

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

О. Є. Кофанов, С. О. Солнцев, О. В. Зозульов

**ПРОГРАМУВАННЯ ІЗ ВИКОРИСТАННЯМ R
У СТАТИСТИЧНИХ ТА МАРКЕТИНГОВИХ
ДОСЛІДЖЕННЯХ:
НАВЧАЛЬНО-МЕТОДИЧНИЙ КОМПЛЕКС ДИСЦИПЛІНИ**

*Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня бакалавра
за спеціальністю 075 «Маркетинг»*

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2023

Рецензенти: *Федорченко Андрій Васильович*, д-р екон. наук, проф., завідувач кафедри маркетингу імені А. Ф. Павленка Київського національного економічного університету імені Вадима Гетьмана

Холковський Юрій Романович, канд. техн. наук, доц., доцент кафедри комп'ютерних технологій дизайну і графіки Національного авіаційного університету

Відповідальний редактор *Кубишина Надія Сергіївна*, канд. екон. наук, доц., доцент кафедри промислового маркетингу КПІ ім. Ігоря Сікорського

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського (протокол № 7 від 27.04.2023 р.) за поданням Вченої ради факультету менеджменту та маркетингу (протокол № 10 від 24 квітня 2023 р.)

Електронне мережне навчальне видання

Кофанов Олексій Євгенович, канд. екон. наук, канд. тех. наук

Солнцев Сергій Олексійович, д-р фіз.-мат. наук, проф.

Зозульов Олександр Вікторович, канд. екон. наук, проф.

ПРОГРАМУВАННЯ ІЗ ВИКОРИСТАННЯМ R У СТАТИСТИЧНИХ ТА МАРКЕТИНГОВИХ ДОСЛІДЖЕННЯХ: НАВЧАЛЬНО-МЕТОДИЧНИЙ КОМПЛЕКС ДИСЦИПЛІНИ

Програмування із використанням R у статистичних та маркетингових дослідженнях: Навчально-методичний комплекс дисципліни [Електронний ресурс] : навч. посіб. для студентів спеціальності 075 «Маркетинг» / О. Є. Кофанов, С. О. Солнцев, О. В. Зозульов; КПІ ім. Ігоря Сікорського. Електронні текстові дані (1 файл: 4,63 Мбайт). Київ : КПІ ім. Ігоря Сікорського, 2023. 204 с.

Навчальний посібник представляє собою навчально-методичний комплекс для розв'язання практичних завдань і закріплення теоретичного матеріалу з дисципліни «Програмування із використанням R» для студентів спеціальності 075 «Маркетинг». У посібнику приділена увага загальним особливостям використання мови програмування R, роз'ясненню способів її застосування для аналізу статистичних даних і вирішення маркетингових завдань. Посібник рекомендовано для самостійної роботи, роботи на практичних і лекційних заняттях, використання в умовах самоосвіти і дистанційного навчання, а також при вивченні таких дисциплін, як «Маркетингові дослідження», «Теорія ймовірностей та математична статистика», «Мови програмування при роботі з великими даними» тощо. Навчальне видання може бути використане під час викладання маркетингових дисциплін студентам інших спеціальностей, а також буде корисним фахівцям з промислового маркетингу.

© О. Є. Кофанов, С. О. Солнцев, О. В. Зозульов, 2023

© КПІ ім. Ігоря Сікорського, 2023

ЗМІСТ

ВСТУП	6
Розділ 1. МОВА ПРОГРАМУВАННЯ R, ЇЇ ОСОБЛИВОСТІ І СФЕРИ ЗАСТОСУВАННЯ	11
1.1 Історичні аспекти появи мови R	11
1.2 Особливості, переваги і недоліки мови R	15
1.3 Процес встановлення R на комп'ютер	16
1.4 Встановлення RStudio на комп'ютер	19
1.5 Знайомство з бібліотеками застосунку R	22
1.6 Базові приклади застосування мови R	26
1.6.1 Використання R як наукового калькулятора, робота з консоллю	27
1.6.2 Отримання допомоги при використанні мови R	30
Розділ 2. ВИКОНАННЯ ОБЧИСЛЕНЬ У R З ВИКОРИСТАННЯМ ЗМІННИХ	35
2.1 Числа, вектори, матриці та масиви	35
2.1.1 Основні правила присвоєння	35
2.1.2 Операції з векторами	37
2.1.3 Робота в R з матрицями	40
2.1.4 Робота з масивами і розмірності	44

2.3 Вибір і вилучення елементів у R	44
2.3.1 Проста індексація	45
2.3.2 Логічні вирази	46
2.3.3 Видалення елементів	49
2.4 Створення функцій та елементи програмування в R	50
2.4.1 Оголошення функцій	50
2.4.2 Обчислення функцій	53
2.4.3 Конструкції if, else, ifelse	55
2.4.4 Організація циклів в R	56
2.4.5 Інтерполяція і згладжування в R	57
 Розділ 3. СТАТИСТИЧНЕ ОПРАЦЮВАННЯ ДАНИХ, ВИКОРИСТАННЯ СТАТИСТИЧНИХ ПАКЕТІВ У МОВІ R	 59
3.1 Теоретичні основи статистичного опрацювання даних	59
3.1.1 Термінологія статистичного опрацювання даних	59
3.1.2 Оцінювання відтворюваності і достовірності результатів	68
3.2 Теоретичні основи кореляційного та регресійного аналізів	80
3.2.1 Статистичний зв'язок між випадковими величинами	80
3.2.2 Оцінка значущості коефіцієнта кореляції	93
3.3 Регресійний аналіз	95
3.4 Алгоритми та приклади виконання розрахунків (лінійна регресія)	100

3.5 Використання статистичних пакетів у R	108
3.5.1 Обробка числових даних	108
3.5.2 Ґраткова графіка	110
3.6 Шкали вимірювань, що застосовуються в економічних та маркетингових дослідженнях	116
Розділ 4. ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ ТА ПРИКЛАДИ ЇХ РОЗВ'ЯЗАННЯ	122
ВИКОРИСТАНІ ЛІТЕРАТУРНІ ДЖЕРЕЛА	140
ДОДАТКИ	144
Додаток А	145
Додаток Б	146
Додаток В	147
Додаток Г	148
Додаток Д	165
Додаток Е	177
Додаток Ж	178

ВСТУП

Останніми роками методи й засоби аналізу даних зазнали принципових змін. З появою персональних комп'ютерів, Інтернету і хмарних технологій різко зросли обсяги даних, а представлення інформації у зручному вигляді взагалі перетворилось на мистецтво. Сьогодні дані є основою для розуміння й налагодження бізнес-процесів, вирішення багатьох складних маркетингових та управлінських проблем підприємств. Тому обробка, аналіз і візуалізація великих масивів даних стає окремою і принципово важливою галуззю діяльності людини.

Експериментальні дані, представлені у комп'ютерному форматі у вигляді взаємопов'язаних таблиць, потребують таких процедур їхньої обробки й аналізу, які, по-перше, дозволяють встановити потенційно можливі закономірності і зв'язки між окремими компонентами, а, по-друге, – надають змогу передбачити нові факти.

На рис. 1 наведено схему основних етапів типового аналізу даних. Зазначимо, що на будь-якому з цих етапів є принципова можливість проводити трансформацію даних, встановлювати і вживати заходи щодо пропущених значень, додавати (видаляти) окремі змінні тощо, після чого процес обробки даних буде продовжений. Тобто процес завершиться лише тоді, коли дослідник (аналітик) буде вважати, що масив даних повністю досліджений, і всі необхідні відповіді знайдено.

Окрім того, сучасні методи обробки даних все частіше спираються на графічні способи представлення результатів дослідження – їх візуалізацію. У свою чергу, для перевірки складних гіпотез і розв'язання різнопланових бізнес-завдань застосовується апарат вилучення знань з великих баз даних (Knowledge Discovery in Databases), основою яких є інтелектуальний аналіз даних – Data Mining.

Data Mining це метафора, яка походить із гірничодобувної промисловості, в якій традиційно при розробці порід добуваються цінні ресурси, але, у даному випадку, йдеться про вилучення знань і знаходження необхідних відповідей на

важливі питання, спираючись на великі масиви даних. На сьогодні термін Data Mining застосовується як збірна назва цілої сукупності методів при розв'язанні конкретних завдань аналізу даних з метою виявлення прихованих або неочевидних закономірностей.

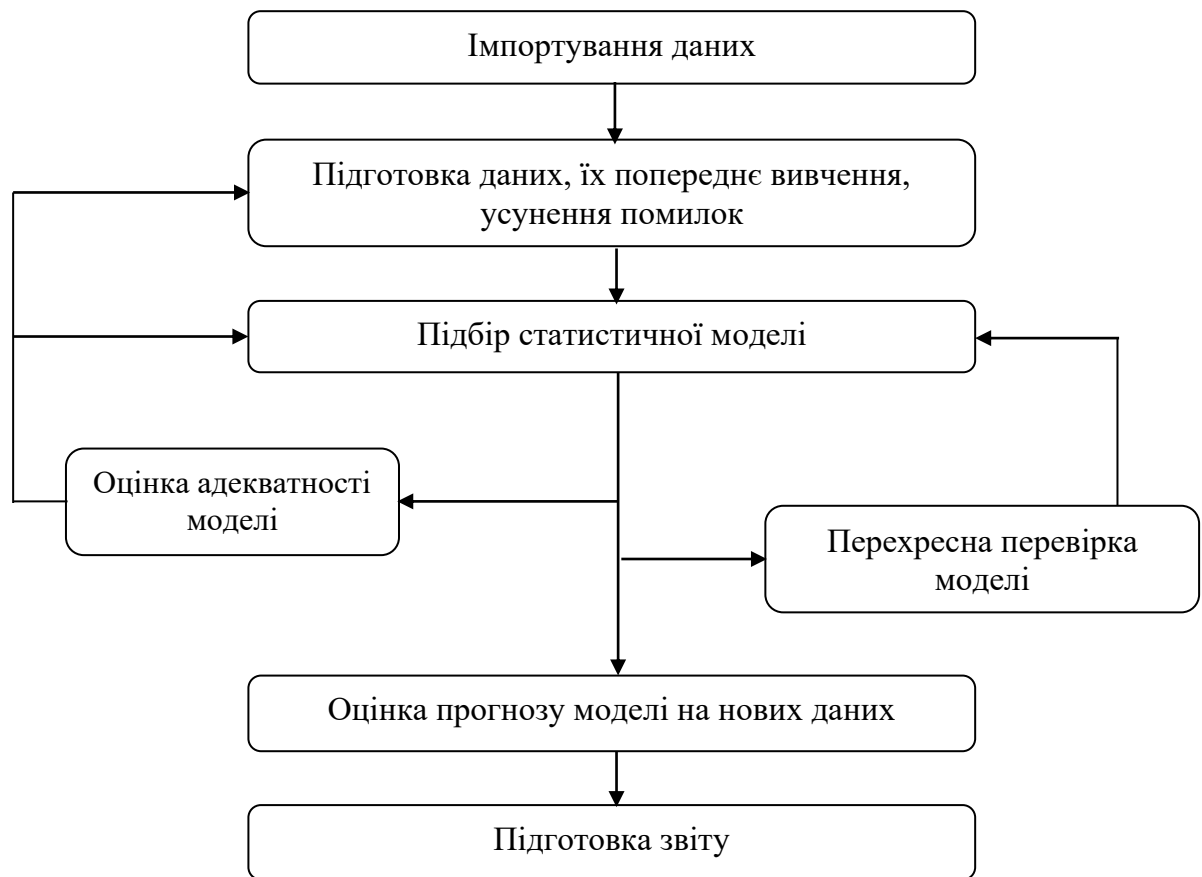


Рис. 1 Основні етапи типового аналізу даних.

Формулювання наукових гіпотез на етапі Data Mining та їх подальший аналіз (Data Analysis) – два комплексні, безпосередньо і невідривно пов'язані процеси. Вони відбуваються за стандартною схемою, а саме:

- збір експериментальних даних,
- організація їх у вигляді таблиць і
- пошук такого способу обробки, який надав би можливість встановити у вихідних даних нові знання про процес (явище), що аналізується.

Проте всі ці знання все одно залишаються дещо наближеними (мають певну невизначеність), оскільки чим глибше аналізується реальна складна система, тим менш визначеними стають наші судження про її поведінку та властивості.

Таким чином, для здобувачів, які пов'язали свою кар'єру з маркетингом, вивчення особливостей роботи з даними та безпосередньо з мовою програмування R є надзвичайно важливим, оскільки вони будуть опановувати методи аналізу даних, які нададуть їм змогу вирішувати складні і різнопланові маркетингові завдання. Зокрема, аналіз даних надає змогу вивчати поведінку споживачів, що, у свою чергу, допомагає зрозуміти їхні потреби і побажання, відстежити ефективність маркетингових кампаній та встановити і вилучити ті елементи, які не дають бажаного результату.

Вивчення мови програмування R дозволяє здобувачам отримати необхідні навички роботи зі статистичними методами аналізу, які допоможуть навчитися оброблювати результати, отримані у ході маркетингових досліджень. Так, знання методів регресійного аналізу дозволить виявити й дослідити залежності між різними змінними, спрогнозувати майбутні тенденції. Вивчення мови програмування R надає можливість виконувати складні обчислення в області промислового маркетингу. Наприклад, в R можна проводити кластерний аналіз, що надасть змогу виокремити групи споживачів за їхніми поведінковими характеристиками. Крім того, R дозволяє здобувачам виконувати моделювання задля прогнозування майбутніх подій і розроблення ефективних маркетингових стратегій. Для більш складних аналітичних задач, таких як прогнозування попиту на продукти або визначення оптимальних шляхів проведення рекламних кампаній, можна використовувати методи машинного навчання.

Отже, предметом навчальної дисципліни «Програмування із використанням R» є вивчення здобувачами основ роботи у програмному середовищі R і оволодіння навичками розв'язання типових завдань економічного й маркетингового спрямування у цьому середовищі, візуалізації даних за допомогою бібліотек мови R.

Метою навчальної дисципліни є формування у здобувачів системи знань про базові поняття, переваги і недоліки застосування, історію виникнення і розвитку мови програмування R, алгоритми проведення аналізу даних у RStudio із використанням R та спеціалізованих пакетів для вирішення маркетингових завдань задля здобуття студентами наступних компетентностей (здатностей):

- здатність застосовувати знання у практичних ситуаціях;
- здатність проведення досліджень на відповідному рівні;
- здатність використовувати інформаційні та комунікаційні технології;
- здатність логічно і послідовно відтворювати отримані знання предметної області маркетингу;
- здатність проводити маркетингові дослідження у різних сферах маркетингової діяльності;
- здатність аналізувати поведінку ринкових суб'єктів та визначати особливості функціонування ринків.

Вивчення дисципліни дозволить сформувати у здобувача наступні програмні результати навчання:

- аналізувати і прогнозувати ринкові явища та процеси на основі застосування фундаментальних принципів, теоретичних знань і прикладних навичок здійснення маркетингової діяльності;
- застосовувати набуті теоретичні знання для розв'язання практичних завдань у сфері маркетингу, у тому числі на промисловому та споріднених ринках;
- збирати та аналізувати необхідну інформацію, розраховувати економічні та маркетингові показники, обґрунтовувати управлінські рішення на основі використання необхідного аналітичного й методичного інструментарію;
- виявляти навички самостійної роботи, гнучкого мислення, відкритості до нових знань, бути критичним і самокритичним.

Задля досягнення зазначених програмних результатів навчання під час проведення лекційних занять із дисципліни заплановане застосування пояснювально-ілюстративного та інтерактивного методів навчання; практичних

занять – проблемно-пошукового, практичного та інтерактивного методів; під час самостійної роботи та підготовки до практичних занять – дослідницького методу навчання.

Передумовою до вивчення дисципліни «Програмування із використанням R» є знання, отримані при вивченні таких дисциплін, як «Математика для економістів», «Основи програмування», «Програмне забезпечення маркетингу», які вивчаються на бакалаврському рівні вищої освіти за освітньою програмою «Промисловий маркетинг». Результати вивчення дисципліни використовуються у дипломному проектуванні, а також при вивченні деяких дисциплін вибіркового циклу.

Важливим для подальшого професійного розвитку здобувачів вищої освіти є те, що вони мають змогу пройти онлайн-курс(и) за однією або декількома темами, передбаченими силабусом навчальної дисципліни. Онлайн-курс(и) здобувач може обрати самостійно або за рекомендацією викладача.

Навчальний посібник представляє собою навчально-методичний комплекс для розв'язання практичних завдань і закріплення теоретичного матеріалу з дисципліни «Програмування із використанням R» для студентів спеціальності 075 «Маркетинг». У посібнику приділена увага загальним особливостям використання мови програмування R, роз'ясненню способів її застосування для аналізу статистичних даних і вирішення багатьох маркетингових завдань.

Посібник рекомендовано для самостійної роботи, роботи на практичних і лекційних заняттях, використання в умовах самоосвіти і дистанційного навчання. Матеріали даного навчального посібника можуть бути використані здобувачами ступеня бакалавра за освітньою програмою «Промисловий маркетинг» спеціальності 075 «Маркетинг» при вивченні таких дисциплін як: «Маркетингові дослідження», «Теорія ймовірностей та математична статистика», «Мови програмування при роботі з великими даними», а також при опануванні маркетингових дисциплін студентами інших спеціальностей.

РОЗДІЛ 1

МОВА ПРОГРАМУВАННЯ R, ЇЇ ОСОБЛИВОСТІ І СФЕРИ ЗАСТОСУВАННЯ

1.1 Історичні аспекти появи мови R

Мова програмування це формалізована мова для опису алгоритмів розв’язування завдань за допомогою комп’ютера.

Мови програмування зазвичай класифікують за такими ознаками:

- машинні (або їх ще називають машинними кодами);
- машинно-орієнтовані (Assembler);
- процедурно-орієнтовані (Pascal, C та ін.);
- об’єктно-орієнтовані (C++, C#, Java та ін.);
- проблемно-орієнтовані (LISP, Prolog, SQL тощо)

Мова R належить до інтерпретованих мов програмування й для роботи з нею використовується командний інтерпретатор. R підтримує концепцію об’єктно-орієнтованого програмування.

R (programming language) мова програмування для статистичної обробки даних (статистичних обчислень) і роботи з графікою (візуалізації). Це також програмне середовище з відкритим кодом, що розвивається у межах проєкту GNU. Отже, файли із розширенням .r належать саме до мови програмування R.

R – проєкт, аналогічний мові програмування S, яка була розроблена у 1976 році в Bell Labs (раніше AT & T, зараз – Lucent Technologies) Джоном Чемберсом (John Chambers) з колегами як внутрішнє середовище для статистичного аналізу (була реалізована у вигляді бібліотеки Fortran і працювала під операційною системою GCOS; у 1980 році реалізація була переписана під UNIX).

R виник як вільний аналог середовища S-PLUS (третя версія, 1988 рік), яке, у свою чергу, є комерційною реалізацією мови S. Можна сказати, що R це альтернативна реалізація мови S. Між ними, звісно, є певні відмінності, але, як правило, код на S буде працювати і для середовища R.

R була розроблена у 1991 році Росом Іхаком (Ross Ihaka) і Робертом Джентлменом (Robert Gentleman), співробітниками Оклендського університету (Нова Зеландія), та названа за першою літерою їхніх імен – R. У 1995 році Мартін Маклер (Martin Makler) переконав Росса і Роберта використати GNU General Public License, щоб R стала безкоштовним програмним забезпеченням. У наш час R підтримується й розвивається R core team.

У 2000 році було випущено версію 1.0.0 R, 31 жовтня 2022 року випущено реліз R version 4.2.2 (Innocent and Trusting), а у квітні 2023 року очікується реліз R version 4.3.0 (Already Tomorrow).

R може бути застосована для роботи з масивами даних. За допомогою R, наприклад, розв'язують складні завдання математичної статистики, роблять первинний аналіз даних, здійснюють математичне моделювання. В освітніх цілях R часто використовують у поєднанні з інтерфейсом RStudio (також безкоштовне, вільно поширюване програмне забезпечення), який має зручний інтерфейс та безліч додаткових опцій, що значно підвищують продуктивність роботи фахівця.

Застосовуючи R, можна підготувати дані для дослідження і обробки результатів експерименту у різних сферах життя, наприклад, медицині, природокористуванні, охороні довкілля, економетриці й фінансовому аналізу, маркетингу, інженерних розрахунках тощо. R не тільки підтримує широкий спектр статистичних і чисельних методів, а й може бути розширений за допомогою програмних пакетів – бібліотек для роботи специфічних функцій або спеціальних сфер застосування, які доступні для завантаження через центральну систему збереження пакетів Comprehensive R Archive Network (CRAN) за посиланням: <https://cran.rstudio.com>.

У базовий дистрибутив R включений основний набір пакетів, тоді як усього їх чисельність наближається до 1000. У наш час у світі доступні більше 60 дзеркал CRAN. Головний вузол – (<http://cran.r-project.org>) розташований у Відні, Австрія. Ще однією перевагою R є його графічні можливості, які включають змогу створення якісних графічних зображень та використання математичних символів.

R доступний під ліцензією GNU GPL (General Public License – загальна публічна ліцензія); розповсюджується у вигляді вихідних кодів, а також відкомпільованих застосунків під низку операційних систем (ОС): деякі дистрибутиви GNU/Linux, FreeBSD, Microsoft Windows, Mac OS X та деякі інші Unix. Два-три рази на рік виходить безкоштовний інформаційний бюлетень, який містить інформацію по статистичній обробці даних і розробках команди. Бюлетень виходить з січня 2001 року.

Код R може бути написаний на мовах високого рівня, наприклад, C, C++, Fortran, або на мові R. За допомогою командного коду на C або Java можна маніпулювати R об'єктами.

У реалізації бібліотек R зазвичай використовуються статистичні методи, а також включено лінійне і нелінійне моделювання.

У R застосовується інтерфейс командного рядка зазвичай разом із наявністю інтерфейсів сторонніх розробників, наприклад, таких як графічний інтерфейс користувача RStudio, який встановлюється вже після установки самого R. Процес встановлення необхідного для подальшої роботи програмного забезпечення буде розглянуто у наступних підрозділах.

Для роботи з R розроблено низку графічних інтерфейсів, у тому числі такі, як:

- JGR;
- RKWard;
- SciViews-R;
- Statistical Lab;
- Rcmdr;
- Rattle тощо.

Більш того, у деяких редакторах передбачено спеціальні режими для роботи з R, наприклад:

- ConTEXT;
- Emacs (Emacs Speaks Statistics);
- jEdit;
- Kate;
- Syn;
- TextMate;
- Tinn-R;
- Vim;
- Bluefish;
- R плагін для Eclipse IDE;
- WinEdt з пакетом RWinEdt.

Функціонал R є доступним і з використанням мови програмування Python за допомогою пакета RPy. У відомому статистичному пакеті SPSS, починаючи з версії 16.0, також з'явилась підтримка функцій R.

Отже, R є і залишається однією з найбільш гнучких і потужних мов програмування, розробленою спеціально для аналізу даних. Мова R здатна оброблювати велику кількість типів різноманітних об'єктів – вектори, матриці, списки, таблиці з даними тощо. Також мова програмування R може працювати із великою кількістю типів даних. Це можуть бути, наприклад, числа з дробовою частиною, цілі числа, текстові записи, значення дати та часу, значення логічних операцій та ін.

Розміри векторів і матриць можуть змінюватися за бажанням користувача; дані не мають жорсткої структури. Це є дуже корисним при прогнозуванні, особливо якщо користувачу потрібно надати прогноз на довільний термін.

Отже, R є не просто платформою для написання програм, вона здатна також імпортувати дані з різних джерел інформації, у тому числі й текстові дані, системи управління базами даних, інші статистичні програми і

спеціалізовані репозиторії даних. Окрім того, R може записувати дані у форматах всіх цих систем.

І все ж, навчитися виконувати всі види аналізу, що реалізовані в R, є непростою задачею. Тому завданнями посібника є розглянути можливості R, показати особливості застосування тих чи інших пакетів для визначених дослідником задач.

1.2 Особливості, переваги і недоліки мови R

Як зазначалось, синтаксис R дуже схожий на S, що робить його легким для переходу користувачів S-PLUS. Семантика зовні схожа на S, але насправді – зовсім різна.

R працює практично на будь-якій стандартній обчислювальній платформі / ОС, а її часті релізи (річний + релізи з виправленням помилок) демонструють її постійний активний розвиток.

R є досить невеликим за розмірами програмним забезпеченням; але графічні можливості дуже розвинуті і набагато кращі за більшість інших статистичних пакетів. Для впевненого користування мовою R та швидкого орієнтування у переліку функцій цього програмного середовища можна використовувати спеціальні довідки збірки, наприклад, «R Reference Card», «R Reference Card by Jonathan Baron», «Time series Reference Card» і «Regression Reference Card» (<https://cran.r-project.org/other-docs.html>).

Інтерфейс є зручним для роботи, R також містить і потужну мову програмування для розробки нових інструментів (користувач → програміст).

Серед недоліків R можна відзначити такі:

- по суті базується на основі більш ніж 40-річної технології;
- невелика вбудована підтримка динамічної або 3-D графіки на початку розробки;

- функціональність заснована на споживчому попиті і власних внесків користувачів (краудсорсингу);
- об'єкти R мають бути збережені, як правило, у фізичній пам'яті;
- на жаль, неідеально підходить для всіх можливих ситуацій (хоча це є недоліком майже всіх програмних пакетів).

1.3 Процес встановлення R на комп'ютер

Для встановлення R на ваш комп'ютер перш за все відвідайте сторінку: <https://cran.rstudio.com> і послідовно виконайте кроки, наведені далі. У даному прикладі програмний продукт встановлено на комп'ютер із операційною системою MS Windows, але R може бути встановлено також і на пристроях із операційними системами Linux та macOS.

Натисніть «Download R for Windows», як показано на рис. 1.1.

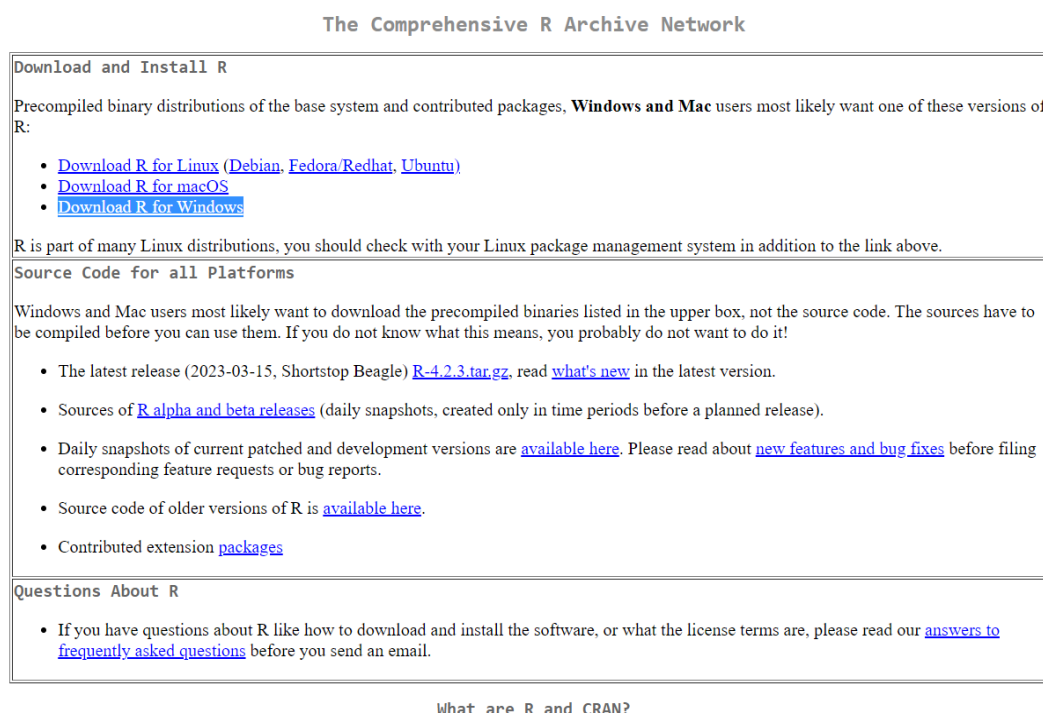


Рис. 1.1 Перший етап встановлення R на комп'ютер.

На сторінці, яка після цього відкриється, натисніть «Install R for the first time» і після цього – «Download R for Windows». Поряд із «R» буде наведено актуальну версію програмного продукту.

Далі запустіть файл, що завантажився, у даному випадку – R-4.2.1-win.exe і розпочніть процедуру встановлення застосунку на комп'ютер (рис. 1.2).

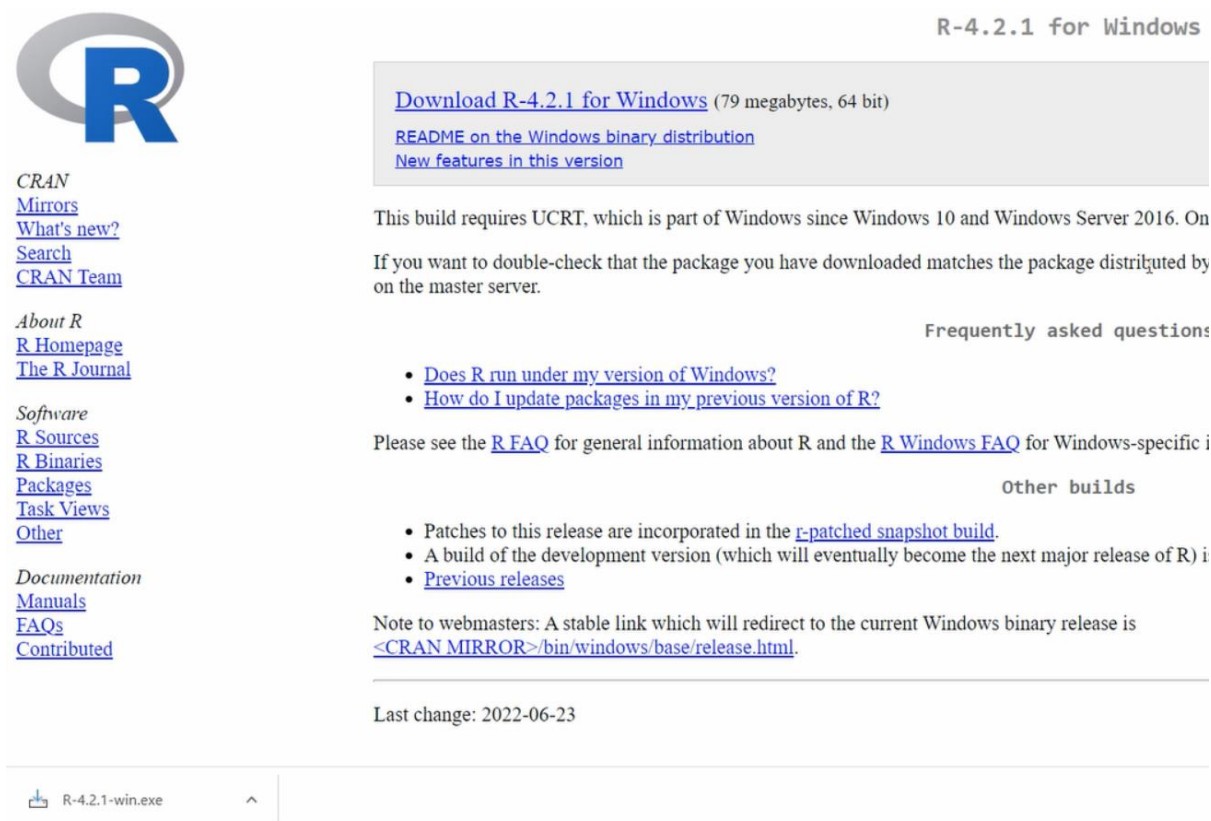
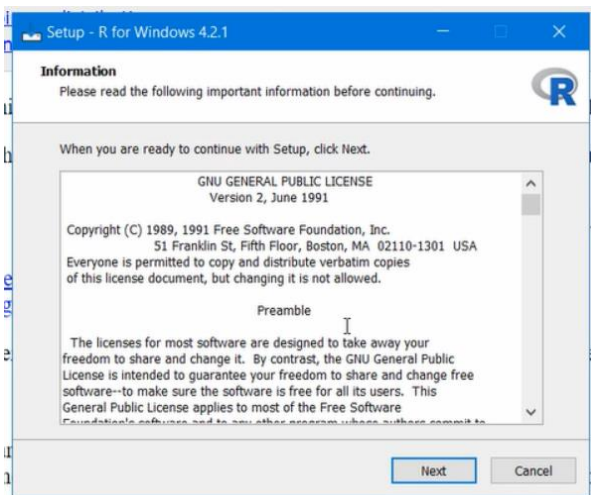
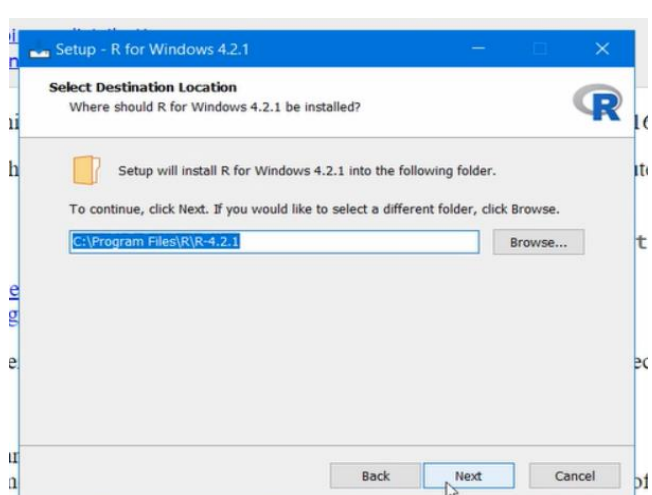


Рис. 1.2 Другий етап встановлення R на комп'ютер.

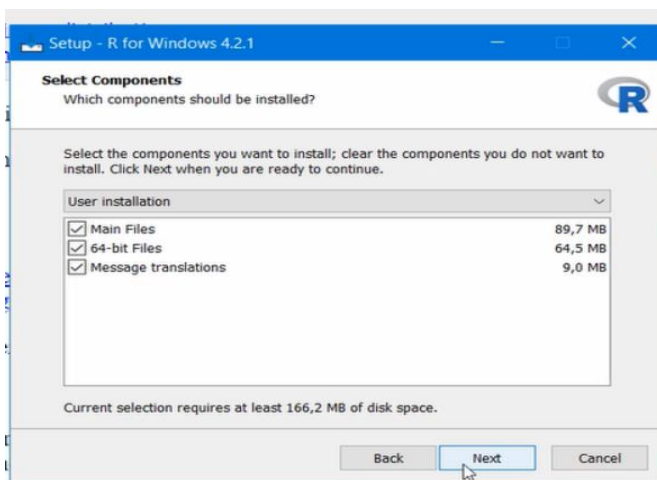
На запит, який висвітиться на екрані комп'ютера, надайте застосунку дозвіл на внесення змін на вашому комп'ютері. Далі оберіть мову встановлення, після чого має відобразитися інтерфейс, показаний на рис. 1.3а. Надалі необхідно слідувати інструкціям, відображеним на екрані. Найважливіші подальші складові третього етапу встановлення R наведено на рис. 1.3б–рис. 1.3е.



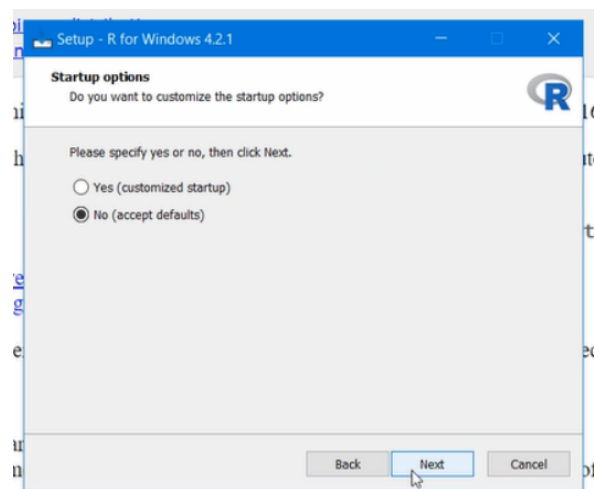
а)



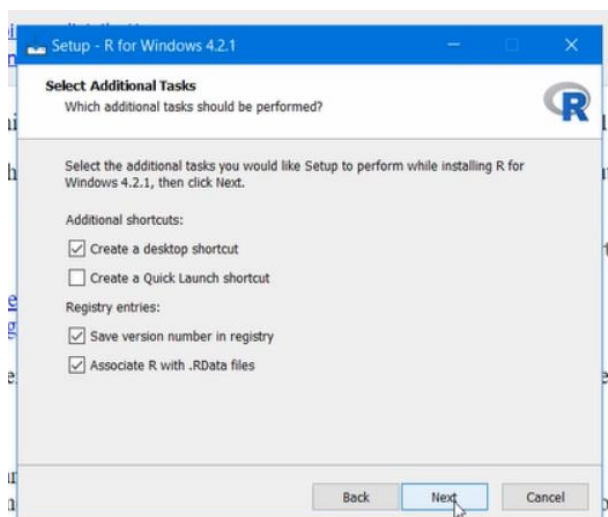
б)



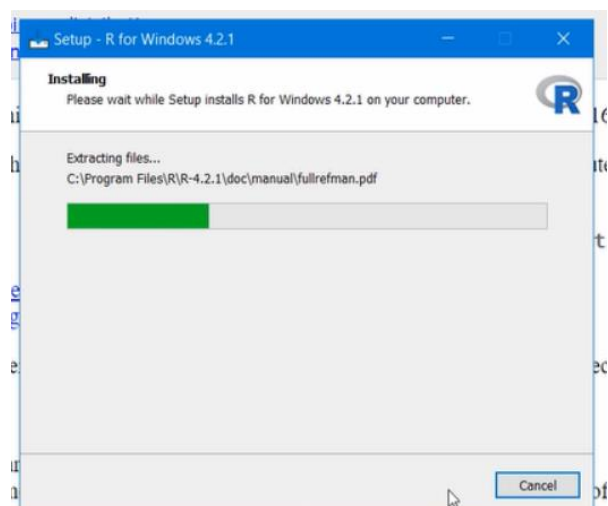
в)



г)



д)



е)

Рис. 1.3 Ключові складові третього етапу встановлення R на комп'ютер (а–е).

Слід зазначити, що не зважаючи на те, що у процесі встановлення програмного забезпечення є можливість обрати інші каталоги, рекомендовано встановлювати R саме у кореневий каталог, наприклад, «C:\R\». Встановлення в інший каталог може у майбутньому призвести до непрацездатності деяких бібліотек та інших технічних проблем.

1.4 Встановлення RStudio на комп'ютер

Для встановлення RStudio на комп'ютер необхідно скористатися ресурсом: https://posit.co/download/rstudio-desktop/?utm_source=downloadrstudio&utm_medium=Site&utm_campaign=home-hero-cta, натиснувши «DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS», як показано на рис. 1.4.

posit PRODUCTS SOLUTIONS LEARN & SUPPORT EXPLORE MORE PRICING

DOWNLOAD

RStudio Desktop

Used by millions of people weekly, the RStudio integrated development environment (IDE) is a set of tools built to help you be more productive with R and Python.

1: Install R

RStudio requires R 3.3.0+. Choose a version of R that matches your computer's operating system.

DOWNLOAD AND INSTALL R

2: Install RStudio

DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS

Size: 208.08 MB | [SHA-256: 885432DB](#) | Version: 2023.03.0+386 | Released: 2023-03-16

Рис. 1.4 Перший етап встановлення RStudio на комп'ютер.

Запустіть файл, що завантажився, у даному випадку – RStudio-2022.07.2....exe і розпочніть процедуру встановлення застосунку на комп'ютер (рис. 1.5).

На запит, який висвітиться на екрані комп'ютера, надайте застосунку дозвіл на внесення змін на вашому комп'ютері. Далі слідує інструкція на екрані аналогічно до процедури, показаної для встановлення R.

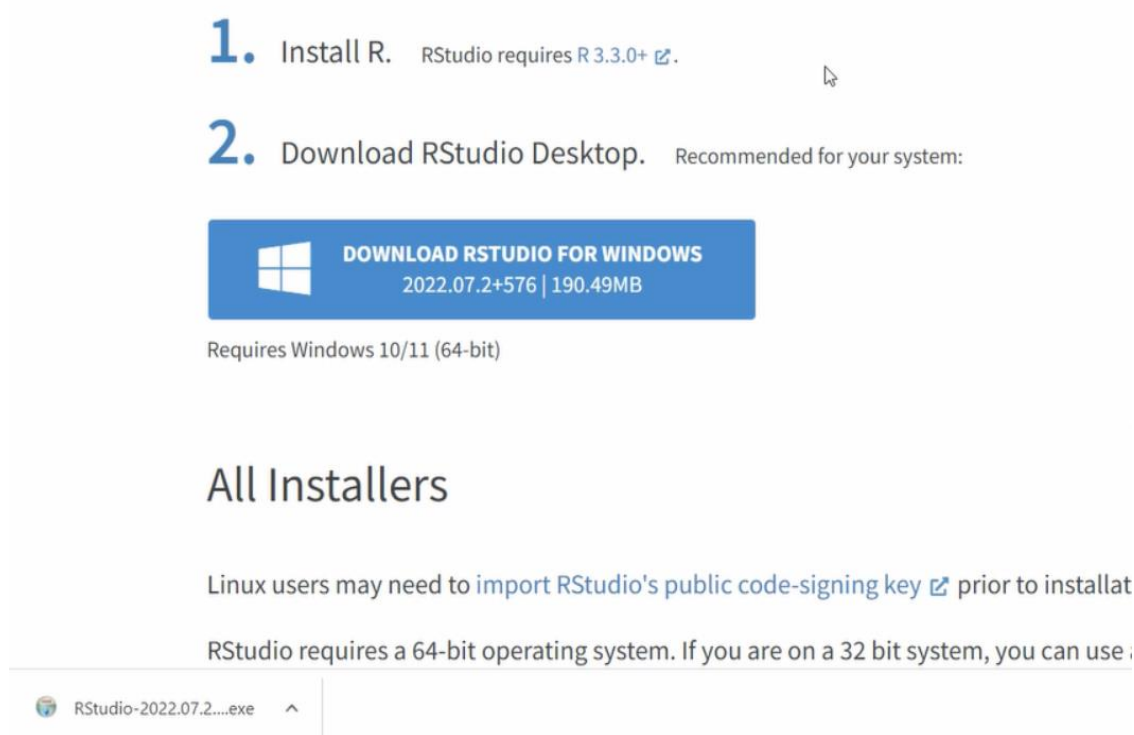


Рис. 1.5 Другий етап встановлення RStudio на комп'ютер.

