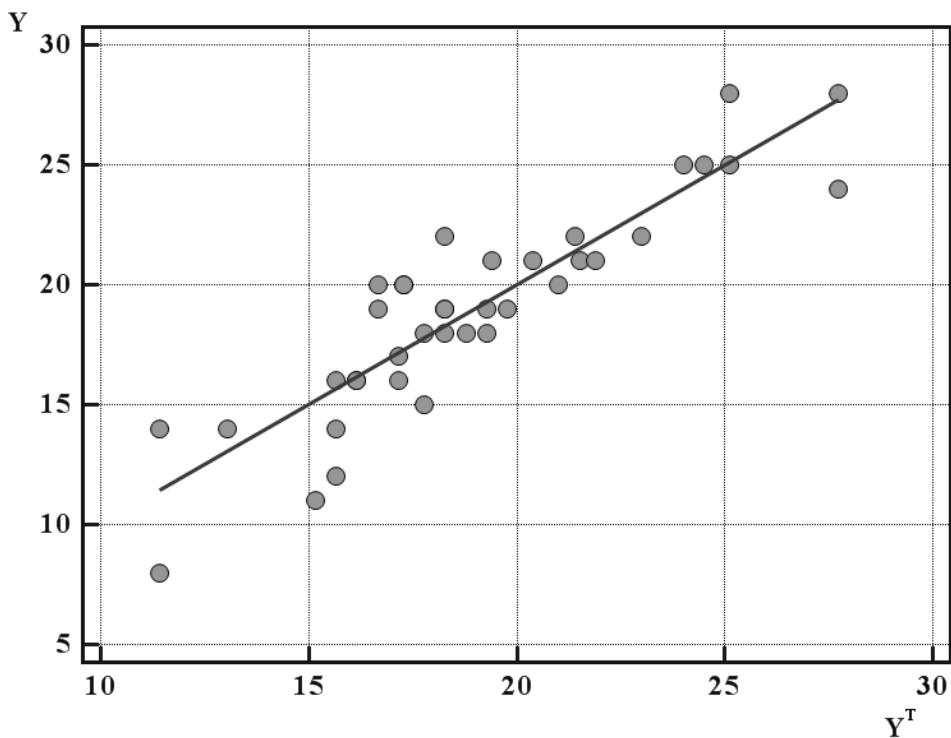


Рис. 13.4. Діаграма розсіювання для двофакторної моделі в координатах Y (експериментальне значення числа правильних відповідей), Y^T (розраховане значення числа правильних відповідей).

На рисунку 13.5 наведена діаграма аналізу залишків двофакторної моделі.

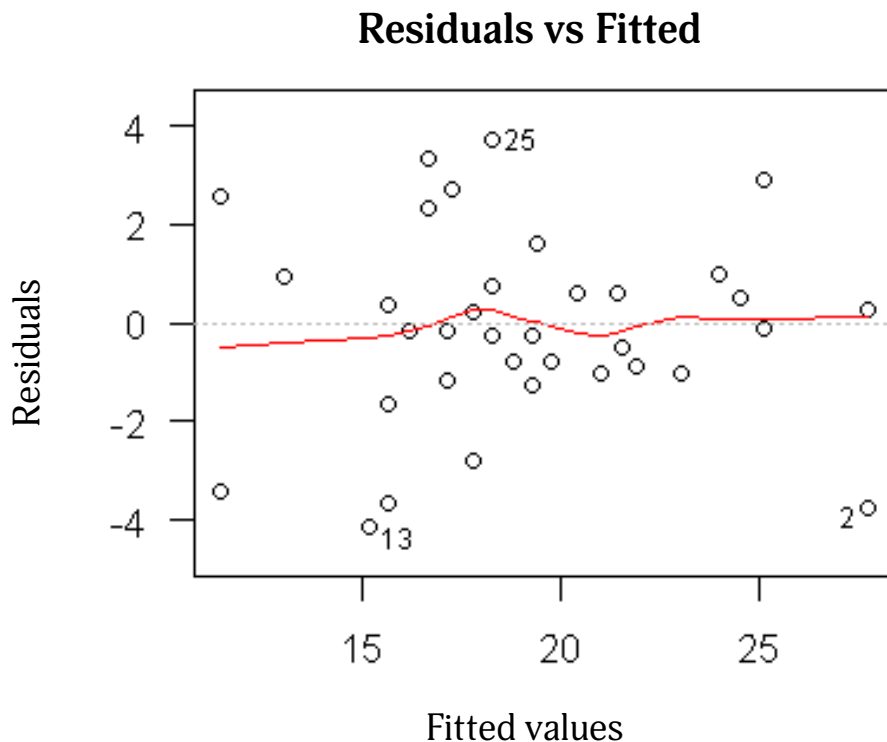


Рис. 13.5. Діаграма аналізу залишків для двофакторної лінійної моделі прогнозування числа правильних відповідей.

Проведений аналіз свідчить про адекватність лінійної двофакторної моделі прогнозу-

вання числа правильних відповідей експериментальним даним.

РОЗДІЛ 14. ПОБУДОВА І АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. ЛОГІСТИЧНА МОДЕЛЬ РЕГРЕСІЇ

При проведенні аналізу результатів медичних досліджень часто виникає задача прогнозування бінарної вихідної (Y) ознаки (наприклад, є ускладнення/немає ускладнень і т. і.) за впливом факторних змінних X_1, X_2, \dots, X_m . У цьому випа-

дку для проведення аналізу може бути використана регресія схожа з багатофакторною лінійною регресією – **логістична модель регресії**. Для логістичної моделі регресії залежність Y від X_1, X_2, \dots, X_m шукають у вигляді рівняння:

$$\ln\left(\frac{Y}{1-Y}\right) = b_0 + b_1 \times x_1 + b_2 \times x_2 + \dots + b_m \times x_m, \quad (14.1)$$

де змінна Y , що приймає значення ($Y=1$ – випадок, $Y=0$ – не випадок), називається вихідною змінною, а змінні X_j – вхідними змінними ($j=1, 2, \dots, m$), b_0 – вільний член регресії, b_1, b_2, \dots, b_m – коефіцієнти регресії при відповідному предикторі. Для такої моделі значення Y може бути інтерпретовано як імовірність випадку (як правило, захворювання, ускладнення і т. і.), а прогнозована в моделі величина $Y/(1-Y)$ – шанси випадку. Тоді величина $\exp(b_j)$ показує, у скільки разів зміняться шанси випадку при зміні X_j на 1 одиницю (при врахуванні «on adjustment» інших факторних ознак).

При побудові логістичної моделі регресії дослідником можуть бути поставлені задачі:

- **створення математичної моделі (формули) за якою, знаючи значення факторних змінних, може бути прогнозоване значення результуючої змінної Y – імовірність випадку;**
- **визначення факторних ознак, які значимо впливають на ризик випадку;**
- **оцінка сили впливу факторної ознаки на ризик випадку (при урахуванні впливу інших факторів).**

14.1. ПОБУДОВА І АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. ЛОГІСТИЧНА МОДЕЛЬ РЕГРЕСІЇ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Нехай ми маємо на меті побудувати мультиплікативну модель (логістичну модель регресії) залежності шансів результуючої бінарної змінної Y від факторних змінних X_1, X_2, \dots, X_m . Побудова й аналіз моделі (як і у випадку лінійної моделі регресії) проводиться за алгоритмом:

розрахунок коефіцієнтів $b_0, b_1, b_2, \dots, b_m$ логістичної моделі регресії на навчальній множині,

оцінка якості й адекватності моделі регресії загалом,

оцінка впливу для кожної окремої факторної ознаки,

підтвердження практичного використання математичної моделі на контрольній множині.

14.1.1. Розрахунок коефіцієнтів логістичної моделі регресії

Нехай у нас є вибірка з N пацієнтів, і для кожного із них проведено вимірювання значення факторних X_1, X_2, \dots, X_m ознак і відоме значення

бінарної вихідної ознаки Y (діагноз). Будемо вважати, що на генеральній сукупності залежність результуючої ознаки від предикторів є багатofакторною логістичною моделлю регресії.

Тоді для розрахунку значень b_i на навчальній множині, як правило, використовується метод найбільшої правдоподібності. Формули для розрахунку коефіцієнтів b_i наводити тут не будемо.

У результаті проведених розрахунків отримаємо точкову оцінку β -коефіцієнтів моделі (b_i) і стандартну похибку оцінки (m_i).

14.1.2. Оцінка якості й адекватності моделі

Для оцінки якості побудованої моделі в пакеті EZR може бути розраховано значення критерію хі-квадрат:

$$X_i - \text{квадрат} = -2 \times \ln \left(\frac{L}{L_0} \right), \quad (14.2)$$

де L_0 – функція правдоподібності для прогнозування вихідної змінної Y без урахування факторних ознак, L – функція правдоподібності для прогнозування вихідної

змінної Y з урахуванням факторних ознак. Якщо включення факторних ознак не змінює значення функції правдоподібності (хі-квадрат статистично значимо не відрізняється від 0, $p > 0.05$), то можна говорити про незалежність вихідної змінної від факторних змінних. Чим більше відмінність L_0 та L (чим більше значення хі-квадрат моделі), тим більший вплив факторних ознак на зміну результуючої ознаки.

Для перевірки відмінності від 0 для кожного j -го коефіцієнту моделі – β_j використовується розраховується значення статистики Вальда, яка відповідає на питання, чи впливає j -а факторна ознака (з урахуванням стандартизації за всіма іншими факторами) на вихідну змінну.

Для оцінки ступеню впливу для кожної j -ої факторної змінної в логістичній моделі регресії розраховується **показник відношення шансів (ВШ)**, що оцінює вплив j -ої факторної ознаки (з урахуванням стандартизації за всіма іншими факторами) на вихідну змінну. Значення показника j -ої

факторної ознаки $ВШ > 1$ ($b_j > 0$ при $p < 0.05$) указує на зростання шансів випадку при зростанні цієї ознаки, якщо ж $ВШ < 1$ ($b_j < 0$ при $p < 0.05$), це свідчить про зниження шансів випадку при зростанні цієї ознаки. Якщо ж значення показника $ВШ$ j -ої факторної ознаки статистично значимо не відрізняється від 1, то зміна цієї ознаки не пов'язана зі зміною ризику випадку.

Для оцінки якості прогнозування в математичній моделі (для оцінки діагностичного тесту) використовуються показники: **чутливість (Sensitivity)** моделі, **специфічність (Specificity)** моделі, **прогностичність позитивного результату (Positive predictive value)**, **прогностичність негативного результату (Negative predictive value)**.

Для розуміння цих показників розглянемо чотирипільну таблицю спряженості (див. Таблицю 14.1), що побудована на основі результатів класифікації в математичній моделі (або деякого тесту).

Таблиця 14.1.

Оцінка прогностичних характеристик

Рішення, прийняте в моделі (тесті)	Фактичний результат	
	Випадок	Не випадок
Позитивне	TP	FP
Негативне	FN	TN

Тут:

TP (*True Positive*) – кількість правильно класифікованих випадків (як правило, негативних для пацієнта наслідків, як то, є захворювання, є рецидив, не досягнуто вилікування, тощо);

TN (*True Negative*) – кількість правильно класифікованих не випадків (як правило, позитивних для пацієнта наслідків, як то, немає захворювання, немає рецидиву, досягнуто вилікування, тощо);

FN (*False Negative*) – хибно негативні приклади, кількість неправильно класифікованих випадків (похибка першого роду), коли випадок хибно не прогнозується;

FP (*False Positive*) – хибно позитивні приклади, кількість неправильно класифікованих не випадків (похибка другого роду), коли хибно прогнозується випадок.

Чутливістю моделі (тесту) називається доля пацієнтів, у яких фактично був випадок і модель для них дала правильний прогноз (доля хворих, для яких модель дала правильний прогноз):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \quad (14.3)$$

Специфічністю моделі (тесту) називається доля пацієнтів, у яких фактично був не випадок і модель для них дала правильний прогноз (доля не хворих, для яких модель дала правильний прогноз):

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \quad (14.4)$$

Прогностичністю позитивного результату (**PPV**) моделі (тесту) називається доля пацієнтів, для яких було дано позитивний прогноз

(є випадок) і у них фактично був випадок (доля правильних прогнозів, для яких модель прогнозує наявність захворювання). Якщо Таблиця 14.1 відповідає розповсюдженості (**Prevalence**) випадку, то **PPV** розраховується за формулою:

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% ,$$

(14.5)

В іншому випадку, для розрахування **PPV** на практиці, потрібно враховувати розповсюдженість:

$$\text{PPV} = \frac{\text{Sensitivity} \times \text{Prevalence}}{\text{Sensitivity} \times \text{Prevalence} + (1 - \text{Specificity}) \times (1 - \text{Prevalence})} \times 100\%$$

(14.6)

Прогностичністю негативного результату (**NPV**) моделі (тесту) називається доля пацієнтів, для яких було дано негативний прогноз (немає випадку) і у них фактично був не випадок (доля вірних прогнозів, для яких модель прогнозувала відсутність захворювання). Якщо Таблиця 14.1 відповідає розповсюдженості (**Prevalence**)

випадку то **NPV** розраховується за формулою:

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \times 100\% ,$$

(14.7)

В іншому випадку, для розрахування **NPV** на практиці, потрібно враховувати розповсюдженість:

$$\text{NPV} = \frac{\text{Specificity} \times (1 - \text{Prevalence})}{(1 - \text{Sensitivity}) \times \text{Prevalence} + \text{Specificity} \times (1 - \text{Prevalence})} \times 100\%$$

(14.8)

Ці чотири показники дають оцінку прогностичним якостям моделі, зрозуміло, що якісна модель повинна давати вірний прогноз і для «хворих» (мати високу, близьку до 100%, **Sensitivity**) і для здорових (мати високу, близьку до 100%,

Specificity). Крім того, для практичного застосування моделі важливим є те, що, якщо модель дала позитивний прогноз, то пацієнт з високою імовірністю буде дійсно «хворим» (високе, близьке до 100% значення **PPV**), якщо ж мо-

дель дала негативний прогноз, то пацієнт з високою імовірністю буде дійсно «здоровим» (високе, близьке до 100% значення NPV).

У той же час, слід зазначити, що три із цих величин є незалежними оцінками, тобто, у загальному випадку:

$$PPV \neq \text{Sensitivity}, \\ NPV \neq \text{Specificity}.$$

14.1.3. Оцінка якості моделі. Побудова та аналіз кривих операційних характеристик

Для оцінки якості логістичної моделі регресії (або деякого тесту) часто використовується метод побудови та аналізу кривої операційних характеристик моделі (*Receiver Operating Characteristic – ROC curve analysis*).

Нехай побудовано модель прогнозування деякого захворювання, або запропоновано виставляти діагноз за результатом деякого тесту. Тоді для кожного із пацієнтів, якому виставляється діагноз, ставиться у відповідність деяке число (результат). При цьому,

якщо це число менше або дорівнює деякому граничному значенню ($Z_{\text{гр}}$) – виставляється діагноз «здоровий», якщо ж це число більше $Z_{\text{гр}}$ – виставляється діагноз «хворий». Прогностичні характеристики моделі залежать від цього граничного значення, наприклад, якщо ми візьмемо дуже низьке значення $Z_{\text{гр}}$, то всім пацієнтам буде виставлено діагноз «хворий» ($\text{Sensitivity}=100\%$ але $\text{Specificity}=0\%$). Якщо ж взяти велике значення $Z_{\text{гр}}$, то всім пацієнтам буде виставлено діагноз «здоровий» ($\text{Sensitivity}=0\%$ але $\text{Specificity}=100\%$). При зміні значення $Z_{\text{гр}}$ Sensitivity і Specificity тесту будуть змінюватись. Криву залежності чутливості моделі від її специфічності при переборі всіх значень $Z_{\text{гр}}$ називають **кривою операційних характеристик моделі (ROC-curve)**. На рисунку 14.1 представлено ROC-криву для деякої моделі.