Альтернативним варіантом є встановлення RStudio як складову іншого програмного продукту – Anaconda Navigator. Цей застосунок доступний для операційних систем Windows, macOS та Linux і може бути встановлений з офіційного веб-ресурсу: <https://docs.anaconda.com/free/navigator/install> та <https://www.anaconda.com> (рис. 1.6).

Anaconda Navigator це GUI, тобто графічний інтерфейс користувача настільних ПК, що входить у дистрибутив Anaconda і надає можливість споживачам встановлювати й запускати пов'язані програми, керувати

пакетами, середовищами і каналами Conda без необхідності використовувати менш зручний командний рядок (рис. 1.7).

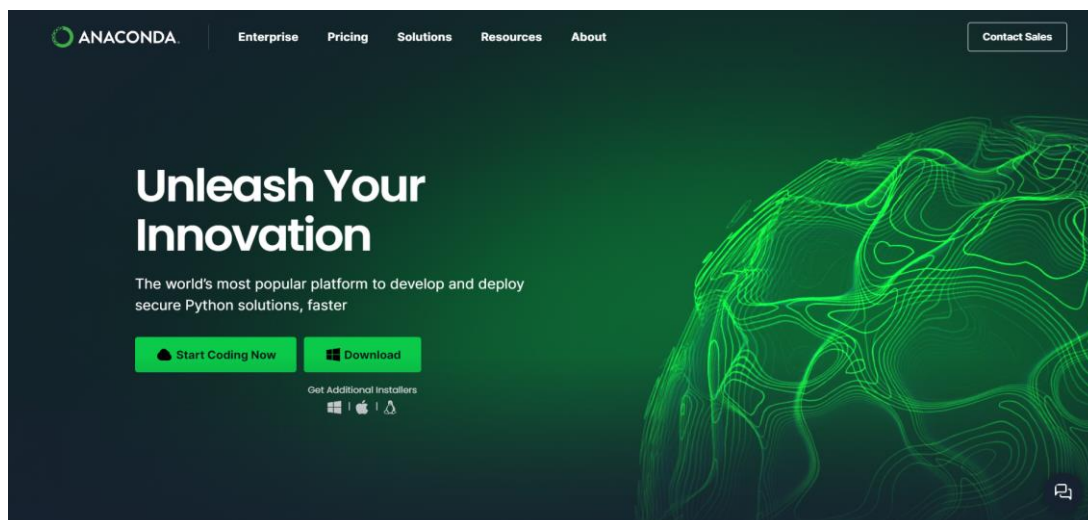


Рис. 1.6 Завантаження і встановлення Anaconda Navigator на комп'ютер.

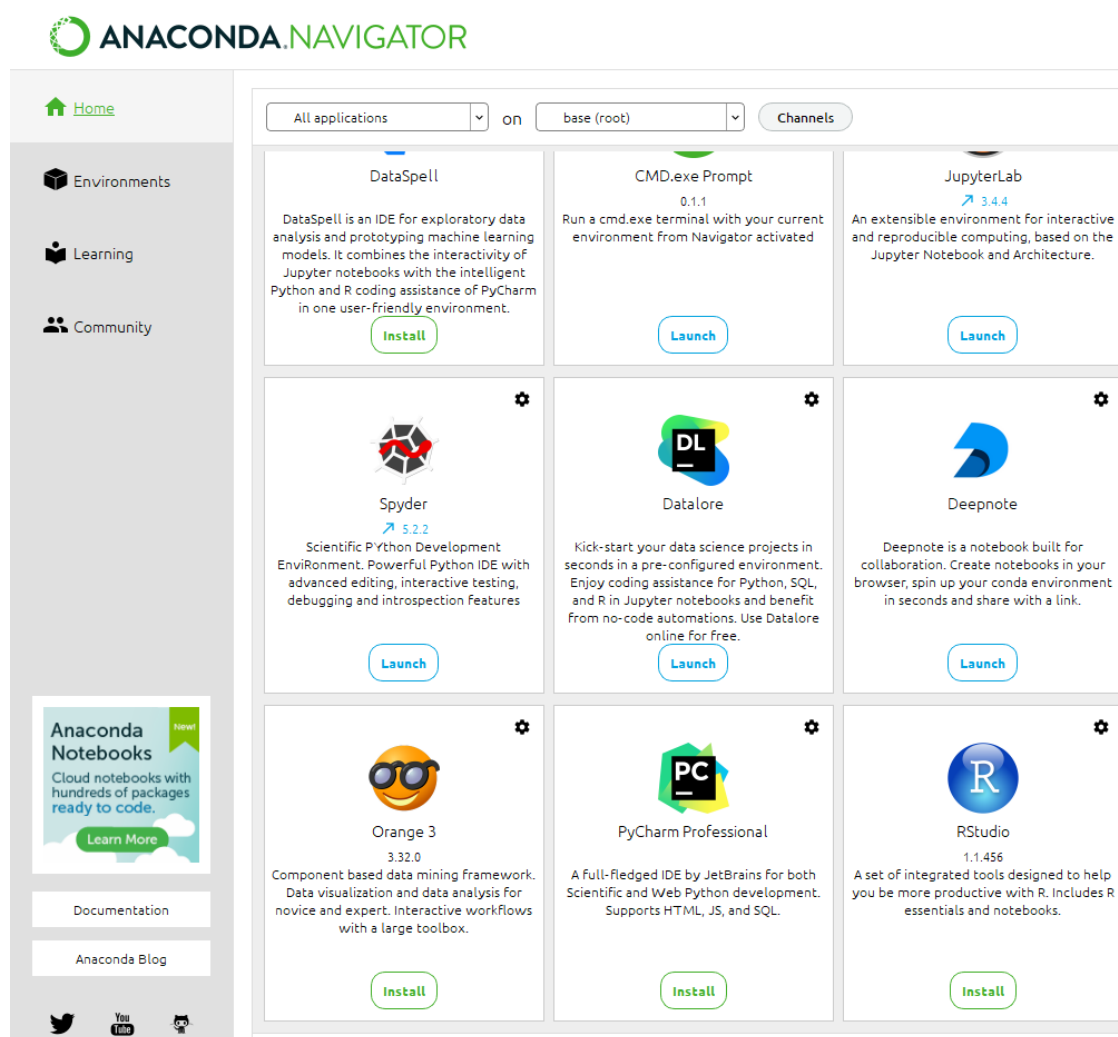


Рис. 1.7 Встановлення RStudio на комп'ютер через Anaconda Navigator.

Встановлення додатків в середовищі Anaconda Navigator є дуже простим. Для цього необхідно натиснути на «Install» під необхідним пакетом (рис. 1.7) і слідувати інструкціям на екрані.

Інтерфейс RStudio складається з декількох вікон, зокрема, з лівої сторони розташоване вікно консолі, у якому пишуться команди після символу «>» й відбувається виконання програмного коду R. Це одне з найважливіших вікон і саме в ньому виконуються всі розрахунки.

В правому лівому куті розташоване вікно робочого простору / історії. У вікні робочого простору можна побачити які дані та значення RStudio тримає в пам'яті. Окрім того, можна продивитись та відредагувати значення, натиснувши на них. Вікно історії показує що було введено раніше.

В правому нижньому куті розташоване вікно файлів / графіків / пакетів / довідки. Тут можна відкривати файли, продивлятися і зберігати графіки (включаючи попередньо побудовані графіки), встановлювати та завантажувати пакети або використовувати функцію довідки. Розмір усіх вікон можна змінювати, перетягуючи сірі смуги між ними.

1.5 Знайомство з бібліотеками застосунку R

Як зазначалось, R має потужні можливості для здійснення статистичного аналізу, у тому числі й з використанням лінійної і нелінійної регресії, а також класичні статистичні тести, аналіз часових рядів, кластерний аналіз та інші види аналізу даних. Окрім того, R легко доповнюється різноманітними додатковими пакетами – бібліотеками, причому, ті бібліотеки, що застосовуються найчастіше, вже вбудовано до R.

Щоб подивитися список усіх встановлених у R пакетів (бібліотек), потрібно перейти у вікно пакетів або набрати в консолі `library()`. Якщо побачите, що навпроти імені пакету стоїть прапорець, то це означає що зазначений пакет вже завантажено в систему і активовано, а отже його можна використовувати (рис. 1.8). Тут і надалі для роботи з R буде застосовано RStudio.

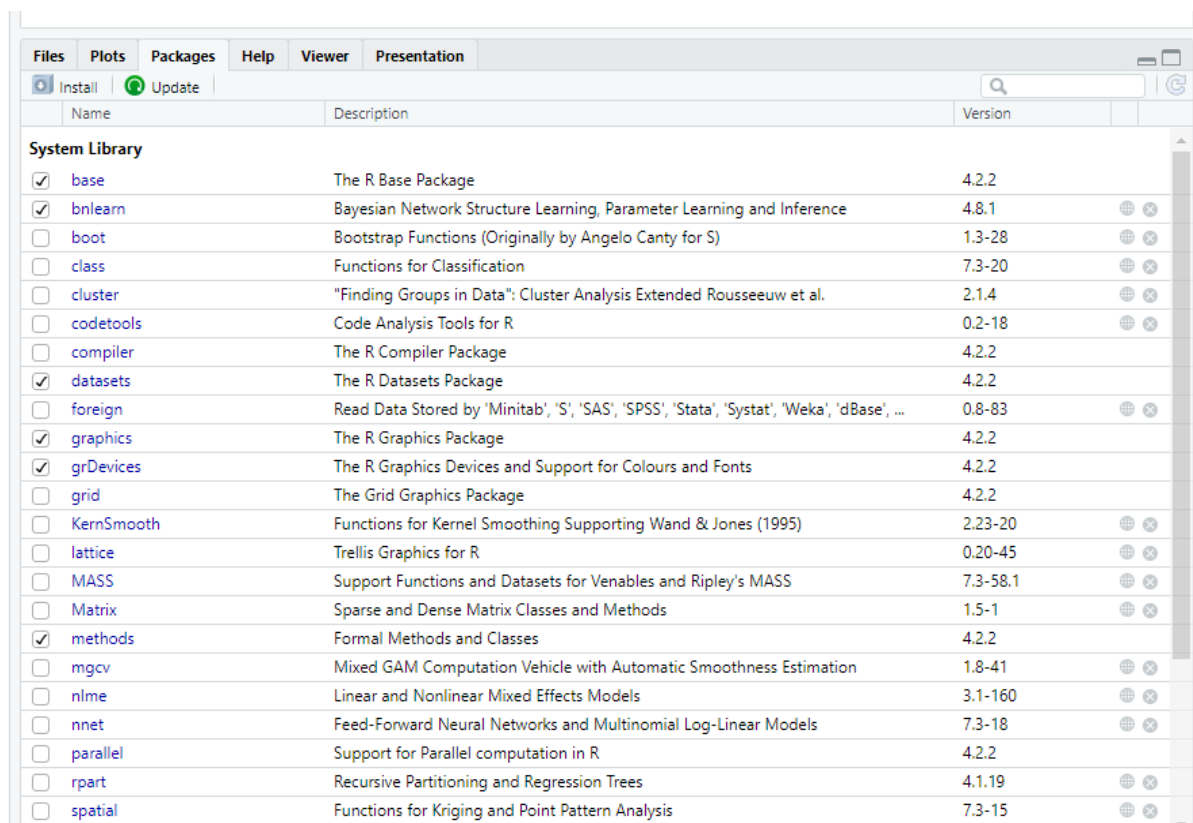


Рис. 1.8 Активовані для роботи пакети в RStudio.

Багато доступних для вільного скачування пакетів є на сайті R. Наприклад, щоб скачати і встановити бажаний пакет «geometry», потрібно (рис. 1.9):

- 1) натиснути «install» у вікні пакетів і ввести geometry або набрати «install.packages» («geometry») у командному вікні;
- 2) для завантаження пакету поставити прапорець навпроти geometry або набрати «library» («geometry») у командному вікні.

При цьому рекомендується завжди обирати опцію «Install Dependencies» для автоматичного встановлення усіх взаємозалежних пакетів, оскільки в іншому випадку доведеться встановлювати їх вручну, що може зайняти доволі значний час. Так, при встановленні одного пакету, часто можна побачити ситуацію, представлену на рис. 1.10, коли разом з ним встановлюються десятки інших.

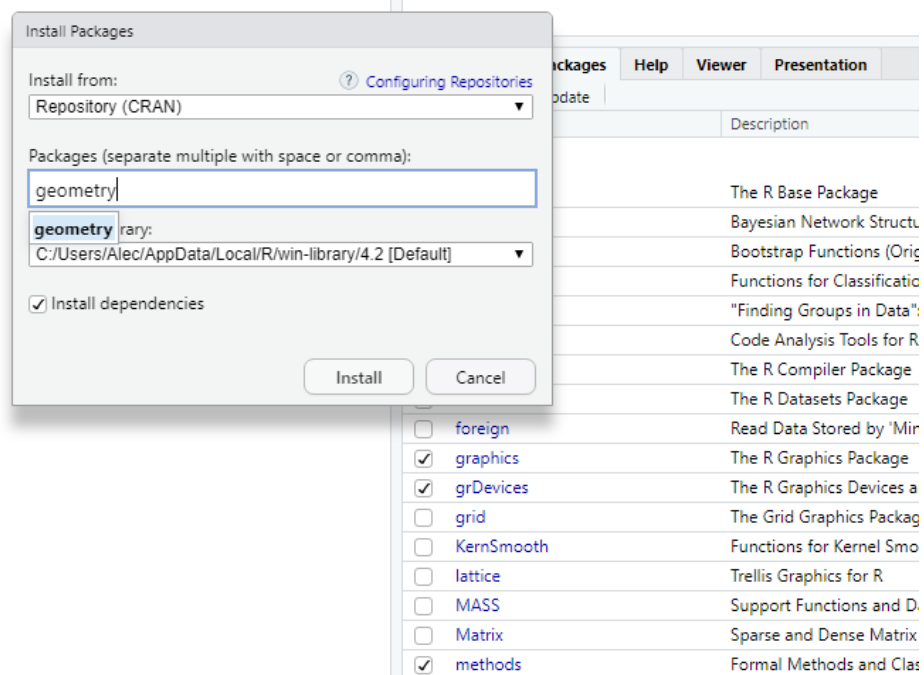


Рис. 1.9 Процес завантаження і встановлення нового пакету в RStudio.

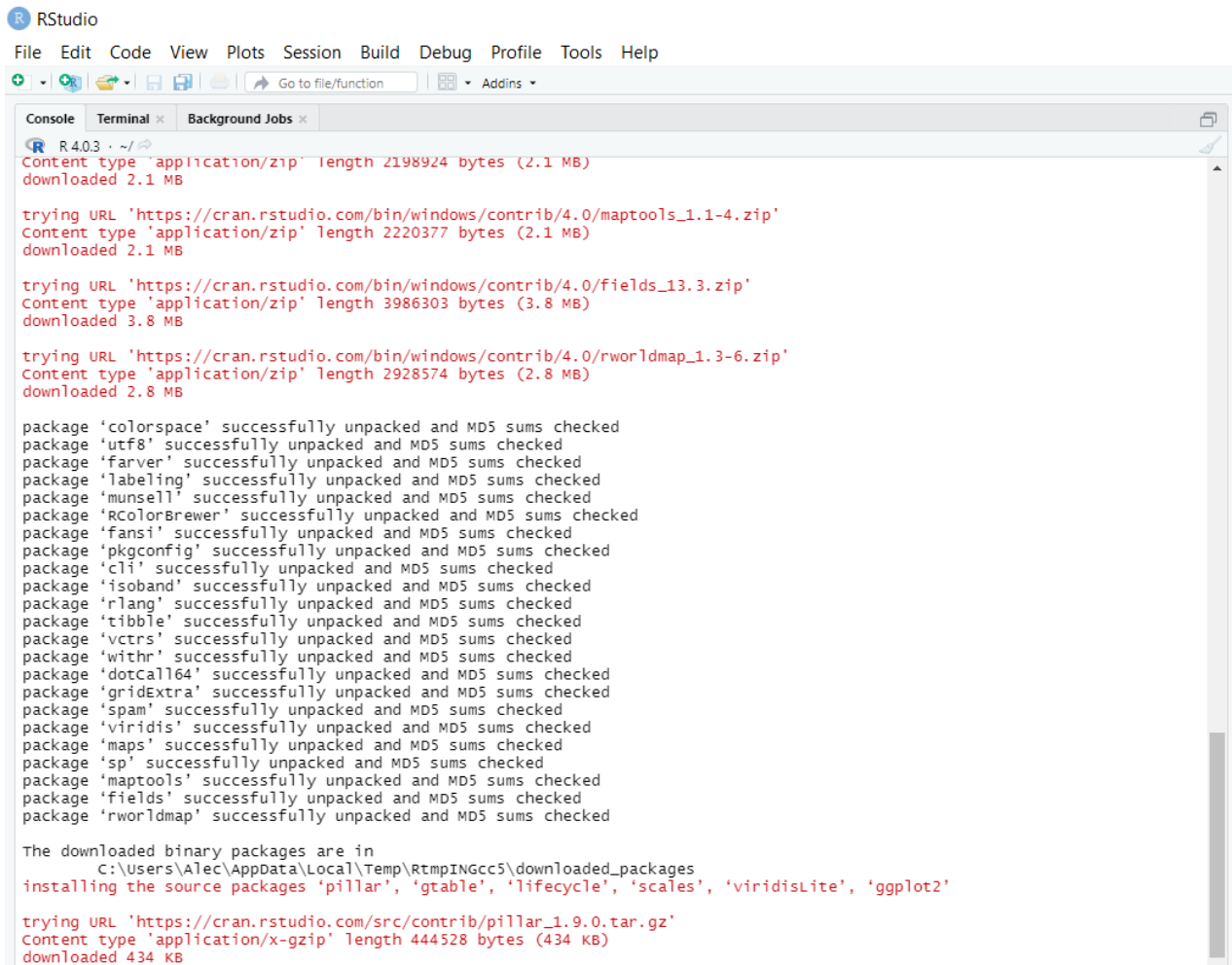
R має широкі графічні можливості, з якими можна ознайомитися, наприклад, набравши у командному рядку:

```
> demo(graphics)
> demo(image)
> demo(persp)
```

для отримання відображення можливостей R в 1-D, 2-D і 3-D зображеннях.

R також має потужні функції для створення зображень і додавання контурів. Наприклад, набір даних «Bathymetry» (сукупність даних про глибини водного об'єкта) з пакету `marelac` може використовуватися для генерації батиметрії (і гіпсометрії) Світового океану (і Землі):

```
> require(marelac)
> image(Bathymetry$x,Bathymetry$y,Bathymetry$z,col=femmecol(
100,asp=TRUE,xlab="",ylab=""))
> contour(Bathymetry$x,Bathymetry$y,Bathymetry$z,add=TRUE)
```



```
R 4.0.3 ~/  
Content type 'application/zip' length 2198924 bytes (2.1 MB)  
downloaded 2.1 MB  
  
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.0/maptools_1.1-4.zip'  
Content type 'application/zip' length 2220377 bytes (2.1 MB)  
downloaded 2.1 MB  
  
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.0/fields_13.3.zip'  
Content type 'application/zip' length 3986303 bytes (3.8 MB)  
downloaded 3.8 MB  
  
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.0/rworldmap_1.3-6.zip'  
Content type 'application/zip' length 2928574 bytes (2.8 MB)  
downloaded 2.8 MB  
  
package 'colorspace' successfully unpacked and MD5 sums checked  
package 'utf8' successfully unpacked and MD5 sums checked  
package 'farver' successfully unpacked and MD5 sums checked  
package 'labeling' successfully unpacked and MD5 sums checked  
package 'munsell' successfully unpacked and MD5 sums checked  
package 'RColorBrewer' successfully unpacked and MD5 sums checked  
package 'fansi' successfully unpacked and MD5 sums checked  
package 'pkgconfig' successfully unpacked and MD5 sums checked  
package 'cli' successfully unpacked and MD5 sums checked  
package 'isoband' successfully unpacked and MD5 sums checked  
package 'rlang' successfully unpacked and MD5 sums checked  
package 'tibble' successfully unpacked and MD5 sums checked  
package 'vctrs' successfully unpacked and MD5 sums checked  
package 'withr' successfully unpacked and MD5 sums checked  
package 'dotCall64' successfully unpacked and MD5 sums checked  
package 'gridExtra' successfully unpacked and MD5 sums checked  
package 'spam' successfully unpacked and MD5 sums checked  
package 'viridis' successfully unpacked and MD5 sums checked  
package 'maps' successfully unpacked and MD5 sums checked  
package 'sp' successfully unpacked and MD5 sums checked  
package 'maptools' successfully unpacked and MD5 sums checked  
package 'fields' successfully unpacked and MD5 sums checked  
package 'rworldmap' successfully unpacked and MD5 sums checked  
  
The downloaded binary packages are in  
C:\Users\Alec\AppData\Local\Temp\RtmpINGcc5\downloaded_packages  
installing the source packages 'pillar', 'gtable', 'lifecycle', 'scales', 'viridisLite', 'ggplot2'  
  
trying URL 'https://cran.rstudio.com/src/contrib/pillar_1.9.0.tar.gz'  
Content type 'application/x-gzip' length 444528 bytes (434 KB)  
downloaded 434 KB
```

Рис. 1.10 Встановлення взаємозалежних пакетів разом із новим пакетом в RStudio.

При цьому команда «asp=TRUE» використовується для підтримки співвідношення сторін.

Аналогічні інструменти можуть бути використані і для візуалізації даних при проведенні маркетингових досліджень, коли необхідно нанести результати на мапу. Так, на мапі можуть бути відображені рівні прибутку компанії чи продажів продукції.

Для вирішення маркетингових завдань часто застосовують пакет `rworldmap`, приклад візуалізації для якого наведено на рис. 1.11.

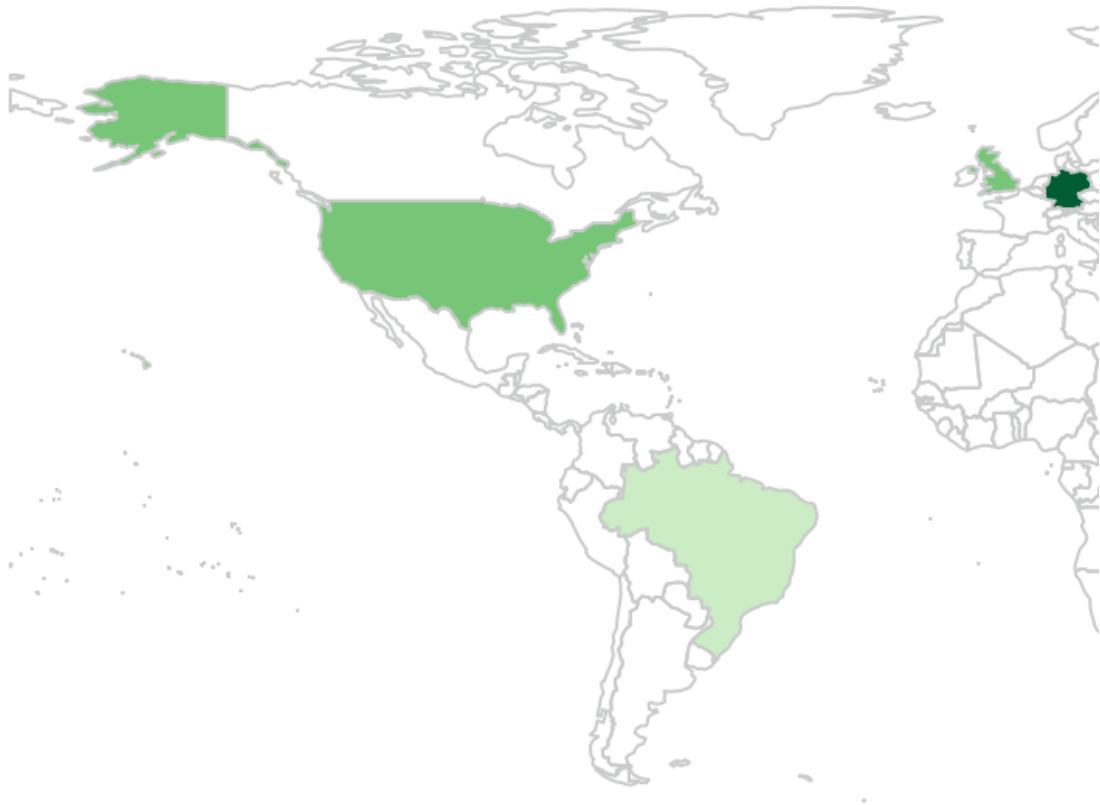


Рис. 1.11 Візуалізація даних в RStudio із використанням пакету `rworldmap`.

1.6 Базові приклади застосування мови R

Вважається, що R-код є досить легким для сприйняття. Проте обов'язково треба пам'ятати, що:

- 1) «<-» це оператор присвоювання;
- 2) вирази, що починаються з «#», трактуються як коментарі.

Наприклад:

```
> x <- c(1,2,3,4,5) # створюємо числовий вектор  
> y <- 2^x # підносимо число до степеню x  
> y # переглядаємо зміст об'єкта y, аналогічно до print(y)  
[1] 2 4 8 16 32
```

```
> b1 <- c("New York","Kyiv","Los Angeles") # створюємо символний  
вектор
```

```
> b1
```

```
[1] "New York" "Kyiv" "Los Angeles"
```

3) мова програмування R є дуже чутливою до регістру, оскільки «a» і «A» є двома різними об'єктами.

1.6.1 Використання R як наукового калькулятора, робота з консоллю

Як правило, при використанні програмного середовища R як наукового калькулятора користувачі повинні записувати команди, використовуючи командний рядок – консоль. Як і у звичайному калькуляторі, у R використовуються символи «+», «-», «*», «/», «^» (ступінь) тощо, а за необхідності дотримуватись необхідного порядку виконання математичних операцій потрібно використовувати дужки **()**.

Це є особливо важливим, якщо користувач не впевнений у пріоритеті певних математичних операцій. У такому разі за допомогою дужок потрібно точно позначити, в якому порядку необхідно здійснити робити обчислення.

Отже, розглянемо детальніше приклади обчислень деяких арифметичних операцій за допомогою мови R.

Приклад 1.

Для виконання наведених в умові (а)–(д) завдань у командному рядку вводимо вираз і отримуємо бажаний результат:

```
а) 41 + 3 = ?
```

```
> 41+3
```

```
[1] 44
```

б) $3 - 2 = ?$

> 3-2

[1] 1

в) $5 \cdot 6 = ?$

> 5*6

[1] 30

г) $99 / 9 = ?$

> 99/9

[1] 11

г) $2^3 = ?$

> 2^3

[1] 8

д) $(2 + 2) \cdot 2 = ?$

> (2+2)*2

[1] 8

У наступних прикладах розглянемо обчислення більш складних арифметичних виразів, що мають визначений порядок проведення математичних операцій.

Приклад 2. Для обчислення виразів, наведених у прикладах (а)–(д), у командному рядку треба набрати відповідні команди, після чого отримуємо відповідь:

а) $(4/6 \cdot 8 - 1)^{2/3}$

> (4/6*8-1)^(2/3)

[1] 2.657958

```
б) ln(20)
> log(20)
[1] 2.995732
```

```
в) log24096
> log2(4096)
[1] 12
```

```
г) 2•π•3
> 2 * pi * 3
[1] 18.84956
```

```
г) e(2 + cos0,5π)
> exp(2+cos(0.5*pi))
[1] 7.389056
```

```
д) (2,32 + 5,42 – 2•2,3•5,4•cos(π/8))1/2
> # length of 3rd side of a triangle with size 2.3 and 5.4 and angle pi/8
> sqrt(2.3^2+5.4^2-2*2.3*5.4*cos(pi/8))
[1] 3.391288 23
```

Якщо у користувача виникають сумніви щодо правильності використання тих чи інших функцій або команд, розробники R і активні користувачі програмного забезпечення підготували файли-довідки (документацію) із їх детальним описом. Зокрема, для виклику довідки потрібно набрати ?«+». Ця команда відкриє файл довідки із загальними арифметичними операторами.

Як зазначалось у попередніх підрозділах, програмне забезпечення R містить безліч пакетів (бібліотек). Зокрема, пакет в R є файлом, що містить багато функцій, які виконують певні специфічні завдання.

Як правило, пакети завантажують з офіційного веб-сайту R. Після установки для завантаження і отримання доступу до всіх доступних пакетів потрібно виконати такі команди, а саме, на прикладі пакету deSolve:

```
> library()
> library(deSolve)
> library(help=deSolve)
> help(package=deSolve)
```

Приклад 3. Введемо у вікні консолі:

```
> pi*2.234^5; 43*5/sqrt(56); log(36)
```

У цьому випадку отримуємо наступний результат:

```
[1] 174.8098
[1] 28.73058
[1] 3.583519
```

У цьому записі:

- sqrt і log є вбудованими в R функціями;
- pi – вбудована стала, а
- крапку з комою (;) використовуємо для відокремлення R-команд.

1.6.2 Отримання допомоги при використанні мови R

У R отримати довідку можна декількома шляхами. Крім звичайного вікна довідки, доступ до неї можна отримати з командного рядка консолі.

Зокрема, щоб отримати необхідну довідку, наприклад, про певну функцію, то потрібно ввести символ «?» перед назвою функції.

Приклад 4. Якщо набрати у вікні консолі код наступного характеру:

```
> ?log  
> ?sqrt  
> ?sin
```

то отримаємо довідку-пояснення про функцію логарифмів (і експонент), а також корінь, тригонометричні функції тощо.

Приклад 5. Якщо набрати:

```
> ?Arithmetic
```

то отримаємо інформацію про арифметичні операції, які є доступними в R.

Дуже зручним є те, що більшість файлів з довідкою містять різноманітні приклади, які можна запустити за допомогою R-команди «example».

Приклад 6.

а) Вводимо наведену вище команду:

```
> example(matrix)
```

У такому разі будуть виконуватися всі приклади з файлу-довідки «matrix».

б) при введенні команди:

```
> example(pairs)
```

викликаємо потужний засіб візуалізації парних відносин. При цьому будуть виконані всі наявні приклади з файлу-довідки «pairs».

Можна також обрати тільки один приклад з файлу-довідки, скопіювати його у буфер обміну і потім вставити його у вікно консолі.

І основне програмне забезпечення R, і багато пакетів R містять якісний демонстраційний матеріал.

Приклад 7. а) При введенні:

```
> demo()
```

отримуємо весь список доступних для перегляду демонстрацій.

б) При введенні:

```
> demo(graphics)
```

отримуємо демонстрацію декількох простих графічних засобів.

Зазначимо, що для більшості команд, функцій та пакетів мови R є детальні описи, які можна подивитися за допомогою команди **help**, наприклад:

```
> help("summary")
```

Іноді у користувачів виникає потреба надати в програмі посилання на певне джерело, пакет тощо, наприклад, «див. офіційне джерело». У такому разі в R передбачена функція **citation**, яка виводить інформацію, що надається розробниками пакету як «офіційне літературне посилання» (але це стосується тільки інформації, що закладена у поточну версію пакету).

Наприклад, якщо для посилань на пакети **R** використано BibTeX-нотацію:

```
> citation("startup")
> ##
> ## To cite startup in publications, please use:
> ##
> ## E. Ries. The Lean Startup: How Today's Entrepreneurs Use
Continuous Innovation to Create Radically Successful Businesses.
> ## Crown Business New York, 2011.
> ##
> ## A BibTeX entry for LaTeX users is
> ##
> ## @Book{
> ##   author = {Eric Ries},
> ##   title = {The Lean Startup: How Today's Entrepreneurs Use
Continuous Innovation to Create Radically Successful Businesses},
> ##   publisher = {Crown Business New York},
> ##   year = {2011},
> ##   isbn = {9780670921607},
> ##   url = {https://cutt.ly/fwqTzfh5},
```

Окрім цього, користувачам потрібно запам'ятати декілька основних правил використання мови R. Наприклад, шляхи в R записуються з косою рисою «/» (нагадуємо, що в ОС Windows використовуються зворотні косі риски «\»).

Отже, щоб встановити робочий каталог в R, потрібно набрати в командному рядку:

```
> setwd(«C:/R_Code/»)
```

Причому, якщо вираз на одній лінії є синтаксично правильним, то R буде виконувати його і у тому випадку, якщо він продовжується і на наступному (наступних) рядку (рядках).

Приклад 8.

Запишемо у командному рядку:

```
> A <- 5 + cos(pi)
> -sqrt(7)
[1] -2.645751
```

У такому разі A буде мати значення $(5 + \cos(\pi))$, і R виведе значення $(-\sqrt{7})$.

Проте при записі:

```
> 5 + cos(pi) -
> sqrt(7)
[1] 1.354249
```

R буде виводити значення $5 + \cos(\pi) - \sqrt{7}$, оскільки вираз на першому рядку є синтаксично незавершеним.

Тобто за такого запису R припускає, що код буде продовжений на наступному рядку. Проте фахівці радять бути дуже уважними і, за можливості, уникати розділення виразів на декілька рядків.

РОЗДІЛ 2

ВИКОНАННЯ ОБЧИСЛЕНЬ У R З ВИКОРИСТАННЯМ ЗМІННИХ

2.1. Числа, вектори, матриці та масиви

2.1.1 Основні правила присвоєння

У мові програмування R при використанні змінних вони мають обов'язково бути ініціалізовані числами.

Приклад 1.

```
> A <- 1  
> B <- 2  
> A + B  
[1] 3
```

Окрім того, як аргументи функцій R може приймати не тільки числа, а й вектори, матриці чи масиви.

Приклад 2.

```
> V <- factorial(10)
```

Отже, у цьому прикладі спочатку виконується обчислення $10!$, тобто:

```
(1*2*3*4*5*6*7*8*9*10)
```

Потім оператор «<—» присвоює результат даного обчислення змінній V.

Зазначимо, що змінна V потім може бути використана у подальших розрахунках, наприклад:

```
> V/10  
[1] 362880
```

Проте присвоєння значення змінній V у вікні його не відображає. Якщо потрібно відобразити значення V, то користувачу треба написати відповідну команду:

```
> V  
[1] 3628800
```

Користувач також може присвоїти результат обчислення змінної і переглянути цей результат, взявши вираз в дужки.

Приклад 3.

```
> (X <- sin(3/2*pi))  
[1] -1
```

Як зазначалось раніше, окрім цілих (integer), дійсних і комплексних (complexnumber) чисел (numeric), мова R здатна розпізнавати нескінченність (Inf) і навіть ті значення, які є невизначеними (NaN).

Приклад 4.

```
a)  
> 1/0  
[1] Inf
```

б)

```
> 0/0
```

```
[1] NaN
```

в)

```
> 1e-8*1000
```

```
[1] 1e-05
```

В останньому прикладі, зазначимо, що «e-8» означає 10^{-8} .

2.1.2 Операції з векторами

Як відомо, вектори складаються з упорядкованого набору чисел і можуть бути створені різними способами. Наприклад:

- можна використати R-функцію «vector»;
- функція «c()» поєднає числа у вектор;
- оператор «:» створить послідовність значень, причому, кожне наступне значення буде більшим (або меншим) за попереднє;
- більш загальну послідовність можна створити за допомогою R-функції «seq».

Приклад 5.

а)

```
> c(0, pi/2, pi, 3*pi/2, 2*pi)
```

```
[1] 0.000000 1.570796 3.141593 4.712389 6.283185
```

б)

```
> seq(from = 0, to = 2*pi, by = pi/2)
```

```
[1] 0.000000 1.570796 3.141593 4.712389 6.283185
```

У цьому прикладі користувачу треба звернути увагу на те, що R-функція «seq» як вихідні параметри приймає «from» (початкове значення), «to» (кінцеве значення) і «by»(крок).

в)

```
> seq(0, 2*pi, pi/2)
[1] 0.000000 1.570796 3.141593 4.712389 6.283185
```

У цьому прикладі зазначимо, що у випадку, коли порядок зберігається, то можна не вказувати ім'я параметрів.

Як можна побачити з прикладу 4 (а)–(в), всі команди нададуть користувачу створити вектор, що складається з $0, \pi, \dots, 2\pi$.

Приклад 6.

Наступна команда надає змогу обчислити синус цього вектора і, відповідно, виводить результат обчислення:

```
> sin(seq(0, 2*pi, pi/2))
[1] 0.000000e+00 1.000000e+00 1.224606e-16 -1.000000e+00
-2.449213e-16
```

Приклад 7.

Розглянемо наступний вираз, який створює послідовність цілих чисел від 1 до 100 і обчислює квадратний корінь з усіх них, а потім відображає результат обчислення на екрані:

```
> V <- 1:100
```

```
> sqrt(V)
```

```
[1] 1.000000 1.414214 1.732051 2.000000 2.236068 2.449490 2.645751  
[8] 2.828427 3.000000 3.162278 3.316625 3.464102 3.605551 3.741657  
[15] 3.872983 4.000000 4.123106 4.242641 4.358899 4.472136 4.582576  
[22] 4.690416 4.795832 4.898979 5.000000 5.099020 5.196152 5.291503  
[29] 5.385165 5.477226 5.567764 5.656854 5.744563 5.830952 5.916080  
[36] 6.000000 6.082763 6.164414 6.244998 6.324555 6.403124 6.480741  
[43] 6.557439 6.633250 6.708204 6.782330 6.855655 6.928203 7.000000  
[50] 7.071068 7.141428 7.211103 7.280110 7.348469 7.416198 7.483315  
[57] 7.549834 7.615773 7.681146 7.745967 7.810250 7.874008 7.937254  
[64] 8.000000 8.062258 8.124038 8.185353 8.246211 8.306624 8.366600  
[71] 8.426150 8.485281 8.544004 8.602325 8.660254 8.717798 8.774964  
[78] 8.831761 8.888194 8.944272 9.000000 9.055385 9.110434 9.165151  
[85] 9.219544 9.273618 9.327379 9.380832 9.433981 9.486833 9.539392  
[92] 9.591663 9.643651 9.695360 9.746794 9.797959 9.848858 9.899495  
[99] 9.949874 10.000000
```

Нагадаємо, що оператор «<-» присвоює послідовність в V.

Розглянемо приклади з використанням оператора «:».

Приклад 8.

а)

```
> (X <- 0.5:10.5)
```

```
[1] 0.5 1.5 2.5 3.5 4.5 5.5 6.5 7.5 8.5 9.5 10.5
```

б)

```
> 6:1
```

```
[1] 6 5 4 3 2 1
```

У мові програмування R можна надати власні назви окремим елементам вектора.

Приклад 9.

а)

```
> (fruit <- c(banana = 1, apple = 2, orange = 3))  
banana apple orange  
      1      2      3
```

б)

```
> names(fruit)  
[1] "banana" "apple" "orange"
```

2.1.3 Робота в R з матрицями

У мові програмування R матриці можна також створити декількома способами, які значно відрізняються один від одного, зокрема:

- за допомогою R-функції «matrix»;
- використовуючи R-функцію «diag», яка побудує діагональну матрицю;
- R-функції «cbind» і «rbind» надають змогу додавати стовпці і рядки до вже існуючої матриці або в інший вектор.

Наприклад, такий вираз, як:

```
> A <- matrix(nrow = 2, data = c(1, 2, 3, 4))
```

створить матрицю A з двома рядками і двома стовпчиками (оскільки є чотири елементи). При цьому дані потрібно вводити як вектор, тобто за допомогою R-функції c().

Приклад 10. Наступні два приклади відображають матрицю **A** і матрицю квадратних коренів з її елементів:

а)

```
> A
```

```
      [,1] [,2]  
[1,]    1    3  
[2,]    2    4
```

б)

```
> sqrt(A)
```

```
      [,1]      [,2]  
[1,] 1.000000 1.732051  
[2,] 1.414214 2.000000
```

При цьому користувачу треба пам'ятати, що за замовчуванням R заповнює матриці по стовпцях. Але це "правило" можна змінити, використовуючи параметр «byrow».

Приклад 11.

```
> (M <- matrix(nrow = 4, ncol = 3, byrow = TRUE, data = 1:12))
```

```
      [,1] [,2] [,3]  
[1,]    1    2    3  
[2,]    4    5    6  
[3,]    7    8    9  
[4,]   10   11   12
```

Одиничну матрицю (**I**) створюємо за допомогою R-функції «diag».

Приклад 12.

```
> diag(1, nrow = 2)
      [,1] [,2]
[1,]    1    0
[2,]    0    1
```

У наступному прикладі показано, як можна встановити назви стовпців і рядків під час формування матриці. При цьому треба пам'ятати, що у цьому випадку ми використовуємо R-функцію «`c()`», оскільки назви рядків і стовпців є векторами, що містять певні рядки.

Приклад 13.

```
> rownames(A) <- c("x","y")
> colnames(A) <- c("a","b")
> A
      a b
x 1 3
y 2 4
```

Матриці в R також можна створювати шляхом об'єднання (зв'язування) векторів, наприклад, по рядках.

Приклад 14.

```
> V <- 0.5:5.5
> rbind(V, sqrt(V))
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
V 0.5000000 1.500000 2.500000 3.500000 4.50000 5.500000
  0.7071068 1.224745 1.581139 1.870829 2.12132 2.345208
```

Приклад 15. У даному випадку $t(A)$ виконує транспонування матриці A , тобто буде здійснено перестановку рядків і стовпців.

```
> t(A)
  x y
a 1 2
b 3 4
```

2.1.4 Робота з масивами і розмірності

Під терміном "масив" ми розуміємо багатовимірні узагальнення матриць. Тобто і матриці, і масиви в R є векторами з атрибутом розмірності.

Отже, багатовимірний масив в R створюється таким чином:

```
> AR <- array(dim = c(2, 3, 2), data = 1)
```

У цьому випадку AR – масив $2 \times 3 \times 2$, а всі його елементи – 1:

Приклад 16.

```
> AR
, , 1
  [,1] [,2] [,3]
[1,]  1   1   1
[2,]  1   1   1

, , 2
  [,1] [,2] [,3]
[1,]  1   1   1
[2,]  1   1   1
```

Приклад 17. При введенні команд:

а)
`> length(A)`
`[1] 4`

б)
`> dim(A)`
`[1] 2 2`

в)
`> ncol(M)`
`[1] 3`

г)
`> nrow(M)`
`[1] 4`

Таким чином, команда у прикладі (а) повертає довжину (або загальну кількість елементів) V (вектора або матриці);

команда у прикладі (б) – розмірність матриці або масиву A ;

команда у прикладі (в) – кількість стовпців і

команда у прикладі (г) – кількість рядків матриці M , відповідно.

2.3 Вибір і вилучення елементів у R

За необхідності вибору підмножини векторів або матриць можна або вказати номери елементів, які нам потрібно (проста індексація), або вказати вектор логічних значень (TRUE / FALSE), щоб вказати, які саме елементи

включити (TRUE) чи які саме елементи не включити (FALSE). Для цього в R передбачено використання спеціальних логічних виразів.

2.3.1 Проста індексація

Елементи векторів, масивів чи матриць індексуються в R, використовуючи оператор «[]».

Приклад 18.

а)

`> M[1, 1]`

б)

`> M[1, 1:2]`

в)

`> M[1:3, c(2,4)]`

Результатом у прикладі (а) буде елемент з першого рядка і першого стовпця матриці M, яка була попередньо задана; у прикладі (б) – вибірка елементів з першого рядка і перших двох стовпців; у прикладі (в) – елементи з перших трьох рядків, другого і четвертого стовпців матриці M.

Якщо індекс пропущено, то у такому разі всі рядки (якщо пропущено перший індекс) або всі стовпці (якщо пропущено другий індекс) обираються.

Приклад 19. У цих прикладах:

а)

`> M[, 2] <- 0`

б)

```
> M[1:3,] <- M[1:3,]*2
```

спочатку всі елементи в другому стовпці (приклад (а)) матриці *M* обнуляються, а потім елементи на перших трьох рядках матриці *M* множаться на 2 (приклад (б)).

Приклад 20. Подібні методи вибору застосовуються і до векторів:

а)

```
> V[1:10]
```

```
[1] 0.5 1.5 2.5 3.5 4.5 5.5 NA NA NA NA
```

б)

```
> V[seq(from = 1, to = 5, by = 2)]
```

```
[1] 0.5 2.5 4.5
```

Отже, у виразі прикладу (а) вибираються перші 10 елементів вектора *V*, а у виразі прикладу (б) – 1-й, 3-й і 5-й елементи вектора *V*.

2.3.2 Логічні вирази

Логічні вирази в *R* використовуються для того, щоб вибрати елементи векторів (матриць), які підпорядковуються певним, наперед заданим, критеріям.

Наприклад, мова *R* розрізняє такі логічні змінні, як *TRUE* (істина) і *FALSE* (хиба), які представлені числами 1 і 0.

Зазначимо, що ці вирази можна (хоча й небажано) замінити на «Т» і «F», проте неможна написати як *True* і *False*.

При цьому операції:

> ?Comparison

> ?Logic

нададуть перелік реляційних і логічних операторів, що є доступними в R.

Команда:

```
> (V <- seq(-2, 2, 0.5))
```

```
[1] -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
```

```
> V > 0
```

```
[1] FALSE FALSE FALSE FALSE FALSE TRUE TRUE TRUE TRUE
```

поверне TRUE для тих значень послідовності V, які є додатними; у той час як команда:

```
> V[V > 0]
```

```
[1] 0.5 1.0 1.5 2.0
```

вибере додатні значення з V.

Команда:

```
> V[V > 0] <- 0
```

обнулить всі додатні елементи в V, а команда:

```
> sum(V < 0)
```

```
[1] 4
```

поверне кількість від'ємних елементів, тобто просумує значення TRUE(=1).

Команда:

```
> V[V != 0]  
[1] -2.0 -1.5 -1.0 -0.5
```

покаже всі ненульові елементи з V (це оператор логічного заперечення («не»)).

Логічні вирази можна об'єднувати, наприклад, за допомогою операторів «|» (логічний оператор «або») та «&» (логічний оператор «і»). Можна порівнювати різні змінні між собою (ставимо «==», наприклад:

```
> a == b  
[1] FALSE
```

Приклад 21.

```
> V[V < (-1) | V > 1 ]  
[1] -2.0 -1.5
```

Ця команда виводить на екран всі значення з V , які є <-1 або >1 . Окрім того, треба звернути увагу на те, що при наборі команди на консолі «-1» подано у дужках $()$.

Приклад 22. Команда

```
> which(V == 0)  
[1] 5 6 7 8 9  
  
> which.min(V)  
[1] 1
```

поверне індекс елемента із значенням 0 і мінімальним значенням.

2.3.3 Видалення елементів

Якщо перед індексом елемента поставити «-», то даний елемент буде видалено, наприклад:

```
> M[,-1]
```

При цьому на екран буде виведено вміст матриці M , за винятком першого стовпчика.

Приклад 23.

а)

```
> x<-x[-1]
```

б)

```
> M<-M[-1,]
```

в)

```
> V<-V[-V=0]
```

У наведених прикладах видалиться перший елемент X (у прикладі (а)); 1-й рядок матриці M (у прикладі (б)) і всі додатні елементи V (у прикладі (в)). За необхідності отримання додаткової інформації потрібно набрати у командному рядку команду:

```
> ?Extract
```

2.4 Створення функцій та елементи програмування в R

2.4.1 Оголошення функцій

У мові програмування R користувачі можуть оголошувати свої власні функції, які можна додати до вже вбудованих функцій R. Як правило, складні функції пишуться у R-script файлах, які потім підключають до платформи R.

Приклад 24. Команда:

```
>Circlesurface<- function (radius) pi*radius^2
```

визначає функцію з назвою «Circlesurface»), яка приймає як вхідний аргумент змінну «radius» і яка потім повертає значення $\pi \cdot \text{radius}^2$ (вираз для обчислення площі кола з радіусом «radius»).

Отже, після підключення цієї функції до середовища R, можна її використовувати для обчислення площі кола із заданим радіусом.

Приклад 25. Команда:

```
> Circlesurface(10)  
[1] 314.1593
```

або команда

```
> Circlesurface(1:20)
```

[1]	3.141593	12.566371	28.274334
[4]	50.265482	78.539816	113.097336
[7]	153.938040	201.061930	254.469005
[10]	314.159265	380.132711	452.389342
[13]	530.929158	615.752160	706.858347
[16]	804.247719	907.920277	1017.876020
[19]	1134.114948	1256.637061	

яка буде обчислювати площі кіл з радіусами 1, 2, 3, ... , 20.

Приклад 26. Більш складні функції можуть повертати більше одного елемента:

```
> Sphere <- function(radius)
+ {
+   volume <- 4/3*pi*radius^3
+   surface <- 4 *pi*radius^2
+   return(list(volume=volume,surface=surface))
}
```

Таким чином, у наведеному прикладі визначаємо:

- заголовок функції (перший рядок прикладу), вказавши ім'я функції («Sphere») і вхідний параметр («radius»);

- тіло функції.

Оскільки функція включає декілька операторів, то тіло функції пишеться всередині фігурних дужок {...};

- значення, що повертаються (останній рядок тіла функції).

Тобто функція «Sphere» поверне об'єм і площу поверхні сфери у вигляді списку.

Наприклад, якщо Земля має радіус близько 6371 км, то її об'єм (у км³) і площа поверхні (у км²) будуть дорівнювати:

```
> Sphere(6371)
$volume
[1] 1.083207e+12
```

```
$surface
[1] 510064472
```

Приклад 27. За такого запису команди на екрані будуть відображатися об'єми сфер з радіусами 1, 2, ... 5:

```
> Sphere(1:5)$volume
[1] 4.18879 33.51032 113.09734 268.08257
[5] 523.59878
```

Під час однотипних розрахунків є необхідність, щоб, наприклад, деякі вхідні параметри мали конкретне значення «за замовчуванням».

Приклад 28. Функція оцінює щільність «стандартної океанської води» (у кг/м³) залежно від температури T (для солоності $\text{salinity} = 0$, тиску $\text{pressure} = 1$ атм) (Міллер і Пуассон, 1981); вхідний параметр T за замовчуванням дорівнює 20 °C:

```
> Rho_W <- function(T=20)
+ {
+   999.842594 + 0.06793952 * T - 0.00909529 * T^2 +
+     0.0001001685 * T^3 - 1.120083e-06 * T^4 + 6.536332e-09
+   * T^5
+ }
```

Як можна побачити, після першого рядка прикладу стоїть «+», тобто цим вказано, що вираз не закінчився і має продовження на наступному рядку.

Виклик функції без вказування температури використовує значення «за замовчуванням»:

```
> Rho_W()  
[1] 998.2063  
> Rho_W(20)  
[1] 998.2063  
> Rho_W(0)  
[1] 999.8426
```

2.4.2 Обчислення функцій

Приклад 29. Знаходження квадратного кореню з числа:

```
> 16^0.5  
[1] 4
```

Можна також скористатися спеціальною функцією `sqrt()`:

```
> sqrt(16)  
[1] 4
```

Приклад 30.

а) Обчислити логарифм $\log(8)$.

Нагадаємо що логарифм – показник степені, в яку потрібно піднести число (основа логарифму), щоб отримати задане число. За замовчуванням у

виразі `log(8)` основа дорівнює числу Ейлера (2,7182818...), тобто у такому вигляді функція обчислить натуральний логарифм числа:

```
> log(8)
[1] 2.079442
```

Отже, у функції `log()` в R є аргумент `base =`.

б)

```
> log(x = 8, base = 2)
[1] 3
```

в)

```
> log(8,2)
[1] 3
```

г) можна результат одних функцій використовувати як вхідні дані для інших, наприклад:

```
> log(8, sqrt(4))
[1] 3
```

д) Якщо експліцитно написати імена аргументів, то порядок їх запису в функції неважливий, наприклад:

```
> log(base = 2, x = 8)
[1] 3
```

е) У R можна недописувати імена аргументів, якщо вони не співпадають з іншими, наприклад:

```
> log(b = 2, x = 8)
```

```
[1] 3
```

Зазначимо, що арифметичні знаки, які було використано у попередніх прикладах, а саме: +, -, /, ^ тощо називаються операторами, і вони також є функціями:

```
> '+'(3,4)
```

```
[1] 7
```

2.4.3 Конструкції if, else, ifelse

Приклад 31. Запишемо фрагмент програми:

```
> Dummy <- function (x)
+ {
+   if ( x<0 ) string <- "x<0" else
+     if ( x<2 ) string <- "0>=x<2" else
+       string <- "x>=2"
+   print(string)
+ }
> Dummy(-1)
[1] "x<0"
> Dummy(1)
[1] "0>=x<2"
> Dummy(2)
[1] "x>=2"
```

У цьому прикладі вираз «else», записаний на тому самому рядку, що і частина «if», показує, що вираз продовжується на наступному рядку.

Проте конструкції «if» і «else» можуть поєднуватися в один оператор «**ifelse**».

Приклад 32.

```
> x<-2 36
```

```
> ifelse (x>0, "positive", "negative,0")
```

```
[1] "positive"
```

2.4.4 Організація циклів в R

У програмуванні цикли надають змогу виконувати певний набір заданих інструкцій кілька разів. Наприклад, цикл «for» перебирає певний набір значень.

Приклад 33.

Змінна «i» приймає значення (1, 2, 3), отже:

```
> for (i in 1:3) print(c(i,2*i,3*i))
```

```
[1] 1 2 3
```

```
[1] 2 4 6
```

```
[1] 3 6 9
```

Цикли «while» і «repeat» будуть виконуватися до тих пір, поки зазначена в завданні умова не буде виконана.

Приклад 34.

```
> i<-1 ; while(i<3) {print(i); i<-i+1}
```

```
[1] 1
```

```
[1] 2
```

Команда «break» реалізує вихід із циклу; команда «next» зупиняє поточну ітерацію і переходить до наступної ітерації.

Приклад 35.

```
> i<-1  
> {print(i)  
+ i <-i+1  
+ if(i>2) break  
+ }  
[1] 1
```

При цьому фігурні дужки «{...}» охоплюють одразу декілька операторів, які виконуються в кожній ітерації.

Проте фахівці вважають, що цикли в R реалізуються досить неефективно, а, отже, якщо це можливо, то їх використання потрібно обмежити. Якщо залишаються питання стосовно використання у R циклів або, наприклад, конструкцій з «if», то можна набрати у командному рядку:

```
> ?Control
```

2.4.5 Інтерполяція і згладжування в R

Інтерполяцію і згладжування в мові R можна робити кількома способами, а саме:

- оператор `approx` лінійно інтерполює через точки;
- `spline` використовує інтерполяцію сплайн, яка є більш «гладкою»;
- `smooth.spline` розгладжує набори даних (цей метод не пов'язує оригінальні точки).

Приклад 36.

```
> x <- 1:10  
> y <- c(9,8,6,7,5,8,9,6,3,5)  
> plot(x,y,pch=16,cex=2,main="interpolation,smoothing")  
> lines (spline(x,y, n=100),lty=1)  
> points(approx(x,y, xout=seq(1,10,0.1)),pch=1)  
> lines (smooth.spline(x,y),lty=2)  
> legend("bottomleft",lty=c(1,NA,2),pch=c(NA,1,NA),  
+ legend=c("spline","approx","smooth.spline"))
```

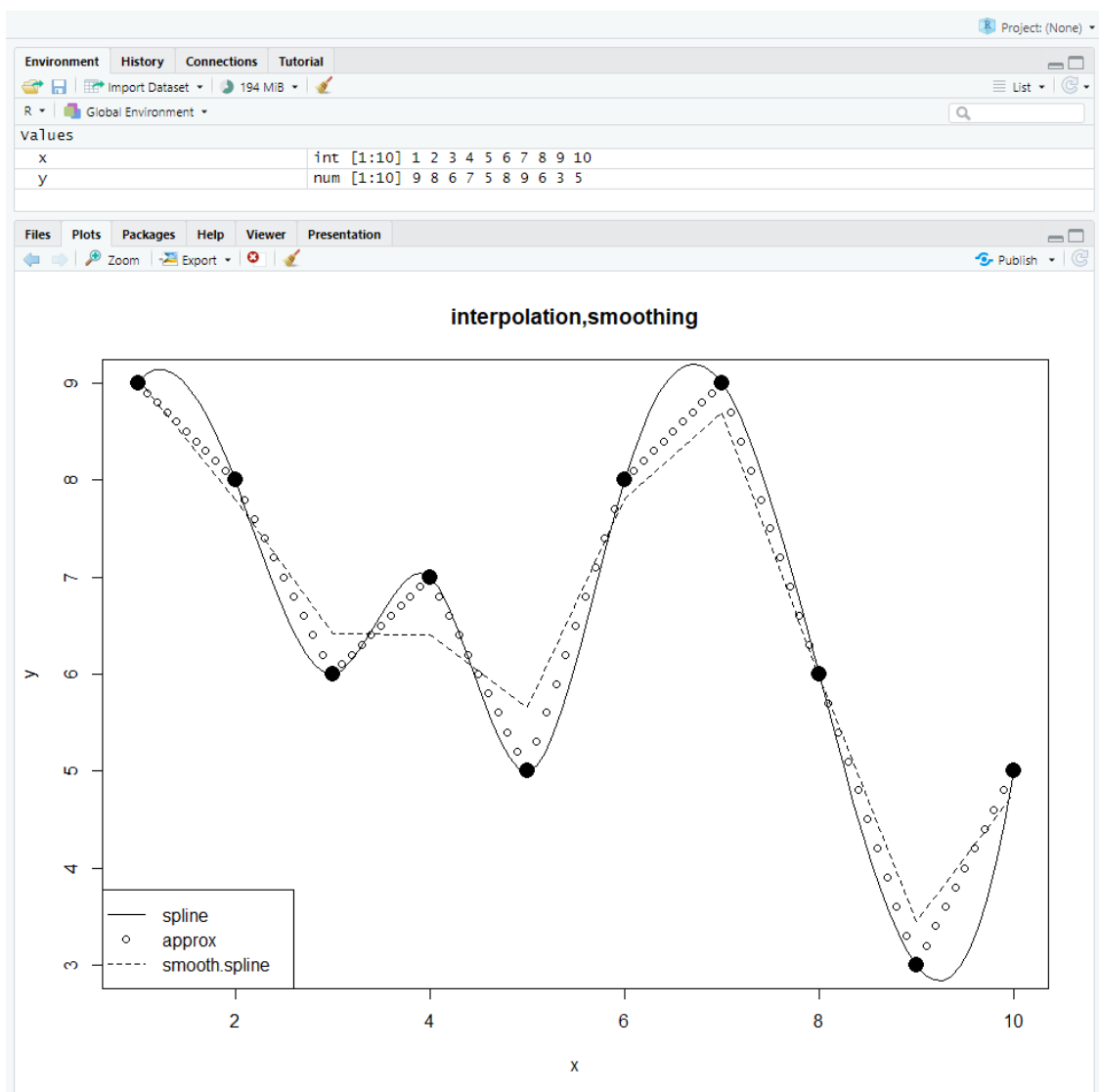


Рис. 2.1 Результат виконання прикладу 36 в RStudio.

РОЗДІЛ 3

СТАТИСТИЧНЕ ОПРАЦЮВАННЯ ДАНИХ, ВИКОРИСТАННЯ СТАТИСТИЧНИХ ПАКЕТІВ У МОВІ R

3.1 Теоретичні основи статистичного опрацювання даних

3.1.1 Термінологія статистичного опрацювання даних

Впровадження в науку і виробництво новітніх інформаційних технологій спричинило математизацію всіх сфер діяльності людини. Як наслідок, методи математичної статистики і теорії імовірностей все більш широко застосовуються при обробці результатів соціально-економічних та інших досліджень, прийнятті різноманітних управлінських рішень в умовах певної невизначеності, для перевірки висунутих наукових гіпотез, побудови математичних моделей різноманітних об'єктів та явищ, що відбуваються в природі та суспільстві. Крім того, коректна математична обробка результатів досліджень є обов'язковим і необхідним етапом наукової та виробничої діяльності.

Результатом спостереження, вимірювання, експерименту тощо може бути число, функція, вектор, певна множина чисел, причому, як іменованих, так і неіменованих. Відповідно, всякому експерименту можуть бути притаманні всі види похибок – і систематичні, і випадкові. Похибки можуть бути зумовлені також неправильним застосуванням методики дослідження, недосконалістю використовуваної математичної моделі, певним класом точності вимірювальної техніки, технічних засобів тощо, впливом зовнішніх і внутрішніх чинників.

Отже, результат вимірювання є випадковою величиною з наперед невідомими характеристиками, які й оцінюють за допомогою статистичного

опрацювання даних. Тобто оцінюється не тільки сам результат вимірювання, а й відбувається оцінювання таких характеристик результату, як набір похибок, невизначеність результату, його вірогідність, надійність і достовірність.

Під час проведення експерименту, спостережень чи вимірювань не завжди можна дослідити всі елементи. Отже, дослідники, як правило, застосовують вибірковий метод досліджень. І основний принцип застосування цього методу дослідження полягає у тому, щоб результат, отриманий при дослідженні вибірки, можна було б розповсюдити на всю генеральну сукупність.

Розглянемо деякі з основних термінів, які найчастіше застосовуються у вибіркових методах дослідження. Усю сукупність об'єктів, які мають певну спільну властивість та підлягають дослідженню, називають **генеральною сукупністю** (ГС). Генеральною сукупністю також можна назвати всю сукупність об'єктів, з яких робиться вибірка.

Випадкова вибірка це сукупність елементів, узятих з генеральної сукупності з метою отримання інформації про неї. Як зазначалось, випадкова вибірка повинна якомога краще відображати особливості всіх складових ГС. Отже, щоб отримувані вибірковим методом оцінки були достовірними, вибірка має бути репрезентативною (представницькою). Це, як правило, забезпечується випадковим і незалежним відбором її окремих елементів.

Коректно взята вибірка буде лише випадково відрізнятися від всієї ГС, і ці випадковості та вірогідність їх появи можна врахувати за допомогою методів математичної статистики.

Припустимо, що експеримент повторено n разів (тобто проведено декілька серій паралельних дослідів) без зміни умов експерименту. Тоді маємо n реалізацій досліджуваної випадкової величини X . Кожна з цих реалізацій x_i ($i = 1, 2, \dots, n$) також є випадковою величиною, а тому й результат даного експерименту можна розглядати як певний набір випадкових величин $X_1, X_2, \dots, X_n$.

Причому, якщо усі значення випадкової величини $X_1, X_2, \dots, X_n$ є незалежними у ГС, то кожна з них має той самий закон розподілу, що й досліджувана випадкова величина X .

Випадковий вектор $(X_1, X_2, \dots, X_n)$ називають **вибірковим вектором**, а результати експерименту $X_1, X_2, \dots, X_n$, одержані при n -кратному повторенні досліду за незмінних умов, названо вибіркою обсягом n .

Таким чином, **вибірка** це будь-яка обмежена частина об'єктів ГС, елементи якої вибираються з ГС випадковим чином (найчастіше із застосуванням таблиць випадкових чисел). Вибіркою також називають сукупність статистично еквівалентних результатів спостережень, одержаних, наприклад, при паралельних визначеннях якої-небудь характеристики системи. Елементи вибірки називають варіантами.

Щоб дослідження, яке планується провести вибірковим методом, було коректним, треба обов'язково проводити рандомізацію вибірки.

Рандомізація (від англ. random – випадково, навмання) це прийом, який дозволяє перевести систематичні похибки експерименту в розряд випадкових. Можливість рандомізації базується на тому, що систематична похибка одиничного явища (приладу, процесу, методу аналізу тощо) при розгляді її в більш широкому аспекті однотипних явищ (серія приладів, група процесів або методів) стає величиною змінної, тобто набуває певні риси випадкової похибки.

Отже, середній результат, отриманий за умови багатофакторної рандомізації (при відповідній обробці та за наявності представницької вибірки), може слугувати найбільш об'єктивною оцінкою досліджуваної характеристики об'єкту дослідження.

Варіаційним рядом $x_1, x_2, \dots, x_n$ (де x_i – елемент варіаційного ряду) називають реалізацію випадкової вибірки, записану, наприклад, в порядку зростання її елементів: $x_1 \leq x_2 \leq \dots \leq x_n$. У разі аналізу вибірки значного обсягу, наприклад, вище 50 варіант, як правило, проводять розбиття варіаційного ряду на окремі класи (інтервали); у такому разі отримують інтервальний варіаційний ряд.

Різницю між найбільшим і найменшим значеннями досліджуваної величини (елементами вибірки) в упорядкованій вибірці називають **розмахом вибірки** або **варіаційним розмахом** (3.1):

$$R = x_n - x_1 \quad \text{або}$$

$$R = x_{\max} - x_{\min}. \quad (3.1)$$

Величина розмаху вибірки R залежить від її обсягу – при сталій випадковій похибці розмах варіювання зазвичай збільшується із зростанням кількості вимірювань.

Вибірковий розмах найбільш вдало можна застосовувати для характеристики розсіювання у вибірці досить малого обсягу (до десяти значень). Розмах вибірки для великих масивів даних (більше десяти) краще характеризувати за допомогою **інтерквартильного розмаху**. Для цього результати експерименту (спостереження) спочатку впорядковують у порядку їх зростання, $x_1 \leq x_2 \leq \dots \leq x_n$, а потім обчислюють інтервал, що охоплює 50 % вимірювань.

Квартиль це таке значення випадкової величини, нижче якого лежить частина розподілу ймовірностей випадкової величини, кратна одній четвертій. Отже, інтерквартильний розмах одержують як різницю між тим значенням досліджуваної величини, що відсікає чверть експериментальних даних з одного боку (вище цього значення досліджуваної величини залишається чверть результатів), та значенням величини, що відсікає чверть даних з іншого боку (нижче цього значення досліджуваної величини лежить ще одна чверть результатів).

При будь-якому симетричному розподілі значень досліджуваної величини **квартиль** надає можливість з одного погляду отримати уявлення про діапазон розсіювання експериментальних даних. Для характеристики діапазону розсіювання

досліджуваної величини можна також застосовувати й інші аналогічні величини $p\%$, які називають **процентилями**.

При проведенні серії паралельних дослідів можливе неодноразове отримання однакових результатів (такий випадок є досить характерним для дискретної генеральної сукупності). Нехай вибірка $(x_1, x_2, \dots, x_n)$ має k різних значень величин $(q_1, q_2, \dots, q_k)$, а деяке число q_i зустрічається p_i -разів. Число p_i називають **частотою** елемента вибірки із значенням q_i . Слід зазначити, що для неперервної ГС частоти усіх елементів однакові й дорівнюють одиниці, а $\sum p_i = p = 1$.

Послідовність пар (q_i, p_i) називають **статистичним рядом** або **статистичним розподілом вибірки**. Цей ряд, наприклад, можна подати за допомогою таблиці виду:

q_i	q_1	q_2	...	q_k
p_i	p_1	p_2	...	p_k

або задати за допомогою формули або графіка. Закон розподілу генеральної сукупності можна представити у вигляді так званої функції розподілу $F_X(x)$, якій у математичній статистиці відповідає емпірична функція розподілу (функція розподілу вибірки).

Отже, **розподілом вибірки** називають розподіл дискретної випадкової величини X , яка приймає значення $x_1, x_2, \dots, x_n$ з імовірностями $1/n$. Відповідну цьому закону функцію розподілу називають емпіричною функцією розподілу і позначають $F_{X^*}(x)$.

Якщо кількість варіант k не дуже велика, то для отримання наочного уявлення про розподіл випадкової величини будують **полігони частот**. Для цього на осі абсцис відкладають значення варіант, а на осі ординат – відповідні значення частот p_i . Полігон відносних частот надає уявлення про розподіл ймовірностей; графік кумулятивних відносних частот називають емпіричною

функцією розподілу. Вона, зокрема, виражає залежність між значеннями кількісної ознаки і накопиченої частоти і містить всю інформацію про результати вимірювань.

Оскільки у більшості випадків у дослідженнях закон розподілу досліджуваної величини та його характеристики невідомі, то необхідним етапом дослідження є спеціальний аналіз первинних експериментальних даних, наприклад, за допомогою полігону чи гістограми частот (або відносних частот).

Зокрема, гістограмою частот групуваної вибірки називають функцію, яка не змінюється всередині інтервалів групування і приймає на кожному з них значення p_i ($i = 1, 2, \dots, k$). По осі абсцис при цьому відкладають обрані частинні інтервали й на цих відрізках будують прямокутники з висотою, яка дорівнює частоті (кількості реалізацій) відповідного інтервалу (вісь ординат). Площа ступінчастої фігури під графіком такої гістограми дорівнюватиме обсягу вибірки n .

Аналогічно гістограмою відносних частот називають функцію, яка не змінюється всередині інтервалів групування й приймає на кожному з них значення p_i/n ($i = 1, 2, \dots, k$). Площа відповідної ступінчастої фігури під графіком гістограми відносних частот дорівнює одиниці.

Полігоном частот називають ламану лінію з вершинами у точках (q_i, p_i) ($i = 1, 2, \dots, k$). Полігон частот є аналогом гістограми для дискретного розподілу випадкової величини. При цьому по осі абсцис відкладають значення досліджуваної величини, а по осі ординат – відповідні їм частоти p_i (або відносні частоти p_i/n). Отримані точки з'єднують ламаною лінією.

Полігон частот можна також побудувати й для безупинного розподілу випадкової величини. Тоді ламаною лінією з'єднують середини верхніх відрізків прямокутників гістограми. Ламана лінія, яка йде по осі абсцис, потім обводить гістограму і знову проходить по осі абсцис є графіком реалізації емпіричної функції щільності розподілу ймовірностей. Гістограми можуть приймати різний вигляд, зокрема, як показано на рис. 3.1 і рис. 3.2.

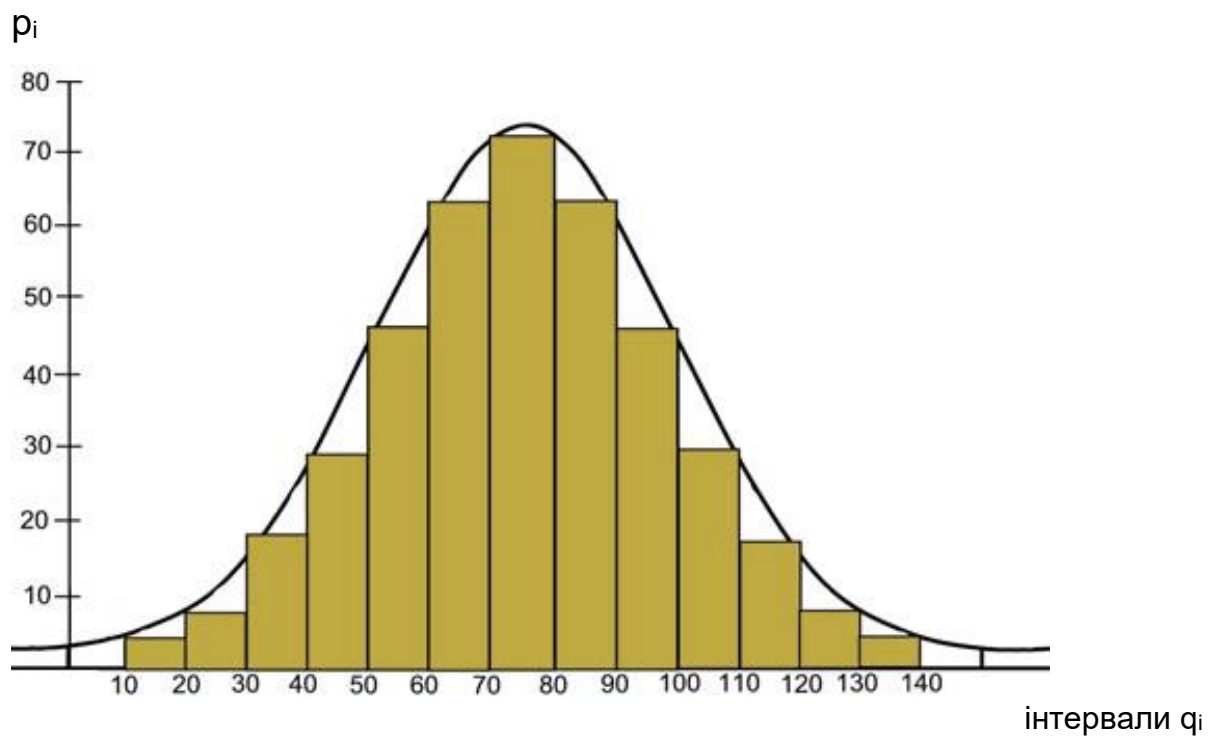


Рис. 3.1 Гістограма 1 (нормальний розподіл).

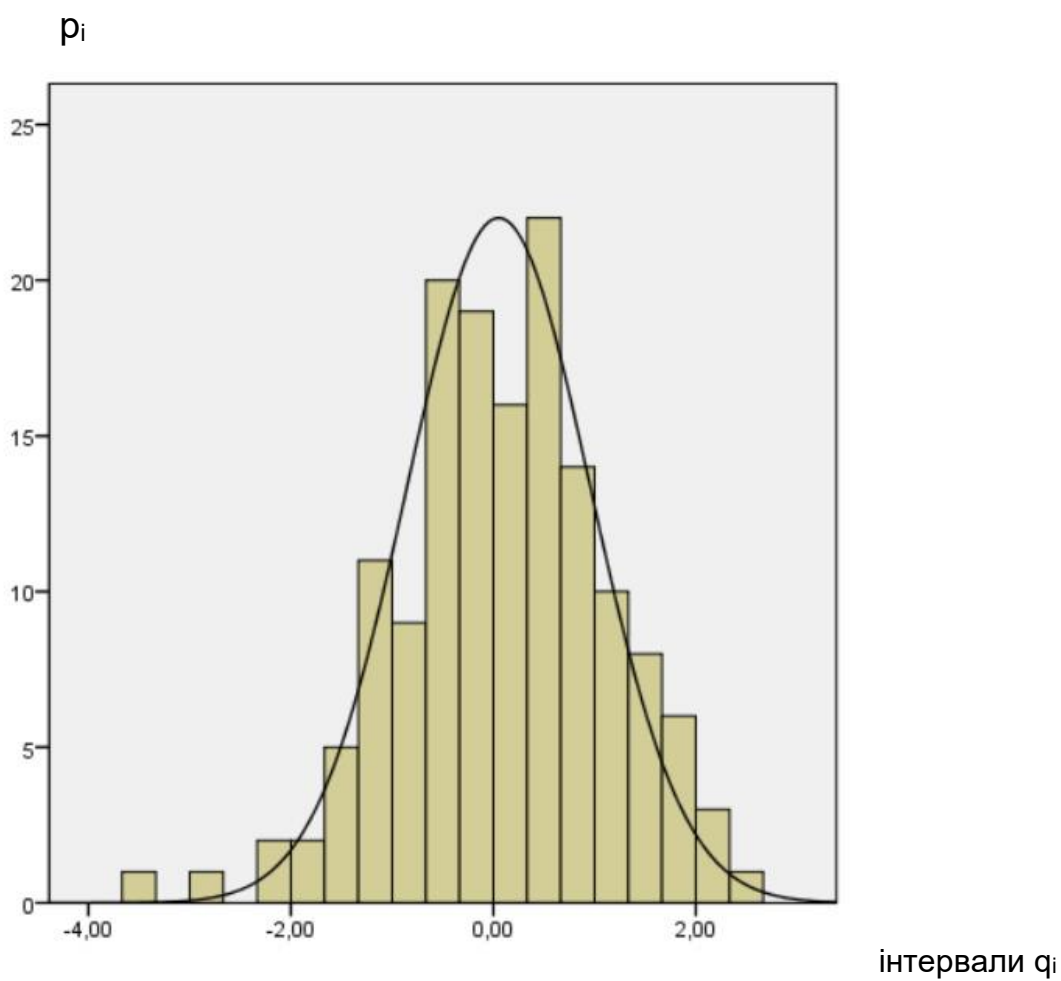


Рис. 3.2 Гістограма 2.

При побудові гістограми особливого значення набуває вибір інтервалів групування, оскільки від цього залежатиме форма розподілу досліджуваної величини. Існують різноманітні формули розрахунку обсягу певного інтервалу гістограми, але спрощено можна вважати, що інтервалів має бути 10...15 залежно від обсягу вибірки, а в кожен з цих інтервалів повинно потрапити не менше 5...6 реалізацій досліджуваної величини (у загальному випадку інтервали можуть мати різні обсяги). При збільшенні до нескінченності обсягу вибірки гістограма перетворюється на графік щільності розподілу випадкової величини.

Ширину інтервалів Δx можна визначити за формулою:

$$\Delta x = (x_n - x_1) / k \quad (3.2)$$

де k – кількість класів. Їх рекомендовані значення вказано у табл. 3.1.

Таблиця 3.1 Рекомендовані кількості класів групуваної вибірки

Обсяг вибірки n	25...40	40...60	60...80.	100...120	200...1000
Кількість класів k	5...6	6...8	7...10	8...12	10...15

Приклад 1.

Таблично задана вибірка з 50 спостережень певної генеральної сукупності X (од.). Подати задану вибірку у вигляді таблиці частот, прийнявши 7 інтервалів групування:

40,67; 56,59; 36,56; 37,34; 35,19; 34,00; 38,31; 30,74; 33,99; 31,81;
 34,51; 26,65; 48,22; 25,22; 18,52; 27,40; 28,64; 21,97; 20,27; 23,06;
 22,65; 37,50; 16,48; 32,05; 33,47; 41,02; 32,63; 34,71; 16,04; 44,10;
 22,27; 54,81; 37,92; 43,21; 21,73; 24,15; 29,81; 12,95; 49,87; 23,76;
 14,47; 35,25; 20,23; 20,55; 25,14; 26,81; 31,50; 42,77; 25,72; 32,39

Розв'язування.

Утворимо варіаційний ряд даної вибірки. Розташуємо елементи вибірки у порядку зростання за стовпцями:

12,95; 20,23; 22,27; 25,14; 27,40; 31,81; 33,99; 35,25; 38,31; 44,10;
14,47; 20,27; 22,65; 25,22; 28,64; 32,05; 34,00; 36,56; 40,67; 48,22;
16,04; 20,55; 23,06; 25,72; 29,81; 32,39; 34,51; 37,34; 41,02; 49,87;
16,48; 21,73; 23,76; 26,65; 30,74; 32,63; 34,71; 37,50; 42,77; 54,81;
18,52; 21,97; 24,15; 26,81; 31,50; 33,47; 35,19; 37,92; 43,21; 56,59

Розраховуємо розмах вибірки

$$R = x_{\max} - x_{\min} = 56,59 - 12,95 = 43,64 \text{ (од.)}$$

Визначаємо обсяг інтервалу групування:

$$a = R / t = 43,64 / 7 \approx 6,23 \text{ (од.)}$$

Для зручності задаємо $a = 7$ од., тоді перший інтервал групування становить від 12 од. до 19 од. Підраховуємо середини інтервалів одночасно з відповідними частотами.

Отримані результати зводимо в таблицю і будуємо полігон і гістограму частот на основі даних прикладу (див. рис. 3.1 і 3.2).

3.1.2 Оцінювання відтворюваності і достовірності результатів

Оцінювання відтворюваності й достовірності отриманих у дослідженні результатів є одним з необхідних етапів статистичної обробки даних.

Оцінювання відтворюваності експерименту полягає у тому, щоб встановити, чи суттєво розрізняються результати двох серій (або декількох серій) визначень однієї й тієї ж досліджуваної величини.

Відтворюваність експерименту характеризують близькістю паралельно одержаних результатів до середнього значення досліджуваної величини. Тобто відтворюваність експерименту визначається, в основному, величиною випадкової складової похибки. Причому, чим більша випадкова похибка, тим сильніше розсіюються значення досліджуваної величини при багаторазовому повторюванні дослідів, і тим менша точність даного методу дослідження.

У загальному випадку точність експерименту визначається наближенням усіх похибок до нуля. Зазначимо, що випадкові похибки, як правило, роблять результати експерименту неточними, а систематичні похибки спричиняють неправильність самого експерименту.

Достовірність експерименту характеризується наближенням результату вимірювання досліджуваної величини до її істинного значення (якщо це значення відоме) або до середнього арифметичного значення результатів вимірювань. Її можна оцінити за допомогою абсолютної складової похибки експерименту (вимірювання) Δx . Якщо експеримент визнано достовірним, це вказує на мінімальну систематичну похибку. У загальному випадку відсутність в експерименті систематичної складової похибки забезпечує його правильність.

При проведенні експерименту (серії вимірювань) необхідно отримати кінцевий результат з похибкою, яка не перевищує заданого наперед допустимого значення.

Тобто похибка експерименту не повинна виходити за межі певного довірчого інтервалу з деякою наперед встановленою довірчою ймовірністю p .

Довірчий інтервал це такий інтервал змінювання досліджуваної величини, який з наперед заданою ймовірністю p містить невідоме значення X даної досліджуваної величини, тобто:

$$P[x_1 < x < x_2] = p = 1 - \alpha, \quad (3.3)$$

де p – довірна ймовірність (надійність), яка зазвичай задається наперед;
 α – рівень значущості.

Іншими словами, ймовірність того, що даний довірчий інтервал "накриває" невідоме значення x досліджуваної величини, називають довірчою ймовірністю (надійністю). На практиці залежно від конкретних умов вимірювання та вимог щодо результатів екологічного експерименту задають такі значення довірчої ймовірності: 0,9; 0,95; 0,99; 0,999 тощо.

При збільшенні кількості вимірювань в досліді (або кількості паралельно проведених аналізів) оцінка досліджуваної величини X зазвичай стає точнішою, а довірчий інтервал при цьому дещо зменшується. Однак це твердження не відноситься до тих випадків, коли вимірювання взаємозалежні або вони містять суттєві систематичні похибки. В цих випадках точність результату експерименту при збільшенні кількості вимірювань не тільки не зростатиме, а й може навіть знизитися.

Рівень значущості α , як правило, задають в інтервалі від 0,001 до 0,05 (рідше – до 0,01). Наприклад, якщо довірна ймовірність $p = 0,95$ (95 %), рівень значущості $\alpha = 1 - 0,95 = 0,05$ (5 %), то можна стверджувати, що вихід результатів вимірювань цієї величини за межі заданого довірчого інтервалу за рахунок випадковості можливий з ймовірністю, що дорівнює 0,05 (5 %). Тобто тільки в 5 % випадків є можливість помилки в висновках щодо результатів проведеного експерименту.

Іншими словами, рівень значущості α характеризує вірогідність випадкового відхилення значень досліджуваної величини від встановлених з певною ймовірністю результатів дослідження. За його допомогою можна встановити, скільки відсотків результатів вимірювань можуть дати випадкові відхилення від істинного значення досліджуваної величини.