Якість побудованої моделі може бути оцінена за площею під ROC-кривою (AUC, заштрихована область на рисунку), чим більше значення AUC, тим кращі прогностичні характеристики моделі (для ідеальної моделі – $\text{Sensitivity}=100\%$ і $\text{Specificity}=100\%$ площа під ROC-кривою, $AUC=1$). Модель буде

адекватною експериментальним даним якщо AUC статистично значимо ($p < 0.05$) перевищує 0.5 (ROC-крива не співпадає з діагональною лінією). Якщо ж AUC не

відрізняється від 0.5, то модель прогнозує не краще ніж при випадковому виборі прогнозу (за частотою випадків), модель не адекватна експериментальним даним.

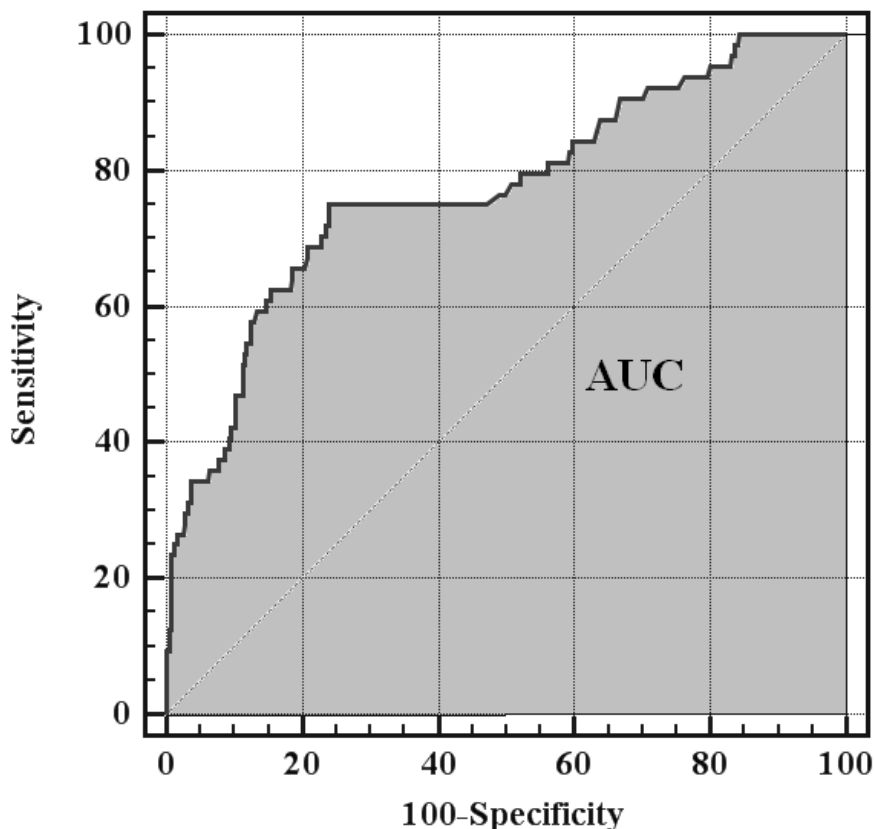


Рис. 14.1. Крива операційних характеристик моделі.

Якість моделі умовно може бути оцінена як:
 відмінна (при $AUC \geq 0.9$),
 дуже добра (при $0.8 \leq AUC \leq 0.9$),
 добра (при $0.7 \leq AUC \leq 0.8$),

*задовільна (при $0.6 \leq AUC \leq 0.7$),
 погана (при $0.5 \leq AUC \leq 0.6$).*

Показник AUC часто використовується для порівняння якості моделей (тестів).

14.1.4. Відбір незалежних ознак

Для відбору мінімальної кількості ознак, що забезпечують якісні прогностичні характеристики логістичної моделі регресії в пакеті EZR може бути використаний *метод покрокового включення/виключення факторних змінних (stepwise method)*. В якості умов включення та виключення, як і для випадку множинної лінійної регресії, можна обранти:

інформаційний критерій Акаїке (Akaike information criterion, AIC),

байєсівський інформаційний критерій (Bayesian information criterion – BIC), критерій досягнення порогу значимості.

Інформаційний критерій Акаїке

При виборі цього методу відбору незалежних ознак, що включаються у модель для оцінки цінності моделі розраховується показник AIC:

$$AIC = 2 \times m + N \times \chi^2, \quad (14.9)$$

де m – кількість факторних змінних, що включені у модель, N –

об'єм вибірки, χ^2 – хі-квадрат статистика моделі. При включенні та виключенні змінних до моделі змінюється перший і другий доданок формули (14.9). За кінцеву обирається модель, що має найменше значення AIC.

Байєсівський інформаційний критерій

При виборі методу BIC розраховується показник BIC:

$$BIC = m \times \ln(N) + N \times \chi^2, \quad (14.10)$$

де m – кількість факторних змінних, що включені у модель, N – об'єм вибірки, χ^2 – хі-квадрат статистика моделі. За кінцеву обирається модель, що має найменше значення BIC.

Критерій досягнення порогу значимості

При виборі цього методу, включення j -ої змінної у модель відбувається тоді, коли рівень значимості відмінності відповідного коефіцієнта логістичної моделі регресії від нуля менше за поріг включення (p_{enter}), виключення j -ої змінної з моделі відбувається тоді, коли рівень значимості відмінно-

сті відповідного коефіцієнта моделі від нуля перевищує поріг виключення (p_{rem}). Після перебору всіх змінних обирається кінцева модель.

14.1.5. Підтвердження практичного використання математичної моделі на контрольній множині

Для упередження перенавчання моделі після її побудови (підгонки коефіцієнтів моделі) її тестують на даних, що не були використані при побудові моделі (контрольній множині) і якщо отримують ту ж якість прогнозування, що і на навчальній множині даних, то можна говорити про відсутність підгонки і про можливість використання результатів прогнозування на практиці.

Слід зауважити, що коли мова йде про логістичні моделі регресії, то питання практичного використання моделі (для прогнозування діагнозу, тощо) часто не являється актуальним тому, що ці моделі найчастіше використовують для вибору факторних ознак, що пов'язані з результуючою змінною, для оцінки сили незалежного впливу цих ознак (при урахуванні інших факторів) на результат, а не для прогнозування результату на практиці.

14.2. ПОБУДОВА І АНАЛІЗ МАТЕМАТИЧНИХ МОДЕЛЕЙ. ЛОГІСТИЧНА МОДЕЛЬ РЕГРЕСІЇ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

14.2.1. Приклад 20. Побудова логістичних моделей регресії

Дані

Вивчалися результати лікування хворих на виразкову хворобу шлунку, що перенесли операційне втручання з приводу кровотечі. В таблиці 20 додатку (Додаток. Приклад 20) наведені результати лікування («Результат» – змінна Y : $Y = \langle 0 \rangle$ – пацієнт виписаний, $Y = \langle 1 \rangle$ – летальний випадок) для вибірки із 30 пацієнтів. Поставлено задачу виявлення факторних ознак, пов'язаних із ризиком летального випадку: («Вік» – змінна X_1 , у роках), («Блювота кров'ю» – змінна X_2 , $\langle 1 \rangle$ – є, $\langle 0 \rangle$ – немає), («Втрата свідомості» – змінна X_3 , $\langle 1 \rangle$ – так, $\langle 0 \rangle$), («Біль у епігастрії» – змінна X_4 , $\langle 1 \rangle$ – так, $\langle 0 \rangle$ – ні), («Наявність згустків» – змінна X_5 , $\langle 1 \rangle$ – є, $\langle 0 \rangle$ – немає).

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 20.xls».

Задачу будемо розв'язувати в рамках побудови та аналізу багатофакторної логістичної моделі регресії.

EZR (R-statistics)

Statistical analysis→

Discrete variables→

Logistic regression →

Обираємо вихідну змінну (Y)

Обираємо факторні змінні (X1, X2, X3, X4, X5)

Виберемо пункт оцінки значимості відмінності коефіцієнтів моделі від 0 (за критерієм Вальда)

Виберемо пункт побудови кривих операційних характеристик моделі

Обираємо метод вибору мінімального набору змінних, що включаються в модель (Stepwise selection based on p-value)

На рисунку 14.3 наведено результати розрахунків, до побудови логістичної моделі регресії, на рисунку 14.4 – ROC-крива для кінцевої двофакторної моделі логістичної регресії.

На рис. 14.3 в пункті (1) наведено результати розрахунку коефіцієнтів оцінка показників ВШ (2), та оцінка якості прогнозування (3) для логістичної моделі регресії, побудованої на всіх п'яти факторних ознаках.

У пунктах (4 та 5) рис. 14.3 наведено результати покрокового аналізу. В пункті (6) наведено результати розрахунку коефіцієнтів, показників ВШ (7) і оцінки якості кінцевої моделі прогнозування імовірності летальних випадків (8), побудованої на двох значимих факторних змінних: «Вік» (X1) та «Наявність згустків» (X5).



Рис. 14.2. Логістична модель регресії з «Приклад 20.xls» у пакеті EZR.

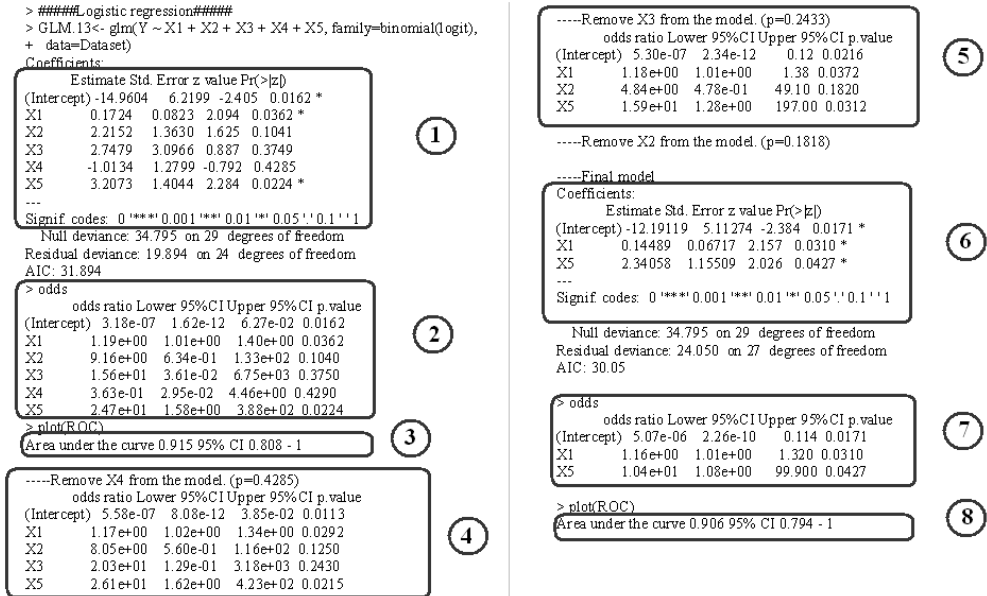


Рис. 14.3. Результати розрахунків для моделі логістичної регресії з «Приклад 20.xls» у пакеті EZR.

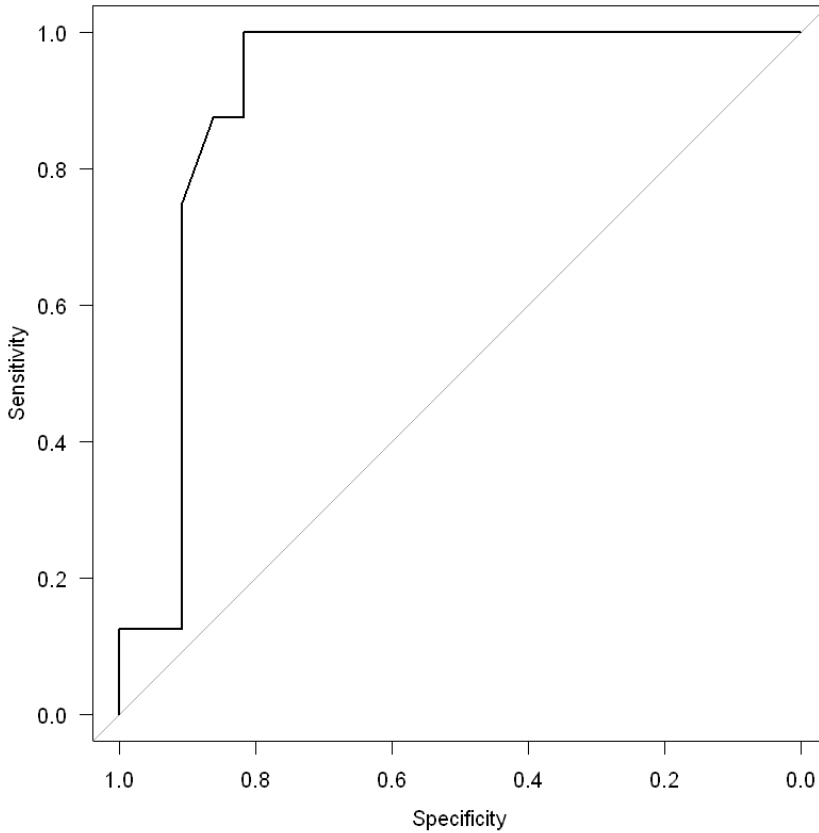


Рис. 14.4. ROC-крива двофакторної логістичної моделі регресії з «Приклад 20.xls» у пакеті EZR.

Аналіз ROC-кривої кінцевої двофакторної моделі прогнозування ризику летального випадку для хворих на виразкову хворобу шлунку, що перенесли операційне втручання з приводу кровотечі (див. Рис.14.3 (8) та Рис.14.4) свідчить про наявність його зв'язку з «Віком» пацієнта та «Наявністю згустків», AUC= 0.91 (95% ВІ 0.79

– 1.00), статистично значимо ($p < 0.05$) більше 0.5.

Результат

Для аналізу зв'язку ризику летального випадку (Y) для хворих на виразкову хворобу шлунку, що перенесли операційне втручання з приводу кровотечі з факторними ознаками: «Вік» (X1), «Блювота

кров'ю» (X2), «Втрата свідомості» (X3), «Біль у епігастрії» (X4), «Наявність згустків» (X5), – використано метод побудови та аналізу моделей логістичної регресії. Аналіз проводився на результатах обстеження 30 хворих (8 летальних випадків).

При побудові п'ятифакторної моделі виявлено залежність ризику летального випадку від факторних ознак, площа під ROC-кривою $AUC = 0.92$ (95% ВІ 0.79 – 1.00), статистично значимо ($p < 0.05$) перевищує 0.5, що є свідченням адекватності побудованої моделі. У таблиці 14.2 наведені результати оцінки коефіцієнтів моделі.

Таблиця 14.2.

Коефіцієнти п'ятифакторної моделі прогнозування ризику летального випадку для хворих на виразкову хворобу шлунку, що перенесли операційне втручання з приводу кровотечі.

Показник	Значення коефіцієнту моделі, $b \pm m_b$	Рівень значимості відмінності коефіцієнту від 0, p	Показник ВІ (95% ВІ)
Const	-15.0 ± 6.2	0.016*	–
X1	0.17 ± 0.08	0.036*	1.19 (1.01 – 1.40)
X2	2.2 ± 1.4	0.104	–
X3	2.7 ± 3.1	0.375	–
X4	-1.0 ± 1.3	0.429	–
X5	3.2 ± 1.4	0.022*	24.7 (1.6 – 388)

Примітка: * – відмінність параметра моделі від 0 статистично значима, $p < 0.05$.