У випадку, коли потрібні дуже точні та достовірні дані, наприклад, у медичних дослідженнях, рівень значущості α має бути не більше 1 % (0,01). Тобто, чим менше значення рівня значущості, тим більш достовірними будуть результати експерименту та висновки, які на них базуються.

Якщо α дорівнює 0,05 – це звичайні вимоги надійності;

$\alpha = 0,01$ – це підвищені вимоги щодо надійності експерименту;

$\alpha = 0,001$ – дуже високі вимоги, а при

$\alpha = 0,1$ – знижені вимоги щодо точності та надійності результату експерименту.

Результати, одержані під час статистичної обробки вибірки експериментальних даних, можна вважати достовірними тільки за умов, якщо сформована вибірка є однорідною. Це означає, що в одержаних результатах експерименту відсутні грубі помилки (промахи).

Оцінку однорідності вибірки, як правило, проводять за так званим контрольним критерієм (або Q-критерієм). Значення контрольного критерію залежать від кількості проведених вимірювань (дослідів) n і величини довірчої ймовірності p ; їх подають у статистичних таблицях (табл. 3.2).

Для визначення однорідності вибірки всі одержані результати експерименту групують у порядку їх зростання:

$$x_1 \leq x_2 \leq \dots \leq x_n .$$

Далі визначають розмах вибірки R (варіаційний інтервал) та розраховують контрольні критерії Q для першого і останнього значень досліджуваної величини. Вибірку вважають неоднорідною, якщо хоча б одне з

цих розрахованих значень контрольного критерію вище за табличне значення $Q_{\text{табл.}}$. Як правило, табличне значення $Q_{\text{табл.}}$ знаходять при довірчій ймовірності p , що дорівнює 0,95 (95%) (див. табл. 3.2 та табл А.1, Додаток А).

Таблиця 3.2 Фрагмент значень Q-критерію (контрольного критерію) за певних значень довірчої ймовірності p

Кількість випробувань (дослідів, спостережень тощо), n	Q-критерій при $p = 0,90$	Q-критерій при $p = 0,95$	Q-критерій при $p = 0,99$
3	0,89	0,94	0,99
4	0,68	0,77	0,89
5	0,56	0,73	0,76

Значення X_1 або X_n , для яких розрахований контрольний критерій, відповідно, Q_1 або Q_n , більше за табличне значення $Q_{\text{табл.}}$ при заданій ймовірності p , не враховують у подальших обчисленнях.

За таких обставин створюють нову вибірку вже зменшеної кількості елементів і знову перевіряють її на однорідність.

Приклад 2.

Перевірити, чи є груба помилка, якщо у спостереженні вибіркоким методом при визначенні певної характеристики X одержано такі результати:

$$\omega(X),\%: 33,56; 31,83; 32,06; 31,71; 31,81.$$

Розв'язування.

Розташовуємо одержані величини у порядку їх зростання, тобто формуємо впорядкований варіаційний ряд:

$\omega(X), \%$: 31,71; 31,81; 31,83; 32,06; 33,56.

Розраховуємо варіаційний розмах:

$$R = x_n - x_1 = 33,56 - 31,71 = 1,85 (\%).$$

Розраховуємо контрольні критерії

$$Q_1 = (31,81 - 31,71) / 1,85 = 0,05;$$

$$Q_2 = (33,56 - 32,06) / 1,85 = 0,81.$$

Визначаємо табличне значення контрольного критерію при довірчій імовірності $p = 0,95$ та кількості проведених дослідів $n = 5$:

$$Q_{\text{табл.}} = 0,73.$$

Порівнюємо розраховані контрольні критерії з табличним значенням:

$$Q_1 < Q_{\text{табл.}}, \text{ а}$$

$$Q_2 > Q_{\text{табл.}}$$

Робимо висновок: оскільки існує груба помилка в результаті експерименту, то значення досліджуваної величини 33,56 необхідно виключити з даної вибірки.

Формуємо нову, зменшену вибірку:

$\omega(X), \%$: 31,71; 31,81; 31,83; 32,06,

і для зменшеної кількості її елементів розраховуємо варіаційний розмах:

$$R = x_n - x_1 = 32,06 - 31,71 = 0,35 (\%)$$

та відповідні контрольні критерії:

$$Q_1 = (x_2 - x_1) / R = (31,81 - 31,71) / 0,35 = 0,28;$$

$$Q_2 = (x_n - x_{n-1}) / R = (32,06 - 31,83) / 0,35 = 0,66.$$

Порівнюємо їх з табличним значенням $Q_{\text{табл.}} = 0,77$ (за умови $p = 0,95$; $n = 4$):

$$Q_1 < Q_{\text{табл.}}, \text{ і}$$

$$Q_2 < Q_{\text{табл.}}.$$

Робимо висновок стосовно принципової можливості використання вибірки для подальшого аналізу: одержана вибірка є однорідною, тобто результати експерименту не містять грубої помилки.

Приклад 3.

Провести статистичну обробку отриманих в експерименті результатів:

$$\omega(X), \%: \quad 24,11; \quad 24,15; \quad 24,18; \quad 24,20.$$

Розв'язування.

Перевіряємо вибірку на однорідність за контрольним критерієм Q (див. табл. 3.2). Для цього розраховуємо варіаційний розмах:

$$R = x_n - x_1 = 24,20 - 24,11 = 0,09 (\%)$$

та контрольні критерії:

$$Q_1 = (x_2 - x_1) / R = (24,15 - 24,11) / 0,09 = 0,44;$$

$$Q_2 = (x_n - x_{n-1}) / R = (24,20 - 24,18) / 0,09 = 0,22.$$

Порівнюємо їх з табличним значенням $Q_{\text{табл.}} = 0,77$ (за умови $p = 0,95$; $n = 4$):

$$Q_1 < Q_{\text{табл.}}, \text{ і } Q_2 < Q_{\text{табл.}}$$

Робимо висновок: надана вибірка однорідна. Далі проводимо статистичну обробку одержаних результатів:

1) розраховуємо середнє арифметичне значення результатів експерименту:

$$\bar{x} = \sum_{i=1}^n x_i / n = (24,11 + 24,15 + 24,18 + 24,20) / 4 = 24,16 (\%)$$

2) визначаємо математичне сподівання, дисперсію та середнє квадратичне відхилення результатів експерименту:

$$M(x) = (x_1 + x_2) / 2 = (24,11 + 24,20) / 2 = 24,155 (\%)$$

$$\sigma^2(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{f} = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{f} = 0,0015;$$
$$\sigma(x) = 0,039 (\%)$$

3) визначаємо стандартне відхилення середнього результату:

$$\sigma(\bar{x}) = \frac{\sigma(x)}{\sqrt{n}} = \frac{0,039}{\sqrt{4}} = 0,019$$

4) розраховуємо довірчий інтервал середнього значення.

$$t(0,95; 4) = 3,18 \text{ (табл. 3.3 та табл. Б.1, Додаток Б).}$$

$$\Delta\bar{x} = t(p, f) \cdot \sigma(\bar{x}) = 3,18 \cdot 0,019 = 0,05 \text{ (\%)},$$

5) визначаємо відносну похибку середнього результату:

$$\delta = \Delta\bar{x} \cdot \frac{100\%}{\bar{x}} = 0,05 \frac{100\%}{24,16} = 0,25 \text{ (\%)}. \quad \bar{x} = 24,16$$

6) результат експерименту подаємо у такому вигляді:

$$\omega(X), \% = 24,16 \pm 0,05.$$

Отже, результат будь-якого експерименту подають у вигляді фізичної величини та показника її точності.

Залежно від складності експерименту та значень досліджуваних величин застосовують такі показники точності:

– довірчі інтервали, в яких із заданою ймовірністю p знаходиться похибка вимірювання Δx (або її систематична складова $\Delta x_{\text{сист.}}$);

– оцінки середніх квадратичних відхилень систематичної і випадкової складових похибки $\sigma(x)_{\text{сист.}}$, $\sigma(x)_{\text{випадк.}}$;

– щільності ймовірностей систематичної та випадкової складових похибки.

Таблиця 3.3 Значення коефіцієнтів Стюдента $t_{\alpha,n}$ для значень довірчої ймовірності $p = 0,90$ і $p = 0,95$ та числа вимірів n (розподіл Стюдента)

n	0,9	0,95	0,99
2	6,31	12,71	63,7
3	2,92	4,30	9,92
4	2,35	3,18	5,84
5	2,13	2,77	4,60
6	2,02	2,57	4,03
7	1,943	2,45	4,71
8	1,895	2,36	3,50
9	1,860	2,31	3,36
10	1,833	2,26	3,25
11	1,812	2,23	3,17
12	1,796	2,20	3,11
13	1,782	2,18	3,06
14	1,771	2,16	3,01
15	1,761	2,14	2,98
16	1,753	2,13	2,95
17	1,746	2,12	2,92
18	1,740	2,11	2,90
19	1,734	2,10	2,88
20	1,729	2,09	2,86
∞	1,645	1,960	2,58

Для представлення показників точності результатів експерименту можна запропонувати такі правила:

1) показники точності мають таку саму розмірність, як і досліджувана величина;

2) показники точності результату експерименту зазвичай містять не більше двох значущих цифр; проте дві цифри залишаються, якщо перша значуща цифра 1, 2 або 3;

3) останні (наймолодші) розряди результату вимірювання та числових показників точності мають бути однаковими.

Наприклад, при вимірюванні часу було одержано такий результат:

$$t = (15,8 \pm 0,1) \text{ с.}$$

У цьому запису:

- 15,8 с це оцінка вимірюваної (досліджуваної) величини;
- 0,1 с це невизначеність результату вимірювання (або показник точності),
- а довірчий інтервал складає – (15,7 ... 15,9) с.

Невизначеність експерименту (дослідження, досліду тощо) це параметр, який характеризує розсіювання можливих значень досліджуваної величини. А оскільки невизначеність зумовлена наявністю похибок експерименту, її не можна подавати з великою точністю.

Отже, як можна зробити висновок з наведених правил, у випадку проведення достатньо точних експериментів зазвичай наводять невизначеності з двома значущими цифрами. У разі не дуже точних вимірювань невизначеність може бути подана й з одною значущою цифрою. Крім того, остання значуща цифра в будь-якому подальшому результаті вимірювання має бути того самого порядку, що й остання значуща цифра невизначеності. Наприклад, результат 67,84 з невизначеністю 0,5 треба подати як $67,8 \pm 0,5$.

Під час проміжних розрахунків числа можуть вміщувати на одну значущу цифру більше, ніж це зазначено вище. Це дещо зменшить похибки обчислень, які виникають при округленні чисел. Однак наприкінці розрахунку остаточний результат варто округлити та позбавитися від цієї додаткової цифри, тобто подати остаточний результат відповідно до наведених вище правил.

Якщо немає спеціальних вимог до вибору рівня довіри, як правило, використовують значення довірчої імовірності $p = 0,95$ (95 %).

При подачі результату експерименту його можна супроводжувати коментарем, у якому наводять дані про кількість дослідів (вимірювань, спостережень), умови проведення експерименту тощо.

У процесі обробки результатів дослідження до розглянутих раніше похибок експерименту можуть додаватися також різноманітні похибки обчислень, зокрема:

1) похибки вихідної інформації – похибки, що виникають внаслідок можливої неточності вимірювань або внаслідок неможливості представлення вихідної величини скінченим десятковим дробом (наприклад, неможливо точно представити число π , оскільки воно є ірраціональним);

2) похибки обмеження – такі похибки, що виникають, наприклад, внаслідок застосування вибіркового методу дослідження;

3) похибки округлення це похибки, що виникають у результаті необхідності представити число у вигляді скінченої послідовності цифр (похибки округлення відносять тільки до дійсних чисел). Враховуючи те, що обчислювальні машини працюють з кінцевою кількістю значущих цифр, завжди виникає необхідність в округленні.

Розрізняють два основні методи округлення результатів обчислень:

–відкидання молодших розрядів та

–симетричне округлення.

Слід зазначити, що похибка у разі застосування методу симетричного округлення суттєво менша за похибку, що виникає при відкиданні молодшого розряду.

Розглянемо деякі рекомендації, що дозволять зменшити можливі похибки обчислень:

1. Якщо необхідно виконати дії складання або віднімання довгої послідовності чисел, то спочатку треба працювати з найменшими числами.

2. Якщо формули обчислень містять операцію віднімання майже рівних чисел, то їх, за можливості, треба перетворювати таким чином, щоб уникати подібних операцій.

Не слід вважати, що похибка обчислень завжди зростає при додаванні та зменшується при відніманні. Наприклад, якщо похибки результатів мають протилежні знаки, то при додаванні похибка зменшиться, а при відніманні – збільшиться.

3. Вираз $a \cdot (b - c)$ у разі необхідності перетворюють на вираз $(a \cdot b - a \cdot c)$, а вираз $(b - c)/a$ можна подати у вигляді $(b/a - c/a)$. Але якщо числа в різниці майже рівні, то найкраще спочатку виконати віднімання, а тільки потім – операції множення або ділення.

4. Якщо це можливо, кількість математичних операцій треба зводити до мінімуму.

5. При застосуванні вибіркового методу треба слідкувати за тим, щоб сформована вибірка була, по-перше, однорідною, а по-друге, вона має бути репрезентативною.

Тому при формуванні певної вибірки з генеральної сукупності, вона має відповідати наступним вимогам:

- її статистичний розподіл має співпадати із розподілом генеральної сукупності;

- величина вибірки має бути достатньою для відображення структури всієї генеральної сукупності;

- у тих випадках, коли порівнюються деякі параметри двох вибірок, необхідно забезпечити однаковість розподілу частот цих параметрів у досліджуваних вибірках;

- умови дослідження для окремих елементів вибірки або для двох (або декількох) порівнювальних вибірок обов'язково мають бути однаковими.
- середнє значення суми двох вибірок дорівнює сумі їх середніх, якщо вибірки мають однакові розміри (аналогічне твердження справедливе й для різниці двох вибірок):

3.2 Теоретичні основи кореляційного та регресійного аналізів

3.2.1 Статистичний зв'язок між випадковими величинами

Одним з основних завдань багатьох наукових досліджень є встановлення зв'язку (зв'язків) між величинами або факторами, зміна яких визначає суть досліджуваного процесу.

Отже, у статистичному аналізі розрізняють такі види зв'язків між чинниками, що діють на систему (об'єкт), а саме:

- функціональні;
- стохастичні;
- статистичні.

Функціональний зв'язок це зв'язок між змінними, при якому кожному значенню однієї величини відповідає чітко визначене значення іншої, тобто

$y = f(x_1, x_2, \dots, x_n)$. Зазначимо, що дослідженнями зв'язків такого виду статистика не займається.

У суспільних процесах, навколишньому середовищі, технологічних процесах тощо зв'язки виявляються настільки складними і багатогранними, що їх неможливо описати за допомогою функціональної залежності. Наприклад, у багатьох експериментальних дослідженнях, а також під час обробки результатів спостережень, експериментів тощо трапляються залежності, коли

одному значенню незалежної ознаки X відповідає сукупність значень залежної ознаки Y або розподіл Y . Тому виникає потреба у встановленні і врахуванні залежності цими ознаками – так званого **СТОХАСТИЧНОГО ЗВ'ЯЗКУ**, що відповідає ситуації, за якої зміна значення однієї змінної спричинює зміну закону розподілу іншої.

У випадку дискретних величин це означає, що певному значенню однієї змінної відповідає набір значень іншої і при цьому кожне значення має власну ймовірність реалізації.

Статистичний зв'язок передбачає, що значення однієї змінної варіюються в середньому в залежності від того, які значення приймає інша змінна. Часто статистичний зв'язок розглядають як функціональну залежність (функціональний зв'язок) з випадковою помилкою ε , зокрема:

$$y = f(x_1, x_2, \dots, x_n) + \varepsilon, \quad (3.4)$$

де $f(x_1, x_2, \dots, x_n)$ – функція, що описує залежність Y від сукупності незалежних змінних $X_1, X_2, \dots, X_n$.

Таким чином, знаходження рівняння зв'язку між залежною ознакою Y і незалежною ознакою X є методом узагальнення статистичних зв'язків, що спостерігаються, наприклад, у спостереженнях, експериментах тощо і одночасно це є методом виявлення і дослідження таких зв'язків.

Оскільки сума константи і випадкової величини є випадковою величиною, то значення y , розраховані за формулою (3.4), також будуть випадковими величинами.

На практиці для одержання за експериментальними даними математичних моделей, що описують поведінку певної характеристики в залежності від зміни декількох факторів, застосовують **регресійний аналіз**. При цьому процеси, об'єкти або природні явища розглядають як «чорний

ящик» («чорну скриню»), що має на «вході» незалежні змінні, якими можна керувати, а на «виході» – змінні, що характеризують досліджуваний процес.

Виокремлюють:

- контрольовані;
- керовані;
- некеровані та
- неконтрольовані фактори (чинники).

Зокрема, некерованими факторами називають незалежні змінні, значення яких можна вимірити, але керувати ними не можливо.

Неконтрольовані фактори це ті, що впливають на досліджуваний процес, але їх неможливо виміряти (або може бути невідомо про їх існування).

Основною ознакою наявності статистичного (кореляційного) зв'язку між випадковими величинами X та Y є відмінність від нуля коефіцієнту кореляції r . За наявності кореляційного зв'язку між змінними необхідно виявити його характер, тобто визначити, як саме у середньому впливає зміна величини X на зміну величини Y .

Отже, під кореляційним аналізом розуміють встановлення закономірностей між явищами і процесами, що залежать від багатьох, іноді невідомих параметрів. Використання методу кореляції і регресії дає змогу розв'язати такі завдання:

- встановити характер і тісноту зв'язку між досліджуваними величинами (явищами);
- визначити й кількісно виміряти ступінь впливу окремих факторів (чинників) та їх комплексної дії на певний об'єкт чи явище;
- на підставі кореляційної моделі розраховувати (прогнозувати) кількісні зміни у стані об'єкта чи явища, надавати об'єктивну оцінку цим змінам.

Таким чином, функція $y = f(x)$ є кореляційною, якщо кожному значенню аргументу X відповідає статистичний ряд розподілу функції Y .

Оскільки кореляційні залежності характеризуються ймовірнісними зв'язками, встановлення залежностей між величинами X та Y можливе лише тоді, коли були проведені статистичні вимірювання.

Для отримання статистично достовірної і надійної характеристики зв'язку у дослідженні (спостереженні) потрібно встановити форму зв'язку й знайти кількісні характеристики – обрати математичне рівняння (модель), яке найбільш повно відображає характер зв'язку між досліджуваними ознаками явища чи об'єкта.

Отже, під формою кореляційного зв'язку розуміється тип аналітичного рівняння, що виражає залежність між певними досліджуваними ознаками.

Зв'язок між корелюючими ознаками за напрямом може бути:

- прямим, коли обидві ознаки змінюються в одному напрямі (тобто зі збільшенням факторної ознаки зростає ознака результативна і навпаки);

- оберненим, коли обидві ознаки змінюються в різних напрямках.

За формою або аналітичним вираженням розрізняють:

- прямолінійні зв'язки (або лінійні);
- нелінійні (криволінійні) зв'язки.

Залежно від кількості досліджуваних ознак кореляція може бути:

- парною (простою) і
- множинною.

Тобто, якщо визначається зв'язок між двома ознаками (результативною і факторною), то таку кореляцію називають простою; а у випадку, коли явище розглядається як результат одночасного впливу декількох чинників – множинною.

Встановлення форми зв'язку надає можливість знайти рівняння регресії, яке найбільш повно відображає характер взаємодії між залежною і незалежною ознаками. Проте, сам метод кореляційного аналізу не встановлює наявності та/або причин виникнення зв'язків (причинно-наслідкового зв'язку) між

досліджуваними величинами (явищами). Його призначенням є тільки встановлення наявності та кількісне описання зв'язку.

Кореляційний аналіз можна поділити на 5 основних етапів:

- встановлення наявності зв'язку між досліджуваними ознаками, для чого потрібно спочатку визначити та відібрати найістотніші чинники для аналізу;
- вивчення характеру зв'язку, його напрямку і форми; вибір математичного рівняння для його описання;
- розрахунок характеристик кореляційного зв'язку, а саме знаходження коефіцієнтів рівняння регресії і показників тісноти зв'язку;
- статистичне оцінювання характеристик (показників) зв'язку.

Таким чином, сутність кореляційного аналізу полягає у знаходженні рівняння регресії між випадковими величинами (так звана лінія найкращого наближення), а також в оцінці виду і тісноти зв'язку між ними, достовірності результатів вимірювань (спостережень) тощо.

Щоб попередньо оцінити наявність кореляційного зв'язку між випадковими величинами X та Y , будують так зване кореляційне поле.

Для цього результати випробувань (емпіричний ряд розподілу) наносять точками на координатну площину XOY (рис. 3.3). Множина таких точок й утворює кореляційне поле. За тіснотою групування експериментальних точок навколо прямої (кривої) лінії, а також за нахилом цієї лінії можна візуально оцінити наявність (чи відсутність) кореляційного зв'язку між досліджуваними величинами (явищами).

Наприклад, з рис. 3.3а випливає, що експериментальні дані мають певний лінійний зв'язок між випадковими величинами X та Y , тоді як на рис. 3.3б кореляційного зв'язку між величинами майже не виявлено (або він виявляється дуже слабким).

Отже, за виглядом кореляційного поля можна передбачити вид зв'язку між досліджуваними величинами та встановити, наприклад, чи є лінійною (або криволінійною) залежність між ними, оцінити величину коефіцієнта кореляції r .

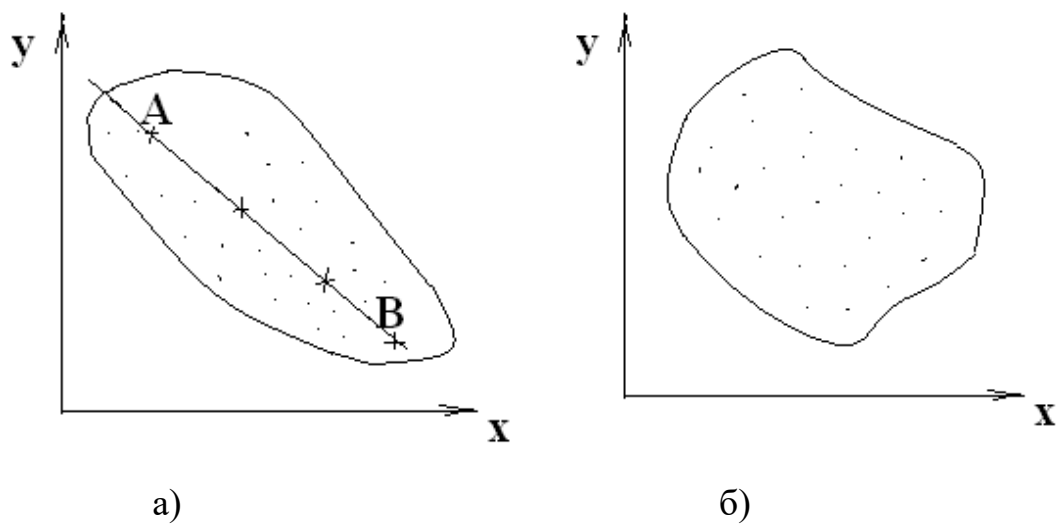


Рис. 3.3 Види кореляційних полів.

При цьому, чим тісніше точки поля групуються навколо деякої лінії (прямої або кривої), тим ближчою є величина коефіцієнта кореляції до одиниці. І навпаки, чим більше кореляційне поле нагадує коло (тобто неможна виділити домінуючого напрямку), тим ближчою є величина коефіцієнта кореляції до нуля (рис. 3.3б).

За величиною коефіцієнта кореляції r можна оцінити не тільки тісноту кореляційного зв'язку, а й напрямок кореляції (пряма чи зворотна). Зокрема, при $r > 0$ кореляційний зв'язок оцінюють як прямий; а при $r < 0$ – як зворотній.

Таким чином, з урахуванням напрямку коефіцієнт кореляції r може змінюватись у межах від -1 до $+1$. Причому, коли $r = 1$, точки лежать на одній прямій. Тоді це означає, що величини x і y зв'язані лінійним функціональним зв'язком.

Якщо r прямує до нуля ($r \rightarrow 0$), то кореляційне поле не має чіткої конфігурації (див. рис. 3.3б). Тому можна стверджувати, що між величинами x та y не існує лінійного кореляційного зв'язку. Проте у такому випадку між досліджуваними величинами може існувати нелінійна регресія.

Оскільки внаслідок статистичного характеру зв'язку одному значенню x може відповідати сукупність значень y , то для описання досліджуваного процесу (явища) оптимальною буде функція, для якої виконується умова:

$$S = \sum_{i=1}^n (y_i - \bar{y})^2 \rightarrow \min, \quad (3.5)$$

де y_i – фактичні ординати кореляційного поля;

$\bar{y}$ – середнє значення ординати з абсцисою x , розрахованою за апроксимуючим рівнянням.

Таким чином, регресійний аналіз ґрунтується на методі найменших квадратів (МНК). Цей метод, як правило, застосовують для знаходження математичного очікування (сподівання) нормально розподіленої випадкової величини.

Якщо на кореляційне поле (див. рис. 3.3а) нанести середні значення $\bar{y}$ (позначені хрестиками), то лінія **AB** буде відповідати функціональній залежності виду:

$$\bar{y} = a + b \cdot x, \quad (\text{або } \bar{y} = a \cdot x + b), \quad (3.6)$$

де **a** – коефіцієнт регресії, початок відліку; за його знаком можна судити про характер коефіцієнта еластичності: в однофакторному рівнянні, якщо $a > 0$, то коефіцієнт еластичності менший за одиницю; якщо $a < 0$, то він більший за одиницю;

b – коефіцієнт регресії, який показує середню зміну залежної змінної при зміні незалежної змінної на одиницю, тобто характеризує міру впливу відповідного фактора на варіацію результативної ознаки;

Середню лінію кореляційного поля, для якої виконується умова (3.5), називають лінією (рівнянням) регресії y на x , а числові коефіцієнти, що входять до виразу цього рівняння, – коефіцієнтами регресії y на x .

Лінія регресії, побудована за рівнянням регресії (3.6), розміщується таким чином, що сума вертикальних відхилень від неї дорівнює нулю. Це означає, що вона збалансовує суму додатніх і від’ємних відхилень від експериментально встановлених значень ознаки y . Ця лінія також має назву лінії найкращого наближення.

Таким чином, рівняння регресії це аналітичне рівняння, за допомогою якого визначають взаємозв’язок між досліджуваними ознаками; це регресійна модель залежності результативної ознаки y від ознаки факторної x .

У загальному випадку рівняння регресії (зв’язку) можуть бути для залежностей як від однієї ознаки при різних формах зв’язку (наприклад, лінійному, криволінійному – параболічному, гіперболічному тощо), так і для множинного зв’язку.

При аналізі кореляційної залежності між величинами x та y , значення яких визначені для елементів певної вибірки, постають такі завдання:

- вибір виду функції регресії;
- визначення коефіцієнтів регресії.

Найпростішою і найбільш розповсюдженою є лінійна кореляція. У цьому випадку кореляційне поле апроксимують за допомогою рівняння першої степені (лінійне рівняння), а лінію регресії розраховують, використовуючи метод найменших квадратів за умови (3.6).

Рівняння лінійної регресії можна представити або за допомогою рівняння прямої (3.6), або, наприклад, таким чином:

$$y = \bar{y} + |r| \cdot \frac{\delta(y)}{\delta(x)} \cdot (x - \bar{x}) \quad (3.7)$$

де $\bar{y}$ – середнє значення результативної ознаки (залежної змінної);

x – значення факторної ознаки (незалежної змінної);

$\bar{x}$ – середнє значення факторної ознаки;

r – коефіцієнт парної кореляції.

За умови нормального розподілу величин x і y лінії регресії як x від y , так і y від x є прямими лініями.

Для визначення коефіцієнтів рівняння регресії « a » і « b » знаходять часткові похідні від $S(a, b)$ по a і b , прирівнюють їх до нуля (похідна функції у точці екстремуму дорівнює нулю).

Отже, запишемо систему рівнянь:

$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum_{i=1}^n (bx_i + a - y_i) = 0 \\ \frac{\partial S}{\partial b} = -2 \sum_{i=1}^n (bx_i + a - y_i)x_i = 0 \end{cases}$$

або

$$\begin{cases} b \sum_{i=1}^n x_i^2 + a \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ b \sum_{i=1}^n x_i + a \cdot n = \sum_{i=1}^n y_i \end{cases}$$

Розв'язуючи систему рівнянь, отримуємо коефіцієнти рівняння регресії « a » і « b » за допомогою виразів (3.8) і (3.9):

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \quad (3.8)$$

$$a = y - b \cdot x = \frac{\sum_{i=1}^n y_i}{n} - b \frac{\sum_{i=1}^n x_i}{n}, \quad (3.9)$$

Досить часто у реальних дослідженнях (досліджуваних системах) кореляційний зв'язок між незалежними ознаками (x) і середнім значенням залежної ознаки ($\bar{y}_x$) не можна задати за допомогою лінійного рівняння виду (3.6). Отже, якщо побудувати графік за точками з координатами ($x, \bar{y}_x$), де $\bar{y}_x$ – умовні середні, та якщо ці точки розмістяться близько від певної кривої, то рівняння цієї кривої виду $\bar{y}_x = f(x)$ буде виражати кореляційну залежність між ознаками x і $\bar{y}_x$. У такому разі кореляційний зв'язок буде називатися криволінійним. При цьому алгоритм побудови рівнянь регресії і встановлення коефіцієнтів регресії криволінійного кореляційного зв'язку не буде суттєво відрізнятися від побудови рівнянь регресії лінійного кореляційного зв'язку.

Отже, під формою кореляційного зв'язку розуміють тип аналітичного рівняння, що виражає залежність між досліджуваними ознаками x та y . Таким чином, встановлення форми зв'язку означає вибір такого рівняння регресії, яке найбільш повно відбиває характер взаємодії між результатом і фактором, за яким проводять дослідження.

Рівняння кореляційного зв'язку можна використовувати, наприклад, для обчислення теоретичної лінії регресії, а також очікуваних (теоретичних) чи прогнозованих значень залежної змінної при тих або інших значеннях аргументів.

Залежно від ступеня адекватності кореляційної моделі буде збільшуватися (або зменшуватися) ступінь наближення розрахункових значень результативної ознаки до її фактичного значення. Тому, якщо кореляційна модель враховує всі основні фактори (чинники), що визначають варіацію результативної ознаки, то її точність буде досить високою.

Зазначимо, що кореляцію можна виявити, досліджуючи достатньо велику кількість спостережень, оскільки кореляційні зв'язки виявляються у формі спряженого варіювання двох або більшої кількості зіставлених ознак. Таким чином, вивчення кореляційних зв'язків є важливим особливо при дослідженні та аналізі явищ, що складаються під впливом великої кількості певних умов та чинників.

Зокрема, регресійна модель лінійної залежності результативної ознаки від декількох факторних ознак має вигляд:

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

У багатфакторному рівнянні, якщо $a > 0$, то сума коефіцієнтів еластичності менша за одиницю; якщо $a < 0$, то їхня сума більша за одиницю). Надійність регресійних рівнянь буде ти вищою, чим вірніше підібрано фактори, що впливають на результативну ознаку.

При проведенні досліджень часто виникає необхідність використання інших кореляційних характеристик. Це, зокрема:

– стандартизованого коефіцієнту регресії (бета-коефіцієнту) β – коефіцієнт регресії, виражений у стандартизованому масштабі і розрахований за формулою (3.10):

$$\beta = b \frac{\sigma_x}{\sigma_y} \quad (3.10)$$

де b – коефіцієнт чистої регресії при i -му факторі;

σ_x і σ_y – середні квадратичні відхилення, відповідно, по i -му фактору і результативній ознаці;

– коефіцієнту еластичності E , що характеризує ступінь і вагомість впливу конкретного фактора на варіацію результативної ознаки та показує, на скільки відсотків у середньому змінюється результативна ознака за умови зміни факторної ознаки на 1 %:

$$E = b \frac{\bar{x}}{\bar{y}} \quad (3.11)$$

де b – коефіцієнт чистої регресії при i -му факторі;

$\bar{x}$ і $\bar{y}$ – середні значення, відповідно, i -го фактора і результативної ознаки.

Для оцінки тісноти або сили зв'язку використовують коефіцієнт кореляції. Причому, для різних форм зв'язку застосовують різні показники його тісноти, наприклад:

- при лінійному зв'язку:
 - парний коефіцієнт кореляції r (при парному зв'язку) та
 - сукупний коефіцієнт кореляції R (при множинному зв'язку);
- при криволінійному зв'язку:
 - індекс кореляції i (при парному зв'язку) та
 - кореляційне відношення η (при множинному зв'язку).

Тіснота зв'язку визначається через відношення двох дисперсій:

$$R(r, i, \eta) = \sqrt{\frac{\sigma_0^2}{\sigma_i^2}} \quad (3.12)$$

$$\sigma_i^2 = \sum (y - \bar{y})^2$$

де σ_0^2 – загальна дисперсія (загальна варіація результативної ознаки, яка не залежить від значень факторної ознаки);

σ_i^2 – дисперсія факторної ознаки – варіація розрахункових значень y з урахуванням лінійного зв'язку, який характеризує досліджувану варіацію.

Як було зазначено, знак коефіцієнту кореляції r – «+» вказує на прямий, а знак «-» – на зворотній зв'язок між досліджуваними величинами. При цьому важливим є те, що знаки для коефіцієнтів регресії і кореляції повинні збігатися. Якщо ці знаки не збігаються, це вказує на недостовірні розрахунки, які можуть бути пов'язані з:

- неправильним урахуванням чинників, що впливають на залежну ознаку;
- неправильним аналітичним вираженням зв'язку;
- недостатньою чисельністю вибірки або її неоднорідністю.

Для розрахунку парного коефіцієнта кореляції, як правило, застосовуються формули (3.13)–(3.15):

$$r = \pm \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}}, \quad (3.13)$$

$$r = \pm \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot \delta(x) \cdot \delta(y)}, \quad (3.14)$$

$$r = \pm \frac{\left[\frac{1}{n} \sum_{i=1}^n (x_i y_i) \right] - \bar{x} \cdot \bar{y}}{\delta(x) \cdot \delta(y)}, \quad (3.15)$$

де

$$\delta(x) = \sqrt{\frac{\sum (x - \bar{x})^2}{n}};$$

$$\delta(y) = \sqrt{\frac{\sum (y - \bar{y})^2}{n}}.$$

Тісноту зв'язку між випадковими величинами X та Y вважають слабкою, якщо $|r| < 0,3$; тіснота зв'язку є задовільною, якщо $|r| \geq 0,5$; якщо $|r|$ лежить у межах від 0,5 до 0,7 – зв'язок між величинами вважають досить тісним; при $|r| > 0,7$ можна стверджувати, що між досліджуваними ознаками існує дуже тісний кореляційний зв'язок.

При проведенні експериментів (досліджень, спостережень тощо) необхідно враховувати, що на отриманий результат впливають безліч різноманітних чинників. Тому важливо визначати істотність впливу кожного з факторів на досліджувану результативну ознаку (величину).

Для визначення відсотка мінливості досліджуваної ознаки Y щодо її середнього значення, яке визначається мінливістю фактора X , розраховують коефіцієнт детермінації:

$$k_g = r^2 \quad (3.16)$$

де r – коефіцієнт парної кореляції.

Для того, щоб оцінити тісноту (силу) зв'язку між досліджуваними величинами, коли між ними існує нелінійний зв'язок (або дані вибірки розподілені не за нормальним законом), рекомендується застосовувати коефіцієнт кореляції Спірмена. При цьому його статистична значущість перевіряється аналогічно до коефіцієнту кореляції Пірсона.

3.2.2 Оцінка значущості коефіцієнта кореляції

Достовірність (значущість) коефіцієнта кореляції оцінюють за допомогою критерію Стюдента, який розраховують за формулою:

$$t = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} \quad (3.17)$$

де n – кількість експериментальних точок (варіант).

Розраховані значення коефіцієнта Стюдента порівнюють з його граничними (табличними) значеннями $t_{05}; t_{01}; t_{001}$, що відповідають рівням значущості $\alpha = 0,05; 0,01; 0,001$ (табл. Б.1, Додаток Б).

При цьому кількість ступенів свободи розраховують як $f = n - 2$.

Отже, за умови $t < t_{05}$ коефіцієнт кореляції вважають незначимим при рівні значущості $\alpha = 0,05$ (достовірність $p \geq 0,95$); за умови $t \geq t_{05}$ коефіцієнт кореляції вважають значимим при рівні значущості $\alpha = 0,05$.

Статистична значущість коефіцієнта кореляції збільшується при $t > t_{01}$ або $t > t_{001}$ (відповідно, з достовірністю $p \geq 0,99$ або $p \geq 0,999$). Таким чином, під час використання кореляційного методу визначають:

- рівняння та коефіцієнти регресії;
- стандартизований коефіцієнт регресії;
- коефіцієнт еластичності та
- коефіцієнти та індекси кореляції і детермінації.

При цьому одержані рівняння регресії застосовують для:

- дослідження характеру зміни ознаки (величини) Y при зміні фактора (величини) X ;
- прогнозування значень величини Y при заданих значеннях X ;
- визначення оптимального значення фактора X для одержання необхідного значення параметра Y або для утримання параметра Y на певному (зазначеному) рівні.

3.3 Регресійний аналіз

Часто на систему чинять вплив різноманітні, іноді навіть взаємопов'язані чинники, які діють з різною силою і можуть бути різними за напрямками впливу. Тому на практиці часто виникає потреба у встановленні зв'язку між величиною (ознакою) y та багатьма параметрами $x_1, x_2, \dots, x_n$.

Як зазначалось, у більшості випадків між змінними існують залежності, при яких кожному значенню однієї величини (аргументу) відповідає безліч можливих значень іншої величини (тобто її розподіл). Така залежність називається стохастичною або ймовірнісною. При цьому окремим випадком стохастичної залежності вважається кореляційна.

Наявність кореляційної залежності між змінними не завжди означає існування між ними безпосереднього зв'язку. Зокрема, зв'язок може існувати і завдяки іншим чинникам, які не було враховано.

Окрім того, досліджувані величини можуть бути пов'язані через латентні, так звані приховані від дослідника змінні. Таке явище часто спостерігається, наприклад, в економічних, педагогічних, психологічних і соціологічних дослідженнях. Можливим також є встановлення випадкової кореляції, яка не обумовлена жодною з причин, виокремлених дослідником.

Отже, кореляційний зв'язок показує тенденцію зміни однієї величини під дією іншої. Тому на підставі кореляції можна стверджувати тільки про **ступінь зв'язку** між досліджуваними змінними, а не про існування між ними причинно-наслідкового зв'язку.

Рівняння зв'язку може бути знайдене не тільки для парних залежностей, а й для множинного зв'язку. Такі завдання зазвичай розв'язують, використовуючи методи **багатофакторної регресії** (багатофакторний кореляційно-регресійний аналіз).

Отже, багатофакторний кореляційний-регресійний аналіз застосовують для таких цілей:

- розрахунку теоретичних (очікуваних) значень результативної ознаки;
- зіставлення й оцінювання фактичного і розрахункового значень результативної ознаки;
- порівняльного аналізу різних вибірок;
- розроблення нормативів і прогнозування розвитку складних техніко-економічних (та інших) систем і явищ.

Основною відмінністю між парною та багатофакторною кореляціями є те, що остання дає змогу оцінити ступінь впливу на результативну ознаку кожного з досліджуваних факторів при фіксованих значеннях решти чинників. Окрім того, як показує практика, парна кореляція не завжди дає правильне уявлення про зв'язок між результативною і факторною ознаками.

Важливою умовою отримання достовірних результатів щодо наявності й виду зв'язку між досліджуваними ознаками є вибір факторів, які найістотніше впливають на результативну ознаку. Це обумовлено тим, що одночасне врахування великої кількості різноманітних факторів (особливо при відносно невеликій чисельності вибірки) може призвести до помилкових результатів та висновків.

Окрім того, при збільшені в рівнянні регресії кількості досліджуваних параметрів значно ускладнюється тлумачення одержаних результатів. Ось чому потрібно оцінити й відібрати найбільш значущі фактори, які чинять найсильніший вплив на результативну ознаку. Для зменшення кількості параметрів рекомендують застосовувати, наприклад, метод статистичного групування.

Головною проблемою при побудові рівняння множинної регресії є вибір адекватного математичного рівняння (функції), яке б повністю відповідало характеру взаємозв'язку між результативною ознакою і включеними до рівняння регресії змінними.

При цьому прийоми вибору виду зв'язку, які використовуються при парній кореляції, як правило, не можуть застосовуватися для множинної регресії. Тому для оцінки тісноти зв'язку між конкретною ознакою y й усіма іншими ознаками $(x_1, x_2, \dots, x_n)$ застосовують множинний коефіцієнт кореляції, який позначається R .

Для його розрахунку потрібно побудувати матрицю парних коефіцієнтів кореляції r_{ij} між досліджуваними ознаками, знайти визначник $|A|$ матриці A та алгебраїчне доповнення A_{ii} елемента r_{ii} цієї матриці. Далі множинний коефіцієнт кореляції розраховують за формулою:

$$R_i = (1 - |A|/A_{ii})^{1/2} . \quad (3.18)$$

Перевірку статистичної значущості множинного коефіцієнта R здійснюють за допомогою F -статистики шляхом порівняння розрахованого значення з критичним значенням $F_{\text{крит.}}$ ($F_{\text{крит.}}$ – табличне значення розподілу Фішера, табл. В.1, Додаток В). Якщо розраховане значення F -статистики за абсолютною величиною більше за критичне значення, то множинний коефіцієнт кореляції вважають значущим на обраному рівні значущості α .

У деяких випадках необхідно оперувати частинними коефіцієнтами кореляції R_{ij} . Для розрахунку цього коефіцієнту також потрібно побудувати матрицю парних коефіцієнтів кореляції A та знайти алгебраїчні доповнення A_{ii} , A_{jj} , A_{ij} елементів r_{ii} , r_{jj} , r_{ij} , відповідно.

Отже, регресійний аналіз проводять за таким алгоритмом:

- встановлюють вид кореляційної залежності результативної ознаки y від факторної ознаки x ;
- будують регресійну модель та
- перевіряють статистичну значущість побудованої моделі.

Для вибору рівняння регресії застосовують такі методи, як побудова комбінаційних групувань, експертні оцінки, вивчення парних зв'язків між

результативною ознакою і кожним з досліджуваних факторів, побудова графіків, перебирання функцій різних типів та ін.

Проте всі ці методи, як правило, призводять до великої кількості додаткових обчислень. При цьому у більшості випадків кореляційні зв'язки досить добре описуються функціями лінійного типу або степеневими функціями, які логарифмуванням або методом заміни змінних можна привести до лінійного вигляду.

У загальному випадку лінійна багатфакторна регресія може бути записана як:

$$y = f(x_1, x_2, \dots, x_n) + \epsilon, \quad (3.19)$$

де y – результативна ознака (залежна змінна);

$x_1, x_2, \dots, x_n$ – факторні ознаки (незалежні змінні);

n – кількість факторних ознак;

ϵ – випадкова похибка моделі.

Таким чином, як зазначалось, при n змінних лінійне рівняння множинної регресії матиме вигляд:

$$\tilde{y}_x = a_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n, \quad (3.20)$$

де $\tilde{y}_x$ – залежна змінна (результативна ознака);

x_i – незалежні змінні (фактори);

a_0 – коефіцієнт регресії, початок відліку;

$b_1, b_2, \dots, b_n$ – інші коефіцієнти регресії.

Таким чином, рівняння множинної регресії це математичне рівняння (модель), яким виражається кореляційний зв'язок між декількома досліджуваними ознаками. Параметри рівняння регресії також знаходять використовуючи метод найменших квадратів.

Зазначимо, що для побудови адекватної багатофакторної регресійної моделі необхідно, щоб випадкова похибка ε мала нормальний розподіл із нульовим математичним очікуванням, а випадкові похибки кожного вимірювання були незалежними і мали однакові дисперсії. Окрім того, кількість спостережень N повинна перевищувати величину $3(n + 1)$.

Найважливішою умовою побудови коректних багатофакторних регресійних моделей і забезпечення їх статистичної значущості є те, що всі досліджувані факторні ознаки повинні бути тісно пов'язані з результативною ознакою, проте бути незалежними одна від одної. Тісноту зв'язку між результативною і факторними ознаками, а також зв'язку факторних ознак між собою визначають аналізом парних і частинних коефіцієнтів кореляції.

Коефіцієнти множинної регресії показують ступінь середньої зміни результативної ознаки за умови зміни відповідної факторної ознаки на одиницю коли всі інші фактори (або ознаки), які включені до рівняння регресії, залишаються сталими на одному середньому рівні.

Отже, коефіцієнти чистої регресії це коефіцієнти множинної регресії, які характеризують зв'язок між результативною ознакою і певним фактором при фіксованих значеннях інших факторів.

Коефіцієнтами повної регресії називають коефіцієнти парної регресії.

3.4 Алгоритми та приклади виконання розрахунків (лінійна регресія)

Розглянемо деякі приклади.

Приклад 4.

У результаті проведення експериментального дослідження отримано сукупність показників x та y (табл. 3.4). Позначимо незалежну ознаку як x , а залежну – як y . Завдання: провести статистичний аналіз даних.

Таблиця 3.4 Експериментальні дані (стовбці 2, 3) і результати розрахунків (стовбці 4–11)

№	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	x^2	y^2	xy	$(x - \bar{x})(y - \bar{y})$
1	2	3	4	5	6	7	8	9	10	11
1	131,72	5,48	-5,12	-1,16	26,17	1,34	17350,16	30,03	721,83	5,93
2	126,08	5,32	-10,76	-1,32	115,68	1,74	15896,17	28,30	670,75	14,19
3	142,95	6,85	6,11	0,21	37,39	0,04	20434,70	46,92	979,21	1,29
4	103,07	3,46	-33,77	-3,18	1140,10	10,11	10623,42	11,97	356,62	107,35
5	101,47	3,41	-35,37	-3,23	1250,71	10,43	10296,16	11,63	346,01	114,21
6	125,03	5,04	-11,81	-1,60	139,37	2,56	15632,50	25,40	630,15	18,88
7	126,93	5,41	-9,91	-1,23	98,12	1,51	16111,22	29,27	686,69	12,18
8	123,81	4,79	-13,03	-1,85	169,66	3,42	15328,92	22,94	593,05	24,09
9	141,14	6,91	4,30	0,27	18,53	0,07	19920,50	47,75	975,28	1,16
10	150,86	7,82	14,02	1,18	196,69	1,39	22758,74	61,15	1179,73	16,56
11	129,62	5,86	-7,22	-0,78	52,06	0,61	16801,34	34,34	759,57	5,62
12	133,66	6,08	-3,18	-0,56	10,08	0,31	17865,00	36,97	812,65	1,78
13	132,81	6,01	-4,03	-0,63	16,20	0,40	17638,50	36,12	798,19	2,53
14	122,26	4,86	-14,58	-1,78	212,44	3,17	14947,51	23,62	594,18	25,94
15	107,92	3,68	-28,92	-2,96	836,10	8,76	11646,73	13,54	397,15	85,57
16	103,36	3,39	-33,48	-3,25	1120,60	10,56	10683,29	11,49	350,39	108,77
17	137,54	6,21	0,70	-0,43	0,50	0,18	18917,25	38,56	854,12	-0,30
18	121,18	4,97	-15,66	-1,67	245,09	2,79	14684,59	24,70	602,26	26,14
19	145,79	7,61	8,95	0,97	80,18	0,94	21254,72	57,91	1109,46	8,69
20	110,52	4,15	-26,32	-2,49	692,50	6,20	12214,67	17,22	458,66	65,51
21	160,03	8,72	23,19	2,08	537,99	4,33	25609,60	76,04	1395,46	48,26

№	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	x^2	y^2	xy	$(x - \bar{x})(y - \bar{y})$
1	2	3	4	5	6	7	8	9	10	11
22	159,08	8,51	22,24	1,87	494,82	3,50	25306,45	72,42	1353,77	41,61
23	130,05	5,71	-6,79	-0,93	46,04	0,86	16913,00	32,60	742,59	6,31
24	128,06	5,82	-8,78	-0,82	77,01	0,67	16399,36	33,87	745,31	7,19
25	126,94	5,03	-9,90	-1,61	97,92	2,59	16113,76	25,30	638,51	15,93
26	111,96	4,11	-24,88	-2,53	618,79	6,40	12535,04	16,89	460,16	62,92
27	143,29	7,49	6,45	0,85	41,66	0,72	20532,02	56,10	1073,24	5,49
28	158,46	9,04	21,62	2,40	467,62	5,76	25109,57	81,72	1432,48	51,91
29	161,03	9,26	24,19	2,62	585,38	6,87	25930,66	85,75	1491,14	63,40
30	165,39	9,91	28,55	3,27	815,37	10,70	27353,85	98,21	1639,01	93,39
31	146,98	8,38	10,14	1,74	102,91	3,03	21603,12	70,22	1231,69	17,66
32	143,27	6,97	6,43	0,33	41,40	0,11	20526,29	48,58	998,59	2,13
33	170,72	10,31	33,88	3,67	1148,17	13,47	29145,32	106,30	1760,12	124,38
34	168,93	10,24	32,09	3,60	1030,06	12,96	28537,34	104,86	1729,84	115,56
35	150,36	7,48	13,52	0,84	182,91	0,71	22608,13	55,95	1124,69	11,37
36	124,93	4,91	-11,91	-1,73	141,74	2,99	15607,50	24,11	613,41	20,59
37	167,78	9,89	30,94	3,25	957,57	10,57	28150,13	97,81	1659,34	100,59
38	163,66	9,57	26,82	2,93	719,56	8,59	26784,60	91,58	1566,23	78,61
39	134,29	6,02	-2,55	-0,62	6,48	0,38	18033,80	36,24	808,43	1,58
40	171,03	10,53	34,19	3,89	1169,27	15,14	29251,26	110,88	1800,95	133,04
41	168,96	10,27	32,12	3,63	1031,99	13,18	28547,48	105,47	1735,22	116,63
42	125,09	4,99	-11,75	-1,65	137,95	2,72	15647,51	24,90	624,20	19,37
43	98,02	3,07	-38,82	-3,57	1506,64	12,74	9607,92	9,42	300,92	138,55
44	97,36	3,09	-39,48	-3,55	1558,31	12,60	9478,97	9,55	300,84	140,11
45	101,73	3,31	-35,11	-3,33	1232,39	11,08	10348,99	10,96	336,73	116,88
46	163,74	10,11	26,90	3,47	723,86	12,05	26810,79	102,21	1655,41	93,38
47	99,53	3,27	-37,31	-3,37	1391,69	11,35	9906,22	10,69	325,46	125,70
48	149,59	8,57	12,75	1,93	162,68	3,73	22377,17	73,44	1281,99	24,62
49	160,48	9,35	23,64	2,71	559,07	7,35	25753,83	87,42	1500,49	64,09
50	173,31	10,73	36,47	4,09	1330,40	16,73	30036,36	115,13	1859,62	149,20
-	Сума	Сума			Сума	Сума	Сума	Сума	Сума	Сума
	6841,77	331,97			25375,82	280,42	961572,16	2484,50	48061,78	2636,53
	$\bar{x}$	$\bar{y}$								
	136,84	6,64								

Проаналізуємо вихідні дані за таким алгоритмом:

- для побудови кореляційного поля, наприклад, у MS Office Excel, IBM SPSS Statistics, RStudio (за умови попередньої підготовки даних) чи іншій програмі до табл. 3.4 (стовбці 2 і 3) вносимо значення змінних x та y ;
- проводимо розрахунки за формулами, наведеними у табл. 3.4;
- підбираємо рівняння лінійної регресії, розраховуємо коефіцієнти регресії та кореляції; перевіряємо значущість за допомогою критерію Стюдента та аналізуємо тісноту кореляційного зв'язку між досліджуваними ознаками;
- будуємо графік залежності між ознаками y та x і порівнюємо його з експериментальними даними.

Отже, вводимо до табл. 3.4 експериментальні значення x та y ($N = 50$) і будуємо кореляційне поле (рис. 3.3). Із розташування точок видно, що вони формують певну залежність, яка вірогідно має лінійний характер.

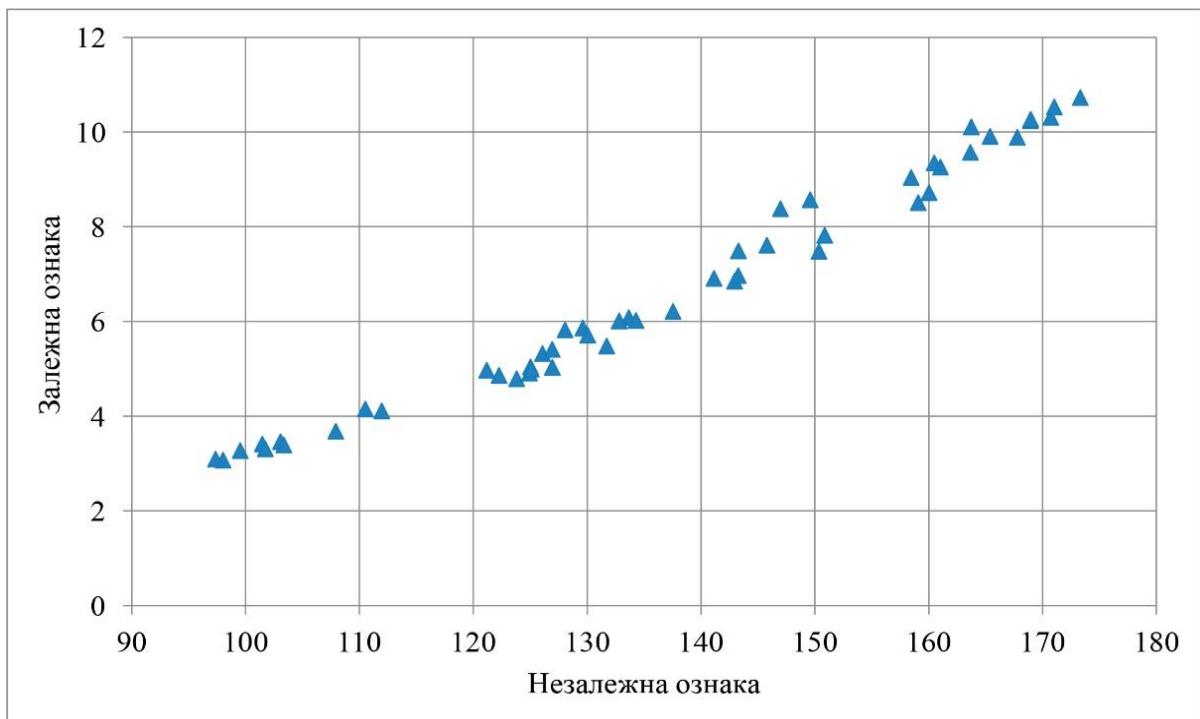


Рис. 3.3 Вид кореляційного поля для прикладу 4.

У результаті проведення розрахунків отримуємо такі коефіцієнти регресії a і b та коефіцієнт кореляції r :

$$b = 0,1039$$

$$a = -7,5777$$

$$r = 0,99$$

Оскільки r наближається до 1, доходимо висновку, що існує дуже тісний кореляційний зв'язок між незалежною ознакою X та залежною ознакою Y .

Таким чином, отримане рівняння регресії має вигляд:

$$y = -7,5777 + 0,1039 \cdot x$$

Для перевірки розрахунків розв'язуємо обернену задачу; результат розрахунку оберненої задачі наведено у табл. 3.5.

Таблиця 3.5 Результат перевірки правильності розрахунків

№	Y	y'	№	Y	y'
1	2	3	4	5	6
1.	5,48	6,11	26.	4,11	4,05
2.	5,32	5,52	27.	7,49	7,31
3.	6,85	7,27	28.	9,04	8,89
4.	3,46	3,13	29.	9,26	9,15
5.	3,41	2,96	30.	9,91	9,61
6.	5,04	5,41	31.	8,38	7,69
7.	5,41	5,61	32.	6,97	7,31
8.	4,79	5,29	33.	10,31	10,16
9.	6,91	7,09	34.	10,24	9,97
10.	7,82	8,10	35.	7,48	8,04
11.	5,86	5,89	36.	4,91	5,40
12.	6,08	6,31	37.	9,89	9,85

№	y	y'	№	y	y'
1	2	3	4	5	6
13.	6,01	6,22	38.	9,57	9,43
14.	4,86	5,13	39.	6,02	6,37
15.	3,68	3,64	40.	10,53	10,19
16.	3,39	3,16	41.	10,27	9,98
17.	6,21	6,71	42.	4,99	5,42
18.	4,97	5,01	43.	3,07	2,61
19.	7,61	7,57	44.	3,09	2,54
20.	4,15	3,91	45.	3,31	2,99
21.	8,72	9,05	46.	10,11	9,43
22.	8,51	8,95	47.	3,27	2,76
23.	5,71	5,93	48.	8,57	7,96
24.	5,82	5,73	49.	9,35	9,10
25.	5,03	5,61	50.	10,73	10,43

Як можна побачити, значних розбіжностей між експериментальними і розрахунковими даними не спостерігається (рис. 3.4).

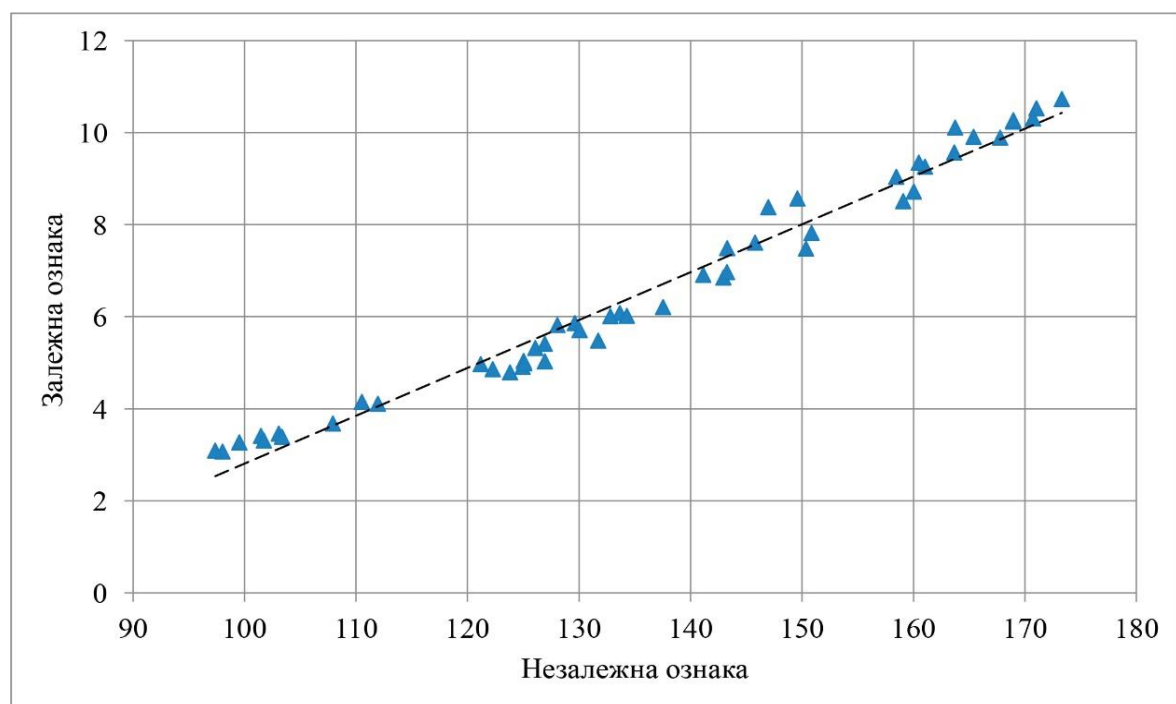


Рис. 3.4 Залежність між ознаками y та x для прикладу 4.

Приклад 5.

У дослідженні отримано дані для x та y (рис. 3.5). Аналіз цих даних виконуємо аналогічно до попереднього прикладу.

Із розташування точок видно, що існує певна залежність між досліджуваними ознаками, однак чіткої лінійної залежності не помітно.

Визначаємо шукані коефіцієнти рівняння регресії та коефіцієнт кореляції, а також перевіряємо значущість коефіцієнта кореляції за допомогою критерію Стьюдента:

$$b = 0,0384$$

$$a = 1,3849$$

$$r = 0,37$$

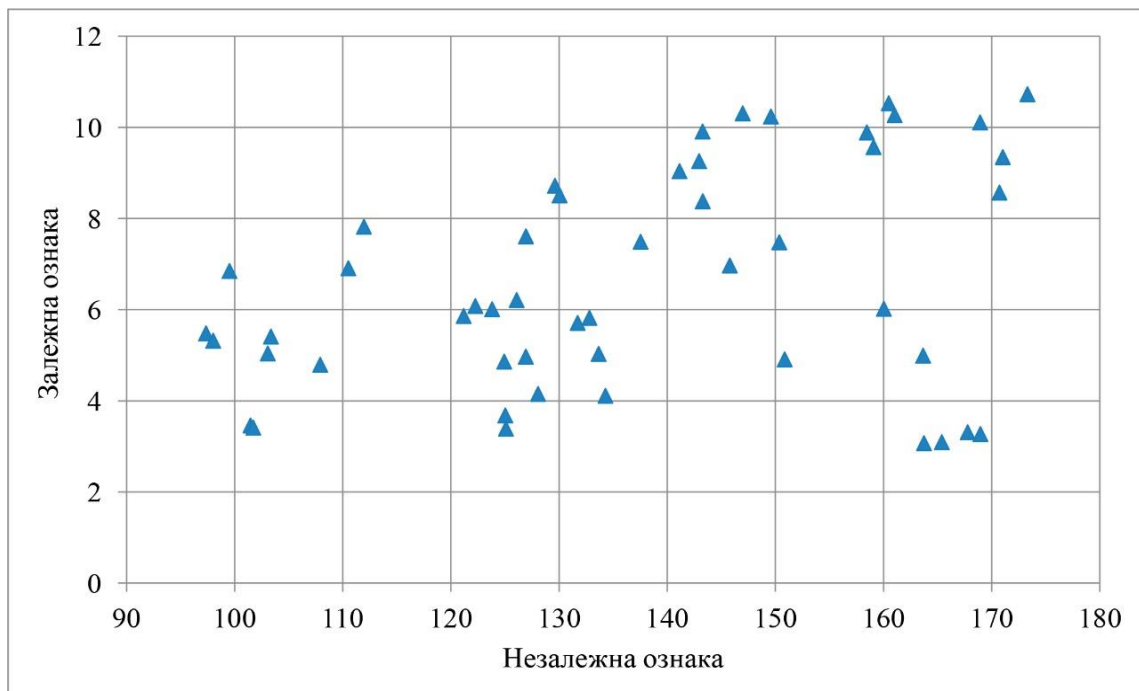


Рис. 3.5 Вид кореляційного поля для прикладу 5.

За величиною коефіцієнта кореляції r доходимо висновку, що зв'язок між ознаками помірний. Отже, рівняння регресії має вигляд:

$$y = 1,3849 + 0,0384 \cdot x$$

У результаті розв'язання оберненої задачі встановлюємо, що значних розбіжностей між експериментальними та розрахунковими значеннями (y та y') немає (рис. 3.6).

Отже, отримані результати свідчать про коректність розрахунків.

Приклад 6.

Шляхом проведення експериментального дослідження отримано набір даних для y та x . Аналіз і статистичну обробку результатів дослідження виконуємо аналогічно до попередніх прикладів.

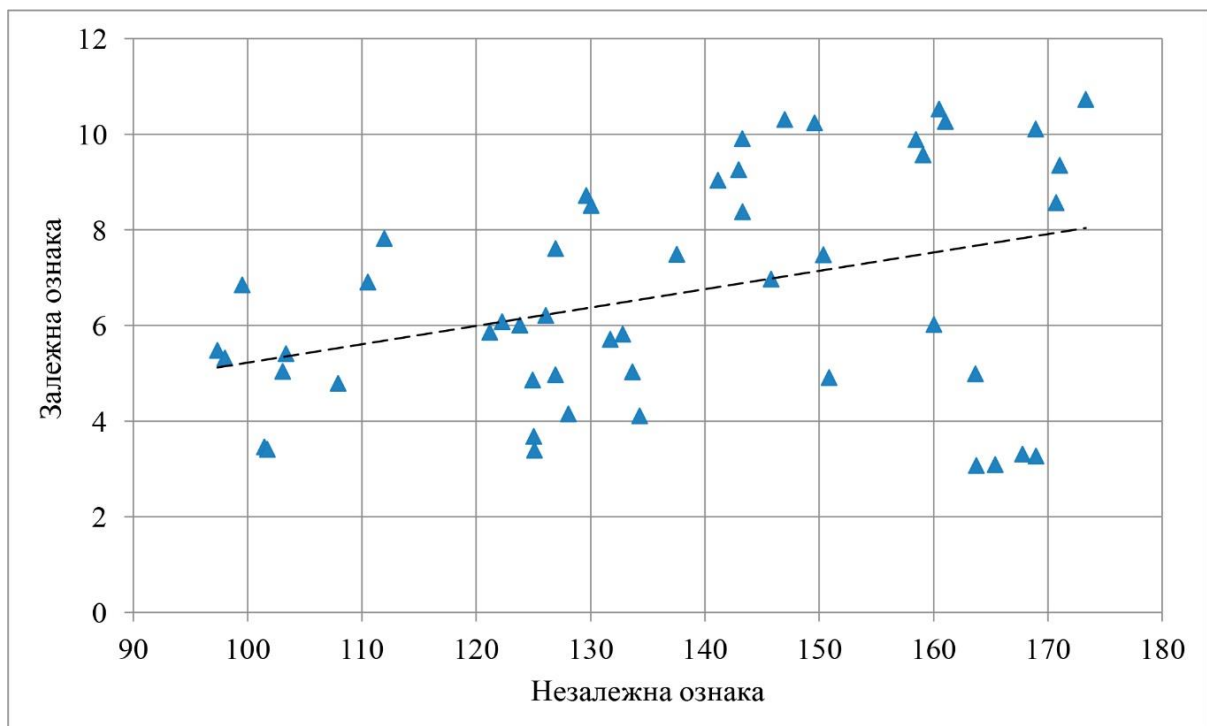


Рис. 3.6 Залежність ознаки y від ознаки x для прикладу 5.

На рис. 3.7 показано побудоване за експериментальними даними кореляційне поле. Із розташування точок видно, що залежності між величинами не простежується.

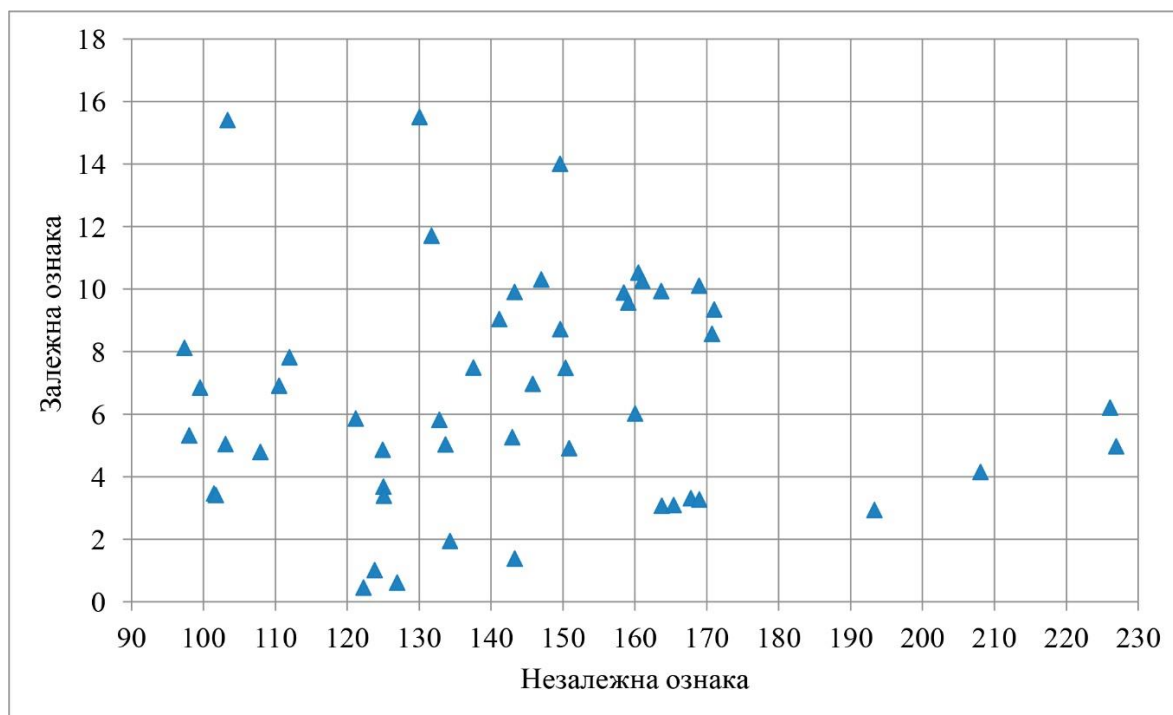


Рис. 3.7 Кореляційне поле, побудовано за даними прикладу 6.

Таким чином, розрахунок показує, що величина коефіцієнту кореляції становить $r = 0,02$. Отже, доходимо висновку, що кореляційний зв'язок між досліджуваними ознаками практично відсутній і у такому випадку визначати коефіцієнти рівняння регресії не потрібно.

Проте у подібних випадках існує вірогідність того, що було невірно обрано факторну ознаку або на результативну ознаку Y одночасно чинять вплив декілька різних факторів.

3.5 Використання статистичних пакетів у R

Серед статистичних пакетів мови програмування R найбільш популярними є:

- fastR (є на CRAN);
- lattice (пакет для графіки);
- mosaic (пакет, призначений для вивчення статистики; є на CRAN);
- Hmisc (є на CRAN).

3.5.1 Обробка числових даних

Як зазначалось, R включає в себе функції, які обчислюють значний спектр чисельних і графічних даних для статистичного аналізу.

Приклад 7.

Для цього прикладу використовуємо інформацію з журналу Motor Trend (США, 1974 р.), а саме набір даних mtcars. Інформація з журналу містить такі статистичні дані, як споживання моторного палива і 10 аспектів автомобільного дизайну і продуктивності для 32 автомобілів (моделей 1973–1974 рр.) (табл. 3.6).

Таблиця 3.6 Набір статистичних даних для прикладу 7

Індекс	Назва змінної	Опис
[,1]	mpg	Міль/галон(США)
[,2]	cyl	Кількість циліндрів
[,3]	disp	Об'єм (в кубічних дюймах)
[,4]	hp	Механічна кінська сила

Продовження таблиці 3.6		
[,5]	drat	Передатне відношення головної передачі
[,6]	wt	Вага (фунт/1000)
[,7]	qsec	Час проходження ¼ милі
[,8]	vs	V/S
[,9]	am	Передача (0 = автоматична, 1 = механічна)
[,10]	gear	Кількість передач
[,11]	carb	Кількість карбюраторів

Отже, розглянемо декілька можливих способів обчислення статистичних значень.

а) Середнє значення набору даних `mtcars` по показнику `mpg`:

```
> mean(mtcars$mpg)
```

```
[1] 20.09062
```

б) Медіана набору даних `mtcars` по показнику `mpg`:

```
> median(mtcars$mpg)
```

```
[1] 19.2
```

в) Квантиль набору даних `mtcars` по показнику `mpg`:

```
> quantile(mtcars$mpg)
```

```
0%    25%   50%   75%   100%
10.400 15.425 19.200 22.800 33.900
```

г) Інтерквантиль набору даних `mtcars` по показнику `mpg`:

```
> IQR(mtcars$mpg)
```

```
[1] 7.375
```

г) Функція `favstats()` з пакету `mosaic` обчислює одразу кілька статистичних показників, наприклад:

```
> favstats(mtcars$mpg)
  min   Q1 median  Q3  max  mean    sd      n missing
10.4 15.425  19.2  22.8 33.9 20.09062 6.026948 32      0
```

д) У пакеті `Hmisc` є функція `summary()`, яка відображає данні у зручному вигляді, зокрема:

```
> summary(mpg ~ am, data = mtcars, fun=favstats)
mpg   N=32
+-----+---+--+---+-----+-----+---+---+-----+-----+---+-----+
|   | |N|min|Q1 |median|Q3 |max|mean |sd  |n|missing|
+-----+---+--+---+-----+-----+---+---+-----+-----+---+-----+
|am  |No |19|10.4|14.950|17.3 |19.2|24.4|17.14737|3.833966|19|0  |
|    |Yes|13|15.0|21.000|22.8 |30.4|33.9|24.39231|6.166504|13|0  |
+-----+---+--+---+-----+-----+---+---+-----+-----+---+-----+
|Overall| |32|10.4|15.425|19.2 |22.8|33.9|20.09062|6.026948|32|0  |
+-----+---+--+---+-----+-----+---+---+-----+-----+---+-----+
```

3.5.2 Ґраткова графіка

У мові програмування R існує низка способів для побудови графіків. Один з цих способів – використання системи ґраткових графіків. І перший крок для використання ґраток («`lattice`») – необхідність завантажити пакет «`lattice`», використовуючи прапорець на вкладинці `Packages` або за допомогою команди:

```
> require(lattice)
```

Граткові графіки використовують формульний інтерфейс:

```
> plotname( y ~ x | z, data=dataname,  
groups=grouping_variable, ...)
```

Наприклад, у списку представлено імена декількох граткових графіків:

- `histogram` (для побудови гістограм);
- `bwplot` (для коробкових графіків або графіків «ящик з вусами»);
- `xyplot` (для діаграм розсіювання);
- `qqmath` (для графіків квантиль-квантиль (Q-Q)).

У цих пакетах:

- `x` це ім'я змінної, яка відкладається по горизонтальній (x) осі;
- `y` це ім'я змінної, яка відкладається по вертикальній (y) осі;

Для деяких графіків, це поле є порожнім, оскільки R обчислює ці значення із значень `x`;

- `z` це змінна стану – використовується для поділу графіка на кілька областей – так званих панелей;

- `grouping_variable` – використовується для відображення різних груп по-різному (наприклад, різні кольори та/або символи) у межах тієї ж панелі.

Зазначимо, що існує вельми багато додаткових аргументів цих функцій, які надають змогу контролювати, як само виглядатиме графік.

Гістограма – один із способів візуалізації, графічного представлення табличних даних; являє собою діаграму, що складається з прямокутників без розривів між ними.

Приклад 8.

Для побудови гістограми стандартного набору даних `mtcars` по показнику `hp` (див. приклад 7), потрібно набрати:

```
> histogram(~ hp, data=mtcars)
```

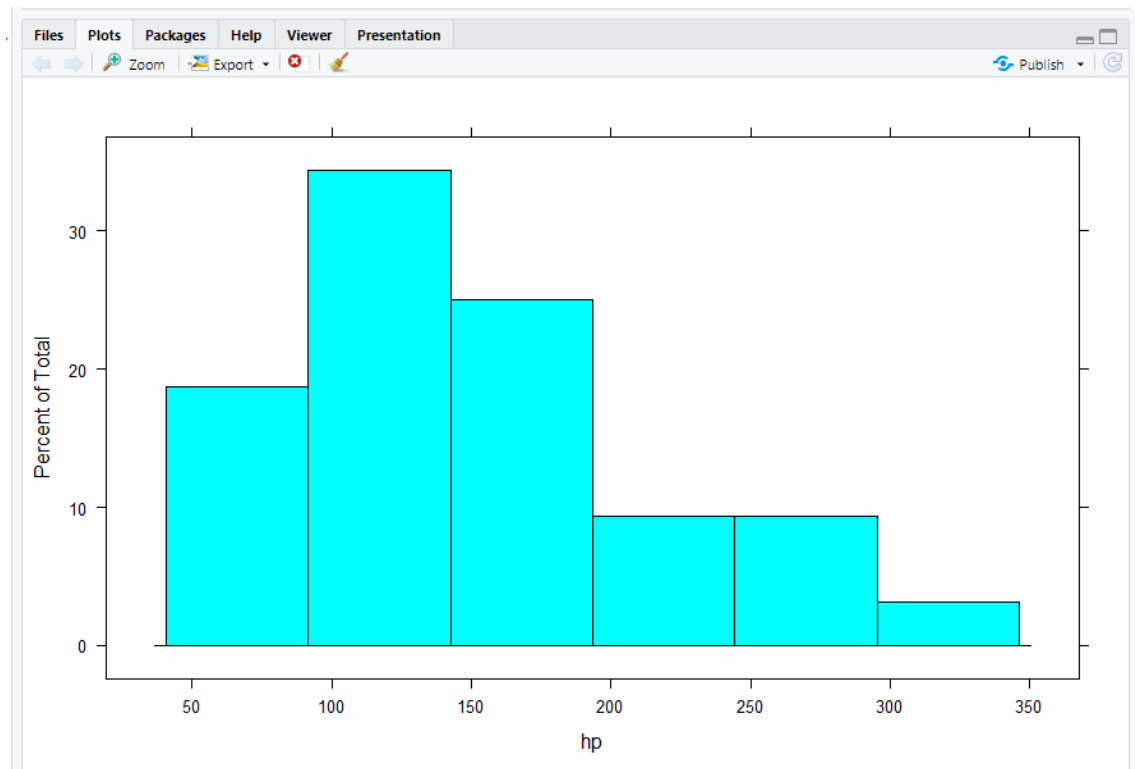


Рис. 3.8 Приклад побудови гістограм в RStudio.

У даному прикладі у компонент формули є порожнім, оскільки R самостійно обчислює висоту стовпців (рис. 3.8).

У R можна використовувати змінні стану, щоб побудувати окремі гістограми для кожного значення цієї змінної. Зокрема, щоб побудувати гістограму стандартного набору даних `mtcars` по показнику `hp` в залежності від значення параметра `am`, потрібно набрати:

```
> histogram(~ hp | factor(am), data=mtcars)
```

Отже, на графіку (рис. 3.9) зображено дві підгрупи, які мають назву панелі, і надписи зверху, які називаються смужками. Смужки можуть бути розміщені на лівому боці за бажанням.

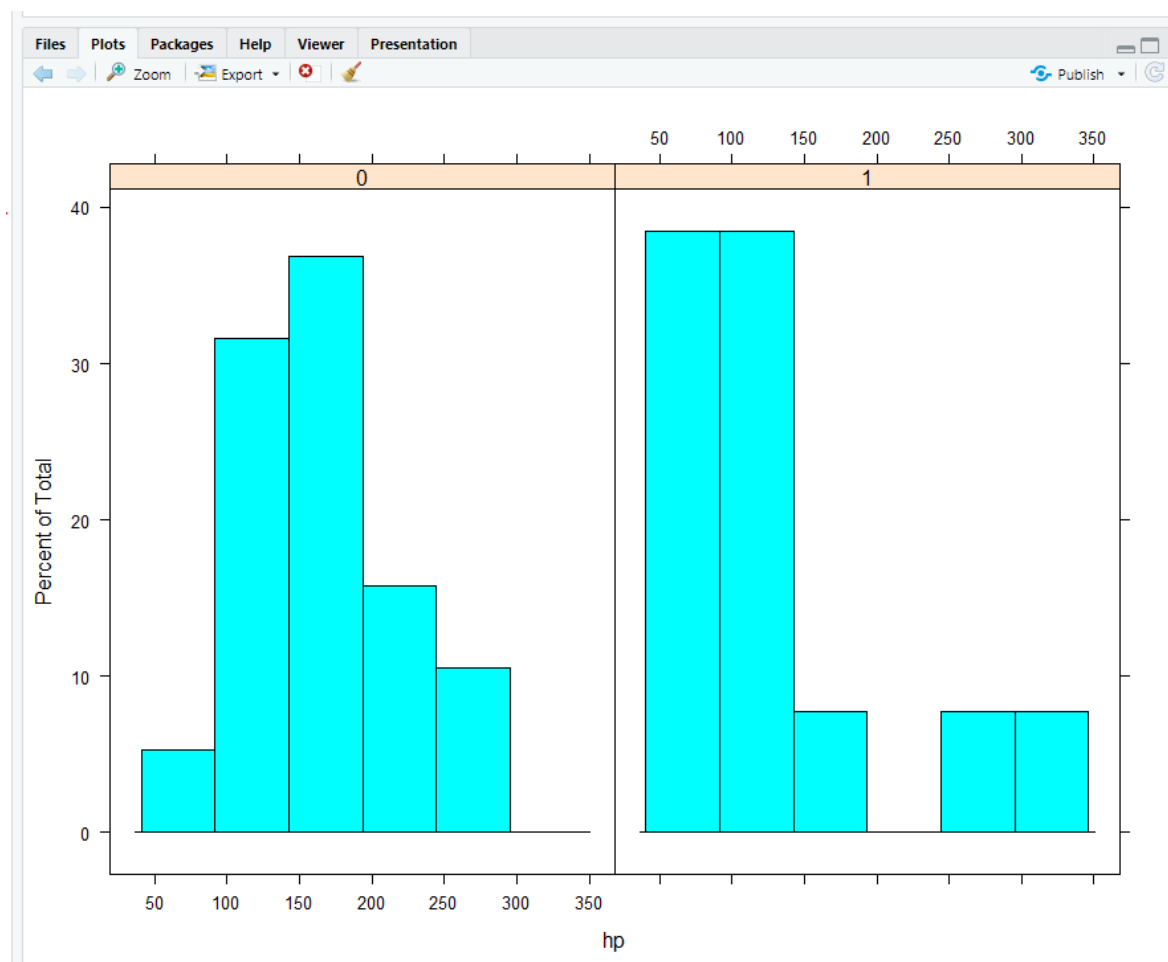


Рис. 3.9 Приклад побудови в RStudio гістограм із набору даних `mtcars` по показнику `hp` в залежності від значення параметра `am`.