Для відбору мінімального набору факторних ознак, пов'язаних з вихідною змінною було використано метод покрокового відкидання/включення змінних (Stepwise). Було виділено дві факторні ознаки, пов'язані з ризиком летального випадку: «Вік» та «Наявність згустків».

На виділених факторних ознаках була побудована логістична модель прогнозування $AUC = 0.91$ (95% ВІ 0.79 – 1.00), статистично значимо ($p < 0.05$) перевищує 0.5, що є свідченням адекватності побудованої моделі. При порівнянні прогностичних характеристик двофакторної моделі з якістю моде-

делі, побудованої на всіх п'яти змінних не виявлено їхнього погіршення. У таблиці 14.3 наведені результати оцінки коефіцієнтів моделі.

Отримана математична модель прогнозування ризику летального випадку може бути виражена формулою:

$$\ln\left(\frac{Y}{1-Y}\right) = -12.2 + 0.14 \times X1 + 2.3 \times X5, \quad (14.11)$$

Таблиця 14.3.

Коефіцієнти двофакторної моделі прогнозування ризику летального випадку для хворих на виразкову хворобу шлунку, що перенесли операційне втручання з приводу кровотечі.

Показник	Значення коефіцієнту моделі, $b \pm m_b$	Рівень значимості відмінності коефіцієнту від 0, p	Показник ВШ (95% ВІ)
Const	-12.2±5.1	0.017*	–
X1	0.14±0.07	0.031*	1.16 (1.01 – 1.32)
X5	2.3±1.2	0.043*	10.4 (1.1 – 100)

Примітка: * – відмінність параметра моделі від 0 статистично значима, $p < 0.05$.

Таким чином, встановлено, що ризик летального випадку пов'язаний ($p < 0.05$) з «Віком» та «Наявністю згустків». Встановлено (див. табл. 14.3), що з підвищенням віку пацієнта ризик летального випадку зростає ($p = 0.031$), ВШ= 1.16 (95% ВІ 1.01 – 1.32) на кожен рік. Також встановлено, що при наявності згустків ризик летального випадку зростає ($p = 0.043$), ВШ= 10.4 (95% ВІ 1.1 – 100).

Дані

Вивчалися результати лікування хворих на виразкову хворобу шлунку, що перенесли операційне втручання з приводу кровотечі. В результаті проведеного аналізу встановлено, що ризик летального випадку (P) цих пацієнтів пов'язаний із віком хворого ($X1$) і наявністю згустків крові ($X5$). Побудовано двофакторну модель прогнозування ризику, що описується формулою:

14.2.2. Приклад 21. Побудова ROC-кривої тесту

$$\ln\left(\frac{P}{1-P}\right) = -12.2 + 0.14 \times X1 + 2.3 \times X5, \quad (14.12)$$

Було розраховано значення показника для кожного із 30 пацієнтів. У таблиці 21 додатку (Додаток. Приклад 21) наведені фактичні результати лікування («Результат» – змінна Y: Y= «0» – пацієнт виписаний, Y= «1» – летальний випадок) та значення прогнозу (P) в моделі (14.12) для вибірки із 30 пацієнтів. Поставлено задачу визначення граничного значення $P_{гр.}$ при якому прогнозується негативний результат лікування та оцінити прогностичні характеристики тесту для вибраного порогу.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 21.xls».

Задачу будемо розв'язувати методом побудови кривої операційних

характеристик запропонованого тесту (P).

EZR (R-statistics)

Statistical analysis →

Accuracy of diagnostic test →

ROC curve analysis for quantitative test → *Обираємо вихідн змінну (Y)*

Обираємо факторну змінну P (результат тесту)

Оптимальну границю будемо визначати за методом максимізації чутливість + специфічність

Цінність прогнозу «випадку» та «не випадку» вважаємо однаковою, тому обираємо Cost of false negative classification= 1

Будемо вважати, що розповсюдженість «випадку» відповідає нашим даним, тому обираємо Prevalence=0.267

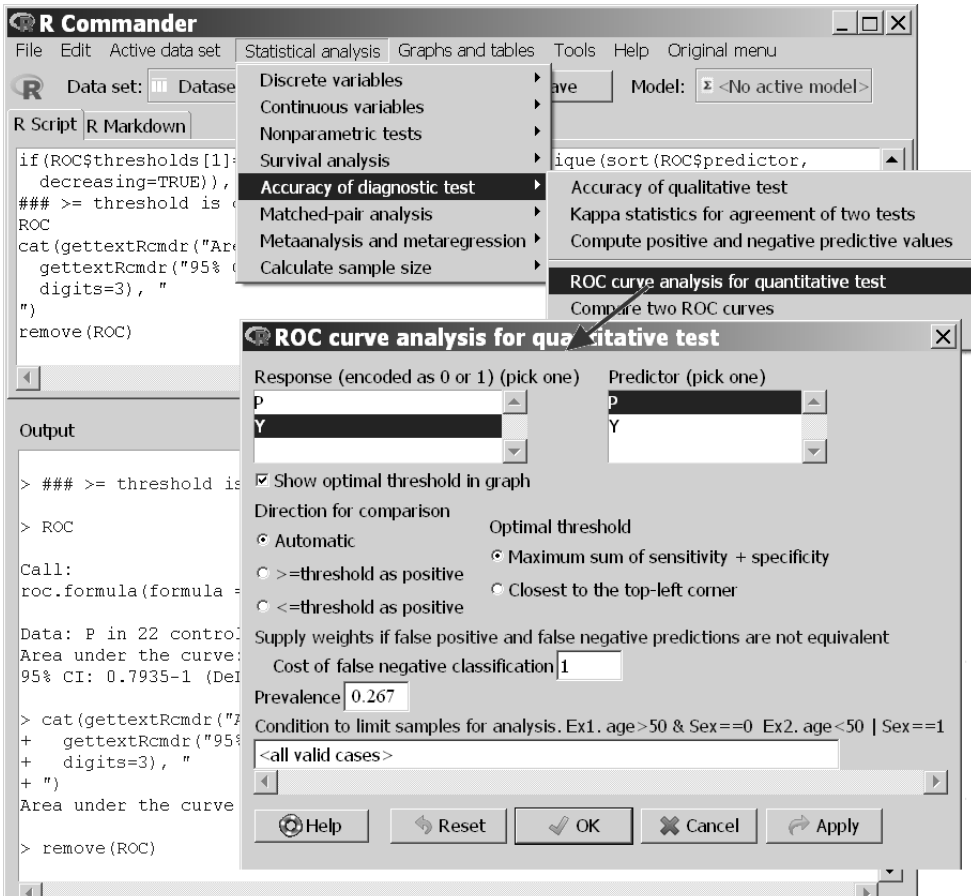


Рис. 14.5. Побудова ROC-кривої з «Приклад 21.xls» у пакеті EZR.

На рисунку 14.6 наведено результати розрахунків перебору порогів прийняття рішення (1), та оці-

нки AUC (2) тесту, на рисунку 14.7 – ROC-крива з визначенням оптимальної границі.

```
> #####ROC curve analysis for quantitative test#####
```

```
> if(ROC$thresholds[1]==-Inf){ coords(ROC, x=c(-Inf,
+ unique(sort(ROC$predictor)), Inf))}
```

```
-Inf 0.005 0.008 0.009 0.013 0.019 0.026
threshold -Inf 0.005 0.00800000 0.00900000 0.0130000 0.0190000 0.0260000
specificity 0 0.000 0.04545455 0.09090909 0.1363636 0.1818182 0.2272727
sensitivity 1 1.000 1.00000000 1.00000000 1.0000000 1.0000000 1.0000000
0.046 0.088 0.09 0.114 0.13 0.132 0.147
threshold 0.0460000 0.0880000 0.0900000 0.1140000 0.1300000 0.132 0.1470000
specificity 0.2727273 0.3181818 0.3636364 0.4090909 0.4545455 0.500 0.5454545
sensitivity 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.000 1.0000000
0.187 0.19 0.239 0.262 0.291 0.322
threshold 0.1870000 0.1900000 0.2390000 0.2620000 0.2910000 0.3220000
specificity 0.6363636 0.7727273 0.8181818 0.8181818 0.8636364 0.9090909
sensitivity 1.0000000 1.0000000 1.0000000 0.8750000 0.8750000 0.7500000
0.567 0.572 0.641 0.734 0.884 0.898 Inf
threshold 0.5670000 0.5720000 0.6410000 0.7340000 0.8840000 0.898 Inf
specificity 0.9090909 0.9090909 0.9090909 0.9090909 0.9545455 1.000 1
sensitivity 0.6250000 0.5000000 0.2500000 0.1250000 0.1250000 0.125 0
```

①

```
> ### >= threshold is considered positive
```

```
> ROC
```

```
Data: P in 22 controls (Y 0) < 8 cases (Y 1).
Area under the curve: 0.9062
95% CI: 0.7935-1 (DeLong)
```

②

Рис. 14.6. Результати розрахунків для побудови ROC-кривої з «Приклад 21.xls» у пакеті EZR.

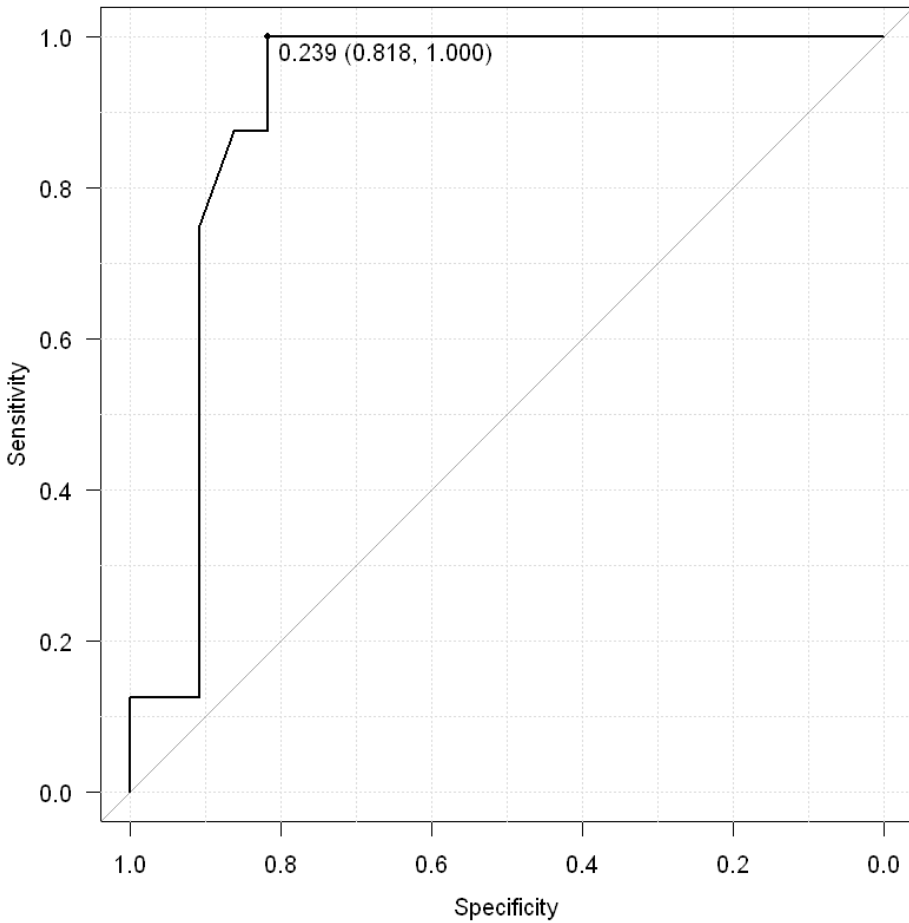


Рис. 14.7. ROC-крива тесту з «Приклад 21.xls» у пакеті EZR.

Таким чином, оптимальне граничне значення для тесту визначено рівним $P_{гр.} = 0.239$: для $P(\text{пацієнта}) \geq P_{гр.}$ – прогнозується ризик «випадку», для $P(\text{пацієнта}) < P_{гр.}$ –

прогнозується «не випадок». За даними з таблиці «Приклад 21.xls» визначимо кількість вірних та неправильних прогнозів при виборі цього граничного значення (Таблиця 14.4).

Таблиця 14.4.

Оцінка прогностичних характеристик тесту з «Приклад 21.xls».

Рішення, прийняте в тесті ($R_{гр}=0.239$)	Фактичний результат	
	Випадок	Не випадок
Позитивне	8	4
Негативне	0	18

Для оцінки прогностичних характеристик розрахуємо:
EZR (R-statistics)
Statistical analysis →

Accuracy of diagnostic test →
Accuracy of qualitative test →
 Заносимо в таблицю дані з
 Таблиці 14.4

The screenshot shows the R Commander interface. The menu path is: **Statistical analysis** → **Accuracy of diagnostic test** → **Accuracy of qualitative test**. A dialog box titled "Accuracy of qualitative test" is open, with the following data entered:

Number	Disease (+)	(-)
Test (+)	8	4
Test (-)	0	18

The Output window shows the following results:

```
> epi.tests(.Table, conf.level = 0.95)
      Disease positive Disease negative Total
Test positive          8             4     12
Test negative          0            18     18
Total                  8            22     30

Point estimates and 95 % CIs:
-----+-----
              Estimation Lower CI Upper CI
Apparent prevalence      0.400  0.227  0.594
True prevalence          0.267  0.123  0.459
Sensitivity               1.000  0.518  1.000
Specificity              0.818  0.597  0.948
Positive predictive value 0.667  0.349  0.901
Negative predictive value 1.000  0.740  1.000
Diagnostic accuracy      0.867  0.693  0.962
Likelihood ratio of a positive test 5.500  2.267  13.345
Likelihood ratio of a negative test 0.000  0.000  NaN
```

Рис. 14.8. Оцінка прогностичних характеристик тесту з «Приклад 21.xls» у пакеті EZR.

На рисунку 14.8 наведено результати розрахунку прогностичних характеристик при вибраній оптимальній границі.

Результат

Для визначення граничного значення $P_{гр}$ запропонованого тесту прогнозування ризику летального випадку для хворих на виразкову хворобу шлунку, що перенесли операційне втручання з приводу

кровотечі при якому прогнозується негативний результат лікування та оцінки його прогностичних характеристик було використано метод побудови та аналізу кривої операційних характеристик. Аналіз проводився на результатах лікування 30 хворих (8 летальних випадків).

На рисунку 14.9 приведено ROC-криву тесту.

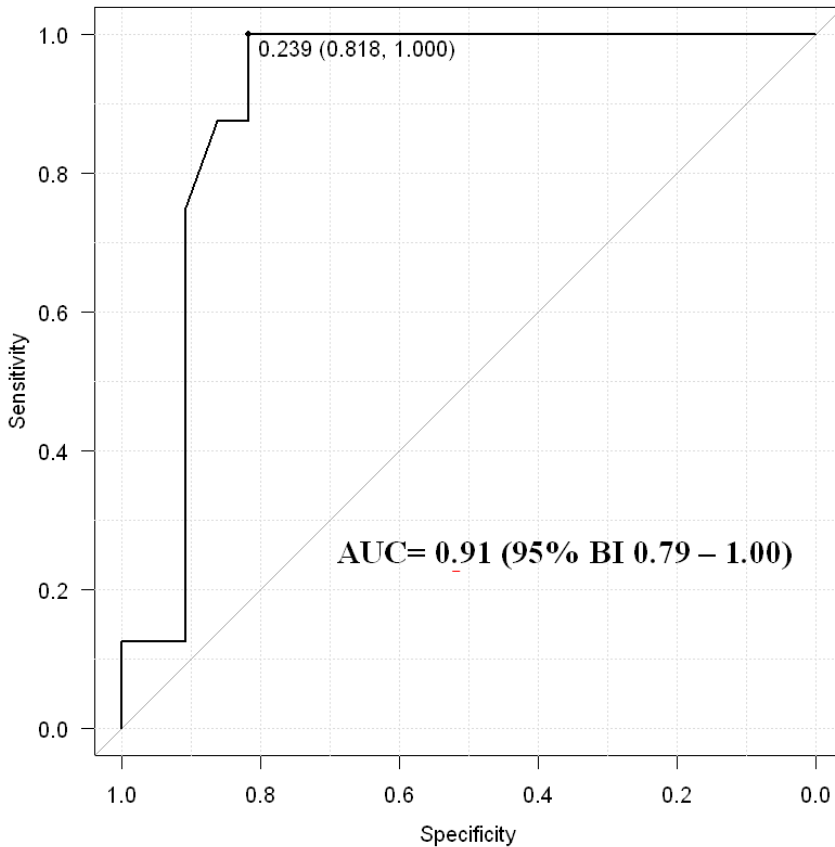


Рис. 14.9. ROC-крива тесту прогнозування ризику летального випадку для хворих на виразкову хворобу шлунку, що перенесли операційне втручання з приводу кровотечі.

Пропонований тест дозволяє прогнозувати ризик летального випадку, $AUC=0.91$ (95% ВІ 0.79 – 1.00), статистично значимо ($p<0.05$) відрізняється від 0.5. Для вибору оптимального граничного значення тесту було використано метод розрахунку Yoden Index (J):

$$J = \max \left\{ \text{чутливість}_{\text{гр.}} + \text{специфічність}_{\text{гр.}} \right\}, \quad (14.13)$$

де чутливість_{гр.} – значення чутливості тесту при даному значенні $P_{\text{гр.}}$, специфічність_{гр.} – значення специфічності тесту при даному значенні $P_{\text{гр.}}$. Оптимальна границя прийняття рішення склала $P_{\text{гр.}}=0.239$: для $P(\text{пацієнта}) \geq P_{\text{гр.}}$ – прогнозується ризик летального випадку, для $P(\text{пацієнта}) < P_{\text{гр.}}$ – прогнозується сприятливий результат лікування.

При обраному порозі прийняття рішення чутливість тесту складає 100% (95% ВІ 51.8% – 100%), специфічність тесту – 81.8% (95% ВІ 59.7% – 94.8%), прогностичність позитивного результату тесту $PPV=66.7\%$ (95% ВІ 34.9% – 90.1%), прогностичність негативного результату тесту $NPV=100\%$ (95% ВІ 74.0% – 100%).

РОЗДІЛ 15. АНАЛІЗ ВИЖИВАНOSTI

Нехай дослідника цікавить проміжок часу, що потрібний пацієнту для досягнення якоїсь кінцевої точки. Для обчислення цього часу нам потрібно знати два моменти часу: момент початку відліку (T_1) і момент досягнення кінцевої точки (T_2), – тоді проміжок часу (ΔT) розраховується як $\Delta T = T_2 - T_1$. Це може бути оцінка ефективності терапевтичного лікування

(де T_1 – дата початку лікування, T_2 – дата досягнення ефекту) або оцінка ефективності оперативного втручання (де T_1 – дата проведення оперативного втручання, T_2 – дата досягнення ефекту).

На рисунку 15.1 наведено приклад аналізу безрецидивного виживання хворих після проведення оперативного втручання.

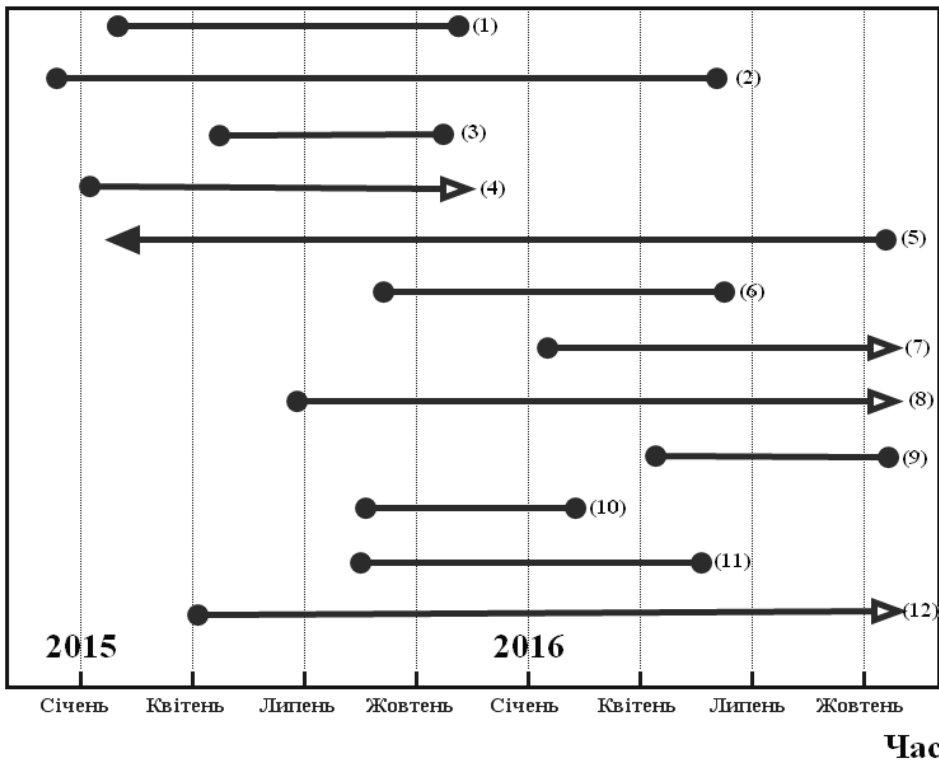


Рис. 15.1. Аналіз безрецидивного виживання хворих після проведення оперативного втручання. Кругами позначені дати операції та рецидиву, трикутниками – цензуровані дані.

Для оцінки ефективності лікування дослідник оцінює проміжок часу від операції до появи рецидиву (дослідження закінчилось 15 листопада 2016 року). Тоді для пацієнта (1) час безрецидивного виживання складе 9.2 місяці, для пацієнта (2) – 18 місяців і т.д. Але, наприклад, для пацієнта (4) останню інформацію про відсутність рецидивів ми мали на 18 листопада 2015 року (далі він виїхав і інформації не маємо), для нього час безрецидивного виживання становить >10.3 місяця, а для пацієнтів (7), (8) та (12) рецидив на момент закінчення дослідження не настав і ми знову не маємо можливості обчислити час безрецидивного виживання. У таких випадках, коли нам невідомий момент досягнення кінцевої точки дослідження (або момент початку відліку), ми говоримо про цензуровані дані. Так, для нашого прикладу для пацієнтів (1), (2), (3), (6), (9), (10), (11) ми маємо dokonані дані; а для пацієнтів (4), (7), (8), (12) маємо *цензуровані справа* і для пацієнта (5) маємо *цензуровані зліва* – недоконані дані. На таких результатах розрахунок оцінки середнього значення часу безрецидивного виживання та його варіабельності із використанням стандартних методів оцінки варіа-

ційних рядів неприпустимі і можуть призвести до хибних висновків. Для аналізу цензурованих даних розроблено ряд спеціальних методів аналізу, відомих як *методи аналізу виживаності*.

Нехай дослідник проводить аналіз цензурованих даних. При цьому можуть бути поставлені задачі:

представлення цензурованих даних,

однофакторний аналіз – порівняння цензурованих даних для двох і більше груп,

багатофакторний аналіз – створення математичної моделі прогнозування результуючої змінної (імовірності випадку) на основі факторних змінних та оцінка сили впливу факторної ознаки на ризик випадку (при урахуванні впливу інших факторів) для цензурованих даних.

15.1. АНАЛІЗ ВИЖИВАНOSTI. ТЕОРЕТИЧНІ ВІДОМОСТІ

15.1.1. Представлення даних

Для представлення цензурованих даних розраховується кумулятивна доля тих, хто вижив на поча-

ток відповідного часового інтервалу (функція виживання). **Функція виживання** ($S(t)$) може бути розрахована (**метод Каплана-Мейєра**) за формулою:

$$S(t) = \prod_{j=1}^{j(t)} \left(\frac{N - j}{N - j + 1} \right)^{d(j)}$$

(15.1)

де t – момент часу для якого розраховується функція; j – хронологічний номер події (добуток береться по всім подіям: dokonаним

і недоконаним, що відбулися на момент часу t); N – загальна кількість подій (спостережень); $d(j)$ дорівнює 1 для dokonаної і 0 недоконаної події j .

$S(t)$ може бути інтерпретовано як долю пацієнтів, що пережили час t , або як імовірність пережити час t для окремого пацієнта. Графічне зображення функції виживання називається **кривою виживання (дожиття)**, тому часто кажуть, що для аналізу цензурованих даних використовується метод побудови і аналізу кривих дожиття.

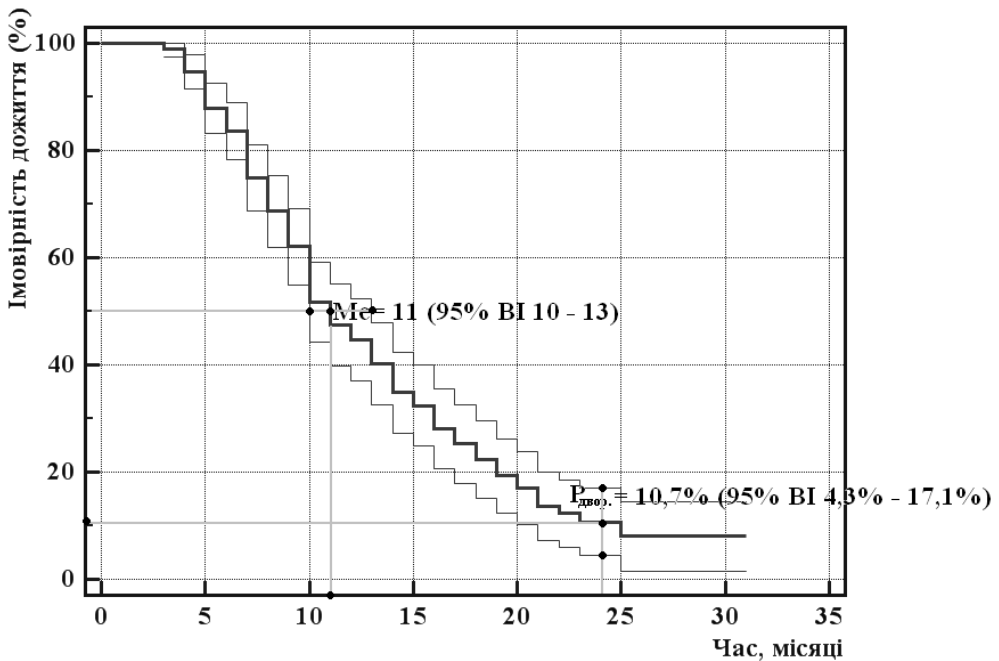


Рис. 15.2. Приклад кривої дожиття (наведено крива та 95% ВІ).

На рисунку 15.2 наведено приклад представлення даних (крива дожиття), при цьому може бути розрахований 95% ВІ для кривої. Крім функції дожиття для представлення цензурованих даних може бути розрахована **медіана виживання** (Me) – мінімальний проміжок часу, для якого значення функції дожиття не перевищує 0.5 (50%). Також може бути розраховано імовірність 1-річного, 2-річного і т.д. виживання – доля пацієнтів, що пережили 1 рік, 2 роки і т.д.

Так, для наведеного прикладу (Рис. 15.2), медіана виживання

становить Me = 11 місяців (95% ВІ 10 місяців – 13 місяців), 2-річне виживання становить P= 10.7% (95% ВІ 4.3% – 17.1%).

15.1.2. Порівняння цензурованих даних

При проведенні досліджень часто виникає необхідність порівняння цензурованих даних для виявлення впливу факторної ознаки, для цього проводиться порівняння кривих виживання (наприклад порівняння двох методів лікування – Рис. 15.3).

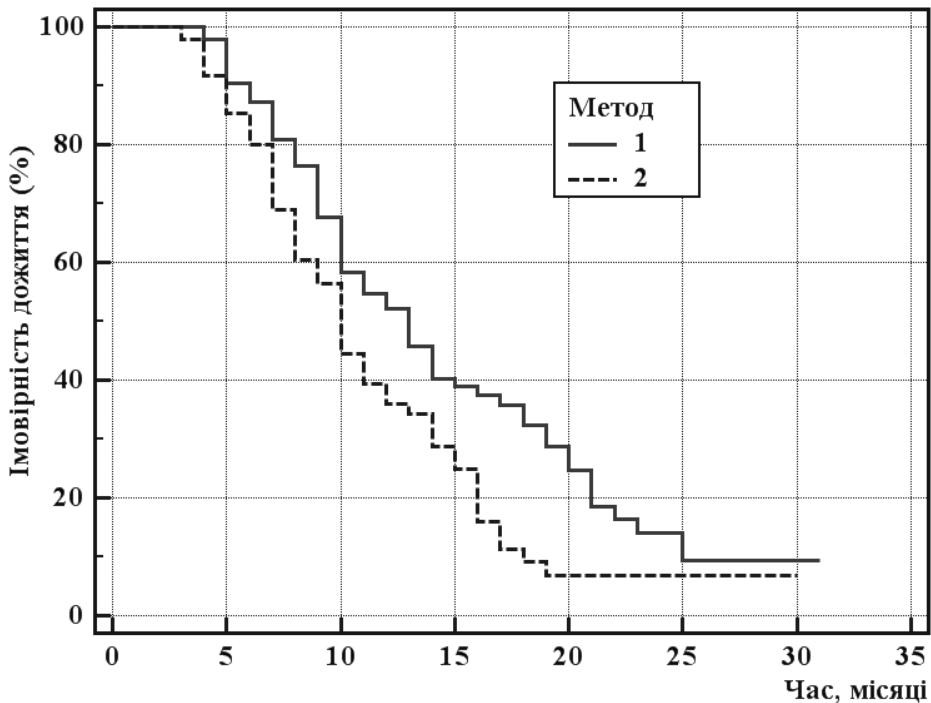


Рис. 15.3. Порівняння кривих виживання.

Нульова гіпотеза полягає в тому, що функції виживання (криві виживання) не відрізняються. Для проведення порівняння використовують **логранговий критерій**, при проведенні розрахунків порівнюють очікувану та експериментальну виживаність для кожного моменту події (при цьому очікувана виживаність буде підпорядковуватись гіпергеометричному закону розподілу). При відкиданні нульової гіпотези на рівні значимості $p \leq \alpha_{кр.}$ робиться висновок про відмінність кривих виживання (відмінність виживання для методів порівняння).

Звернімо увагу, що висновок про відмінність робиться не для 1-річного, 2-річного і т.д. виживання, а за відмінністю всієї кривої виживання.

15.1.3. Вплив факторних ознак на виживання. Модель пропорційних інтенсивностей Кокса

При проведенні аналізу впливу кількох (двох і більше) факторних ознак на функцію виживання виникає необхідність проведення багатофакторного аналізу, при цьому звичайні методи побудови багатофакторної моделі регресії у зв'язку з цензурованістю вихідних даних. Для проведення такого аналізу використовується метод побудови та аналізу **моделі пропорційних інтенсивностей Кокса**.