У R можна використовувати одразу декілька змінних стану (рис. 3.10):

```
> histogram(~ hp | factor(am) + factor(gear), data=mtcars)
```

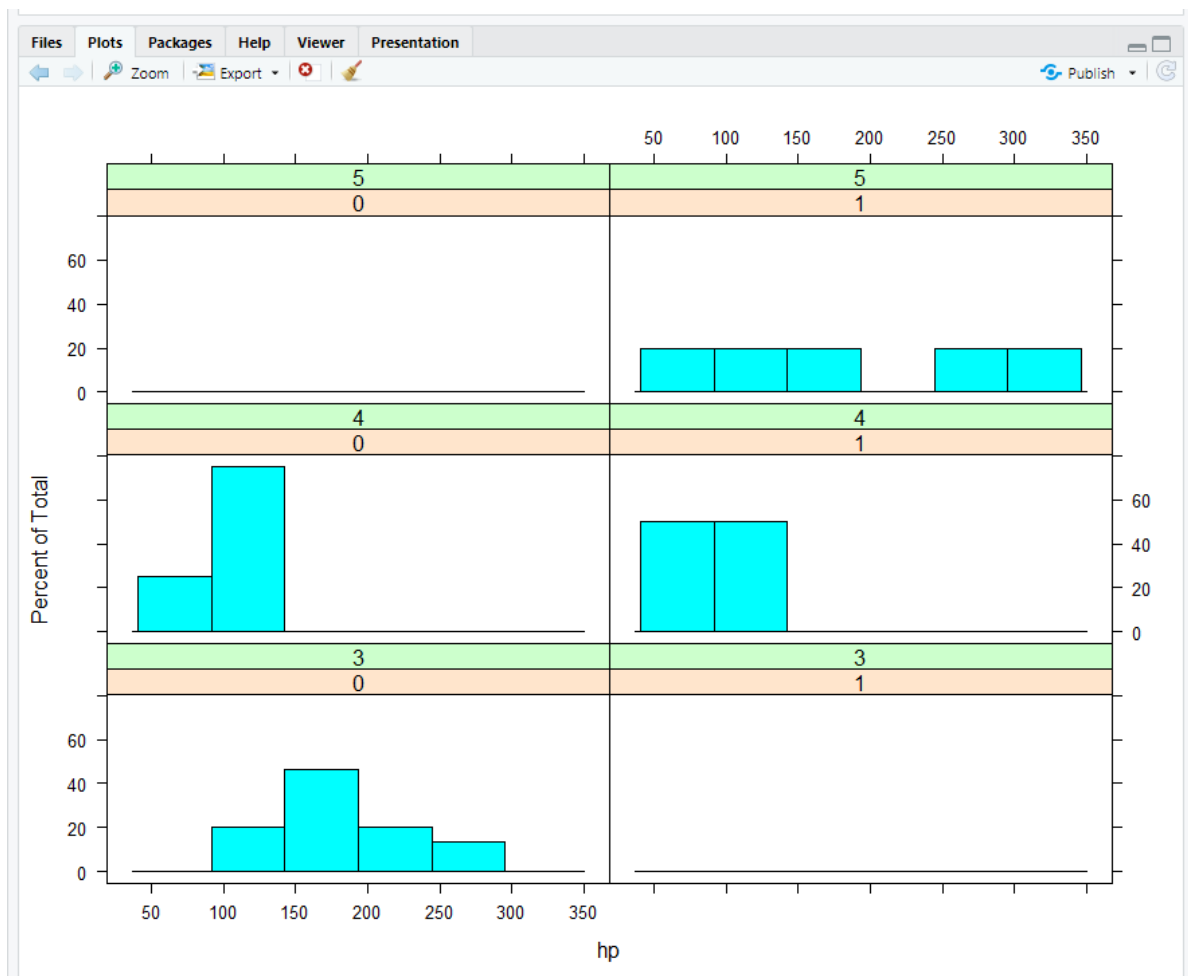


Рис. 3.10 Приклад побудови в RStudio гістограм із набору даних `mtcars` із використанням декількох змінних стану.

Ще одним із типів гістограм є стовпчикова діаграма (рис. 3.11). Щоб її побудувати, потрібно використати наступний код:

```
>barchart(row.names(mtcars) ~ hp, data = mtcars)
```

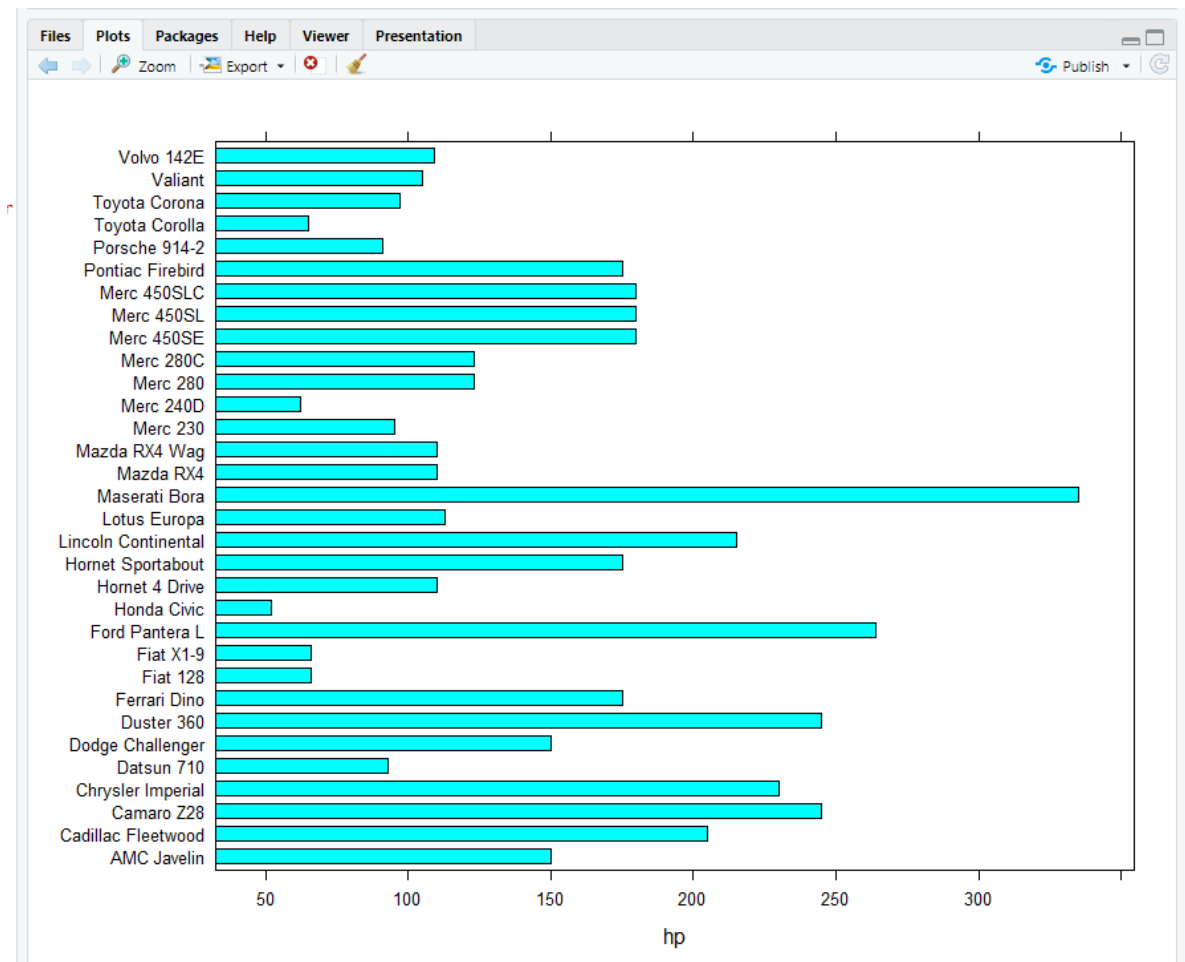


Рис. 3.11 Приклад побудови в RStudio стовпчикової діаграми.

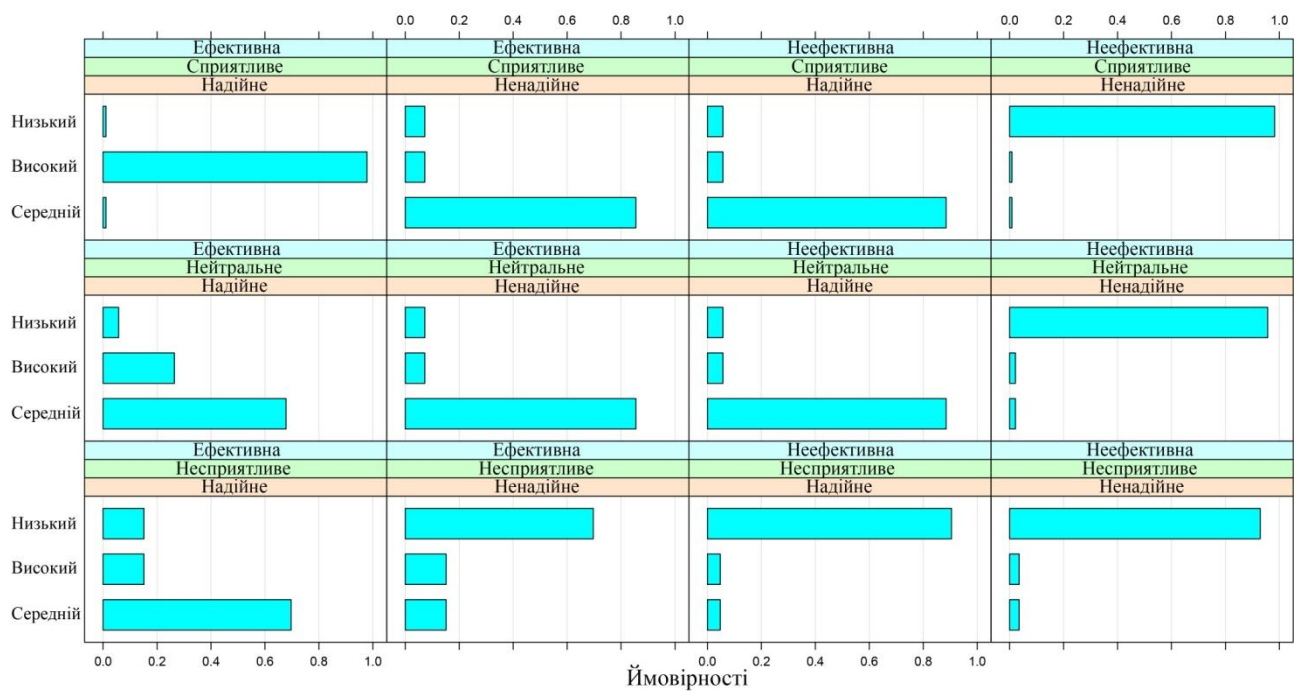


Рис. 3.12 Розподіл умовних ймовірностей успішності реалізації стартапів.

Слід зауважити, що пакет «lattice», часто використовується як додатковий компонент інших більш складних пакетів, які застосовуються у маркетингових дослідженнях. Зокрема, у пакеті «Bnlearn», який застосовується для створення Баєсових мереж для оцінювання успішності ринкового провадження інноваційних продуктів або комерційної діяльності інноваційних стартапів у цілому, прогнозування рівнів продажів тощо. У результаті застосування цього пакету можна отримати, наприклад, розподіл умовних ймовірностей успішності реалізації стартапів (рис. 3.12).

3.6 Шкали вимірювань, що застосовуються в економічних та маркетингових дослідженнях

В економічних та маркетингових дослідженнях можуть бути застосовані 4 типи вимірювальних шкал, а саме:

- номінальна,
- порядкова,
- інтервальна та
- шкала відношень.

Номінальна шкала (тобто шкала найменувань) це шкала, при використанні якої досліджувані ознаки розрізняють за назвою (ім'ям, найменуванням).

Зокрема, як приклад номінальних шкал можна привести такі шкали, як:

- «чоловік» — «жінка»;
- «доросла людина» — «дитина»;
- «тест виконано» — «тест не виконано»,
- «за» — «проти» (шкали, в яких є тільки два з можливих значення, називають дихотомічними);

- кольори у спектрі: «червоний», «помаранчевий», «жовтий», «зелений» тощо;

- варіанти вибору: «перший кандидат», «другий кандидат», «третій кандидат» і так далі.

Звісно, самі найменування не оброблюються. Окрім того, між значеннями номінальної шкали немає відношень виду «більше», «менше» чи «дорівнює».

Проте під час експерименту чи спостереження дослідник може виявити певні числові характеристики, пов'язані з кожним із найменувань. Наприклад, дослідник може підрахувати кількість учасників, які надали (не надали) відповіді на питання анкети; отримати дані щодо вподобань споживачів та встановити частку клієнтів, які лояльно відносяться до певних брендів; підрахувати кількість голосів, що було віддано за певного кандидата на виборах.

Одним з основних методів досліджень при використанні номінальних шкал є метод анкетування. За його результатами дослідник може встановити абсолютні кількості зустрічаємості кожного з наявних найменувань у досліджуваній вибірці або підрахувати їх відносні кількості (частоти). Далі, вже, маючи кількісні показники, можна проводити їх статистичну обробку.

Порядкова шкала забезпечує розташування значень досліджуваної ознаки у певній монотонній послідовності, наприклад, у порядку зростання чи у порядку спадання), але без зазначення ступеню відмінності показників, приписаних до різних градацій (класів). При цьому ознаки можуть бути якісними або кількісними.

Наприклад:

- ґатунок продукції: «вищий ґатунок», «перший ґатунок», «другий ґатунок» і так ділі, тобто якісна ознака;

- певні категорії працівників або посади, зокрема, в університеті це «асистент», «викладач», «старший викладач», «доцент» і «професор» (також якісна ознака);

- групи опитуваних, які розрізняються за віком, наприклад, «молодь», «середня вікова група», «люди похилого віку» (якісна ознака);
- словесне оцінювання рівня засвоєння певного матеріалу: «відмінно», «дуже добре», «добре», «задовільно», «незадовільно» (якісна ознака);
- кількість балів, отриманих, наприклад, працівниками при тестуванні їхніх здібностей 1–2–3–4–5 чи за 100-бальною шкалою (кількісна ознака).

Зазначимо, що при побудові порядкових шкал потрібно дотримуватися таких умов:

- по-перше, кількість градацій ознаки повинна бути не менше, ніж три;
- по-друге, градації мають бути впорядковані за значенням ознаки;
- по-третє, для застосування цієї шкали повинна бути чітко визначена процедура віднесення певної ознаки до тієї чи іншої градації.

На практиці це означає, що при побудові порядкової шкали повинні бути встановлені критерії, за якими однозначно можна віднести певну ознаку до тієї чи іншої градації.

Від якісних та/або кількісних градацій досить часто є необхідність перейти до чисельних значень, що характеризують порядок слідування певних градацій та які мають назву «ранг».

Отже, присвоєння градаціям відповідних їм рангів відбувається за такими правилами:

- найменше значення рангу дорівнює 1; найбільше – кількості ранжованих значень N , наприклад, кількості вимірювань у вибірці або кількості респондентів при опитуванні;
- якщо кількість ранжованих значень співпадає з кількістю градацій, то нижчому рівню приписують ранг 1, наступному – 2 і так далі. Відповідно, найвищій рівень отримує ранг N ;
- якщо серед ранжованих значень декілька потрапляють до однієї градації, то тоді всім їм приписують однаковий ранг, який розраховують за формулою:

$$R_i = \sum_{k=0}^{i-1} n_k + \frac{n_i + 1}{2},$$

де:

i – номер градації;

R_i – ранг кожного значення ознаки, що потрапив до градації i ;

n_i – кількість значень, що потрапили до градації i (n_0 приймається таким, що дорівнює 0).

Операція присвоєння градаціям рангів називається ранжуванням. Перевірка правильності ранжування здійснюється таким чином: необхідно просумувати всі ранги і порівняти з перевіркою сумою, яку визначають за формулою:

$$S_R^T = \frac{N(N+1)}{2}$$

Отже, операція ранжування надає змогу перейти від порядкової шкали з якісними ознаками до шкали з кількісними ознаками, які піддаються потім статистичній обробці. Ця процедура значно розширює кількість ознак, які можна використати як критеріальні в економічних і, зокрема, маркетингових дослідженнях. Багато статистичних методів базуються саме на рангових показниках шкали, тому саме з процедури ранжування починається їх застосування для статистичної обробки даних.

Зазначимо, що ранги встановлюють тільки порядок слідування градацій ознаки, а не інтервали між ними. Якщо, наприклад, ознаки X , Y і Z мають ранги, відповідно, 5, 6 і 7, це не означає, що інтервал між X і Y такий самий, як і між Y та Z , а інтервал між X і Z у два рази більше, ніж між Y і Z .

Інтервальна шкала (шкала рівних інтервалів) це шкала, на якій встановлено однакову кількісну відмінність між сусідніми градаціями ознаки. Тобто процедура вимірювання побудована таким чином, що можна не тільки виявити, яка з 2-х виміряних величин більша, а й на скільки саме одиниць одна з них відрізняється від іншої.

Наприклад, за фіксований інтервал часу працівник А виконав 5 завдань, працівник Б – 3 завдання, а працівник В – 6 завдань. Отже, це означає, що показник працівника А на 1 одиницю гірший за показник працівника В і на 2 одиниці кращий за показник працівника Б.

Проте у інтервальній шкали не встановлена «0» – нульова точка відліку, що не надає змоги встановити відношення досліджуваних показників. Наприклад, на основі наведених вище даних неможливо стверджувати, що працівник Б працює у 2 рази гірше за працівника В. Отже, існує думка, що в економічних науках недоцільно користуватися вимірювальною шкалою. Виключення становлять так звані процентильні шкали і шкали, побудовані в одиницях стандартного відхилення. Але ці шкали доцільно застосовувати тільки у випадках, коли доведено, що вимірювана випадкова величина підкоряється нормальному (гаусовому) закону розподілу.

Наступна шкала – шкала рівних відношень (пропорційна шкала). Це така шкала, що надає змогу встановити пропорції між вимірюваними величинами. Обов'язковою умовою застосування цієї шкали є існування нульової точки відліку, причому, єдиної для всіх вимірювань (так званий абсолютний нуль). Як приклад такої шкали можна навести температурну шкалу Кельвіна, хоча в економічних дослідженнях такі показники зустрічаються доволі рідко.

Наприклад, якщо IQ одного працівника дорівнює 100, а іншого – 150, то це не означає, що цей другий працівник у півтора рази розумніший за першого. Всім відомо, що відсутній «абсолютний нуль» порядності, чесності, розуму або здатностей людини. Отже, в економічних та маркетингових дослідженнях ця шкала також застосовується досить рідко, хоча виключенням є, наприклад, ситуації, коли співставляються відносні показники.

Наприклад, за бренд А свої голоси віддали 20 осіб з 60 тих, хто брав участь у голосуванні; за бренд Б – 40 осіб, а за бренд В – жодного. За цими даними можна досить легко визначити відносні частоти вибору досліджуваних брендів, а саме: $1/3$, $2/3$ і нуль.

Отже, провівши таке маркетингове дослідження, можна стверджувати, що вибір бренду Б відбувався у 2 рази частіше за вибір бренду А.

Отже, на нашу думку, основною шкалою в економічних і, зокрема, у маркетингових дослідженнях є порядкова шкала (рангова шкала). Досить зручною може бути також і шкала відношень (пропорційна), побудована на відносних частотах, оскільки її можна застосовувати у випадках, коли первинні дані отримано за допомогою інших шкал, зокрема, номінальної або порядкової.

РОЗДІЛ 4

ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ ТА ПРИКЛАДИ ЇХ РОЗВ'ЯЗАННЯ

Розглянемо деякі з прикладів виконання завдань для самостійної роботи, зокрема, проведення маркетингового дослідження для з'ясування споживчих мотивацій на певному ринку. Здобувачам пропонується за заданим набором даних провести аналіз і зробити на його основі висновки та надати певні обґрунтовані рекомендації.

Отже, здобувачам потрібно надати на перевірку:

- файл з кодом на мові R із розширенням «.R»;
- текстовий файл з описом виконаних дій та аналізом отриманих результатів. Особлива увага під час перевірки роботи звертається на аналітичну частину та висновки, які робить здобувач на основі отриманих візуалізацій, а також на пропозиціях, наданих умовному замовнику дослідження щодо вирішення певної маркетингової управлінської проблеми;
- скріншоти введеного здобувачем в консолі коду і отриманих результатів на кожному етапі виконання завдання;
- скріншоти отриманих візуалізацій або їх збережені зображення.

Відмітимо, що набори маркетингових даних досить часто мають багато змінних, кількість яких завжди вигідно скоротити для подальшого аналізу. Наприклад, під час проведення маркетингового дослідження можемо мати багато пунктів в опитуванні споживачів, які відображають значно меншу кількість основних питань, таких як задоволеність клієнта послугою або лідерство бренду в категорії.

Якщо звести дані до їх базових вимірів, ми можемо більш чітко визначити базові зв'язки між досліджуваними поняттями. У цьому й буде полягати серія наведених далі завдань.

Отже, задано набір даних із рейтинговими балами споживачів для брендів кави. Бренди кави є умовними, а набір даних тренувальний; ці дані не мають відношення до реальних брендів кави, представлених на ринку України.

Подібний набір даних є типовим для опитувань щодо сприйняття бренду споживачами. Дані відображають рейтинги брендів за споживчими вподобаннями за перцептивними прикметниками (прикметниками сприйняття), які зазначаються в опитуваннях у такій формі:

«За шкалою від 1 до 10, де 1 це найменше значення, а 10 – найбільше, оцініть, наскільки [ПРИКМЕТНИК] відповідає [БРЕНДУ «А»]?»

Отже, у заданому наборі даних містяться змодельовані рейтинги 10 умовних брендів (від «а» до «j») за 9 прикметниками для 100 умовних респондентів.

Під час обробки даних рекомендується використовувати латинські літери, оскільки текст, написаний українською мовою, не завжди буде сприйматися RStudio коректно, що, у свою чергу, може призвести до помилок, неправильних результатів або неправильної візуалізації, а, як наслідок, і до неправильних висновків і наданих маркетологом рекомендацій.

Приклад відображення результатів українською та англійською мовою наведено на рис. 4а і рис. 4б, відповідно.

Окрім того, щоб RStudio відображав інформацію про потенційні помилки англійською мовою, і легше було знайти відповіді на питання у мережі Інтернет, радимо починати обробку даних із введення коду:

```
Sys.setenv(LANG = "en")
```

Окрім того, виконайте перевірку простим кодом, наприклад:

```
1+x
```

```
Console Terminal x Jobs x
L:/MCR/MCR_1/
> brand.ratings <- read.csv("L:/MCR/rMCR1.csv")
> head (brand.ratings)
оперційна_перевага...лідер_ринку...сучасний...веселий...серйозний...вигідно_купувати...дорогий...трендовий...ку
плю_повторно...бренд.
1 2,4,8,8,2,9,7,4,6,"a"
2 1,1,4,7,1,1,1,2,2,"a"
3 2,3,5,9,2,9,5,1,6,"a"
4 1,6,10,8,3,4,5,2,1,"a"
5 1,1,5,8,1,9,9,1,1,"a"
6 2,8,9,5,3,8,7,1,2,"a"
> tail (brand.ratings)
```

Приклад 1.

Перш за все, необхідно зберегти набір даних на локальний диск комп'ютера. Шлях до файлу має бути прописаний латиницею та він не повинен бути занадто довгим.

Отже, для завантаження даних для роботи у середовище RStudio та їх перевірки використовуємо наступний код:

```
brand.ratings <- read.csv("L:/MCR/r8.csv")  
head (brand.ratings)  
tail (brand.ratings)
```

Нагадуємо, що шлях до файлу на комп'ютері має бути записаний у форматі: L:/MCR/r8.csv

Зверніть увагу на те, що використовуються саме символи «/», як було зазначено у попередніх розділах. Шлях, прописаний вами, може відрізнятися від наданого у прикладі.

Приклад 2.

Кожен зі 100 респондентів надав оцінки кожному із 10 брендів кави, тому загальна кількість рядків становить 1000. Виконуємо перевірку за допомогою `summary()` і `str()`, щоб перевірити якість і структуру даних. У результаті в консолі буде отримано результат, аналогічний до представленого на рис. 4.3.

```
summary (brand.ratings)  
str (brand.ratings)
```

На панелі зверху справа буде відображено інформацію, аналогічну представленій на рис. 4.4.

```

Console Terminal Jobs
L:/MCR/MCR_1/
>
> brand.ratings <- read.csv("L:/MCR/r8.csv")
> head (brand.ratings)
  perform leader latest fun serious bargain value trendy rebuy brand
1      2      4      8  8      2      9      7      4      6      a
2      1      1      4  7      1      1      1      2      2      a
3      2      3      5  9      2      9      5      1      6      a
4      1      6     10  8      3      4      5      2      1      a
5      1      1      5  8      1      9      9      1      1      a
6      2      8      9  5      3      8      7      1      2      a
> tail (brand.ratings)
  perform leader latest fun serious bargain value trendy rebuy brand
995      4      2      8  7      1      3      3      5      2      j
996      2      2      3  6      4      8      5      1      2      j
997      3      2      6  7      1      3      3      2      1      j
998      1      1     10 10      1      6      5      5      2      j
999      1      1      7  5      1      1      2      5      1      j
1000     7      4      7  8      4      1      2      5      1      j
> summary (brand.ratings)
      perform      leader      latest      fun      serious      bargain
Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
1st Qu.: 1.000   1st Qu.: 2.000   1st Qu.: 4.000   1st Qu.: 4.000   1st Qu.: 2.000   1st Qu.: 2.000
Median : 4.000   Median : 4.000   Median : 7.000   Median : 6.000   Median : 4.000   Median : 4.000
Mean   : 4.488   Mean   : 4.417   Mean   : 6.195   Mean   : 6.068   Mean   : 4.323   Mean   : 4.259
3rd Qu.: 7.000   3rd Qu.: 6.000   3rd Qu.: 9.000   3rd Qu.: 8.000   3rd Qu.: 6.000   3rd Qu.: 6.000
Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000
      value      trendy      rebuy      brand
Min.   : 1.000   Min.   : 1.00   Min.   : 1.000   Length:1000
1st Qu.: 2.000   1st Qu.: 3.00   1st Qu.: 1.000   Class :character
Median : 4.000   Median : 5.00   Median : 3.000   Mode  :character
Mean   : 4.337   Mean   : 5.22   Mean   : 3.727
3rd Qu.: 6.000   3rd Qu.: 7.00   3rd Qu.: 5.000
Max.   :10.000   Max.   :10.00   Max.   :10.000
> str (brand.ratings)
'data.frame': 1000 obs. of 10 variables:
 $ perform: int  2 1 2 1 1 2 1 2 2 3 ...
 $ leader : int  4 1 3 6 1 8 1 1 1 1 ...
 $ latest : int  8 4 5 10 5 9 5 7 8 9 ...
 $ fun    : int  8 7 9 8 8 5 7 5 10 8 ...
 $ serious: int  2 1 2 3 1 3 1 2 1 1 ...
 $ bargain: int  9 1 9 4 9 8 5 8 7 3 ...
 $ value  : int  7 1 5 5 9 7 1 7 7 3 ...
 $ trendy : int  4 2 1 2 1 1 1 7 5 4 ...
 $ rebuy  : int  6 2 6 1 1 2 1 1 1 1 ...
 $ brand  : chr  "a" "a" "a" "a" ...
>

```

Рис. 4.3 Результат, отриманий в RStudio під час виконання прикладу 2.

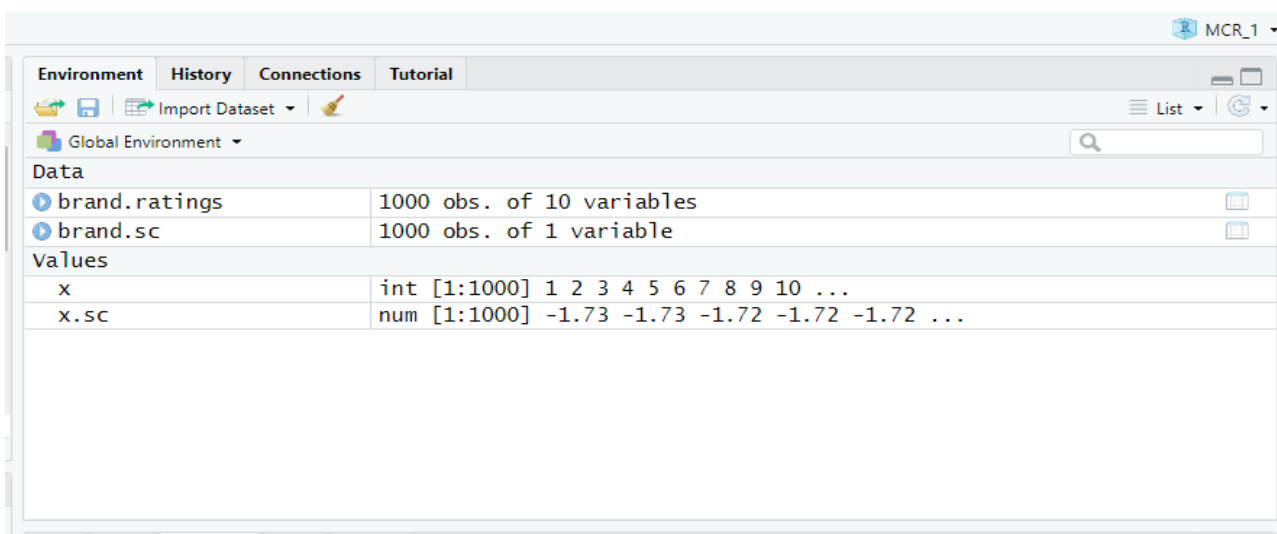


Рис. 4.4 Вкладка «Environment» у RStudio після виконання прикладу 2.

Приклад 3.

Приклад присвячено ремасштабуванню (rescaling) вихідних даних.

Зазначимо, що використовувані характеристики (прикметники) мають такі значення:

Perform – бренд має хороші операційні показники;

Leader – бренд є лідером у даній категорії;

Latest – бренд має продукцію, створену із використанням найновіших технологій, рішень і т.д.;

Fun – бренд сприймається як веселий, кмітливий і т.д.;

Serious – бренд сприймається як серйозний гравець на ринку;

Bargain – товари бренду вигідні для покупки;

Value – продукти бренду мають значну цінність;

Trendy – бренд модний;

Rebuy – споживач готовий повторно купувати продукцію бренду.

Ремасштабування (rescaling) вихідних даних досить часто є корисною практикою. Воно може бути досягнуто за рахунок центрування кожної змінної шляхом віднімання її середнього значення з кожного спостереження, а потім ремасштабування цих центрованих значень як одиниць середньоквадратичного відхилення. Це зазвичай називають **стандартизацією** або **нормалізацією даних**.

У RStudio дані можна стандартизувати таким чином за допомогою математичних виразів із використанням `mean()` і `sd()`:

```
x <- 1:1000  
x.sc <- (x - mean(x)) / sd(x)  
summary(x.sc)
```

Результат виконання наведеного вище коду представлено на рис. 4.5.

```

Console Terminal x Jobs x
L:/MCR/MCR_1/ ↗
Mean : 4.488 Mean : 4.417 Mean : 6.195 Mean : 6.068 Mean : 4.323 Mean : 4.259
3rd Qu.: 7.000 3rd Qu.: 6.000 3rd Qu.: 9.000 3rd Qu.: 8.000 3rd Qu.: 6.000 3rd Qu.: 6.000
Max. :10.000 Max. :10.000 Max. :10.000 Max. :10.000 Max. :10.000 Max. :10.000
value trendy rebuy brand
Min. : 1.000 Min. : 1.00 Min. : 1.000 Length:1000
1st Qu.: 2.000 1st Qu.: 3.00 1st Qu.: 1.000 Class:character
Median : 4.000 Median : 5.00 Median : 3.000 Mode :character
Mean : 4.337 Mean : 5.22 Mean : 3.727
3rd Qu.: 6.000 3rd Qu.: 7.00 3rd Qu.: 5.000
Max. :10.000 Max. :10.00 Max. :10.000
> str (brand.ratings)
'data.frame': 1000 obs. of 10 variables:
 $ perform: int 2 1 2 1 1 2 1 2 2 3 ...
 $ leader : int 4 1 3 6 1 8 1 1 1 1 ...
 $ latest : int 8 4 5 10 5 9 5 7 8 9 ...
 $ fun : int 8 7 9 8 8 5 7 5 10 8 ...
 $ serious: int 2 1 2 3 1 3 1 2 1 1 ...
 $ bargain: int 9 1 9 4 9 8 5 8 7 3 ...
 $ value : int 7 1 5 5 9 7 1 7 7 3 ...
 $ trendy : int 4 2 1 2 1 1 1 7 5 4 ...
 $ rebuy : int 6 2 6 1 1 2 1 1 1 1 ...
 $ brand : chr "a" "a" "a" "a" ...
> x <- 1:1000
> x.sc <- (x - mean(x)) / sd(x)
> summary(x.sc)
 Min. 1st Qu. Median Mean 3rd Qu. Max.
-1.7295 -0.8647 0.0000 0.0000 0.8647 1.7295
>

```

Рис. 4.5 Стандартизація даних у RStudio.

Проте більш простим способом є використання `scale()` для ремасштабування всіх змінних одночасно. Щоб не змінювати (і потенційно випадково не зіпсувати) вихідні дані, спочатку призначаємо оригінальні значення новому дата-фрейму `brand.sc` (рис. 4.6):

```

brand.sc <- brand.ratings
brand.sc[, 1:9] <- data.frame(scale(brand.ratings[, 1:9]))
summary(brand.sc)

```

```

Console Terminal x Jobs x
L:/MCR/MCR_1/ ↗
> brand.sc <- brand.ratings
> brand.sc[, 1:9] <- data.frame(scale(brand.ratings[, 1:9]))
> summary(brand.sc)
perform leader latest fun serious
Min. : -1.0888 Min. : -1.3100 Min. : -1.6878 Min. : -1.84677 Min. : -1.1961
1st Qu.: -1.0888 1st Qu.: -0.9266 1st Qu.: -0.7131 1st Qu.: -0.75358 1st Qu.: -0.8362
Median : -0.1523 Median : -0.1599 Median : 0.2615 Median : -0.02478 Median : -0.1163
Mean : 0.0000 Mean : 0.0000 Mean : 0.0000 Mean : 0.00000 Mean : 0.0000
3rd Qu.: 0.7842 3rd Qu.: 0.6069 3rd Qu.: 0.9113 3rd Qu.: 0.70402 3rd Qu.: 0.6036
Max. : 1.7206 Max. : 2.1404 Max. : 1.2362 Max. : 1.43281 Max. : 2.0434
bargain value trendy rebuy brand
Min. : -1.22196 Min. : -1.3912 Min. : -1.53897 Min. : -1.0717 Length:1000
1st Qu.: -0.84701 1st Qu.: -0.9743 1st Qu.: -0.80960 1st Qu.: -1.0717 Class:character
Median : -0.09711 Median : -0.1405 Median : -0.08023 Median : -0.2857 Mode :character
Mean : 0.00000 Mean : 0.0000 Mean : 0.00000 Mean : 0.0000
3rd Qu.: 0.65279 3rd Qu.: 0.6933 3rd Qu.: 0.64914 3rd Qu.: 0.5003
Max. : 2.15258 Max. : 2.3610 Max. : 1.74319 Max. : 2.4652
>

```

Рис. 4.6 Ремасштабування всіх змінних одночасно.

Приклад 4.

Використовуємо бібліотеку `corrplot()` для початкової перевірки зв'язків між змінними:

```
library(corrplot)
corrplot(cor(brand.sc[, 1:9]), order="hclust")
```

У результаті у секції графіків (Plots) буде відображено візуалізацію, представлену на рис. 4.7.

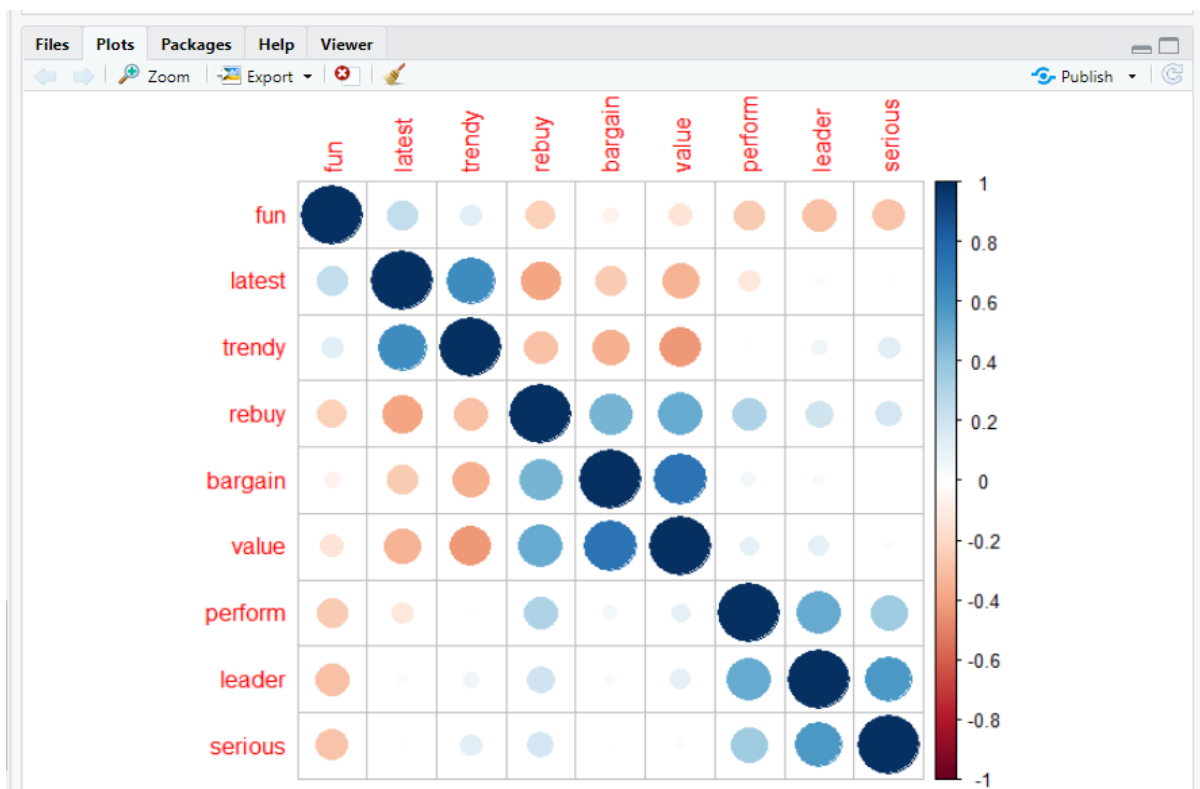


Рис. 4.7 Візуалізація зв'язків між змінними за допомогою бібліотеки `corrplot()`.

На рис. 4.7 видно, що рейтинги групуються в три кластери подібних змінних. Здобувачам-маркетологам необхідно проаналізувати отриману візуалізацію і з'ясувати:

- за якими ознаками відбувається групування,
- чому, на їхню думку, групування відбулося саме за цими ознаками,
- яким чином отримані результати можуть бути використані замовником маркетингового дослідження для формування позиціонування власного бренду на ринку тощо.

Загальний вигляд середовища RStudio після виконання даного прикладу представлено на рис. 4.8.

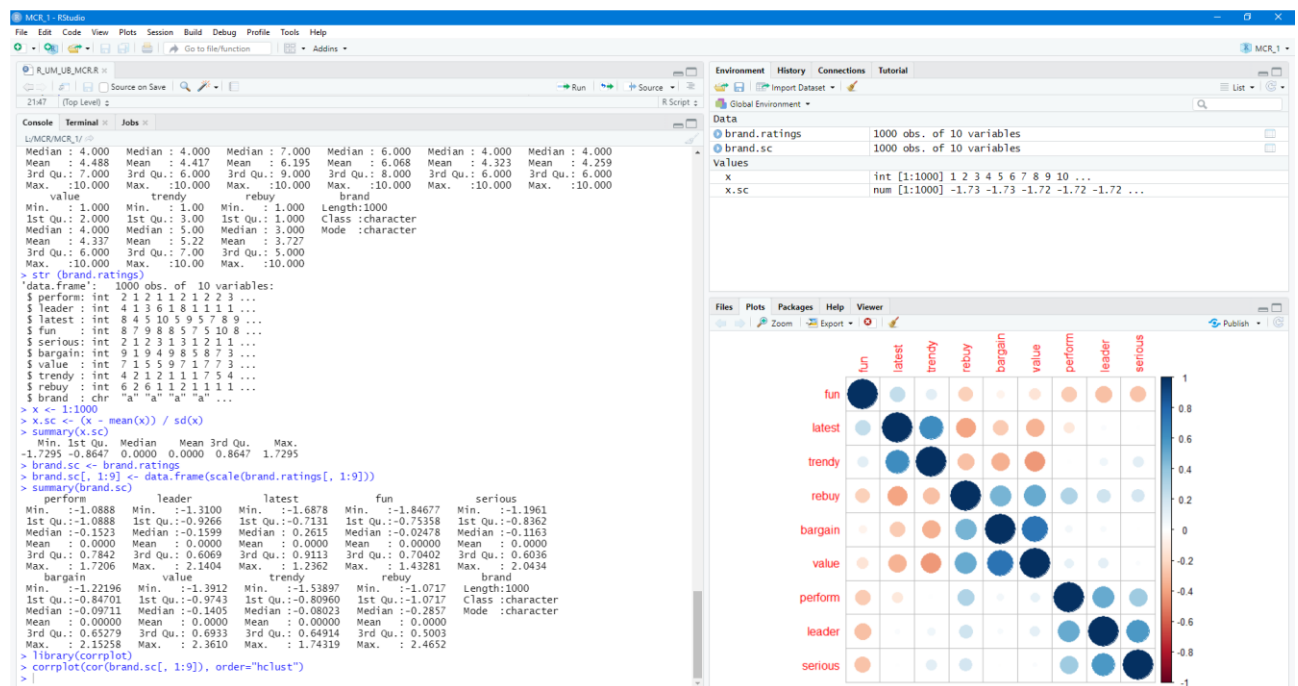


Рис. 4.8 Результат виконання прикладу №4 в RStudio.

Приклад 5.

У цьому завданні необхідно відповісти на питання, яка саме буде середня позиція бренду за кожною із аналізованих характеристик.