У моделі Кокса вважають, що ризик (імовірність) події (Hazard – $H(t)$) є мультиплікативною функцією незалежних змінних у будь-який момент часу, що може виражено формулою:

$$H\{(t), (x_1, x_2, \dots, x_m)\} = H_0(t) \times \exp(\beta_1 \times x_1 + \dots + \beta_m \times x_m), \quad (15.1)$$

де X_1, X_2, \dots, X_m – факторні ознаки, t – час, $H_0(t)$ – базова функція інтенсивності події, що дорівнює імовірності події для випадку, коли значення всіх факторних ознак дорівнює нулю, j -ий коефіцієнт моделі β_j

вказує вплив відповідної факторної ознаки на функцію $H(t)$.

Після логарифмічного перетворення модель може бути зведена до простої лінійної:

$$\ln[H\{(t), (x_1, x_2, \dots, x_m)/H_0(t)\}] = \beta_1 \times x_1 + \dots + \beta_m \times x_m \quad (15.2)$$

і далі розраховані коефіцієнти β_j . Відношення $H(t)/H_0(t)$ має назву

показник відношення ризиків (*Hazard Ratio* – *HR*), при збільшенні факторної ознаки на одиницю ризик події збільшується у $\exp(\beta_j)$ разу.

Для оцінки якості побудованої моделі може бути розраховано значення критерію χ^2 -квадрат. Якщо включення факторних ознак не змінює значення функції правдоподібності (χ^2 -квадрат статистично значиме не відрізняється від 0, $p > 0.05$), то можна говорити про незалежність функції інтенсивності $H(t)$ від факторних змінних. Для виявлення ступеню впливу кожної із факторних ознак проводиться оцінка β -коефіцієнтів моделі (b_j), розраховується стандартна похибка оцінки m_{b_j} і рівень значимості відмінності коефіцієнту від 0.

Для відбору мінімальної кількості факторних ознак, що забезпечують якісні прогностичні характеристики моделі Кокса в пакеті EZR може бути вибраний: інформаційний критерій AIC, байєсівський інформаційний критерій BIC чи критерій досягнення порогу значимості.

15.2. АНАЛІЗ ВИЖИВАНOSTI. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

15.2.1. Приклад 22. Побудова і аналіз кривих виживання

Дані

Вивчалися результати лікування хворих з раком молочної залози (71 пацієнтка) з використанням двох методів хіміотерапії: стандартного (Contr) і запропонованого (New). У таблиці 22 додатку (Додаток. Приклад 22) наведені результати лікування: «Результат» – змінна Y ($Y = \langle 0 \rangle$ – пацієнт живий на момент останнього спостереження, $Y = \langle 1 \rangle$ – летальний випадок), «Час» (час у місяцях від початку лікування до моменту останнього спостереження). Поставлено задачу порівняння ефективності методик лікування.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 22.xls».

Задачу будемо розв'язувати методом побудови та аналізу кривих виживання.

EZR (R-statistics)

Statistical analysis →

Survival analysis →

Kaplan-Meier survival curve and logrank test →

Обираємо змінну часу спостереження (Time)

Обираємо змінну результату (Res)

Обираємо змінну групування (Meth)

Обираємо логранговий критерій порівняння кривих виживання

Для постеріорних порівнянь вибираємо (<none>, поправки на множинність груп не використовуємо)



Рис. 15.4. Побудова та аналіз кривих виживання з «Приклад 22.xls» у пакеті EZR.

На рисунку 15.5 наведено результати аналізу кривих виживання: в пункті (1) наведено результати розрахунку кривої виживання для хворих, що лікувалися з використанням стандартного методу хіміотерапії (Contr); в пункті (2) наведено результати розрахунку кри-

вої виживання для хворих, що лікувалися з використанням пропонованого методу хіміотерапії (New); у пункті (3) наведено результати порівняння кривих виживання за логранговим критерієм; у пункті (4) наведено результати розрахунку медіани виживання.

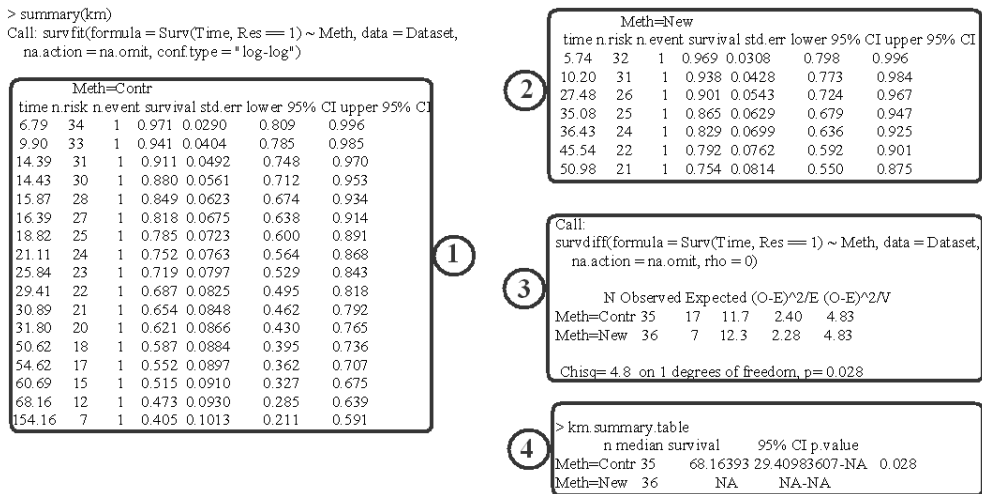


Рис. 15.5. Результати аналізу кривих виживання з «Приклад 22.xls» у пакеті EZR.

Проведений аналіз свідчить про статистично значиму ($p=0.028$) відмінність виживання пацієнтів двох груп. Медіана виживання в групі хворих, що лікувалися з використанням стандартного методу хіміотерапії склала 68.2 міс. (95% ВІ 29.4 міс. – ∞), а в групі хворих, що лікувалися з використанням пропонованого методу хіміотерапії не була досягнута (на протязі >150 міс.).

Результат

Для аналізу ефективності лікування хворих з раком молочної залози з використанням запропонованого методу хіміотерапії було використано метод порівняння кривих виживання Каплан-Мейєра. Аналіз проводився за результатами спостереження виживання хворих лікованих з використанням

стандартного (Contr) і запропонованого (New) методів хіміотерапії на протязі 20 років.

На рисунку 15.6 представлені криві виживання двох груп.

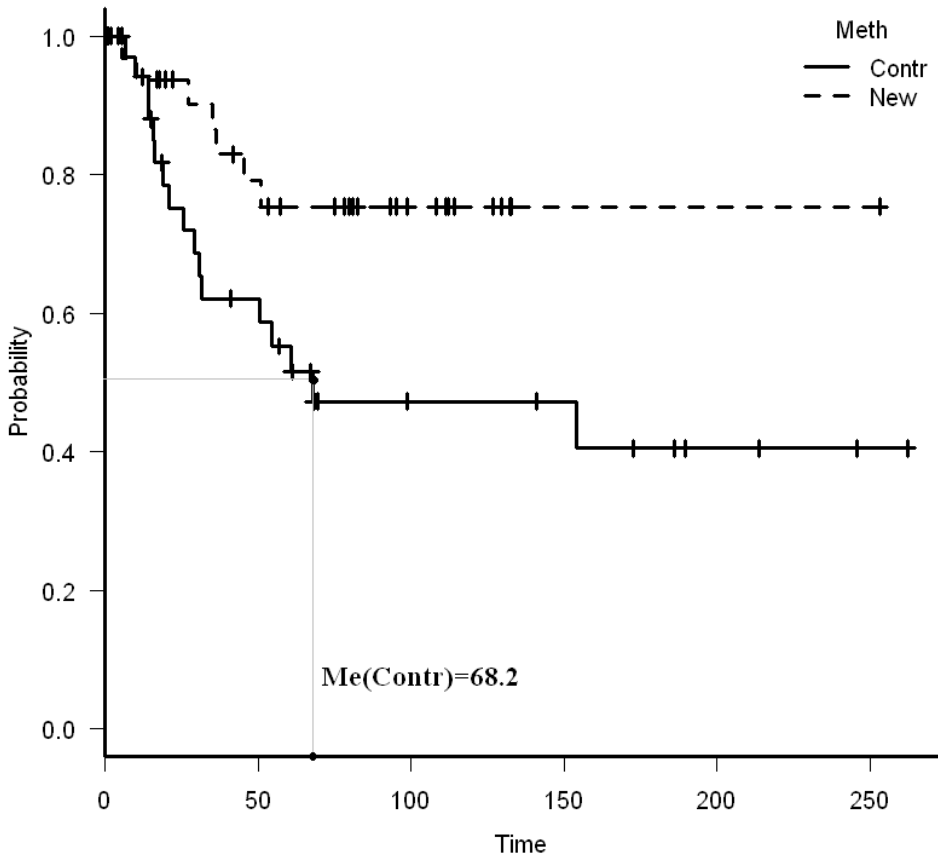


Рис. 15.6. Криві виживання хворих лікованих з використанням стандартного (Contr) і запропонованого (New) методів хіміотерапії.

При проведенні порівняння виявлено, що запропонований метод хіміотерапії при лікуванні хворих з раком молочної залози є більш ефективним, встановлено, що виживання хворих лікованих з вико-

ристанням запропонованого методу вище ($p=0.028$), ніж для стандартного методу хіміотерапії. Так, трирічне виживання хворих групи (Contr), склало $62.1 \pm 8.7\%$ ($\% \pm SE$), 15-річне виживання – $40.5 \pm 10.1\%$ (медіана виживання 68.2 міс. 95%

Ві 29.4 міс. – ∞). Для хворих групи (New) трирічне виживання склало $86.5 \pm 6.3\%$, 15-річне виживання – $75.4 \pm 8.1\%$ (медіана виживання на протязі періоду спостереження не була досягнута).

15.2.2. Приклад 23. Модель пропорційних інтенсивностей Кокса

Дані

Вивчалися результати лікування хворих з діагнозом недрібноклітинного раку легені III-IV стадій. Вивчався вплив препарату ТТТ на результати лікування. Першу досліджувану групу склали 25 пацієнтів, які перорально отримували препарат ТТТ починаючи за 2–3 дні до початку хіміотерапії, протягом усього циклу та ще 2–3 дні після її закінчення. До другої, контрольної групи, було включено 35 хворих, яким проводилася ПХТ без додавання до схеми лікування препарату ТТТ. У таблиці 23 додатку (Додаток. Приклад 23) наведені результати лікування: «Результат» – змінна Y ($Y = \langle 0 \rangle$ – пацієнт живий на момент останнього спостереження, $Y = \langle 1 \rangle$ – летальний випадок), «Час» (час у місяцях від початку лікування до моменту

останнього спостереження). У таблиці вказано вік пацієнта, локалізацію пухлини ($v1$ – нижня частка легені, $v2$ – верхня частка легені, $v3$ – середня частка легені) та метод лікування (1 – із додаванням препарату ТТТ, 0 – без додавання препарату ТТТ). Поставлено задачу виявлення факторних ознак, пов'язаних з виживанням хворих.

Розв'язання в пакеті EZR (R-statistics)

Введемо отримані результати в таблицю Excel та збережемо в файлі «Приклад 23.xls».

Задачу будемо розв'язувати методом побудови моделі пропорційних інтенсивностей Кокса.

EZR (R-statistics)

Statistical analysis →

Survival analysis →

Cox proportional hazard regression →

Обираємо змінну часу спостереження (*Time*)

Обираємо змінну результату (*Res*)

Обираємо факторні змінні (*Meth, Loc, Age*)

Обираємо метод відбору значимих факторних ознак (*AIC*)

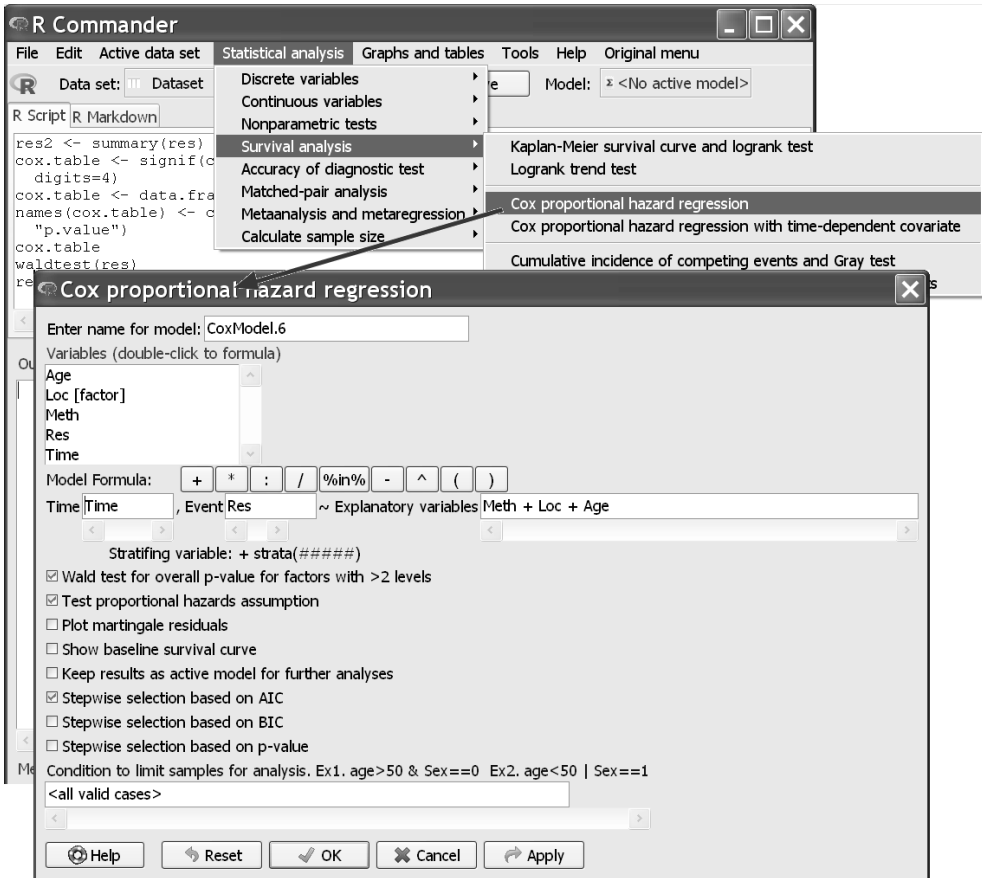


Рис. 15.7. Побудова та аналіз моделі пропорційних інтенсивностей Кокса з «Приклад 23.xls» у пакеті EZR.