Для цього використовуємо `aggregate()` і знаходимо середнє значення кожної змінної за брендом (рис. 4.9):

```
brand.mean <- aggregate(. ~ brand, data=brand.sc, mean)
brand.mean
```

```
Console Terminal Jobs
L:/MCR/MCR_1/
> brand.mean <- aggregate(. ~ brand, data=brand.sc, mean)
> brand.mean
  brand perform leader latest fun serious bargain value trendy
1 a -0.88591874 -0.5279035 0.4109732 0.6566458 -0.91894067 0.21409609 0.18469264 -0.52514473
2 b 0.93087022 1.0707584 0.7261069 -0.9722147 1.18314061 0.04161938 0.15133957 0.74030819
3 c 0.64992347 1.1627677 -0.1023372 -0.8446753 1.22273461 -0.60704302 -0.44067747 0.02552787
4 d -0.67989112 -0.5930767 0.3524948 0.1865719 -0.69217505 -0.88075605 -0.93263529 0.73666135
5 e -0.56439079 0.1928362 0.4564564 0.2958914 0.04211361 0.55155051 0.41816415 0.13857986
6 f -0.05868665 0.2695106 -1.2621589 -0.2179102 0.58923066 0.87400696 1.02268859 -0.81324496
7 g 0.91838369 -0.1675336 -1.2849005 -0.5167168 -0.53379906 0.89650392 1.25616009 -1.27639344
8 h -0.01498383 -0.2978802 0.5019396 0.7149495 -0.14145855 -0.73827529 -0.78254646 0.86430070
9 i 0.33463879 -0.3208825 0.3557436 0.4124989 -0.14865746 -0.25459062 -0.80339213 0.59078782
10 j -0.62994504 -0.7885965 -0.1543180 0.2849595 -0.60218870 -0.09711188 -0.07379367 -0.48138267
rebuy
1 -0.59616642
2 0.23697320
3 -0.13243776
4 -0.49398892
5 0.03654811
6 1.35699580
7 1.36092571
8 -0.60402622
9 -0.20317603
10 -0.96164748
>
```

Рис. 4.9 Визначення середнього значення кожної змінної за брендом.

Далі, перш ніж продовжити, виконаємо певну обробку даних, використовуючи `brand.mean`. При цьому називаємо рядки «мітками» бренду, які `aggregate()` поміщає в стовпець `brand`, а потім видаляємо цей стовпець як зайвий.

Отриману матрицю тепер чітко відформатовано з брендами за рядками і значеннями характеристик за стовпцями (рис. 4.10):

```
rownames(brand.mean) <- brand.mean[, 1]
brand.mean <- brand.mean[, -1]
brand.mean
```

```

Console Terminal Jobs
L:/MCR/MCR_1/
1 -0.59616642
2 0.23697320
3 -0.13243776
4 -0.49398892
5 0.03654811
6 1.35699580
7 1.36092571
8 -0.60402622
9 -0.20317603
10 -0.96164748
> rownames(brand.mean) <- brand.mean[, 1] # використовуємо бренд для призначення назв рядків
> brand.mean <- brand.mean[, -1] # видаляємо стовпець з назвами брендів
> brand.mean
  perform leader latest fun serious bargain value trendy rebuy
a -0.88591874 -0.5279035 0.4109732 0.6566458 -0.91894067 0.21409609 0.18469264 -0.52514473 -0.59616642
b 0.93087022 1.0707584 0.7261069 -0.9722147 1.18314061 0.04161938 0.15133957 0.74030819 0.23697320
c 0.64992347 1.1627677 -0.1023372 -0.8446753 1.22273461 -0.60704302 -0.44067747 0.02552787 -0.13243776
d -0.67989112 -0.5930767 0.3524948 0.1865719 -0.69217505 -0.88075605 -0.93263529 0.73666135 -0.49398892
e -0.56439079 0.1928362 0.4564564 0.2958914 0.04211361 0.55155051 0.41816415 0.13857986 0.03654811
f -0.05868665 0.2695106 -1.2621589 -0.2179102 0.58923066 0.87400696 1.02268859 -0.81324496 1.35699580
g 0.91838369 -0.1675336 -1.2849005 -0.5167168 -0.53379906 0.89650392 1.25616009 -1.27639344 1.36092571
h -0.01498383 -0.2978802 0.5019396 0.7149495 -0.14145855 -0.73827529 -0.78254646 0.86430070 -0.60402622
i 0.33463879 -0.3208825 0.3557436 0.4124989 -0.14865746 -0.25459062 -0.80339213 0.59078782 -0.20317603
j -0.62994504 -0.7885965 -0.1543180 0.2849595 -0.60218870 -0.09711188 -0.07379367 -0.48138267 -0.96164748
>

```

Рис. 4.10 Подальша обробка даних при виконанні прикладу №5.

Приклад 6.

На завершення побудуємо «теплову карту». Теплова карта (heatmap) є зручним способом репрезентації результатів, оскільки вона «розфарбовує» точки даних за інтенсивністю їх значень. Задля цього використовуємо `heatmap.2()` з пакету `gplots` із кольорами з пакету `RColorBrewer`:

```

library(gplots)
library(RColorBrewer)
heatmap.2(as.matrix(brand.mean),
  + col=brewer.pal(9, "GnBu"), trace="none", key=FALSE, dend="none",
  + main="\n\n\n\n\nBrand attributes")

```

Звернемо увагу на останні три рядки коду. Код може не працювати у консолі, якщо запустити його весь одразу. Отже, якщо ви працюєте саме із консоллю, то краще кожний рядок запускати послідовно. Символ «+» при цьому буде з'являтися у консолі автоматично.

Нагадаємо, що зазначені вище пакети необхідно додатково встановити. Для цього необхідно перейти на вкладнику пакетів у правій нижній частині

екрану, як показано на рис. 4.11, а далі натиснути на **Install**, ввести назву пакету і слідувати інструкціям на екрані (рис. 4.12).

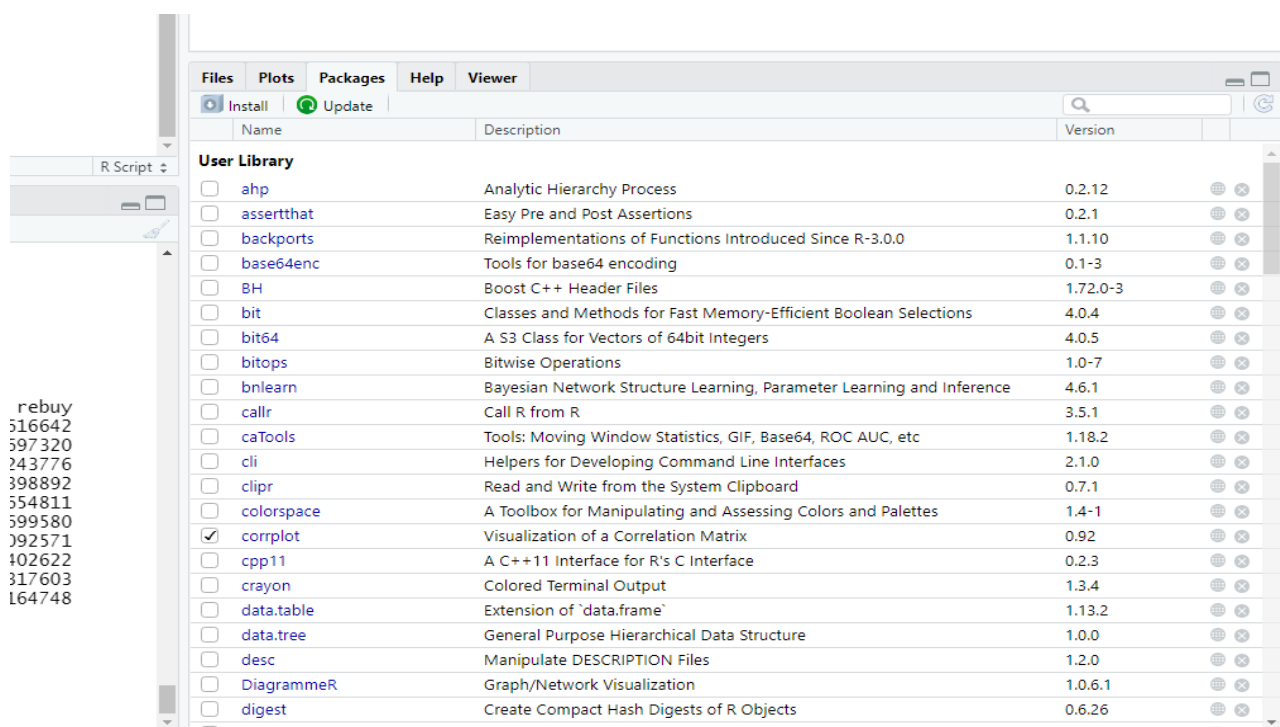


Рис. 4.11 Перший етап встановлення нового пакету в RStudio.

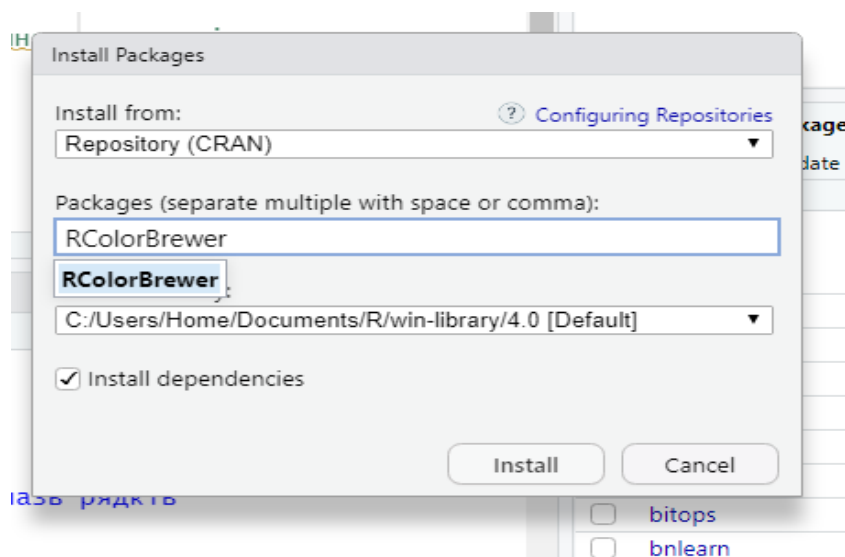


Рис. 4.12 Другий етап встановлення нового пакету в RStudio на прикладі пакету **RColorBrewer**, застосованого у прикладі 6.

У результаті виконання прикладу 6 у консолі буде отримано результат, представлений на рис. 4.13, тоді як результат із секції «Environment» представлений на рис. 4.14.

```

6 1.35699580
7 1.36092571
8 -0.60402622
9 -0.20317603
10 -0.96164748
> rownames(brand.mean) <- brand.mean[, 1] # використовуємо бренд для призначення назв рядків
> brand.mean <- brand.mean[, -1] # видаляємо стовпець з назвами брендів
> brand.mean
      perform      leader      latest      fun      serious      bargain      value      trendy      rebuy
a -0.88591874 -0.5279035  0.4109732  0.6566458 -0.91894067  0.21409609  0.18469264 -0.52514473 -0.59616642
b  0.93087022  1.0707584  0.7261069 -0.9722147  1.18314061  0.04161938  0.15133957  0.74030819  0.23697320
c  0.64992347  1.1627677 -0.1023372 -0.8446753  1.22273461 -0.60704302 -0.44067747  0.02552787 -0.13243776
d -0.67989112 -0.5930767  0.3524948  0.1865719 -0.69217505 -0.88075605 -0.93263529  0.73666135 -0.49398892
e -0.56439079  0.1928362  0.4564564  0.2958914  0.04211361  0.55155051  0.41816415  0.13857986  0.03654811
f -0.05868665  0.2695106 -1.2621589 -0.2179102  0.58923066  0.87400696  1.02268859 -0.81324496  1.35699580
g  0.91838369 -0.1675336 -1.2849005 -0.5167168 -0.53379906  0.89650392  1.25616009 -1.27639344  1.36092571
h -0.01498383 -0.2978802  0.5019396  0.7149495 -0.14145855 -0.73827529 -0.78254646  0.86430070 -0.60402622
i  0.33463879 -0.3208825  0.3557436  0.4124989 -0.14865746 -0.25459062 -0.80339213  0.59078782 -0.20317603
j -0.62994504 -0.7885965 -0.1543180  0.2849595 -0.60218870 -0.09711188 -0.07379367 -0.48138267 -0.96164748
> library(gplots)
> library(RColorBrewer)
> heatmap.2(as.matrix(brand.mean),
+ col=brewer.pal(9, "GnBu"), trace="none", key=FALSE, dend="none",
+ main="\n\n\n\nBrand attributes")
>

```

Рис. 4.13 Результат виконання прикладу №6 у консолі RStudio.

Global Environment	
Data	
brand.mean	10 obs. of 9 variables
brand.ratings	1000 obs. of 10 variables
brand.sc	1000 obs. of 10 variables
Values	
x	int [1:1000] 1 2 3 4 5 6 7 8 9 10 ...
x.sc	num [1:1000] -1.73 -1.73 -1.72 -1.72 -1.72 ...

Рис. 4.14 Результат виконання прикладу №6 у секції «Environment» в RStudio.

Відповідно, у секції «Plots» буде отримано результат, представлений на рис. 4.15. Цю візуалізацію здобувачам-маркетологам необхідно буде проаналізувати і надати відповіді на питання маркетингового дослідження, аналогічно до попередніх прикладів.

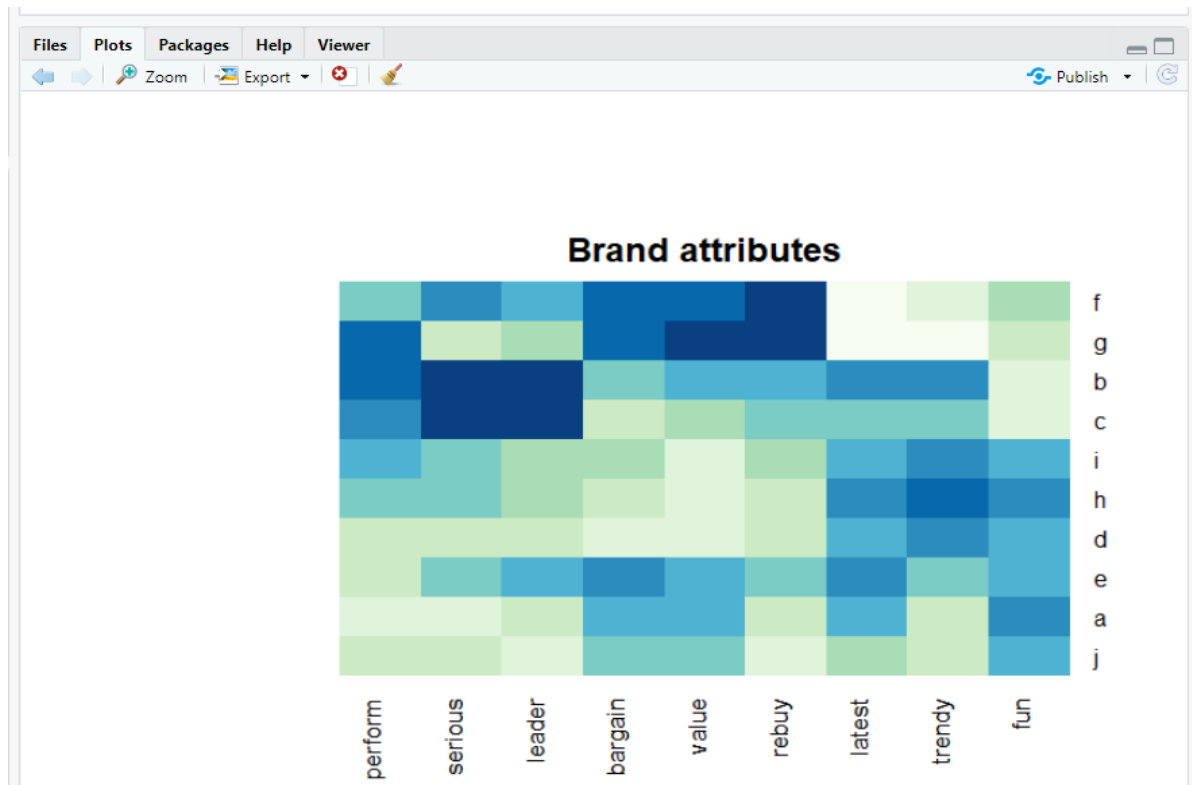


Рис. 4.15 Візуалізація, тримана у результаті виконання прикладу 6.

При цьому головним питанням є те, яким чином отримані результати можуть бути використані замовником маркетингового дослідження для вирішення маркетингової управлінської проблеми, визначеної на початкових етапах проведення дослідження.

Під час виконання роботи здобувачам рекомендується ознайомитися з офіційною документацією щодо використаних пакетів у довідці RStudio і дещо змінити код, щоб отримати унікальні візуалізації чи додаткові результати унаслідок переосмислення представлених у даному розділі прикладів.

Зокрема, зразок результату-візуалізації, отриманий під час виконання одного з прикладів здобувачем-маркетологом, представлено на рис. 4.16. У зручному для використання форматі програмний код наведено у Додатку Г.

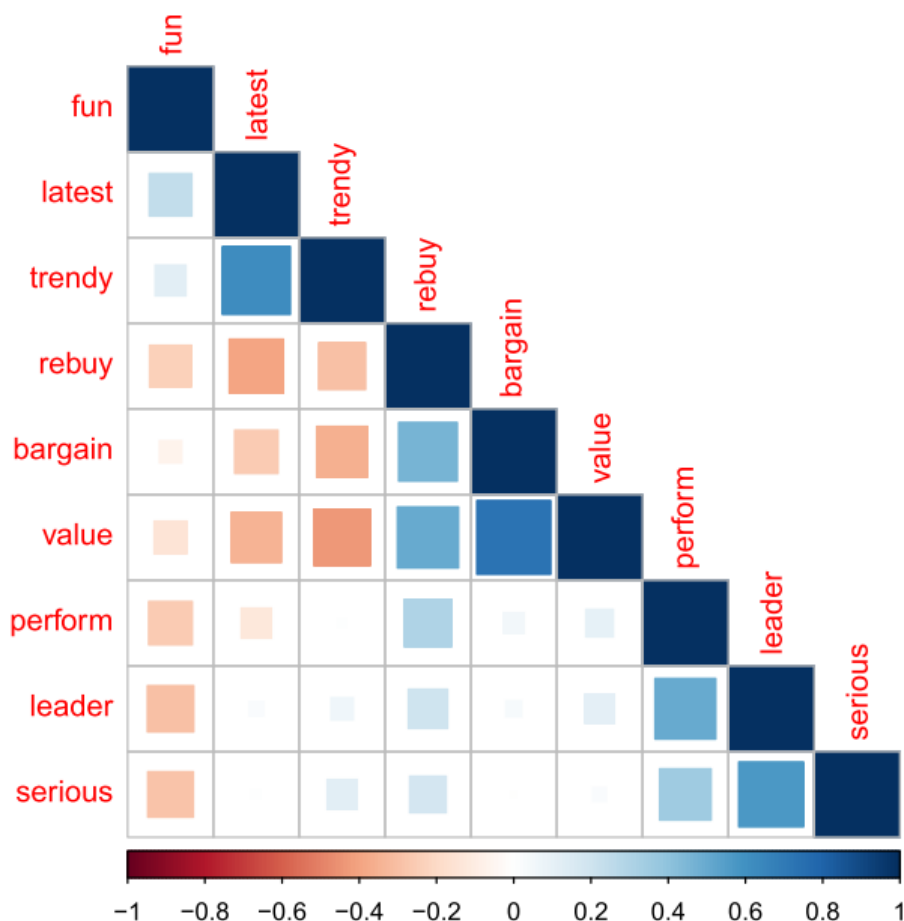


Рис. 4.16 Зразок результату-візуалізації, отриманий здобувачем-маркетологом.

Таким чином, розглянуті приклади завдань і кодів на мові програмування R надають здобувачам-маркетологам можливість на основі даних маркетингового дослідження щодо з'ясування споживчих мотивацій на певному ринку провести аналіз великої кількості первинних даних і на їх основі надати обґрунтовані рекомендації щодо маркетингових стратегій підприємства-замовника і тим самим підвищити його репутацію, рейтингову оцінку на ринку та конкурентоспроможність.

Приклад 7.

Здобувачам пропонується вирішити завдання, у якому необхідно не лише правильно написати програмний код і здійснити розрахунки, але і продумати, яким чином за допомогою запропонованої методики може бути вирішена конкретна проблема інноваційного бізнесу.

Однією із найбільш популярних форм інноваційного підприємництва сьогодні є стартап-проект. Проте зважаючи на високу ризикованість реалізації стартапів актуалізується потреба визначення доцільності їх реалізації та фінансування, починаючи з етапу, коли у проекті є лише гіпотеза, яка вже пройшла попереднє тестування із використанням інструменту Business Model Canvas (Додаток Г).

Для подальшої оцінки перспективності стартап-проекту доцільно застосувати методику на основі методу аналізу ієрархій Т. Сааті. Для цього перш за все здобувачам пропонується виокремити комплекс критеріїв для оцінки перспективності конкретного проекту, який функціонує на обраному ними ринку. Після цього здобувачі повинні сформулювати із виокремленого ними комплексу критеріїв ієрархію, кінцевим висновком якої є вибір на основі експертних думок однієї з двох альтернатив «Розпочати реалізацію стартап-проекту» або «Відмовити у реалізації проекту». Надалі здобувачі визначають групи експертів (зазвичай дві групи), які будуть проводити оцінку стартапу за запропонованими критеріями.

Інформація щодо доцільності реалізації проекту є необхідною інвесторам, засновникам стартапів, а також іншим стейкхолдерам. Тому це тренувальне завдання максимально наближене до реальної бізнес-ситуації.

Аналіз проводиться у програмному комплексі R Studio на базі пакету *ahr*, який надає можливість створювати програмні компоненти на основі методу аналізу ієрархій Т. Сааті. Інструмент *ahr* це не єдиний пакет (бібліотека), який може бути застосований у RStudio для реалізації методу Т. Сааті. Проте, як показує практика, саме *ahr* є найлегшим у використанні і надає змогу отримати в результаті зручні візуалізації у графічній і табличній формах.

Приклад результату, одержаного здобувачем під час виконання даного завдання, наведено на рис. 4.17.

	Weight	Відмовити у реалізації проєкту	Розпочати реалізацію проєкту
Встановити доцільність реалізації проєкту відкриття коворкінгу	100.0%	57.2%	42.8%
Загальна привабливість	66.8%	37.5%	29.3%
Місцезнаходження	27.8%	10.4%	17.3%
Вартість споруди	24.2%	21.1%	3.1%
Орендна плата	11.4%	3.0%	8.3%
Можливість перепродажу	3.4%	2.9%	0.5%
Маркетингова привабливість	20.8%	17.2%	3.6%
Мікросередовище	14.7%	12.8%	2.0%
Конкуренти	8.6%	7.3%	1.3%
Клієнти	6.1%	5.5%	0.6%
Макросередовище	6.1%	4.4%	1.7%
Демографічне становище	5.4%	4.3%	1.1%
Економічне становище	0.6%	0.1%	0.5%
Інвестиційна привабливість	12.4%	2.5%	10.0%
Загальна інвестиційна привабливість	9.7%	1.6%	8.1%
Термін окупності	7.0%	1.1%	5.9%
IRR	1.8%	0.4%	1.5%
NPV	0.9%	0.1%	0.7%
Привабливість для вітчизняного капіталу	2.0%	0.3%	1.7%
Політична ситуація	1.2%	0.2%	0.9%
Річний дохід	0.8%	0.1%	0.7%
Привабливість для іноземного капіталу	0.7%	0.6%	0.2%
Курс валюти	0.6%	0.5%	0.1%
Податкова політика	0.1%	0.0%	0.1%

Рис. 4.17 Зразок результату, отриманого здобувачем під час виконання прикладу 7.

Зразок візуалізованої здобувачем ієрархії, а також написаного ним програмного коду наведено у додатках Д та Е.

Окрім того, в реальних умовах рішення зазвичай приймаються на основі думок кількох груп експертів. У такому разі результати доцільно представляти не тільки у загальному вигляді, а й по кожній групі експертів окремо, як показано в Додатку Ж.

Так, під час використання на практиці авторської методики оцінювання перспективності інноваційних проєктів було отримано результати для 5-ти груп експертів, а саме:

- фахівців-науковців (група № 1);
- представників інвесторів (група № 2);
- виробників (група № 3);
- практиків стартап-підприємництва (група № 4) та
- представників бізнес-інкубаторів і стартап-акселераторів (група № 5).

ВИКОРИСТАНІ ЛІТЕРАТУРНІ ДЖЕРЕЛА

1. Василенко О. А., Сенча І. А. Математично-статистичні методи аналізу у прикладних дослідженнях: навч. посіб. Одеса: ОНАЗ ім. О. С. Попова, 2011. 166 с.
2. Грицевич В. С. Кореляційний та регресійний аналіз в суспільній географії: тексти лекцій. Львів : Малий видавничий центр. Лабораторія тематичного картографування географічного факультету, 2016. 24 с.
3. Грицевич В. С. Статистичні ознаки та характеристики їхньої центральної тенденції: тексти лекцій. Львів : Видавничий центр ЛНУ імені Івана Франка, 2008. 52 с.
4. Захожай В. Б., Попов І. І., Коваленко О. В. Практикум з основ статистики. К. : МАУП, 2001. 172 с.
5. Кофанова О. В., Перельот Т. М. Основи статистичного обліку в екології. Модуль 1: Елементи теорії імовірностей та математичної статистики: Електронне навчальне видання. 2010. URL: <http://ela.kpi.ua/handle/123456789/1850>.
6. Кофанова О. В. Кофанов О. Є. Методичні вказівки для самостійної роботи з дисципліни "Кількісний та якісний аналіз стану екологічних систем". Обробка результатів експериментальних досліджень": Електронне навчальне видання. 2017. URL: <http://ela.kpi.ua/handle/123456789/24574>
7. Лапач С. Н., Чубенко А. В., Бабич П. Н. Статистика в науці і бізнесі. К.: МОРИОН, 2002. 640 с.
8. Математичні обчислення засобами пакету R-програмування: навчально-методичний посібник для студентів всіх спеціальностей / укл.: В. В. Кальченко, В. П. Мурашківська, Ю. М. Ткач Чернігів: ЧНТУ, 2017, 86 с.
9. Матковський С. О., Марець О. Р. Теорія статистики: навч. посіб. К.: Знання, 2009. 534 с.

- 10.Тарасова В. В. Екологічна статистика (з блочно-модульною формою контролю знань): підручник. К. : Центр учбової літератури, 2008. 392 с.
- 11.Турчин В. М. Теорія ймовірностей і математична статистика. Основні поняття, приклади, задачі: підручник. Дніпропетровськ : ІМА-прес, 2014. 555 с.
- 12.Швець В. Т. Теорія ймовірностей і математична статистика: навч. посіб. Одеса : ВМБ, 2014. 200 с.
- 13.Analytic Hierarchy Process ahp. The Comprehensive R Archive Network. 2016. URL: <https://cran.microsoft.com/snapshot/2016-08-05/web/packages/ahp/index.html>
- 14.Blank S., Dorf B. The startup owner's manual: The step-by-step guide for building a great company. Pescadero : K&S Ranch, 2012. 571 p.
- 15.Chapman C., McDonnell Feit E. R For Marketing Research and Analytics. Philadelphia : Springer Cham, 2019. 487 p.
- 16.Chen Y., Tsai C., Liu H. Applying the AHP Model to Explore Key Success Factors for High-Tech Startups Entering International Markets. *International Journal of E-Adoption (IJEА)*, 2019, no. 11(1), pp. 45-63. doi: 10.4018/IJEА.2019010104
- 17.Chhibber M. Startup Profit Predictor Using Machine Learning Techniques. In: Dutta, P., Chakrabarti, S., Bhattacharya, A., Dutta, S., Shahnaz, C. (eds) *Emerging Technologies in Data Mining and Information Security. Lecture Notes in Networks and Systems*, 2023, vol. 490. Singapore, Springer. doi: 10.1007/978-981-19-4052-1_6
- 18.Data Publisher for Earth & Environmental Science – Pangaea [Електронний ресурс]. URL : <https://www.pangaea.de>.
- 19.EarthChem [Електронний ресурс]. URL: <http://www.earthchem.org>.
- 20.Global Biodiversity Information Facility [Електронний ресурс]. URL: <https://www.gbif.org>.
- 21.Kofanov O., Zozul'ov O. Successful development of startups as a global trend of innovative socio-economic transformations. *International and*

- Multidisciplinary Journal of Social Sciences*, 2018, no. 7 (2), pp. 191-217.
doi: 10.17583/rimcis.2018.3576
22. National Centers for Environmental Information (NCEI) [Электронный ресурс]. URL: <https://www.ncei.noaa.gov>.
23. NCBI Taxonomy [Электронный ресурс]. URL: <https://www.ncbi.nlm.nih.gov/taxonomy>.
24. Ofungwu J. Statistical Applications for Environmental Analysis and Risk Assessment (Statistics in Practice). Hackettstown: John Wiley & Sons, Inc., 2014. 648 с.
25. ORNL DAAC for Biogeochemical Dynamics [Электронный ресурс]. URL: <https://daac.ornl.gov>.
26. Ries E. The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses. New York : Crown Business, 2011. 296 p.
27. Romans A. The entrepreneurial bible to venture capital: Inside secrets from the leaders in the startup game. New York, NY : McGraw-Hill Education. 2013. 233 p.
28. Saaty T. L., Vargas L. G. Decision making with the analytic network process. Economic, political, social and technological applications with benefits, opportunities, costs and risks. New York : Springer, 2013. 363 p. doi: 10.1007/978-1-4614-7279-7
29. Sarkar D. Lattice multivariate data visualization with R. New York : Springer-Verlag New York, 2008. 268 p. DOI: 10.1007/978-0-387-75969-2.
30. Schuenemeyer J. H., Drew L. J. Drew Statistics for Earth and Environmental Scientists. Hoboken : John Wiley & Sons, Inc., 2011. 384 с.
31. Scutari M., Denis J.-B. Bayesian networks with examples in R. Boca Raton : CRC Press Taylor & Francis Group, 2015. 221 p.
32. Software project risk analysis using Bayesian networks with causality constraints / Yong Hu et al. *Decision Support Systems*. 2013. № 56. P. 439–449. DOI: 10.1016/j.dss.2012.11.001.

33. The Knowledge Network for Biocomplexity (KNB) [Электронный ресурс].
URL: <https://knb.ecoinformatics.org>.
34. The power of reciprocal knowledge sharing relationships for startup success /
Allen T. J. et al. *Journal of Small Business and Enterprise Development*. 2016.
№ 23 (3). P. 636–651. DOI: 10.1108/JSBED-08-2015-0110.
35. The rise and fall of startups : Creation and destruction of revenue and jobs by
young companies / Davila A. et al. *Australian Journal of Management*. 2015.
№ 40 (1). P. 6–35. DOI: 10.1177/0312896214525793.
36. Thiel P., Masters B. Zero to one : notes on startups, or how to build the future.
New York, NY : Crown Business, 2014. 210 p.

ДОДАТКИ

Таблиця А.1 Значення критерію Q для встановлення
однорідності вибірки (вилучення грубих помилок)¹

Кількість випробувань (дослідів, спостережень тощо), n	Q-критерій при $p = 0,90$	Q-критерій при $p = 0,95$	Q-критерій при $p = 0,99$
3	0,89	0,94	0,99
4	0,68	0,77	0,89
5	0,56	0,64	0,76
6	0,48	0,56	0,70
7	0,43	0,51	0,64
8	0,48	0,55	0,68
9	0,44	0,51	0,64
10	0,41	0,48	0,60

Примітка: ¹Для вибірок, в яких більше ніж 26 елементів, критичні значення критерію Q приймаються рівними 8 при $p = 0,95$ і 10 при $p = 0,99$.

ДОДАТОК Б

Таблиця Б.1 Розподіл Стюдента $t_{\alpha,n}$ для різних значень довірчої імовірності p та числа вимірів n

n	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99
2	1.000	1,376	1,963	3,08	6,31	12,71	31,8	63,7
3	0,816	1,061	1,336	1,886	2,92	4,30	6,96	9,92
4	0,765	0,978	1,250	1,638	2,35	3,18	4,54	5,84
5	0,741	0,941	1,190	1,533	2,13	2,77	3,75	4,60
6	0,727	0,920	1,156	1,476	2,02	2,57	3,36	4,03
7	0,718	0,906	1,134	1,440	1,943	2,45	3,14	4,71
8	0,711	0,896	1,119	1,415	1,895	2,36	3,00	3,50
9	0,706	0,889	1,108	1,397	1,860	2,31	2,90	3,36
10	0,703	0,883	1,110	1,383	1,833	2,26	2,82	3,25
11	0,700	0,879	1,093	1,372	1,812	2,23	2,76	3,17
12	0,697	0,876	1,088	1,363	1,796	2,20	2,72	3,11
13	0,695	0,873	1,083	1,356	1,782	2,18	2,68	3,06
14	0,694	0,870	1,079	1,350	1,771	2,16	2,65	3,01
15	0,692	0,868	1,076	1,345	1,761	2,14	2,62	2,98
16	0,691	0,866	1,074	1,341	1,753	2,13	2,60	2,95
17	0,690	0,865	1,071	1,337	1,746	2,12	2,58	2,92
18	0,689	0,863	1,069	1,333	1,740	2,11	2,57	2,90
19	0,688	0,862	1,067	1,330	1,734	2,10	2,55	2,88
20	0,688	0,861	1,066	1,328	1,729	2,09	2,54	2,86
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,33	2,58

ДОДАТОК В

Критичні значення F-критерію (при $\alpha = 0,05$)

Ступені свободи, k_1	Ступені свободи, k_2						
	1	2	3	4	5	6	50
9	5,12	4,26	3,86	3,63	3,48	3,37	2,8
10	4,96	4,10	3,71	3,48	3,33	3,22	-
11	4,84	3,98	3,59	3,36	3,20	3,09	2,5
12	4,75	3,88	3,49	3,26	3,11	3,00	-
13	4,67	3,80	3,41	3,18	3,02	2,92	2,32
14	4,60	3,74	3,34	3,11	2,96	2,85	-
15	4,54	3,68	3,29	3,06	2,90	2,79	2,18
16	4,49	3,63	3,24	3,01	2,85	2,47	-
17	4,45	3,59	3,20	2,96	2,81	2,70	2,08
18	4,41	3,56	3,16	2,93	2,77	2,66	-
19	4,33	3,52	3,13	2,90	2,74	2,63	2,08
20	4,35	3,49	3,10	2,87	2,71	2,90	-
21	4,32	3,47	3,07	2,84	2,68	2,57	1,93
22	4,30	3,44	3,05	2,82	2,66	2,55	-
23	4,28	3,42	3,03	2,80	2,64	2,53	1,88
24	4,26	3,40	3,01	2,78	2,62	2,51	-
25	4,24	3,38	2,99	2,76	2,60	2,49	1,84
26	4,22	3,37	2,98	2,74	2,59	2,47	-
27	4,21	3,35	2,96	2,73	2,57	2,46	1,80
28	4,20	3,34	2,95	2,71	2,56	2,44	-
29	4,18	3,38	2,93	2,70	2,54	2,43	1,77
30	4,17	3,32	2,92	2,69	2,53	2,42	-
35	4,11	3,26	2,86	2,63	2,48	2,36	-
40	4,08	3,23	2,84	2,61	2,45	2,34	-
50	4,03	3,18	2,79	2,56	2,40	2,29	-
60	4,00	3,15	2,76	2,52	2,37	2,25	-
∞	3,84	2,99	2,60	2,37	2,21	2,09	-

**ЗРАЗКИ ПРОГРАМНИХ КОДІВ ТА ДОДАТКОВІ МАТЕРІАЛИ
ДЛЯ ВИКОНАННЯ ЗАВДАНЬ**

**ЗРАЗКИ ПРОГРАМНИХ КОДІВ ДЛЯ ВИКОНАННЯ ПРИКЛАДІВ 1–6,
РОЗДІЛ 4**

```
Sys.setenv(LANG = "en")
```

```
1+x
```

```
brand.ratings <- read.csv("L:/MCR/r8.csv")  
head (brand.ratings)  
tail (brand.ratings)
```

```
summary (brand.ratings)  
str (brand.ratings)
```

```
x <- 1:1000  
x.sc <- (x - mean(x)) / sd(x)  
summary(x.sc)
```

```
brand.sc <- brand.ratings  
brand.sc[, 1:9] <- data.frame(scale(brand.ratings[, 1:9]))  
summary(brand.sc)
```

```
library(corrplot)
corrplot(cor(brand.sc[, 1:9]), order="hclust")
```

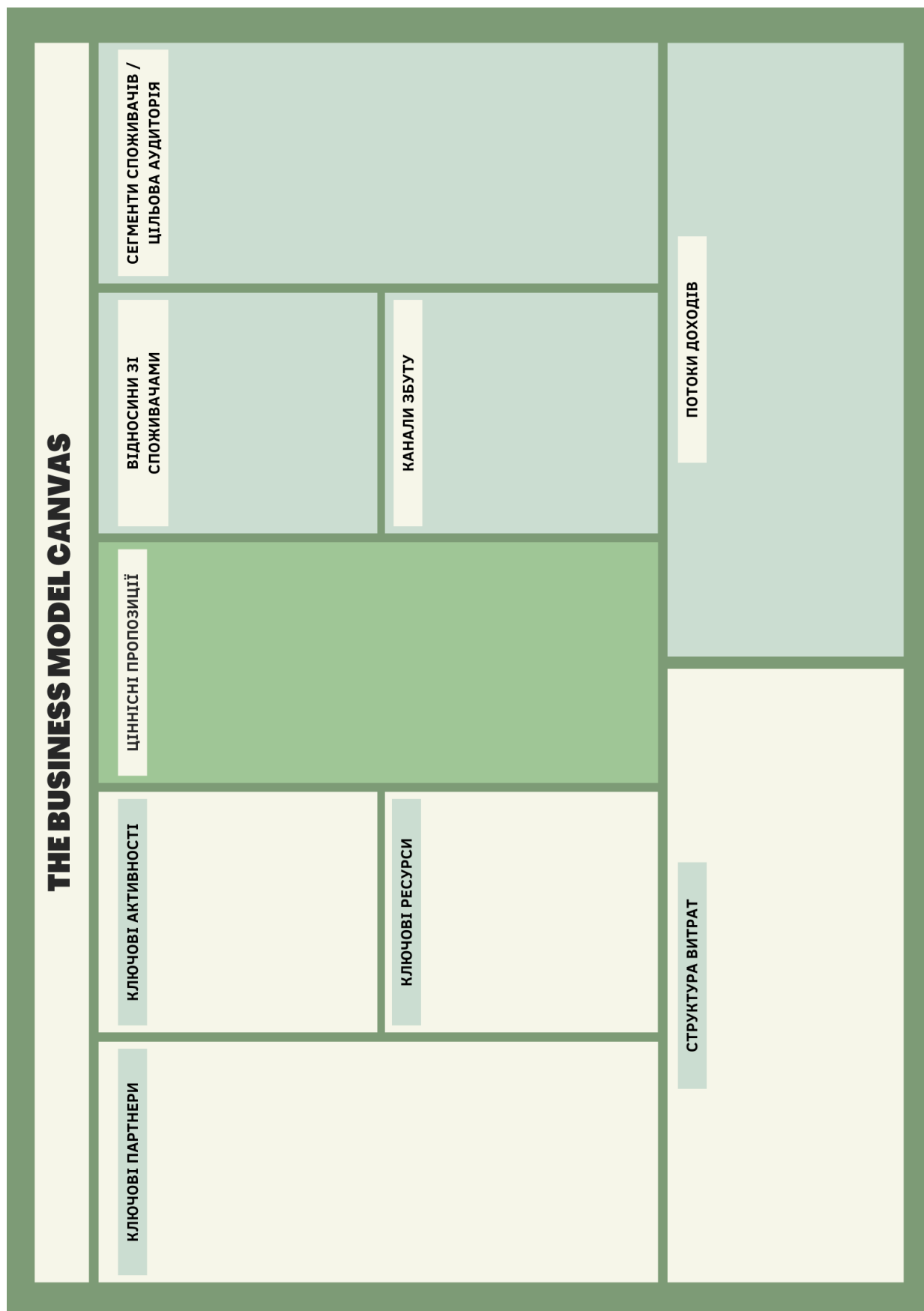
```
brand.mean <- aggregate(. ~ brand, data=brand.sc, mean)
brand.mean
```

```
rownames(brand.mean) <- brand.mean[, 1]
brand.mean <- brand.mean[, -1]
brand.mean
```

```
library(gplots)
library(RColorBrewer)
heatmap.2(as.matrix(brand.mean),
          + col=brewer.pal(9, "GnBu"), trace="none", key=FALSE,
dend="none",
          + main="\n\n\n\n\nBrand attributes")
```

ПРОДОВЖЕННЯ ДОДАТКУ Г

**БІЗНЕС-МОДЕЛЬ CANVAS ДЛЯ ВИКОНАННЯ ПРИКЛАДУ 7,
РОЗДІЛ 4**



ЗРАЗКИ ПРОГРАМНИХ КОДІВ ДЛЯ ВИКОНАННЯ ПРИКЛАДУ 7,

РОЗДІЛ 4

Version: 2.0

#####

Alternatives Section

#

Alternatives: &alternatives

Here, we list all the alternatives, together with their attributes.

We can use these attributes later in the file when defining

preferenceFunctions. The attributes can be quantitative or

qualitative.