На рисунку 15.8 наведено результати покрокового відбору факторів ризику та аналізу отриманої моделі пропорційних інтенсивностей Кокса. У пункті (1) наведено

результати розрахунку коефіцієнтів та оцінки якості моделі, побудованої на всіх факторних ознаках. У пункті (2) покроковий відбір значимих факторних ознак.

```

> #####Cox proportional hazard regression#####
n= 60, number of events= 37
              coef exp(coef) se(coef)      z Pr(>|z|)
Meth        -0.79209  0.45290  0.36942 -2.144  0.0320 *
Loc[T.v2]   -0.42627  0.65294  0.41341 -1.031  0.3025
Loc[T.v3]   -0.92777  0.39544  0.43051 -2.155  0.0312 *
Age          0.01400  1.01410  0.02225  0.629  0.5292
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
Meth          0.4529    2.2080    0.2196    0.9342
Loc[T.v2]     0.6529    1.5315    0.2904    1.4681
Loc[T.v3]     0.3954    2.5289    0.1701    0.9194
Age           1.0141    0.9861    0.9708    1.0593

Concordance= 0.664 (se = 0.057 )
Rsquare= 0.155 (max possible= 0.986 )
Likelihood ratio test= 10.14 on 4 df,  p=0.03821
Wald test              = 9.51 on 4 df,  p=0.04948
Score (logrank) test = 9.92 on 4 df,  p=0.04174

              Hazard ratio Lower 95%CI Upper 95%CI p.value
Meth          0.4529    0.2196    0.9342  0.03202
Loc[T.v2]     0.6529    0.2904    1.4680  0.30250
Loc[T.v3]     0.3954    0.1701    0.9194  0.03116
Age           1.0140    0.9708    1.0590  0.52920

> res <- stepwise(CoxModel.6, direction="backward/forward", crite

Direction: backward/forward
Criterion: AIC

Start: AIC=253.55
Surv(Time, Res == 1) ~ Meth + Loc + Age
              Df      AIC
- Age  1 251.95
<none> 253.55
- Loc  2 254.31
- Meth 1 256.53

Step: AIC=251.95
Surv(Time, Res == 1) ~ Meth + Loc
              Df      AIC
<none> 251.95
- Loc  2 252.76
+ Age  1 253.55
- Meth 1 254.61

> summary(res)
Call:
coxph(formula = Surv(Time, Res == 1) ~ Meth + Loc, data = TempDF,
      method = "breslow")

n= 60, number of events= 37
              coef exp(coef) se(coef)      z Pr(>|z|)
Meth        -0.7579  0.4687  0.3650 -2.077  0.0378 *
Loc[T.v2]   -0.4719  0.6238  0.4087 -1.155  0.2482
Loc[T.v3]   -0.9367  0.3919  0.4296 -2.180  0.0292 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Concordance= 0.656 (se = 0.056 )
Rsquare= 0.15 (max possible= 0.986 )
Likelihood ratio test= 9.74 on 3 df,  p=0.02095
Wald test              = 9.12 on 3 df,  p=0.02776
Score (logrank) test = 9.57 on 3 df,  p=0.02255

> cox.table
              Hazard ratio Lower 95%CI Upper 95%CI p.value
Meth          0.4687    0.2292    0.9583  0.03784
Loc[T.v2]     0.6238    0.2800    1.3900  0.24820
Loc[T.v3]     0.3919    0.1689    0.9097  0.02923

```

Рис. 15.8. Результати аналізу моделі пропорційних інтенсивностей Кокса з «Приклад 23.xls» у пакеті EZR.

У пункті (3) наведено результати розрахунку коефіцієнтів та оцінки якості моделі, побудованої на відібраних факторних ознаках. У пункті (4) наведено таблицю оцінки впливу відібраних факторних ознак на ризик летального випадку.

При проведенні встановлено вплив локалізації пухлини та методу лікування на виживання хворих з діагнозом недрібноклітинного раку легені III-IV.

Результат

Для аналізу впливу препарату ТТТ та зв'язку виживання хворих з діагнозом недрібноклітинного раку легені III-IV стадій з факторними ознаками (вік пацієнта та локалізація пухлини) використано метод побудови та аналізу моделей пропорційних інтенсивностей Кокса. Аналіз проводився на результатах обстеження 60 хворих (17 летальних випадків), спостереження проводилося на протязі трьох років.

Модель, побудована на всіх трьох факторних ознаках адекватна, показник узгодженості склав

0.66 ± 0.06 , за показником відношення правдоподібності $\chi^2 = 10.1$ (при чотирьох ступенях вільності, $p = 0.038$).

Для відбору мінімального набору факторних ознак, пов'язаних з виживанням метод покрокового відкидання/включення змінних (з використанням інформаційного критерію Акаїке). Було виділено дві факторні ознаки, пов'язані з виживанням пацієнтів: локалізація пухлини та метод лікування.

На виділених факторних ознаках була побудована модель пропорційних інтенсивностей Кокса. Модель, побудована двох виділених факторних ознаках адекватна, показник узгодженості 0.66 ± 0.06 , за показником відношення правдоподібності $\chi^2 = 9.7$ (при трьох ступенях вільності, $p = 0.021$). Таким чином виключення віку не привело до погіршення якості моделі, що є свідченням того, що виживання хворих з діагнозом недрібноклітинного раку легені III-IV стадій пов'язано з локалізацією пухлини та методом лікування.

У таблиці 15.1 наведені результати оцінки коефіцієнтів моделі.

Таблиця 15.1.

Коефіцієнти двофакторної моделі пропорційних інтенсивностей Кокса прогнозування виживання хворих з діагнозом недрібноклітинного раку легені III-IV стадій.

Показник	Значення коефіцієнту моделі, $b \pm m_b$	Рівень значимості відмінності коефіцієнту від 0, p	Показник HR (95% ВІ)
Локалізація, «v2» vs «v1»	-0.47 ± 0.41	0.248	–
Локалізація, «v3» vs «v1»	-0.94 ± 0.43	0.029*	0.39 (0.16 – 0.91)
Використання препарату ТТТ	-0.76 ± 0.37	0.038*	0.47 (0.22 – 0.96)

Примітка: * – відмінність параметра моделі від 0 статистично значима, $p < 0.05$.

Таким чином, встановлено, що виживання хворих з діагнозом недрібноклітинного раку легені III-IV стадій пов'язано ($p=0.021$) з локалізацією пухлини та методом лікування. Встановлено (табл. 14.1), що при локалізації пухлини в середній частці легені ризик летального випадку менше ($p=0.029$), HR= 0.39 (95% ВІ 0.16 – 0.91) у порівнянні з локалізацією у нижній

частці. Показано, що використання перорально препарату ТТТ починаючи за 2–3 дні до початку хіміотерапії, протягом усього циклу та ще 2–3 дні після її закінчення при лікуванні цих хворих дозволяє знизити ($p=0.038$) ризик летального випадку, HR= 0.47 (95% ВІ 0.22 – 0.96) у порівнянні з лікуванням без препарату (при стандартизації за локалізацією пухлини).

РОЗДІЛ 16. ОЦІНКА ОБ'ЄМУ ВИБІРКИ

Питання визначення об'єму вибірки, достатнього для розв'язання задачі медичного дослідження, є надзвичайно важливим при його плануванні. Очевидно, що із збільшенням об'єму вибірки висновки, отримані в результаті аналізу результатів медичного дослідження, будуть більш точними (буде зменшуватися імовірність помилки). У той же час будь-яке дослідження обмежене ресурсами (часом, фінансами і т. і.), що вимагає зменшення об'єму вибірки. Крім того, часто збільшення об'єму вибірки у медичному дослідженні пов'язане з виникненням етичних проблем (наприклад, порівняння плацебо і нового методу лікування – проблема наразити на ризик використання неефективного методу). Таким чином, тут виникає проблема оптимізації, що потребує визначення **мінімального об'єму вибірки**, яка забезпечить задовільну точність висновку.

Від чого залежить мінімальний об'єм вибірки?

Потрібно чітко усвідомлювати, що не існує універсального розв'язання цієї проблеми, що буде справедливе для будь-якого медичного дослідження (чи типу ме-

дичного дослідження). Кожне дослідження потребує розрахунку свого мінімального об'єму вибірки. Мінімальний об'єм вибірки, що буде достатнім для розв'язання задачі медичного дослідження, залежить, по-перше, від величини ефекту, наявності (чи відсутності) якого бажає визначити дослідник – **клінічно значимий ефект**. Чим менше клінічно значимий ефект, тим більше досліджень потрібно провести для виявлення його наявності (чи відсутності). По-друге, мінімальний об'єм вибірки залежить від надійності висновку про наявність ефекту, імовірності не допустити помилку першого роду – **критичного рівня значимості**. Чим нижче критичний рівень значимості, тим більше досліджень потрібно провести для його досягнення. По-третє, мінімальний об'єм вибірки залежить від імовірності не пропустити ефект, якщо він насправді існує, від імовірності не допустити помилку другого роду – **потужності** висновку. Чим більша потужність, тим більше досліджень потрібно провести для досягнення такої високої імовірності не пропустити наявності ефекту, якщо він існує.

Таким чином, для розрахунку мінімального об'єму вибірки, достатнього для розв'язання поставленої задачі, досліднику потрібно задати величину мінімального клінічно значимого ефекту, оцінити варіабельність ознаки, визначити статистичний критерій, використаний для прийняття рішення, задати критичний рівень значимості критерію при вказаній потужності.

16.1. ОЦІНКА ОБ'ЄМУ ВИБІРКИ. ТЕОРЕТИЧНІ ВІДОМОСТІ

Розрахунок мінімального об'єму вибірки проводиться за різними методами залежно від статистичного критерію, який використовується для прийняття рішення. Розрахунки, як правило, проводяться для **5% рівня значимості** і **80% потужності**.

Звернімо увагу, що при плануванні дослідження порівняння більш ніж двох груп слід використовувати розрахунки мінімального об'єму вибірки саме для множинних порівнянь (грубо можна врахувати зниженням критичного рівня значимості з урахуванням поправки Бонферроні).

Часто при плануванні медичного дослідження виникає необхідність

планування відмінних за об'ємом вибірок, що потрібно вказати при проведенні розрахунків.

16.2. ОЦІНКА ОБ'ЄМУ ВИБІРКИ. ПРИКЛАДИ РЕАЛІЗАЦІЇ В ПАКЕТІ EZR (R-STATISTICS)

16.2.1. Приклад 24. Порівняння частоти для двох незалежних вибірок

Дані

Планується порівняти ефективність нової методики лікування з традиційною. За попередніми результатами традиційна методика досягає ефекту в $P1 = 50\%$ випадків. Ми хочемо визначити, чи відрізняється ефективність нової методики від традиційної, відмінність будемо вважати значимою при різниці у ефективності більше $\Delta P = 20\%$. З урахуванням можливих ризиків при використанні нової методики та її витратності плануємо, що в групі контролю буде у чотири рази більше пацієнтів, ніж у групі дослідження. Визначити мінімальні об'єми вибірок, що достатні для розв'язання поставленої задачі.

Розв'язання в пакеті EZR (R-statistics)

Задачу будемо розв'язувати методом розрахунку об'єму вибірки.

EZR (R-statistics)

Statistical analysis →

Calculate sample size →

Calculate sample size for comparisons between two proportions →

Вказуємо частоти в групах порівняння

Обираємо рівень значимості та потужність

Вказуємо задане співвідношення об'ємів вибірок

Обираємо двосторонню критичну область та критерій порівняння

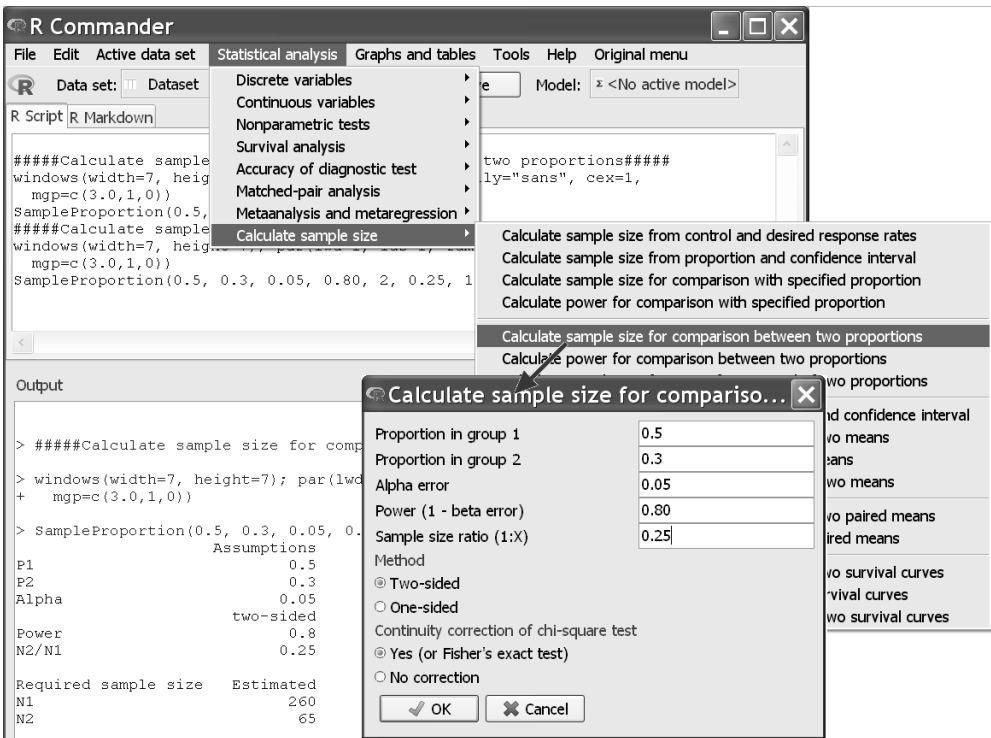


Рис. 16.1. Розрахунок об'єму вибірки для Прикладу 24 у пакеті EZR.

Розрахунки дають загальну кількість у 325 пацієнтів.

Звернімо увагу, що при рівних об'ємах вибірки (1:1) для

розв'язання цієї задачі було б достатньо 206 пацієнтів (по 103 у кожній групі).

Результат

При плануванні дослідження було отримано, що 325 пацієнтів у групах лікування: 260 пацієнтів у групі контролю та 65 пацієнтів у групі дослідження, – буде достатньо для виявлення відмінності у ефективності методик у 20%, при умові 50% ефективності лікування у групі контролю. Розрахунки проводилися для 5% рівня значимості та 80% потужності.

16.2.2. Приклад 25. Порівняння середніх для двох незалежних вибірок

Дані

Планується з'ясувати, чи є відмінність показника ТТТ для праворуких та ліворуких студентів медичного вишу. Відмінність клінічно значима при величині більше ніж 1 нмоль/л. Попередні дослідження на 5 студентах показали, що середнє квадратичне відхилення показника ТТТ становить

2 нмоль/л. Визначити мінімальні об'єми вибірок, що достатні для розв'язання поставленої задачі.

Розв'язання в пакеті EZR (R-statistics)

Задачу будемо розв'язувати методом розрахунку об'єму вибірки.