Розпочати реалізацію проєкту:

Costs in UAH: 1000000

Відмовити у реалізації проєкту:

Costs in UAH: 10000

#

End of Alternatives Section

#####

#####

Goal Section

#

Goal:

The goal spans a tree of criteria and the alternatives

name: Встановити доцільність реалізації проєкту

description: >

This is a multiple decision maker problem.

author: Name

preferences:

preferences are defined pairwise

1 means: A is equal to B

9 means: A is highly preferable to B

1/9 means: B is highly preferable to A

Група №1:

pairwise:

- [Загальна привабливість, Інвестиційна привабливість, 1/4]
- [Загальна привабливість, Маркетингова привабливість, 3]
- [Інвестиційна привабливість, Маркетингова привабливість, 4]

Група №2:
pairwise:
- [Загальна привабливість, Інвестиційна привабливість, 1/3]
- [Загальна привабливість, Маркетингова привабливість, 1/2]
- [Інвестиційна привабливість, Маркетингова привабливість, 1/2]
children:
Загальна привабливість:
preferences:

Група №1:
pairwise:
- [Критерій П1, Критерій П2, 3]
- [Критерій П1, Критерій П3, 4]
- [Критерій П1, Критерій П4, 3]
- [Критерій П2, Критерій П3, 5]
- [Критерій П2, Критерій П4, 6]
- [Критерій П3, Критерій П4, 2]

Група №2:
pairwise:
- [Критерій П1, Критерій П2, 2]
- [Критерій П1, Критерій П3, 4]
- [Критерій П1, Критерій П4, 1]
- [Критерій П2, Критерій П3, 1]
- [Критерій П2, Критерій П4, 3]
- [Критерій П3, Критерій П4, 1]
children:
Критерій П1:
preferences:

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 3]

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1]

children: *alternatives

Критерій П2:

preferences:

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/6]

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/7]

children: *alternatives

Критерій П3:

preferences:

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/6]

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 1/8]

children: *alternatives

Критерій П4:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 1/3]

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 1/5]

children: *alternatives

Інвестиційна привабливість:

preferences:

Група №1:

pairwise:

- [Критерій В1, Критерій В2, 1/3]

- [Критерій В1, Критерій В3, 1/2]

- [Критерій В2, Критерій В3, 1]

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №2:

pairwise:

- [Критерій В1, Критерій В2, 1/2]
- [Критерій В1, Критерій В3, 1]
- [Критерій В2, Критерій В3, 1]

children:

Критерій В1:

preferences:

Група №1:

pairwise:

- [Критерій В1.1, Критерій В1.2, 5]
- [Критерій В1.1, Критерій В1.3, 7]
- [Критерій В1.2, Критерій В1.3, 1/3]

Група №2:

pairwise:

- [Критерій В1.1, Критерій В1.2, 1]
- [Критерій В1.1, Критерій В1.3, 7]
- [Критерій В1.2, Критерій В1.3, 1/6]

children:

Критерій В1.1:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 1/5]

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/3]

children: *alternatives
Критерій В1.2:
preferences:

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 7]

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 4]

children: *alternatives
Критерій В1.3:
preferences:

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/6]

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/3]

children: *alternatives
Критерій B2:
preferences:

Група №1:
pairwise:
- [Критерій B2.1, Критерій B2.2, 5]

Група №2:
pairwise:
- [Критерій B2.1, Критерій B2.2, 1/2]
children:
Критерій B2.1:
preferences:

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/6]

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/5]

children: *alternatives

Критерій В2.2:

preferences:

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 8]

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 9]

children: *alternatives

Критерій В3:

preferences:

Група №1:

pairwise:

- [Критерій В3.1, Критерій В3.2, 2]

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №2:

pairwise:

- [Критерій В3.1, Критерій В3.2, 1/2]

children:

Критерій В3.1:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 5]

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 3]

children: *alternatives

Критерій В3.2:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 7]

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 8]

children: *alternatives

Маркетингова привабливість:

preferences:

Група №1:

pairwise:

- [Критерій Г1, Критерій Г2, 1/3]

Група №2:

pairwise:

- [Критерій Г1, Критерій Г2, 1/2]

children:

Критерій Г1:

preferences:

Група №1:

pairwise:

- [Критерій Г1.1, Критерій Г1.2, 9]

Група №2:

pairwise:

- [Критерій Г1.1, Критерій Г1.2, 8]

children:

Критерій Г1.1:

preferences:

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/3]

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/5]

children: *alternatives
Критерій Г1.2:
preferences:

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 4]

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 5]

children: *alternatives
Критерій Г2:
preferences:

Група №1:
pairwise:
- [Критерій Г2.1, Критерій Г2.2, 1]

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №2:

pairwise:

- [Критерій Г2.1, Критерій Г2.2, 2]

children:

Критерій Г2.1:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/5]

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/6]

children: *alternatives

Критерій Г2.2:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/8]

ПРОДОВЖЕННЯ ДОДАТКУ Г

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/9]
children: *alternatives

End of Goal Section
#####

ЗРАЗОК ПРОГРАМНИХ КОДІВ, НАПИСАНИХ ЗДОБУВАЧЕМ

ПІД ЧАС ВИКОНАННЯ ПРИКЛАДУ 7,

РОЗДІЛ 4

Version: 2.0

#####

Alternatives Section

#

Alternatives: &alternatives

Here, we list all the alternatives, together with their attributes.

We can use these attributes later in the file when defining

preferenceFunctions. The attributes can be quantitative or

qualitative.

Розпочати реалізацію проєкту:

Costs in UAH: 1000000

Відмовити у реалізації проєкту:

Costs in UAH: 10000

#

End of Alternatives Section

#####

#####

Goal Section

#

Goal:

The goal spans a tree of criteria and the alternatives

name: Встановити доцільність реалізації проєкту відкриття коворкінгу

description: >

This is a multiple decision maker problem.

author: ПІБ здобувача

preferences:

preferences are defined pairwise

1 means: A is equal to B

9 means: A is highly preferable to B

1/9 means: B is highly preferable to A

Група №1:

pairwise:

- [Загальна привабливість, Інвестиційна привабливість, 5]
- [Загальна привабливість, Маркетингова привабливість, 3]
- [Інвестиційна привабливість, Маркетингова привабливість,

1/2]

Група №2:

pairwise:

- [Загальна привабливість, Інвестиційна привабливість, 4]
- [Загальна привабливість, Маркетингова привабливість, 5]
- [Інвестиційна привабливість, Маркетингова привабливість,

1/2]

children:

Загальна привабливість:

preferences:

ПРОДОВЖЕННЯ ДОДАТКУ Д

Група №1:

pairwise:

- [Місцезнаходження, Орендна плата, 3]
- [Місцезнаходження, Вартість споруди, 1]
- [Місцезнаходження, Можливість перепродажу, 7]
- [Орендна плата, Вартість споруди, 1/3]
- [Орендна плата, Можливість перепродажу, 5]
- [Вартість споруди, Можливість перепродажу, 7]

Група №2:

pairwise:

- [Місцезнаходження, Орендна плата, 2]
- [Місцезнаходження, Вартість споруди, 2]
- [Місцезнаходження, Можливість перепродажу, 6]
- [Орендна плата, Вартість споруди, 1/2]
- [Орендна плата, Можливість перепродажу, 3]
- [Вартість споруди, Можливість перепродажу, 8]

children:

Місцезнаходження:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проекту, Відмовити у реалізації

проекту, 3]

ПРОДОВЖЕННЯ ДОДАТКУ Д

- Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1]
children: *alternatives
Орендна плата:
preferences:
Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 4]
Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 2]
children: *alternatives
Вартість споруди:
preferences:
Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/6]

ПРОДОВЖЕННЯ ДОДАТКУ Д

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 1/8]

children: *alternatives

Можливість перепродажу:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 1/5]

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 1/7]

children: *alternatives

Інвестиційна привабливість:

preferences:

Група №1:

pairwise:

- [Загальна інвестиційна привабливість, Привабливість для іноземного капіталу, 9]

- [Загальна інвестиційна привабливість, Привабливість для вітчизняного капіталу, 9]

- [Привабливість для іноземного капіталу, Привабливість для вітчизняного капіталу, 1/5]

Група №2:

pairwise:

- [Загальна інвестиційна привабливість, Привабливість для іноземного капіталу, 7]

ПРОДОВЖЕННЯ ДОДАТКУ Д

- [Загальна інвестиційна привабливість, Привабливість для вітчизняного капіталу, 7]

- [Привабливість для іноземного капіталу, Привабливість для вітчизняного капіталу, 1/4]

children:

Загальна інвестиційна привабливість:

preferences:

Група №1:

pairwise:

- [Термін окупності, NPV, 5]

- [Термін окупності, IRR, 7]

- [NPV, IRR, 1/3]

Група №2:

pairwise:

- [Термін окупності, NPV, 4]

- [Термін окупності, IRR, 7]

- [NPV, IRR, 1/5]

children:

Термін окупності:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації

проєкту, 6]

ПРОДОВЖЕННЯ ДОДАТКУ Д

проєкту, 5]	Група №2:
	pairwise:
	- [Розпочати реалізацію проєкту, Відмовити у реалізації
	children: *alternatives
	NPV:
проєкту, 7]	preferences:
	Група №1:
	pairwise:
	- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 4]	Група №2:
	pairwise:
	- [Розпочати реалізацію проєкту, Відмовити у реалізації
	children: *alternatives
	IRR:
проєкту, 6]	preferences:
	Група №1:
	pairwise:
	- [Розпочати реалізацію проєкту, Відмовити у реалізації

ПРОДОВЖЕННЯ ДОДАТКУ Д

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 3]

children: *alternatives

Привабливість для іноземного капіталу:

preferences:

Група №1:

pairwise:

- [Курс валюти, Податкова політика, 5]

Група №2:

pairwise:

- [Курс валюти, Податкова політика, 7]

children:

Курс валюти:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 1/6]

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації проєкту, 1/5]

children: *alternatives

Податкова політика:

preferences:

ПРОДОВЖЕННЯ ДОДАТКУ Д

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 8]

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 9]

children: *alternatives

Привабливість для вітчизняного капіталу:

preferences:

Група №1:

pairwise:

- [Політична ситуація, Річний дохід, 2]

Група №2:

pairwise:

- [Політична ситуація, Річний дохід, 1]

children:

Політична ситуація:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації

проєкту, 5]

ПРОДОВЖЕННЯ ДОДАТКУ Д

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 3]

children: *alternatives

Річний дохід:

preferences:

Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 7]

Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 8]

children: *alternatives

Маркетингова привабливість:

preferences:

Група №1:

pairwise:

- [Макросередовище, Мікросередовище, 1/3]

Група №2:

pairwise:

- [Макросередовище, Мікросередовище, 1/2]

children:

Макросередовище:

preferences:

Група №1:

pairwise:

- [Демографічне становище, Економічне становище, 9]

ПРОДОВЖЕННЯ ДОДАТКУ Д

Група №2:

pairwise:

- [Демографічне становище, Економічне становище, 8]

children:

Демографічне становище:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації

проєкту, 1/3]

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації

проєкту, 1/5]

children: *alternatives

Економічне становище:

preferences:

Група №1:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації

проєкту, 4]

Група №2:

pairwise:

- [Розпочати реалізацію проєкту, Відмовити у реалізації

проєкту, 5]

children: *alternatives

Мікросередовище:

preferences:

Група №1:

pairwise:

- [Конкуренти, Клієнти, 1]

ПРОДОВЖЕННЯ ДОДАТКУ Д

Група №2:
pairwise:
- [Конкуренти, Клієнти, 2]
children:
Конкуренти:
preferences:
Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/5]
Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/6]
children: *alternatives
Клієнти:
preferences:
Група №1:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/8]
Група №2:
pairwise:
- [Розпочати реалізацію проєкту, Відмовити у реалізації
проєкту, 1/9]
children: *alternatives

End of Goal Section
#####

**ЗРАЗОК ІЄРАРХІЇ, ОТРИМАНОЇ ЗДОБУВАЧЕМ
ПІД ЧАС ВИКОНАННЯ ПРИКЛАДУ 7, РОЗДІЛ 4**



ЗРАЗОК ПРЕДСТАВЛЕННЯ РЕЗУЛЬТАТІВ РОЗРАХУНКІВ ЗА

ОКРЕМИМИ ГРУПАМИ ЕКСПЕРТІВ

ПІД ЧАС ВИКОНАННЯ ПРИКЛАДУ 7, РОЗДІЛ 4

Таблиця Ж.1 Результати аналізу перспективності інноваційного проекту «3D-On» на основі урахування думок групи експертів № 1

Рівень ієрархії та критерій оцінювання	Вагове значення	Продовжити реалізацію проєкта	Припинити реалізацію проєкта
1	2	3	4
Встановити привабливість стартап-проекту 3D-On	100,0 %	80,7 %	19,3 %
<i>Маркетингова привабливість</i>	55,8 %	45,3 %	10,6 %
<i>Товар</i>	24,8 %	20,7 %	4,1 %
Рівень інноваційності товару	14,4 %	12,4 %	2,1 %
Рівень технологічної складності товару	8,6 %	6,9 %	1,7 %
Рівень готовності науково-технічних розробок	1,7 %	1,4 %	0,3 %
<i>Маркетингове середовище</i>	21,6 %	17,6 %	4,0 %
Наявність сформованої та деталізованої у часі маркетингової стратегії	11,4 %	9,5 %	1,9 %
Відповідність визначених у стратегії цілей правилу SMART	5,6 %	4,7 %	0,9 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.1

Рівень розуміння ринку	4,6 %	3,5 %	1,2 %
<i>Ресурси</i>	9,4 %	6,9 %	2,5 %
Перспективи масштабування бізнесу	5,6 %	4,7 %	0,9 %
Наявність власних ресурсів	1,9 %	0,6 %	1,2 %
Доступність виробничих потужностей для запуску виробництва продукції	0,9 %	0,8 %	0,1 %
Можливість залучення ресурсів	0,5 %	0,4 %	0,1 %
Можливості використання краудсорсингу	0,3 %	0,3 %	0,1 %
Можливість передачі частини робіт на аутсорсинг	0,2 %	0,2 %	0,1 %
<i>Ринкова привабливість</i>	32,0 %	25,5 %	6,4 %
Рівень платоспроможності цільової аудиторії (ЦА)	9,5 %	7,1 %	2,4 %
Наявність чітко визначеної ЦА	8,2 %	6,8 %	1,4 %
Оціночна ємність ринку	3,8 %	3,3 %	0,5 %
Доступ до ринків збуту	3,3 %	2,8 %	0,6 %
Доступ до джерел фінансування	2,9 %	2,2 %	0,7 %
Наявність товарів-субститутів і товарів-аналогів та рівень їх схожості з прототипом	1,7 %	1,3 %	0,4 %
Ступінь актуалізації потреб з боку цільової аудиторії	0,9 %	0,7 %	0,2 %
Рівень сформованості ринкового попиту на товарному ринку	0,9 %	0,8 %	0,2 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.1

Наявність досвіду і технічних знань з використання подібної продукції	0,6 %	0,5 %	0,1 %
Інвестиційна привабливість	12,2 %	10,0 %	2,2 %
<i>Фінансово-економічні аспекти</i>	3,2 %	2,6 %	0,6 %
Рентабельність продажу одиниці продукції	0,7 %	0,6 %	0,1 %
Фінансові ресурси для масштабування бізнесу	0,6 %	0,5 %	0,1 %
Регулярність фінансових надходжень	0,5 %	0,4 %	0,1 %
Час виходу на точку беззбитковості	0,4 %	0,3 %	0,1 %
Наявність чіткого бізнес-плану адаптованого до місцезнаходження стартапу	0,4 %	0,3 %	0,0 %
Величина зібраних фінансових ресурсів	0,4 %	0,3 %	0,1 %
Необхідна сума інвестицій	0,3 %	0,3 %	0,0 %
<i>Організаційно-економічні чинники</i>	3,1 %	2,7 %	0,4 %
Технологічна складність виробництва	0,9 %	0,8 %	0,1 %
Інфраструктурна підтримка	0,7 %	0,7 %	0,1 %
Рівень кваліфікації персоналу	0,5 %	0,5 %	0,1 %
Наявність виробничо-економічних потужностей з досвідом використання подібних технологій	0,4 %	0,3 %	0,1 %
Рівень мотивації учасників команди	0,2 %	0,2 %	0,0 %
Досвід реалізації попередніх проектів	0,2 %	0,1 %	0,0 %
Гарантії і рекомендації від авторитетних компаній і організацій-партнерів стартапу	0,1 %	0,1 %	0,0 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.1

Командний дух і система цінностей	0,1 %	0,1 %	0,0 %
<i>Споживчі чинники</i>	2,2 %	1,9 %	0,3 %
Стабільність ринкового попиту	1,1 %	0,9 %	0,2 %
Можливості створення бренду	0,6 %	0,5 %	0,1 %
Наявність чітко вираженої цінності товару для споживача	0,5 %	0,4 %	0,1 %
Оціночна тривалість ЖЦ товару	0,1 %	0,1 %	0,0 %
<i>Енвайронменталістські чинники</i>	2,0 %	1,5 %	0,4 %
Економічна і політична обстановка в країні	1,0 %	0,8 %	0,3 %
Загальний рівень ризикованості бізнесу	0,3 %	0,3 %	0,1 %
Наявність у країні глобальних конкурентних переваг у сфері реалізації стартапу	0,3 %	0,3 %	0,0 %
Рівень легкості ведення бізнесу	0,2 %	0,2 %	0,0 %
Легкість отримання кредитів	0,1 %	0,0 %	0,0 %
<i>Підтримка</i>	1,0 %	0,8 %	0,2 %
Ступінь висвітлення у ЗМІ	0,4 %	0,3 %	0,1 %
Контентна підтримка товарної категорії	0,3 %	0,2 %	0,0 %
Наявність збутових мереж і посередників	0,2 %	0,2 %	0,1 %
Підтримка з боку органів влади	0,1 %	0,0 %	0,0 %
Наявність спеціальних інвестиційних фондів	0,1 %	0,0 %	0,0 %
<i>Організаційно-правове забезпечення</i>	0,7 %	0,5 %	0,2 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.1

Захищеність прав інтелектуальної власності	0,4 %	0,3 %	0,1 %
Наявність договорів на аренду виробничих потужностей	0,1 %	0,1 %	0,0 %
Наявність поручителів за кредитними позиками	0,1 %	0,0 %	0,0 %
Наявність договорів на офісні приміщення	0,0 %	0,0 %	0,0 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Таблиця Ж.2 – Результати аналізу перспективності інноваційного проекту «3D-On» на основі урахування думок групи експертів № 2

Рівень ієрархії та критерій оцінювання	Вагове значення	Продовжити реалізацію проєкта	Припинити реалізацію проєкта
1	2	3	4
Встановити привабливість стартап-проєкту 3D-On	100,0 %	80,2 %	19,8 %
<i>Інвестиційна привабливість</i>	50,0 %	39,6 %	10,4 %
<i>Організаційно-економічні чинники</i>	12,3 %	10,4 %	1,9 %
Технологічна складність виробництва	3,6 %	3,0 %	0,5 %
Інфраструктурна підтримка	2,9 %	2,5 %	0,4 %
Рівень кваліфікації персоналу	2,1 %	1,8 %	0,2 %
Наявність виробничо-економічних потужностей з досвідом використання подібних технологій	1,4 %	1,1 %	0,3 %
Рівень мотивації учасників команди	0,9 %	0,8 %	0,1 %
Досвід реалізації попередніх проєктів	0,6 %	0,4 %	0,2 %
Гарантії і рекомендації від авторитетних компаній і організацій-партнерів стартапу	0,5 %	0,3 %	0,2 %
Командний дух і система цінностей	0,4 %	0,3 %	0,0 %
<i>Фінансово-економічні аспекти</i>	12,0 %	9,4 %	2,6 %
Регулярність фінансових надходжень	2,7 %	1,8 %	0,9 %
Рентабельність продажу одиниці продукції	2,7 %	2,2 %	0,4 %
Час виходу на точку беззбитковості	2,0 %	1,6 %	0,4 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.2

Необхідна сума інвестицій	1,6 %	1,4 %	0,2 %
Фінансові ресурси для масштабування бізнесу	1,2 %	0,9 %	0,2 %
Наявність чіткого бізнес-плану адаптованого до місцезнаходження стартапу	1,1 %	1,0 %	0,2 %
Величина зібраних фінансових ресурсів	0,7 %	0,5 %	0,2 %
<i>Енвіронменталістські чинники</i>	11,6 %	8,6 %	3,0 %
Економічна і політична обстановка в країні	4,9 %	3,3 %	1,6 %
Загальний рівень ризикованості бізнесу	3,1 %	2,3 %	0,8 %
Наявність у країні глобальних конкурентних переваг у сфері реалізації стартапу	1,9 %	1,7 %	0,3 %
Рівень легкості ведення бізнесу	1,4 %	1,2 %	0,2 %
Легкість отримання кредитів	0,4 %	0,3 %	0,1 %
<i>Споживчі чинники</i>	8,8 %	7,6 %	1,3 %
Стабільність ринкового попиту	4,1 %	3,4 %	0,7 %
Наявність чітко вираженої цінності товару для споживача	2,3 %	2,0 %	0,3 %
Можливості створення бренду	1,9 %	1,7 %	0,2 %
Оціночна тривалість ЖЦ товару	0,5 %	0,4 %	0,1 %
<i>Організаційно-правове забезпечення</i>	2,8 %	1,8 %	1,0 %
Захищеність прав інтелектуальної власності	2,0 %	1,3 %	0,7 %
Наявність договорів на аренду виробничих потужностей	0,5 %	0,2 %	0,2 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.2

Наявність поручителів за кредитними позиками	0,2 %	0,1 %	0,1 %
Наявність договорів на офісні приміщення	0,2 %	0,1 %	0,0 %
<i>Підтримка</i>	2,5 %	1,9 %	0,6 %
Ступінь висвітлення у ЗМІ	1,2 %	0,9 %	0,3 %
Контентна підтримка товарної категорії	0,5 %	0,5 %	0,1 %
Наявність збутових мереж і посередників	0,5 %	0,3 %	0,2 %
Підтримка з боку органів влади	0,2 %	0,1 %	0,1 %
Наявність спеціальних інвестиційних фондів	0,1 %	0,1 %	0,1 %
<i>Ринкова привабливість</i>	25,0 %	20,2 %	4,8 %
Рівень платоспроможності цільової аудиторії (ЦА)	7,0 %	5,3 %	1,8 %
Наявність чітко визначеної ЦА	6,7 %	5,8 %	0,8 %
Доступ до ринків збуту	3,5 %	2,9 %	0,6 %
Оціночна ємність ринку	2,5 %	2,2 %	0,3 %
Доступ до джерел фінансування	2,2 %	1,6 %	0,5 %
Наявність товарів-субститутів і товарів-аналогів та рівень їх схожості з прототипом	1,5 %	1,0 %	0,5 %
Рівень сформованості ринкового попиту на товарному ринку	0,7 %	0,6 %	0,1 %
Ступінь актуалізації потреб з боку ЦА	0,5 %	0,4 %	0,1 %
Наявність досвіду і технічних знань з використання подібної продукції	0,4 %	0,4 %	0,1 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.2

Маркетингова привабливість	25,0 %	20,4 %	4,6 %
<i>Товар</i>	14,0 %	11,6 %	2,3 %
Рівень технологічної складності товару	6,8 %	5,7 %	1,1 %
Рівень інноваційності товару	6,2 %	5,2 %	1,0 %
Рівень готовності науково-технічних розробок	0,9 %	0,8 %	0,2 %
<i>Маркетингове середовище</i>	8,0 %	6,3 %	1,7 %
Наявність сформованої та деталізованої у часі маркетингової стратегії	5,6 %	4,4 %	1,1 %
Відповідність визначених у стратегії цілей правилу SMART	1,3 %	1,1 %	0,3 %
Рівень розуміння ринку	1,1 %	0,8 %	0,3 %
<i>Ресурси</i>	3,0 %	2,4 %	0,7 %
Перспективи масштабування бізнесу	1,8 %	1,5 %	0,3 %
Доступність виробничих потужностей для запуску виробництва продукції	0,4 %	0,4 %	0,1 %
Наявність власних ресурсів	0,4 %	0,2 %	0,2 %
Можливість залучення ресурсів	0,2 %	0,2 %	0,0 %
Можливості використання краудсорсингу	0,1 %	0,1 %	0,0 %
Можливість передачі частини робіт на аутсорсинг	0,1 %	0,0 %	0,0 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Таблиця Ж.3 – Результати аналізу перспективності інноваційного проекту «3D-On» на основі урахування думок групи експертів № 3

Рівень ієрархії та критерій оцінювання	Вагове значення	Продовжити реалізацію проєкта	Припинити реалізацію проєкта
1	2	3	4
Встановити привабливість стартап-проєкту 3D-On	100,0 %	81,4 %	18,6 %
<i>Ринкова привабливість</i>	65,7 %	53,4 %	12,3 %
Рівень платоспроможності цільової аудиторії (ЦА)	16,6 %	13,9 %	2,8 %
Доступ до ринків збуту	14,3 %	12,3 %	2,0 %
Наявність чітко визначеної ЦА	13,8 %	11,1 %	2,8 %
Доступ до джерел фінансування	6,5 %	4,9 %	1,6 %
Наявність товарів-субститутів і товарів-аналогів та рівень їх схожості з прототипом	4,8 %	3,2 %	1,6 %
Оціночна ємність ринку	4,5 %	3,9 %	0,6 %
Рівень сформованості ринкового попиту на товарному ринку	2,0 %	1,8 %	0,3 %
Наявність досвіду і технічних знань з використання подібної продукції	1,7 %	1,4 %	0,3 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.3

Ступінь актуалізації потреб з боку ЦА	1,4 %	1,1 %	0,4 %
<i>Маркетингова привабливість</i>	26,1 %	21,2 %	4,9 %
<i>Товар</i>	10,8 %	9,3 %	1,4 %
Рівень технологічної складності товару	6,5 %	5,7 %	0,8 %
Рівень інноваційності товару	3,5 %	3,0 %	0,5 %
Рівень готовності науково-технічних розробок	0,8 %	0,6 %	0,1 %
<i>Маркетингове середовище</i>	8,5 %	6,9 %	1,6 %
Наявність сформованої та деталізованої у часі маркетингової стратегії	3,9 %	3,1 %	0,8 %
Рівень розуміння ринку	2,7 %	2,2 %	0,5 %
Відповідність визначених у стратегії цілей правила SMART	1,9 %	1,6 %	0,3 %
<i>Ресурси</i>	6,8 %	5,0 %	1,8 %
Перспективи масштабування бізнесу	3,5 %	3,0 %	0,5 %
Наявність власних ресурсів	1,5 %	0,5 %	1,0 %
Доступність виробничих потужностей для запуску виробництва продукції	0,9 %	0,8 %	0,1 %
Можливість залучення ресурсів	0,6 %	0,5 %	0,1 %
Можливості використання краудсорсингу	0,2 %	0,2 %	0,0 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.3

Можливість передачі частини робіт на аутсорсинг	0,1 %	0,1 %	0,1 %
<i>Інвестиційна привабливість</i>	8,3 %	6,7 %	1,5 %
<i>Організаційно-економічні чинники</i>	2,5 %	2,2 %	0,4 %
Технологічна складність виробництва	0,8 %	0,7 %	0,1 %
Інфраструктурна підтримка	0,6 %	0,5 %	0,1 %
Рівень кваліфікації персоналу	0,4 %	0,4 %	0,0 %
Наявність виробничо-економічних потужностей з досвідом використання подібних технологій	0,3 %	0,2 %	0,0 %
Рівень мотивації учасників команди	0,2 %	0,1 %	0,0 %
Досвід реалізації попередніх проектів	0,1 %	0,1 %	0,0 %
Гарантії і рекомендації від авторитетних компаній і організацій-партнерів стартапу	0,1 %	0,1 %	0,0 %
Командний дух і система цінностей	0,1 %	0,1 %	0,0 %
<i>Фінансово-економічні аспекти</i>	2,4 %	1,9 %	0,5 %
Регулярність фінансових надходжень	0,6 %	0,4 %	0,2 %
Фінансові ресурси для масштабування бізнесу	0,4 %	0,3 %	0,1 %
Рентабельність продажу одиниці продукції	0,4 %	0,3 %	0,1 %
Величина зібраних фінансових ресурсів	0,3 %	0,3 %	0,1 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж			
Продовження таблиці Ж.3			
Необхідна сума інвестицій	0,3 %	0,3 %	0,0 %
Час виходу на точку беззбитковості	0,2 %	0,2 %	0,0 %
Наявність чіткого бізнес-плану адаптованого до місцезнаходження стартапу	0,2 %	0,2 %	0,0 %
<i>Енвайронменталістські чинники</i>	1,1 %	1,0 %	0,2 %
Економічна і політична обстановка в країні	0,5 %	0,5 %	0,1 %
Загальний рівень ризикованості бізнесу	0,2 %	0,2 %	0,0 %
Наявність у країні глобальних конкурентних переваг у сфері реалізації стартапу	0,2 %	0,2 %	0,0 %
Рівень легкості ведення бізнесу	0,2 %	0,1 %	0,0 %
Легкість отримання кредитів	0,0 %	0,0 %	0,0 %
<i>Споживчі чинники</i>	0,9 %	0,8 %	0,1 %
Стабільність ринкового попиту	0,4 %	0,4 %	0,1 %
Наявність чітко вираженої цінності товару для споживача	0,3 %	0,2 %	0,0 %
Можливості створення бренду	0,2 %	0,2 %	0,0 %
Оціночна тривалість ЖЦ товару	0,0 %	0,0 %	0,0 %
<i>Підтримка</i>	0,8 %	0,6 %	0,2 %
Наявність збутових мереж і посередників	0,2 %	0,1 %	0,1 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.3

Контентна підтримка товарної категорії	0,2 %	0,2 %	0,0 %
Ступінь висвітлення у ЗМІ	0,2 %	0,2 %	0,1 %
Підтримка з боку органів влади	0,1 %	0,0 %	0,0 %
Наявність спеціальних інвестиційних фондів	0,0 %	0,0 %	0,0 %
<i>Організаційно-правове забезпечення</i>	0,5 %	0,4 %	0,1 %
Захищеність прав інтелектуальної власності	0,3 %	0,2 %	0,1 %
Наявність договорів на аренду виробничих потужностей	0,1 %	0,1 %	0,0 %
Наявність договорів на офісні приміщення	0,0 %	0,0 %	0,0 %
Наявність поручителів за кредитними позиками	0,0 %	0,0 %	0,0 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Таблиця Ж.4 – Результати аналізу перспективності інноваційного проекту «3D-On» на основі урахування думок групи експертів № 4

Рівень ієрархії та критерій оцінювання	Вагове значення	Продовжити реалізацію проекту	Припинити реалізацію проекту
1	2	3	4
Встановити привабливість стартап-проекту 3D-On	100,0 %	83,8 %	16,2 %
<i>Інвестиційна привабливість</i>	74,7 %	62,6 %	12,1 %
<i>Фінансово-економічні аспекти</i>	29,4 %	24,3 %	5,1 %
Регулярність фінансових надходжень	5,9 %	4,4 %	1,5 %
Фінансові ресурси для масштабування бізнесу	5,9 %	5,0 %	0,8 %
Необхідна сума інвестицій	5,1 %	4,5 %	0,6 %
Величина зібраних фінансових ресурсів	5,0 %	4,0 %	1,0 %
Час виходу на точку беззбитковості	4,0 %	3,3 %	0,7 %
Рентабельність продажу одиниці продукції	2,0 %	1,7 %	0,3 %
Наявність чіткого бізнес-плану адаптованого до місцезнаходження стартапу	1,5 %	1,3 %	0,2 %
<i>Організаційно-економічні чинники</i>	15,4 %	13,4 %	2,0 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.4

Технологічна складність виробництва	4,2 %	3,7 %	0,5 %
Інфраструктурна підтримка	3,7 %	3,3 %	0,4 %
Рівень кваліфікації персоналу	2,7 %	2,4 %	0,3 %
Наявність виробничо-економічних потужностей з досвідом використання подібних технологій	1,9 %	1,5 %	0,4 %
Рівень мотивації учасників команди	1,1 %	1,0 %	0,1 %
Досвід реалізації попередніх проектів	0,8 %	0,6 %	0,2 %
Гарантії і рекомендації від авторитетних компаній і організацій-партнерів стартапу	0,6 %	0,5 %	0,2 %
Командний дух і система цінностей	0,5 %	0,5 %	0,1 %
<i>Енвіронменталістські чинники</i>	12,8 %	10,8 %	2,0 %
Економічна і політична обстановка в країні	6,1 %	5,0 %	1,0 %
Загальний рівень ризикованості бізнесу	3,2 %	2,6 %	0,5 %
Наявність у країні глобальних конкурентних переваг у сфері реалізації стартапу	1,8 %	1,6 %	0,2 %
Рівень легкості ведення бізнесу	1,4 %	1,2 %	0,2 %
Легкість отримання кредитів	0,4 %	0,3 %	0,1 %
<i>Споживчі чинники</i>	7,6 %	6,6 %	0,9 %
Стабільність ринкового попиту	3,3 %	2,8 %	0,5 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.4

Можливості створення бренду	2,3 %	2,1 %	0,2 %
Наявність чітко вираженої цінності товару для споживача	1,5 %	1,3 %	0,2 %
Оціночна тривалість життєвого циклу товару	0,4 %	0,4 %	0,1 %
<i>Підтримка</i>	6,3 %	4,9 %	1,3 %
Ступінь висвітлення у ЗМІ	2,9 %	2,4 %	0,6 %
Контентна підтримка товарної категорії	1,4 %	1,2 %	0,1 %
Наявність збутових мереж і посередників	1,2 %	0,8 %	0,4 %
Наявність спеціальних інвестиційних фондів	0,4 %	0,3 %	0,1 %
Підтримка з боку органів влади	0,3 %	0,3 %	0,1 %
<i>Організаційно-правове забезпечення</i>	3,3 %	2,5 %	0,7 %
Захищеність прав інтелектуальної власності	2,2 %	1,8 %	0,4 %
Наявність договорів на аренду виробничих потужностей	0,6 %	0,4 %	0,1 %
Наявність поручителів за кредитними позиками	0,2 %	0,2 %	0,1 %
Наявність договорів на офісні приміщення	0,2 %	0,2 %	0,0 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.4

<i>Маркетингова привабливість</i>	13,4 %	11,4 %	1,9 %
<i>Товар</i>	6,6 %	5,7 %	0,9 %
Рівень технологічної складності товару	4,0 %	3,4 %	0,6 %
Рівень інноваційності товару	2,1 %	1,8 %	0,3 %
Рівень готовності науково-технічних розробок	0,5 %	0,5 %	0,1 %
<i>Ресурси</i>	4,2 %	3,5 %	0,7 %
Перспективи масштабування бізнесу	2,5 %	2,2 %	0,3 %
Наявність власних ресурсів	0,7 %	0,5 %	0,2 %
Доступність виробничих потужностей для запуску виробництва продукції	0,4 %	0,4 %	0,1 %
Можливість залучення ресурсів	0,3 %	0,2 %	0,0 %
Можливості використання краудсорсингу	0,2 %	0,1 %	0,0 %
Можливість передачі частини робіт на аутсорсинг	0,1 %	0,1 %	0,0 %
<i>Маркетингове середовище</i>	2,6 %	2,3 %	0,4 %
Наявність сформованої та деталізованої у часі маркетингової стратегії	1,9 %	1,7 %	0,2 %
Відповідність визначених у стратегії цілей правила SMART	0,4 %	0,3 %	0,1 %
Рівень розуміння ринку	0,3 %	0,3 %	0,1 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.4

Ринкова привабливість	11,9 %	9,8 %	2,1 %
Рівень платоспроможності цільової аудиторії (ЦА)	3,8 %	3,2 %	0,6 %
Доступ до джерел фінансування	2,1 %	1,4 %	0,7 %
Наявність чітко визначеної ЦА	1,8 %	1,6 %	0,2 %
Доступ до ринків збуту	1,3 %	1,1 %	0,2 %
Оціночна ємність ринку	1,1 %	1,0 %	0,1 %
Наявність товарів-субститутів і товарів-аналогів та рівень їх схожості з прототипом	0,7 %	0,6 %	0,1 %
Ступінь актуалізації потреб з боку ЦА	0,4 %	0,4 %	0,1 %
Рівень сформованості ринкового попиту на товарному ринку	0,4 %	0,4 %	0,1 %
Наявність досвіду і технічних знань з використання подібної продукції	0,3 %	0,2 %	0,0 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Таблиця Ж.5 – Результати аналізу перспективності інноваційного проекту «3D-On» на основі урахування думок групи експертів № 5

Рівень ієрархії та критерій оцінювання	Вагове значення	Продовжити реалізацію проекту	Припинити реалізацію проекту
1	2	3	4
Встановити привабливість стартап-проекту 3D-On	100,0 %	83,8 %	16,2 %
<i>Маркетингова привабливість</i>	55,0 %	46,6 %	8,4 %
<i>Товар</i>	30,2 %	26,1 %	4,1 %
Рівень технологічної складності товару	17,4 %	15,0 %	2,5 %
Рівень інноваційності товару	10,3 %	9,1 %	1,3 %
Рівень готовності науково-технічних розробок	2,5 %	2,1 %	0,4 %
<i>Маркетингове середовище</i>	13,2 %	10,9 %	2,3 %
Рівень розуміння ринку	6,0 %	4,8 %	1,2 %
Наявність сформованої та деталізованої у часі маркетингової стратегії	5,5 %	4,7 %	0,8 %
Відповідність визначених у стратегії цілей правилу SMART	1,7 %	1,4 %	0,3 %
<i>Ресурси</i>	11,5 %	9,6 %	2,0 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж			
<i>Продовження таблиці Ж.5</i>			
Перспективи масштабування бізнесу	6,7 %	5,8 %	0,8 %
Наявність власних ресурсів	2,3 %	1,5 %	0,8 %
Доступність виробничих потужностей для запуску виробництва продукції	1,2 %	1,1 %	0,1 %
Можливість залучення ресурсів	0,7 %	0,6 %	0,1 %
Можливості використання краудсорсингу	0,4 %	0,4 %	0,1 %
Можливість передачі частини робіт на аутсорсинг	0,2 %	0,2 %	0,1 %
<i>Ринкова привабливість</i>	24,0 %	19,9 %	4,1 %
Наявність чітко визначеної ЦА	7,0 %	6,0 %	1,0 %
Рівень платоспроможності ЦА	5,6 %	4,5 %	1,1 %
Оціночна ємність ринку	3,6 %	3,2 %	0,5 %
Доступ до ринків збуту	2,9 %	2,4 %	0,5 %
Доступ до джерел фінансування	1,7 %	1,3 %	0,4 %
Наявність товарів-субститутів і товарів-аналогів та рівень їх схожості з прототипом	1,5 %	1,1 %	0,4 %
Рівень сформованості ринкового попиту на товарному ринку	0,7 %	0,6 %	0,1 %
Наявність досвіду і технічних знань з використання подібної продукції	0,5 %	0,5 %	0,1 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.5

Ступінь актуалізації потреб з боку ЦА	0,5 %	0,4 %	0,1 %
<i>Інвестиційна привабливість</i>	21,0 %	17,3 %	3,7 %
<i>Фінансово-економічні аспекти</i>	8,2 %	6,6 %	1,6 %
Фінансові ресурси для масштабування бізнесу	1,8 %	1,5 %	0,3 %
Регулярність фінансових надходжень	1,7 %	1,1 %	0,6 %
Необхідна сума інвестицій	1,4 %	1,2 %	0,2 %
Час виходу на точку беззбитковості	1,2 %	1,0 %	0,2 %
Величина зібраних фінансових ресурсів	1,2 %	0,9 %	0,3 %
Рентабельність продажу одиниці продукції	0,5 %	0,4 %	0,1 %
Наявність чіткого бізнес-плану адаптованого до місцезнаходження стартапу	0,4 %	0,4 %	0,1 %
<i>Організаційно-економічні чинники</i>	4,4 %	3,9 %	0,6 %
Технологічна складність виробництва	1,3 %	1,1 %	0,2 %
Інфраструктурна підтримка	1,0 %	0,9 %	0,1 %
Рівень кваліфікації персоналу	0,8 %	0,7 %	0,1 %
Наявність виробничо-економічних потужностей з досвідом використання подібних технологій	0,5 %	0,4 %	0,1 %
Рівень мотивації учасників команди	0,3 %	0,2 %	0,0 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.5

Досвід реалізації попередніх проектів	0,3 %	0,2 %	0,1 %
Гарантії і рекомендації від авторитетних компаній і організацій-партнерів стартапу	0,2 %	0,1 %	0,0 %
Командний дух і система цінностей	0,1 %	0,1 %	0,0 %
<i>Споживчі чинники</i>	3,9 %	3,4 %	0,5 %
Стабільність ринкового попиту	1,9 %	1,6 %	0,3 %
Можливості створення бренду	1,2 %	1,1 %	0,1 %
Наявність чітко вираженої цінності товару для споживача	0,6 %	0,5 %	0,1 %
Оціночна тривалість ЖЦ товару	0,2 %	0,2 %	0,0 %
<i>Енвіронменталістські чинники</i>	2,5 %	2,0 %	0,4 %
Економічна і політична обстановка в країні	1,0 %	0,8 %	0,2 %
Загальний рівень ризикованості бізнесу	0,6 %	0,5 %	0,1 %
Наявність у країні глобальних конкурентних переваг у сфері реалізації стартапу	0,4 %	0,4 %	0,1 %
Рівень легкості ведення бізнесу	0,3 %	0,3 %	0,0 %
Легкість отримання кредитів	0,1 %	0,0 %	0,0 %
<i>Підтримка</i>	1,0 %	0,8 %	0,2 %
Ступінь висвітлення у ЗМІ	0,4 %	0,3 %	0,1 %

ПРОДОВЖЕННЯ ДОДАТКУ Ж

Продовження таблиці Ж.5

Контентна підтримка товарної категорії	0,2 %	0,2 %	0,0 %
Наявність збутових мереж і посередників	0,2 %	0,2 %	0,1 %
Наявність спеціальних інвестиційних фондів	0,1 %	0,0 %	0,0 %
Підтримка з боку органів влади	0,1 %	0,0 %	0,0 %
<i>Організаційно-правове забезпечення</i>	0,9 %	0,7 %	0,3 %
Захищеність прав інтелектуальної власності	0,6 %	0,5 %	0,2 %
Наявність договорів на аренду виробничих потужностей	0,2 %	0,1 %	0,1 %
Наявність договорів на офісні приміщення	0,1 %	0,1 %	0,0 %
Наявність поручителів за кредитними позиками	0,1 %	0,0 %	0,0 %

ДЛЯ НОТАТОК

ДЛЯ НОТАТОК

ДЛЯ НОТАТОК