EZR (R-statistics)

Statistical analysis→

Calculate sample size→

Calculate sample size for comparisons between two means→

Вказуємо величину клінічно значимого ефекту та середнє квадратичне відхилення показника

Обираємо рівень значимості та потужність

Вказуємо задане співвідношення об'ємів вибірок

Обираємо двосторонню критичну область

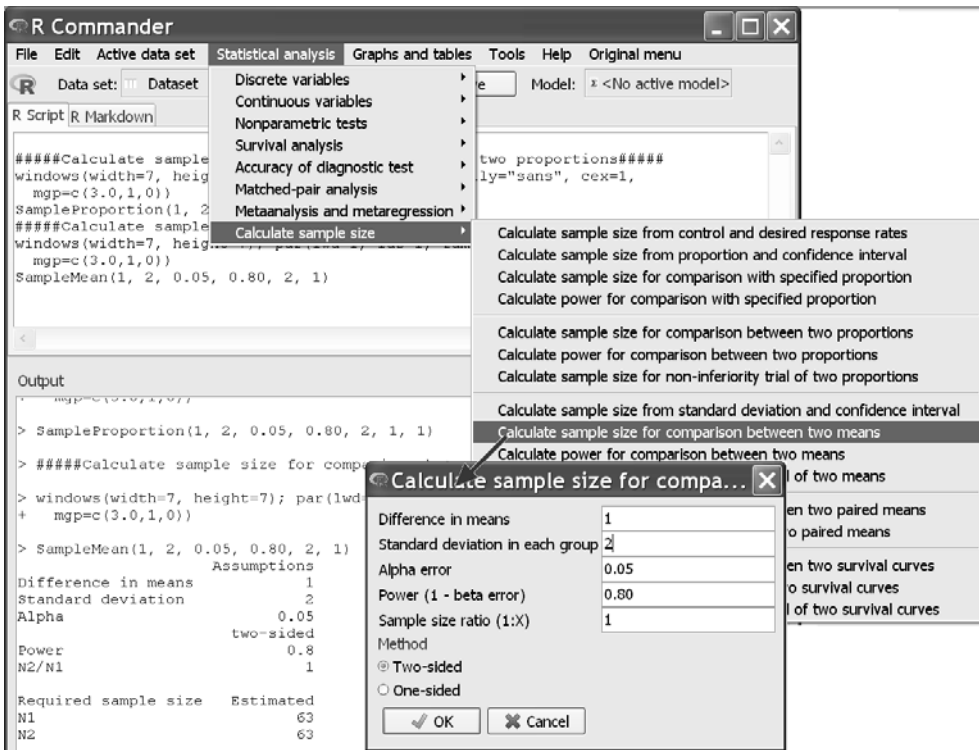


Рис. 16.2. Розрахунок об'єму вибірки для Прикладу 25 у пакеті EZR.

Розрахунки дають загальну кількість у 126 студентів.

Результат

При плануванні дослідження було отримано, що 126 студентів, по 63 у кожній групі, буде достатньо для виявлення відмінності показника у 1 нмоль/л, при варіабельності цього показника у 2 нмоль/л (середнє квадратичне відхилення). Розрахунки проводилися для 5% рівня значимості та 80% потужності.

16.2.3. Приклад 26. Порівняння середніх для двох пов'язаних вибірок

Дані

Планується з'ясувати, чи впливає здача іспиту «Крок -2» на показник ТТТ студентів медичного вишу. Показник планується виміряти до та після проведення іспиту. Відмінність вважається клінічно значимою при зміні більше ніж на 1 нмоль/л. Попередні дослідження на 5 студентах показали, що сере-

дне квадратичне відхилення показника ТТТ становить 2 нмоль/л. Визначити мінімальний об'єм вибірки, що достатній для розв'язання поставленої задачі.

Розв'язання в пакеті EZR (R-statistics)

Задачу будемо розв'язувати методом розрахунку об'єму вибірки.

EZR (R-statistics)

Statistical analysis→

Calculate sample size→

Calculate sample size for comparisons between two proportions→

Вказуємо частоти в групах порівняння

Обираємо рівень значимості та потужність

Вказуємо задане співвідношення об'ємів вибірок

Обираємо двосторонню критичну область та критерій порівняння

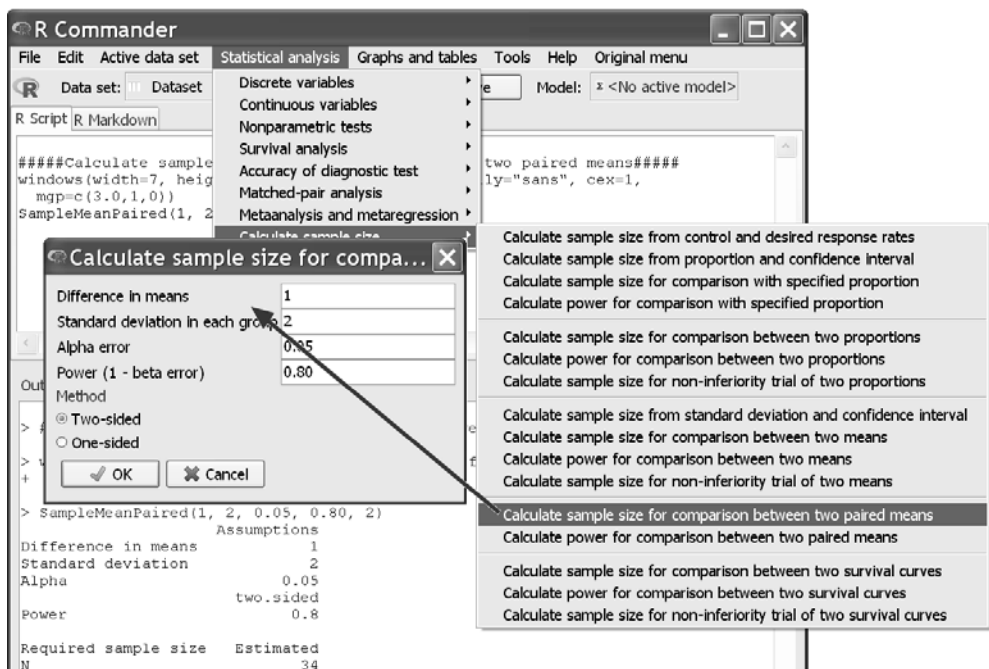


Рис. 16.3. Розрахунок об'єму вибірки для Прикладу 26 у пакеті EZR.

Розрахунки дають кількість у 34 студенти.

Результат

При плануванні дослідження було отримано, що вимірювання до та після іспиту «Крок-2» для 34 сту-

дентів буде достатньо для виявлення відмінності показника у 1 нмоль/л, при його варіабельності у 2 нмоль/л (середнє квадратичне відхилення). Розрахунки проводилися для 5% рівня значимості та 80% потужності.

СПИСОК ЛІТЕРАТУРИ

1. <http://minato.sip21c.org/grad/infop-text2012.pdf>
2. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics// Bone Marrow Transplant. – 2013. – 48. – P. 452–458.
3. <http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/functions.html>
4. <https://www.medcalc.org/index.php>
5. Лях Ю. Е. Основы компьютерной биостатистики. Анализ информации в биологии, медицине и фармации статистическим пакетом MedStat/ Лях Ю. Е., Гурьянов В. Г., Хоменко В. Н., Панченко О. А. – Д. : Папакица Е. К., 2006. – 214 с.: ил.
6. Свердан П. Л. Вища математика. Аналіз інформації у фармації та медицині: Підручник/ Свердан П. Л. – Львів : Світ, 1998. – 332 с.
7. Ланг Т. А. Описание статистики в медицине : руководство для авторов, редакторов и рецензентов / Т. А. Ланг, М. Сесик; пер. с англ. под ред. В. П. Леонова. – М. : Практическая медицина, 2014. – 480 с. : ил.
8. Faul F., Erdfelder E., Lang A.-G., Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences// Behavior Research Methods.– 2007.– 39. – P. 175–191.
9. Гайдышев И. Анализ и обработка данных: специальный справочник/ Гайдышев И. – СПб : Питер, 2001. – 752 с. : ил.
10. Гланц С. Медико-биологическая статистика / Гланц С.; пер. с англ. под ред. Н. Е. Бузикашвили и Д. В. Самойлова. – М. : Практика, 1998. – 459 с. : ил.
11. Петри А. Наглядная статистика в медицине/ Петри А., Сэбин К.; пер. с англ. В. П. Леонова. – М. : ГЭОТАР-МЕД, 2003. – 144 с.

12. Платонов А. Е. Статистический анализ в медицине и биологии : задачи, терминология, логика, компьютерные методы/ Платонов А. Е. – М. : Издательство РАМН, 2000. – 52 с.
13. Сергиенко В. И. Математическая статистика в клинических исследованиях/ Сергиенко В. И., Бондарева И. Б. – М. : ГЭОТАР МЕД., 2000. – 256 с.
14. Statistics with confidence. Confidence intervals and statistical guidelines./ Edited by Altman D. G., Machin D., Bryant T. N., Gardner M. J. Bristol.– BMJ Books, 2003. – 240 p.
15. Medical informatics. Computer application in health care and biomedicine./ Edited by Shortliffe E. H., Perrault L. E., Wiederhold G., Fagan L. M. – New York. : Springer-Verlag, 2001. – 856 p.
16. Bland M. An introduction to medical statistics. – New York. : Oxford University Press, 2003. – 405 p.
17. Introduction to biostatistics. A guide to design, analysis, and discovery/ Edited by Forthofer R. N., Lee S. E. – London : Academic Press, 1995. – 567 p.
18. Hill A B. A short textbook of medical statistics. – London : Hodderand Stouglston, 1977. – 325 p.
19. James F. Jekel, David L. Katz, Joann G. Elmore, Dorothea M. G. Wild Epidemiology, biostatistics, and preventive medicine. – Philadelphia : Saunders Elsevier, 2007. – 421 p.

ДОДАТКИ. ТАБЛИЦІ ДАНИХ

Таблиця. Приклад 1

Група 1		Група 2	
SBP, мм рт.ст.	DBP, мм рт.ст.	SBP, мм рт.ст.	DBP, мм рт.ст.
115	75	170	90
108	76	135	75
143	95	105	60
165	105	100	70
97	65	110	85
125	85	115	73
102	74	90	60
158	97	130	90
160	80	148	100
118	68	180	110
122	63	125	75
134	79	135	85
140	95	140	90
118	73	123	74
127	95	155	100
122	87	132	90
150	90	100	60
		115	85
		145	105
		170	95

Таблиця. Приклад 2

Сечовина, ммоль/л	Альфа-амілаза, ОД/л
9.3	33
24.7	50
8.3	24
9.3	33
16.7	40.9
7.7	17.3
17.4	48.2
13.3	34
16.7	40.9
8.3	24
21.5	49
21.2	41
8.1	22.6
18.9	43
21.5	49
7.8	25.9
19.2	48.2
9.3	33
7.9	25.9
16.7	40.9
21.5	49
21.5	17
7.7	17.3

Таблиця. Приклад 3

Резул	Місто			
	А.	Д.	К.	С.
Патологія	168	141	739	249
Без патології	12591	6504	16762	12933

Таблиця. Приклад 4

Група 1	Група 2
ИЛ-1, пг/мл	ИЛ-1, пг/мл
27.77	26.14
30.42	27.54
20.47	24.39
27.26	22.19
27.72	21.67
34.58	24.06
34.17	17.29
22.49	25.74
30.47	
32.16	
27.17	
28.39	
27.26	
30.68	
20.47	
30.16	
20.06	
21.6	
25.92	
30.47	

Таблиця. Приклад 5

Група 1	Група 2
ИЛ-8, пг/мл	ИЛ-8, пг/мл
10.8	12.52
2.997	10.56
12.38	10.8
11.18	14.67
4.025	10.93
6.707	12.38
13.25	9.977
13.25	3.718
16.81	3.312
16.62	3.514
5.657	
11.43	
11.96	
5.988	
12.38	
11.96	
7.951	
3.718	
13.1	
9.977	

Таблиця. Приклад 6

До лікування	Через 6 місяців
ВМІ, кг/м ²	ВМІ, кг/м ²
28.7	26.8
26.7	22.2
28.7	30
32.5	32.9
24.2	22.7
28.2	27.1
27.9	25.3
29.4	28.4
27.5	27.1
26.7	24
27.2	27.1
24.7	22.1
27.2	27.8
28	28.3
27.9	26.3
27.2	27.5
29	24.7
27.1	25.9
26.1	26.5
31	31.9

Таблиця. Приклад 7

На момент надходження	На момент виписки
NO, ммоль/л	NO, ммоль/л
5	5.2
4	4.5
3.9	4.9
5.4	5
5.4	5.4
4	5
3.8	4.9
5.1	5.4
3.5	4.6
4.9	5.4
4.8	5.2
5	5.4
4.6	4.8
5.2	5.6
5	5.4
5	5.1
4.8	5.4
5	5
4.2	4.6
3.6	4.9
5	5

Таблиця. Приклад 8

Ускладнення: так – «1», ні – «0»	
Метод 1	Метод2
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
0	0
0	
0	
0	
0	
0	

Таблиця. Приклад 9

Головний біль: так – «1», ні – «0»	
До	Після
1	1
1	1
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	1
1	1
1	0
1	0
1	1
0	1
0	1
0	1
0	1
0	1
0	1
0	1

Таблиця. Приклад 10

Результат	Група контролю	Група дослідження
Ускладнення	19	8
Без ускладнення	6	12
Усього прооперовано	25	20

Таблиця. Приклад 11

Загальний білірубін, мкмоль/л		
Форма А	Форма В	Форма С
12.3	10.4	12.3
11.6	14.3	14.5
14.2	19.5	18.4
11.2	21.2	17.2
17	21.2	12.2
22	17.1	19.2
7.5	25	15.1
7.2	11.9	20.3
11.7	14.2	13.8
20	15.8	18
4.8	20.6	12.2
8.6	12.1	22.1
9.4	13.6	14.5
12.2	13.7	27.7
8.6	16	19.7
9.4	9.4	20.4
11.2	13.4	17.7
9.3	11.3	18.3
6.2		
6.6		

Таблиця. Приклад 12

Вік хворого, років		
Форма А	Форма В	Форма С
25	18	24
24	25	27
23	25	48
30	20	49
24	20	35
37	18	34
25	22	44
29	20	43
29	19	36
28	19	26
32	23	39
25	23	26
26	21	28
27	17	51
30	19	36
26	19	44
33	21	33
30	17	27
29		
24		

Таблиця. Приклад 13

Гострота зору пацієнтів		
visus_1_	visus_7_	visus_month
0.02	0.07	0.07
0.2	0.45	0.6
0.2	0.35	0.4
0.25	0.25	0.5
0.3	0.5	0.55
0.3	0.5	0.6
0.3	0.55	0.6
0.3	0.35	0.2
0.35	0.4	0.3
0.5	0.75	0.75
0.5	0.6	0.65
0.6	0.5	0.75
0.7	0.75	0.75
0.8	1	1
0.7	0.75	1
0.9	1	1
0.4	0.4	0.6
0.45	0.5	0.6
0.7	0.75	1
0.8	0.9	0.85
1	1	1
0.9	0.9	0.95
0.7	0.75	0.8
0.8	0.9	0.9
0.5	0.55	0.6

Таблиця. Приклад 14

Гострота зору пацієнтів					
Before	After	V_1month	V_3month	V_6month	V_12month
0.01	0.02	0.08	0.09	0.09	0.09
0.01	0.06	0.2	0.3	0.3	0.3
0.01	0.05	0.07	0.08	0.08	0.08
0.01	0.03	0.6	0.7	0.7	0.7
0.01	0.05	0.08	0.08	0.08	0.08
0.01	0.09	0.04	0.03	0.03	0.03
0.01	0.04	0.09	0.09	0.1	0.1
0.01	0.09	0.6	0.5	0.5	0.5
0.01	0.06	0.3	0.3	0.3	0.3
0.02	0.05	0.08	0.08	0.08	0.08
0.02	0.06	0.06	0.07	0.08	0.08
0.02	0.06	0.08	0.09	0.07	0.08
0.02	0.01	0.2	0.3	0.2	0.3
0.02	0.06	0.08	0.08	0.08	0.08
0.02	0.05	0.5	0.5	0.5	0.5
0.02	0.09	0.8	0.9	0.9	0.9
0.02	0.09	0.08	0.08	0.08	0.08
0.02	0.09	0.06	0.08	0.06	0.04
0.03	0.1	0.09	0.09	0.08	0.08
0.03	0.4	0.3	0.3	0.2	0.2
0.03	0.1	0.3	0.3	0.3	0.3
0.03	0.09	0.6	0.5	0.5	0.5
0.03	0.09	0.09	0.08	0.08	0.08
0.03	0.05	0.06	0.06	0.06	0.06

Таблиця. Приклад 15

Meth	Result
M1	1
M1	1
M1	1
M1	1
M1	1
M1	1
M1	1
M1	1
M1	1
M1	0
M1	1
M1	1
M1	1
M1	1
M1	1
M1	1
M1	1
M1	1
M1	0
M1	1
M1	1
M1	0
M1	1
M1	0
M2	1
M2	1
M2	0
M2	1
M2	1
M2	1
M2	0
M2	1
M2	1
M2	0
M2	0
M2	1
M2	1
M2	0
M2	0
M2	1
M2	1

Meth	Result
M2	0
M2	1
M2	1
M2	1
M2	0
M2	1
M2	0
M2	0
M2	0
M2	0
M2	1
M3	1
M3	0
M3	0
M3	0
M3	0
M3	0
M3	0
M3	0
M3	0
M3	0
M3	0
M3	0
M3	1
M3	0
M3	0
M3	0
M3	0
M3	0
M3	1
M3	0

Таблиця. Приклад 16

Алергенна дія речовини: так – «1», ні – «0»		
P1	P2	P3
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
1	1	1
1	0	0
1	0	0
1	0	0
1	0	0
1	0	0
1	0	0
1	0	0
1	0	0
1	0	0
1	1	0
1	1	0

Таблиця. Приклад 17

Показники цитокінів	
ІЛ-6, пг/л	ІЛ-8, пг/л
18.84	3.00
20.99	13.25
21.28	15.35
22.46	12.38
19.62	13.25
17.61	4.03
22.15	12.52
22.15	12.38
18.59	6.71
15.81	5.66
17.15	10.56
21.86	11.18
14.58	12.10
18.34	12.24
18.84	3.72
22.46	9.01
13.09	11.05
14.19	8.23
17.61	11.43
19.10	9.98
21.28	9.86
21.86	9.01
19.10	10.93
19.62	11.96
18.59	16.81
23.70	16.62
18.10	14.67
20.99	10.80
21.56	10.80

Таблиця. Приклад 18

УО, см ³	Вага, кг
61.00	87
98.00	75
96.00	75
65.00	90
105.00	70
107.00	83
115.00	89
39.00	92
58.00	85
88.00	90
38.00	80
104.00	82
119.00	70
96.00	85
96.00	83
95.00	90
107.00	85
110.00	89
112.00	84
129.00	69
106.00	67
98.00	70
54.00	85
96.00	92
98.00	87

Таблиця. Приклад 19

X1	X2	X3	X4	X5	Y
10	8	7	3	21	8
6	6	8	5	20	11
5	6	4	5	19	12
4	6	7	4	21	14
6	8	5	5	19	14
10	7	8	3	21	14
4	8	7	6	18	15
4	7	5	5	16	16
6	8	5	5	18	16
6	6	7	5	18	16
6	7	7	5	19	16
7	6	10	5	16	17
6	7	5	6	17	18
8	8	7	6	15	18
8	7	8	6	16	18
9	6	7	6	18	18
4	5	8	6	17	19
5	5	8	5	17	19
6	8	6	6	15	19
6	8	8	6	17	19
6	7	6	6	14	19
4	4	6	8	18	20
4	4	4	5	17	20
10	8	9	6	19	20
10	6	8	6	19	20
5	8	7	7	16	21
7	4	5	8	17	21
8	2	7	7	18	21
10	6	6	7	13	21
4	2	6	8	14	22
7	7	7	6	17	22
8	6	8	7	14	22
9	4	6	10	11	24
8	7	6	9	13	25
8	5	9	8	11	25
10	6	6	8	12	25
6	9	6	9	13	28
8	6	7	10	11	28

Таблиця. Приклад 20

X1	X2	X3	X4	X5	Y
70	1	0	1	1	1
72	0	0	1	0	0
59	0	0	1	0	0
77	1	0	1	0	0
78	1	0	1	0	1
83	0	0	0	1	1
70	1	0	1	0	0
72	0	0	0	0	0
72	0	0	0	1	1
75	0	0	1	1	0
71	0	0	0	0	0
74	1	0	1	0	0
54	0	1	1	0	0
79	0	0	1	0	1
57	0	0	1	0	0
58	0	0	1	1	0
52	0	1	0	0	0
74	0	0	1	0	0
74	0	0	1	0	0
52	0	0	0	1	0
60	1	0	0	1	1
86	0	1	0	0	1
78	0	0	0	0	0
68	1	0	0	0	0
51	0	1	1	0	0
55	0	0	0	1	0
70	0	0	1	1	1
82	0	0	1	1	0
47	0	0	0	1	0
47	0	0	1	0	0

Таблиця. Приклад 21

Y	P
1	0.572
0	0.147
0	0.026
0	0.262
1	0.291
1	0.898
0	0.114
0	0.147
1	0.641
0	0.734
0	0.130
0	0.187
0	0.013
1	0.322
0	0.019
0	0.190
0	0.009
0	0.187
0	0.187
0	0.090
1	0.239
1	0.567
0	0.291
0	0.088
0	0.008
0	0.132
1	0.572
0	0.884
0	0.046
0	0.005

Таблиця. Приклад 22

Метод	Час спостереження (міс.)	Результат	Метод	Час спостереження (міс.)	Результат
New	5.4	0	New	1.0	0
New	126.8	0	New	27.5	1
New	18.0	0	Contr	9.9	1
Contr	15.9	1	New	10.2	1
Contr	69.7	0	New	114.1	0
Contr	14.4	1	New	0.7	0
Contr	185.9	0	New	42.1	0
New	17.0	0	New	5.7	1
New	132.7	0	New	108.4	0
Contr	16.4	1	New	81.0	0
New	112.2	0	New	82.5	0
Contr	18.6	0	New	93.2	0
New	45.5	1	New	19.8	0
Contr	25.8	1	Contr	60.7	1
New	253.1	0	Contr	15.0	0
New	132.3	0	Contr	57.0	0
New	95.3	0	Contr	67.3	0
New	126.7	0	Contr	18.8	1
New	35.1	1	Contr	41.0	0
Contr	30.9	1	Contr	50.6	1
New	79.9	0	Contr	21.1	1
New	53.2	0	Contr	12.1	0
New	57.4	0	Contr	2.0	0
Contr	29.4	1	Contr	31.8	1
Contr	189.5	0	Contr	154.2	1
New	4.4	0	Contr	61.2	0
New	129.5	0	Contr	68.2	1
New	98.9	0	Contr	262.1	0
Contr	98.7	0	Contr	141.0	0
New	51.0	1	Contr	14.4	1
Contr	68.2	0	Contr	6.8	1
New	111.5	0	Contr	172.8	0
New	78.4	0	Contr	213.8	0
Contr	245.8	0	Contr	54.6	1
New	75.1	0			
New	36.4	1			
New	22.1	0			

Таблиця. Приклад 23

Вік (років)	Локалізація	Метод	Результат	Час спостереження
75	v2	0	0	8
50	v2	1	1	9
69	v2	1	1	13
39	v2	0	1	15
53	v2	0	0	21
59	v2	1	1	8
56	v2	0	1	7
64	v2	1	1	10
74	v2	1	0	13
54	v2	0	1	4
52	v2	0	0	15
59	v2	1	0	21
49	v2	0	1	8
56	v2	0	1	16
64	v2	0	1	3
56	v2	0	0	12
59	v2	0	1	5
50	v2	0	1	4
65	v2	0	1	14
58	v1	0	0	14
60	v1	1	1	13
65	v1	1	1	9
72	v1	1	1	5
58	v1	0	1	6
57	v1	0	0	8
73	v1	1	0	9
48	v1	1	1	10
63	v1	0	1	4
60	v1	0	1	6
67	v1	0	0	9

Вік (років)	Локалізація	Метод	Результат	Час спостереження
56	v1	1	0	16
52	v1	1	1	7
56	v1	1	0	10
70	v1	0	1	10
67	v1	0	1	4
59	v1	0	1	8
56	v1	0	1	12
67	v1	0	1	10
59	v3	1	0	24
68	v3	1	0	31
63	v3	1	1	5
53	v3	1	0	18
74	v3	1	0	20
62	v3	1	0	19
39	v3	1	0	15
61	v3	1	1	9
51	v3	1	0	7
45	v3	1	0	9
65	v3	1	0	9
57	v3	0	1	16
54	v3	0	1	8
60	v3	0	1	19
70	v3	0	1	11
55	v3	0	1	15
64	v3	0	1	12
69	v3	0	1	10
59	v3	0	1	10
49	v3	0	1	21
66	v3	0	0	6
65	v3	0	0	6

ПРЕДМЕТНИЙ ПОКАЖЧИК

<i>Абсолютний ризик</i>	82, 83
<i>Альтернативна гіпотеза</i>	48
<i>Апостеріорні порівняння</i>	93, 95, 99, 102, 104, 106, 110, 112
<i>Байєсівський інформаційний критерій</i>	132, 148, 168
<i>Вибірка</i>	19
<i>Відношення ризиків</i>	83, 168
<i>Відношення шансів</i>	84, 143
<i>Вірогідний інтервал</i>	13, 31, 32, 33, 44, 45
<i>Генеральна сукупність</i>	22, 32, 44, 51
<i>Гіпергеометричний закон розподілу</i>	42, 168
<i>Двостороння критична область</i>	49, 61, 63
<i>Дискретні числові змінні</i>	20
<i>Дисперсія</i>	13, 26, 91
<i>Дихотомічні змінні</i>	42, 44, 45, 73, 74, 78, 81, 109, 111, 114
<i>Досягнутий рівень значимості</i>	53, 55
<i>Зниження абсолютного ризику</i>	83
<i>Інформаційний критерій Акаїке</i>	132, 148
<i>Кількість хворих, яких потрібно лікувати</i>	83
<i>Клінічна значимість</i>	54, 55, 56
<i>Клінічно значимий ефект</i>	177, 178
<i>Коефіцієнт детермінації</i>	137
<i>Коефіцієнт лінійної кореляції Пірсона</i>	119, 123, 130
<i>Коефіцієнт множинної кореляції</i>	130
<i>Кореляційний аналіз</i>	16, 122, 124, 126
<i>Крива виживання</i>	165
<i>Крива операційних характеристик</i>	146, 147, 150
<i>Критерій Бартлета</i>	93, 95

<i>Критерій Вілкоксона</i>	59, 66, 107
<i>Критерій Данна</i>	94
<i>Критерій Даннета</i>	92, 93
<i>Критерій досягнення порогу значимості</i>	132, 148
<i>Критерій знаків</i>	70
<i>Критерій Кохрена</i>	114, 115, 116
<i>Критерій Крускала-Уолліса</i>	99, 100, 101
<i>Критерій Мак-Немара</i>	79, 80
<i>Критерій перевірки</i>	53, 54
<i>Критерій Стіла</i>	94
<i>Критерій Стьюдента</i>	55, 58, 59, 60, 63, 66, 68
<i>Критерій Тьюкі</i>	92, 93, 95, 96, 97
<i>Критерій Уелча</i>	58
<i>Критерій Фрідмана</i>	102, 106
<i>Критерій Шеффе</i>	92
<i>Критичний рівень значимості</i>	53, 63, 67, 92, 98, 111, 123
<i>Критерій χ^2-квадрат</i>	75, 110, 111, 112, 113
<i>Логістична модель регресії</i>	141, 142, 149, 151
<i>Логранговий критерій</i>	167, 169
<i>Математичне сподівання</i>	23, 24, 25, 32
<i>Медіана</i>	12, 13
<i>Медіана виживання</i>	166, 170, 171, 172
<i>Метод Каплана-Мейєра</i>	165
<i>Метод найменших квадратів</i>	129
<i>Метод покрокового включення /виключення змінних</i>	131
<i>Мінімальний об'єм вибірки</i>	177
<i>Множинні порівняння</i>	92, 93, 94, 97, 100, 102
<i>Мода</i>	12, 13, 24, 25, 27

<i>Модель пропорційних інтенсивностей Кокса</i>	167, 168, 172, 173, 174, 175, 176
<i>Надійність критерію</i>	50
	58, 59, 62, 65, 73, 74, 76, 90, 91, 94, 97, 100, 109,
<i>Незалежні сукупності</i>	110, 111, 112
<i>Непараметричні критерії</i>	54, 70, 93, 102, 124
<i>Неперервні змінні</i>	20
<i>Номинальні змінні</i>	21
	23, 27, 29, 30, 31, 33, 58,
<i>Нормальний закон розподілу</i>	59, 65, 67, 91, 101, 119
	48, 50–53, 58, 61, 66, 68,
	71, 73, 76, 77, 99, 104,
<i>Нульова гіпотеза</i>	112, 116, 167
<i>Одностороння критична область</i>	49
<i>Однофакторний дисперсійний аналіз для повторних вимірювань</i>	101
<i>Однофакторний дисперсійний аналіз</i>	91
<i>Оперативна характеристика критерію</i>	50
<i>Параметричні критерії</i>	54, 63
<i>Перший квантиль</i>	26
<i>Площа під ROC-кривою</i>	146, 154
<i>Пов'язані сукупності</i>	57, 65, 67, 73, 77, 78, 90
<i>Показник рангової кореляції Спірмена</i>	121, 125
	92, 93, 102, 105, 107,
	110, 112–114, 116, 117,
<i>Поправка Бонферроні</i>	178
<i>Порядкові змінні</i>	21
<i>Потужність критерію</i>	92
<i>Похибка другого роду</i>	50
<i>Похибка першого роду</i>	50
<i>Прогностичність негативного результату</i>	143, 162

<i>Прогностичність позитивного результату</i>	143, 162
<i>Процентиль</i>	26
<i>Рандомізація</i>	19
<i>Регресійний аналіз</i>	126, 127
<i>Рівень значимості критерію</i>	53
<i>Розмах</i>	26
<i>Розповсюдженість</i>	145
<i>Специфічність</i>	144, 145
<i>Стандартна похибка</i>	30, 43, 44
<i>Стандартне відхилення</i>	26
<i>Статистична значимість</i>	15, 54
<i>Точний метод Фішера</i>	73
<i>Третій квантиль</i>	26
<i>Функція правдоподібності</i>	142, 143
<i>Цензуровані дані</i>	163–166
<i>Частковий коефіцієнт кореляції</i>	131
<i>Чутливість</i>	143
<i>Шанс</i>	82, 84

<i>Absolute Risk Reduction</i>	83
<i>Akaike information criterion</i>	132, 148
<i>Analysis of variance</i>	91
<i>Area under curve</i>	146
<i>Bayesian information criterion</i>	132, 148
<i>Confidence interval</i>	31
<i>Hazard Ratio</i>	168
<i>Negative predictive value</i>	143
<i>Number Nedeed to Treat</i>	83
<i>Odds Ratio</i>	84
<i>Population</i>	19
<i>Positive predictive value</i>	143
<i>Post hoc tests</i>	92, 102, 110, 114
<i>Prevalence</i>	145
<i>Receiver Operating Characteristic</i>	146
<i>Risk Ratio</i>	83
<i>Sample</i>	19
<i>Sensitivity</i>	143, 145, 146
<i>Specificity</i>	143, 145, 146
<i>Standard deviation</i>	26
<i>Standard error</i>	30
<i>Stepwise method</i>	131, 148
<i>Variance</i>	26

Формат 70x100/16.
Гарнітура Adonis. Друк офс. Папір офс.
Ум.-др. арк. 16.9. Нклад 300 прим.

Видавництво: ТОВ «Вістка»
Адреса: вул. Рибальська, 2, м. Київ, 01011, Україна
Свідоцтво про внесення до Державного реєстру України видавців,
виготовлювачів і розповсюджувачів видавничої продукції:
серія ДК №2280 від 05.09.2005